Available online at www.sciencedirect.com
SCIENCE DIRECT®
COGNITION
ELSEVIER
Cognition 92 (2004) B13–B23
www.elsevier.com/locate/COGNIT

Brief article

Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task

Salvador Soto-Faraco*, Jordi Navarra, Agnès Alsius

SPPB Group, Departament de Psicologia Bàsica, Universitat de Barcelona, Pg. Vall d'Hebrón, 171, 08035 Barcelona, Spain

Received 9 July 2003; accepted 28 October 2003

Abstract

The McGurk effect is usually presented as an example of fast, automatic, multisensory integration. We report a series of experiments designed to directly assess these claims. We used a syllabic version of the *speeded classification* paradigm, whereby response latencies to the first (target) syllable of spoken word-like stimuli are slowed down when the second (irrelevant) syllable varies from trial to trial. This interference effect is interpreted as a failure of selective attention to filter out the irrelevant syllable. In Experiment 1 we reproduced the syllabic interference effect with bimodal stimuli containing auditory as well as visual lip movement information, thus confirming the generalizability of the phenomenon. In subsequent experiments we were able to produce (Experiment 2) and to eliminate (Experiment 3) syllabic interference by introducing 'illusory' (McGurk) audiovisual stimuli in the irrelevant syllable, suggesting that audiovisual integration occurs prior to attentional selection in this paradigm.

Keywords: Audiovisual speech; Multisensory integration; McGurk illusion; Attention; Speeded discrimination

1. Introduction

The integration of audiovisual cues to speech not only helps language comprehension when the auditory signal is degraded (e.g. Sumby & Pollack, 1954), but it has an important role under good listening conditions, as long as lip movement information is available

* Corresponding author. Tel.: +34-93-3125158; fax: +34-93-4021363.
E-mail address: ssoto@psico.psi.ub.es (S. Soto-Faraco).

0022-2860/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2003.10.005

(e.g. McGurk & MacDonald, 1976; Reisberg, McLean, & Goldfield, 1987). This audiovisual integration process is well illustrated by the classic McGurk effect, whereby combining the sound of /ba/ with the sight of the lip movements of [ga] often produces the illusion of hearing "da". There seems to be agreement among researchers that the integration of auditory and visual speech cues is a fast, pre-attentive process that occurs without the observer's awareness (e.g. Colin et al., 2002; Rosemblum & Saldaña, 1996; Summerfield & McGrath, 1984). However, several theories of attention do actually challenge the idea of automatic cross-modal binding (Treisman & Gelade, 1980; Wickens, 1984; though see Massaro, 1998, for an attention-free model of audiovisual integration). For instance, the feature integration theory predicts that binding (including cross-modal features) will not occur without attention (Treisman & Gelade, 1980). Moreover, according to some researchers, the selection of information in one sensory modality can occur without interference from information in other modalities (Wickens, 1984; see Duncan, Marteen, & Ward, 1997, for a more recent example).

From an empirical perspective, direct demonstrations of automaticity in the audiovisual speech integration are neither numerous nor robust. The original support for the unavoidable, automatic, nature of audiovisual speech integration relied on the phenomenological experience of the McGurk illusion (McGurk & MacDonald, 1976). While this type of evidence is valuable, its utility is clearly limited by the problems associated with the introspective method. Another argument for the automaticity of audiovisual speech integration is based on the effects of explicit instructions during audiovisual tasks (e.g. Easton & Basala, 1982; Massaro, 1987). In Massaro's study (see also Massaro, 1998, for an extension), participants were shown bimodal stimuli that could contain congruent or incongruent information (i.e. auditory /ba/ and visual [da]). The pattern of identification responses was similar regardless of whether the instructions to the observers were to focus attention to one or the other modality, or to use both sources of information, thus supporting the automaticity hypothesis. However, manipulating explicit instructions in an intersensory conflict situation is often susceptible to confounds due to cognitive biases (see De Gelder & Bertelson, 2003, p. 462, for a recent discussion), and therefore the interpretation of these findings in terms of automaticity must remain inconclusive. In a recent study using event related potentials (ERP), Colin et al. (2002) obtained a difference in ERP signals, identified as the mismatch negativity (MMN), attributed to the effects of the auditory illusion produced by McGurk stimuli. As the MMN is usually attributed to early sensory processing, this result suggests pre-attentive integration. However, it is unclear whether the ERP correlates observed by Colin et al. were produced by the consequences of audiovisual integration on auditory processing or directly by the consequences of a discrepancy between the auditory and the visual stimuli.

In the present study we address the automaticity of audiovisual speech integration using an adaptation of the speeded classification paradigm (e.g. Garner, 1974; Maddox, 1992). This paradigm is based on the idea that if two dimensions of a stimulus can be attended to independently, then irrelevant variations along one of the stimulus dimensions will not affect response latencies in a discrimination task regarding the other dimension (a classical example is color and shape). If, on the contrary, selective attention cannot be focused to one dimension in isolation, irrelevant variations in one of the stimulus dimensions will slow down response latencies in a discrimination task to the other dimension. In this case,

the two dimensions are said to be processed integrally (an example would be color and hue). There have been some applications of this paradigm to the study of speech perception (e.g. Green & Kuhl, 1991; Pallier, 1994; Tomiak, Mullennix, & Sawusch, 1987). Here we used the syllabic interference task, where participants are asked to make a speeded classification of the first syllable of a disyllabic stimulus (a nonword presented auditorily), while attempting to ignore the second syllable (Pallier, 1994). In one condition (homogenous) people are presented with lists of stimuli where the second syllable always remains constant (i.e. a list containing /toku/ and /paku/). In the other condition (orthogonal) the second syllable of the stimulus varies unpredictably from trial to trial (i.e. a list containing /toku/, /paku/, /togi/ and /pagi/). Typically, reaction times to the orthogonal lists are slower than in the homogenous lists, indicating that the processing of the second syllable is obligatory even if participants attempt to focus their attention on the first syllable alone (Pallier, 1994).

The hypothesis under consideration here is that if audiovisual integration of speech is automatic, the McGurk illusion should occur before selective attention can be allocated. Therefore, by manipulating the identity of the second (irrelevant) syllable by virtue of McGurk fusions, we should be able to produce the syllabic interference effect in lists that are auditorily homogenous, or to cancel out such interference in lists that are auditorily orthogonal. It is important to note that in these experiments participants never made explicit judgments about the McGurk event, and therefore the consequences of audiovisual integration were measured indirectly. In Experiment 1 we tested if the syllabic interference effects can be obtained with audiovisual materials. In Experiments 2 and 3 we attempted to induce and to eliminate, respectively, the syllabic interference effect introducing McGurk combinations in the irrelevant syllable.

2. General method

2.1. Participants

One hundred and twenty undergraduate students from the University of Barcelona participated in this study (40 in each experiment) in exchange for course credit. All reported normal hearing and normal or corrected vision.

2.2. Apparatus and materials

The video-track was presented on a computer monitor (17 inch) and the audio-track was presented through two loudspeaker cones placed on either side of the monitor. We used digital video recordings of a female speaker pronouncing several tokens of the nonwords “tabi”, “tobi”, “tadi”, and “todi” (a close-up of the mouth area, from the upper part of the nose to 3 cm below the chin). In Experiment 1, all audiovisual clips contained matching visual and auditory information but, for consistency with the subsequent experiments, they were cross-dubbed (i.e. the audio-track of one exemplar of “tabi” was cross-dubbed with the video-track of a different exemplar of “tabi”), and carefully synchronized. We created ten different clips of each stimulus, all 1000 ms in duration.

Table 1
Summary of the stimulus lists used in Experiments 1–3

Experiment	Homogenous lists		Orthogonal lists	
Experiment 1	Matching "bi" list	/tabi/+[tabi]	Both lists	/tabi/+[tabi]
		/tobi/+[tobi]		/tobi/+[tobi]
	Matching "di" list	/tadi/+[tadi]		/tadi/+[tadi]
		/todi/+[todi]		/todi/+[todi]
Experiment 2	Matching "bi" list	/tabi/+[tabi]	Both lists	/tabi/+[tabi]
		/tobi/+[tobi]		/tobi/+[tobi]
	McGurk "di" list	/tabi/+[tagi]		/tadi/+[tagi]
		/tobi/+[togi]		/todi/+[togi]
Experiment 3	Matching "di" list	/tadi/+[tadi]	Both lists	/tadi/+[tadi]
		/todi/+[todi]		/todi/+[todi]
	McGurk "di" list	/tabi/+[tagi]		/tabi/+[tagi]
		/tobi/+[togi]		/tobi/+[togi]

The auditory component of each stimulus is denoted within slashes, and the visual (lip-read) part of each stimulus is denoted within square brackets. The auditory and the visual components were cross-dubbed in all experiments. When the visual and the auditory components of a stimulus matched, two different exemplars of the same stimulus were cross-dubbed. Each list contained 80 stimuli. Homogeneous lists included four repetitions of the 10 exemplars of each stimulus. There were two equivalent versions of each orthogonal list. Each contained four repetitions of a different set of 5 exemplars of each stimulus.

On average, the auditory stimulus started 175 ms (SD = 9) after the onset of the clip, and the onset of the second syllable occurred 181 ms (SD = 14) later.

We built four different lists containing 80 stimuli pseudo-randomly ordered (see Table 1). Two of the lists were homogeneous: one contained each exemplar of matching "tabi" and "tobi" repeated four times, while the other contained each exemplar of matching "tadi" and "todi" repeated four times. The other two lists were orthogonal: one contained four repetitions of five randomly selected exemplars of each stimulus ("tabi", "tobi", "tadi" and "todi"), while the other contained four repetitions of each of the remaining five exemplars of each stimulus. Each list contained an additional ten trials at the beginning that were used for warm-up and training purposes. These were not included in the analyses.

2.3. Procedure

Participants sat at a table, facing the video monitor, in a dimly-lit sound-attenuated room. Each trial started with a fixation cue (the " + " symbol) presented during 900 ms, followed by a 250 ms blank interval, and the presentation of an audiovisual clip. The task was to respond as fast and accurately as possible (pressing one of two buttons) to the identity of the first syllable of the stimulus (/ta/ or /to/). A new trial began 150 ms after the response or after the 2200 ms no-response deadline. Participants were tested on each of

the four lists, with list order counterbalanced across participants according to a Latin square.

3. Experiment 1

The goal of this experiment was to reproduce the syllabic interference effect using audiovisual stimuli. This test is necessary in order to ensure that the effect occurs with audiovisual materials and with the particular phonemic variation to be used here (/bi/ vs. /di/).

3.1. Results and discussion

We assessed the participant's mean reaction time for correct responses (Fig. 1A) and accuracy for each type of stimulus (matching "bi" or matching "di") and condition (homogenous vs. orthogonal). Reaction times 2 SD above or below the participant's mean RT for a given condition were eliminated (3%). An analysis of variance (ANOVA) was performed including the within-participants factors stimulus type and condition. Response times did not differ as a function of stimulus type [657 ms for matching "di", and 653 ms for matching "bi", $F(1,39) = 1.9$, $P = 0.173$]. The effect of condition was significant because responses to orthogonal lists were 18 ms slower than responses to homogeneous lists [$F(1,39) = 4.5$, $P < 0.05$]. Finally, the interaction between stimulus type and condition was not significant [$F < 1$], indicating that syllabic interference was robust across both types of stimuli. Accuracy was high (99%) and the error analysis did not reveal any significant effects (all $F < 1$).

This result confirms the syllabic interference effect (Pallier, 1994) and extends it to the audiovisual case. This demonstration is important because previous studies using the syllabic interference effect have typically employed larger phonological variations in the irrelevant syllable (/ku/ vs. /gi/) and a presentation restricted to the auditory modality. Here, participants had the support of additional matching visual information and the irrelevant syllable contained smaller phonological variation (/di/ vs. /bi/), and yet syllabic interference was observed.

4. Experiment 2

The goal of this manipulation was to test if the syllabic interference effect can be induced on the basis of the McGurk illusion. The four types of items used in this experiment were: two audiovisual matching stimuli (/tabi/+[tabi] and /tobi/+[tobi]) and two McGurk stimuli (/tabi/+[tagi] and /tobi/+[togi]).[1] Based on previous literature, we expected that, with the McGurk stimuli, people would perceive "tadi" and "todi" (respectively). Note that the McGurk combination occurred in the irrelevant stimulus

[1] The clips corresponding to "tagi" and "togi" were recorded and cross-dubbed at the same time as the ones described for Experiment 1, and under the same conditions.

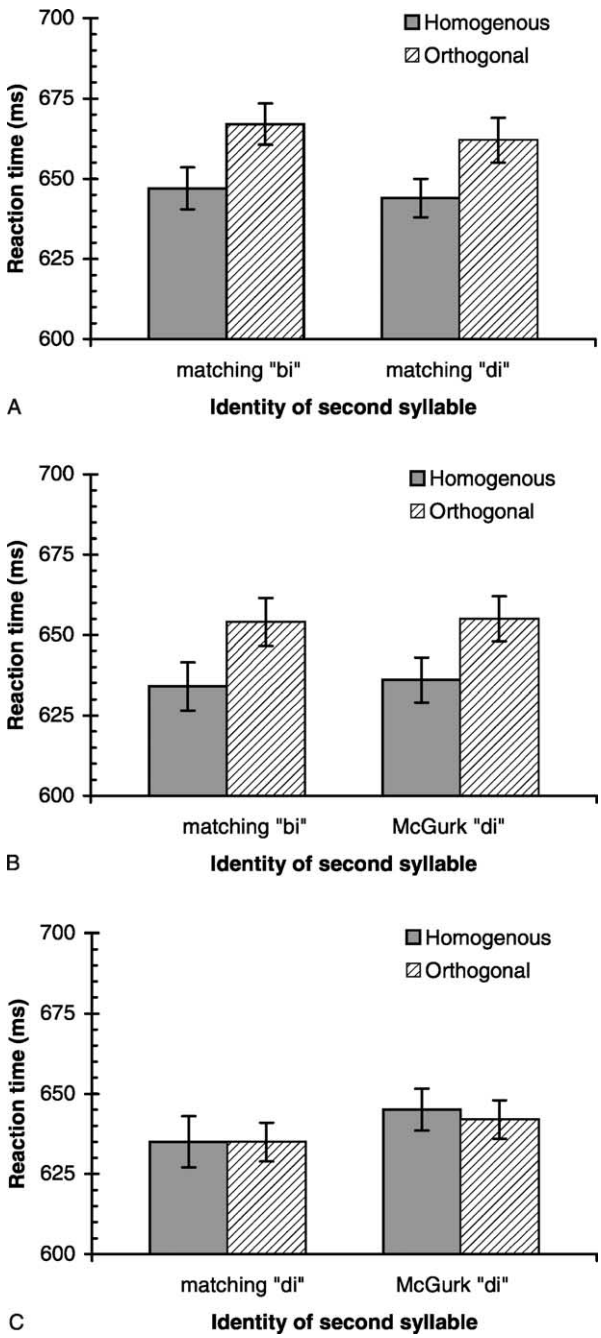

Fig. 1. Average reaction times in Experiments 1–3 (graphs A–C, respectively). Results are shown as a function of the identity of the irrelevant syllable (*x*-axis) and condition (filled bars represent the homogeneous condition and hatched bars represent the orthogonal condition). Error bars represent the standard error of the mean.

dimension (the second syllable) and it was completely unrelated to the task ("ta"–"to" classification). One of the homogenous lists contained the matching stimuli and the other homogenous list contained McGurk stimuli alone. In the "illusory" orthogonal lists, both matching and McGurk stimuli were mixed (see Table 1).

If audiovisual integration occurs in an automatic fashion, then participants will presumably experience variation[2] in the orthogonal lists (containing the matching "bi" stimuli and the McGurk "di" stimuli) and, therefore, reaction times should be slower than in the homogenous lists (that contained either the matching stimuli or the McGurk stimuli). If, on the contrary, people are able to focus their attention to the auditory component alone and ignore the visual influence in the irrelevant syllable, then RTs should be equivalent in both conditions.

4.1. Results and discussion

The data were submitted to the same type of analysis as in Experiment 1. The effect of stimulus type was not significant [640 ms for matching "bi", and 641 ms for McGurk "di", $F < 1$]. The effect of condition was significant [$F(1, 39) = 8.2$, $P < 0.01$], with orthogonal lists leading to slower responses than homogeneous lists (651 vs. 630 ms). The interaction was far from significance [$F < 1$], indicating that the syllabic interference was robust across the two types of stimuli. Overall, accuracy was high (98%), and the error analysis did not reveal any significant effect (all $F < 1$).

The present data clearly show that participants could not ignore the auditory variation produced by the McGurk fusion in the irrelevant syllable while making speeded judgments on the first syllable of the stimulus. Because the actual auditory component of the stimuli was in fact identical in the orthogonal and the homogenous lists, this result supports the hypothesis that the McGurk illusion occurred automatically, before attentional selection of the first syllable was possible. However, an alternative account for the data must be considered. It is possible that even if the McGurk illusion did not occur before attentional selection, the variation in the visual information alone induced the slow-down in reaction times in the orthogonal condition. We address this alternative explanation in Experiment 3.

5. Experiment 3

The goal of Experiment 3 was to test if syllabic interference can be eliminated by virtue of the McGurk illusion. We created orthogonal lists where both the auditory and the visual components of the stimuli contained irrelevant variation, but in which audiovisual integration should result in "illusory" homogeneous lists. Here, both the audiovisual matching stimuli (/tadi/+[tadi] and /todi/+[todi]) as well as the McGurk stimuli

[2] It is important to point out that this phenomenological description of the anticipated participants' experience is used here to clarify the rationale of our hypothesis. However, it is neither part of our measure of choice (the objective reaction times in the discrimination task) nor an integral part of our prediction (slower RTs in the orthogonal lists than in the homogenous lists).

(/tabi/+[tagi] and /tobi/+[togi]) were expected to result in the perception of "tadi" and "todi", respectively. One of the homogenous lists contained matching "di" stimuli while the other homogenous list contained McGurk "di" stimuli. In the orthogonal lists, both matching "di" and McGurk "di" stimuli were mixed (see Table 1).

If audiovisual integration occurs in an automatic fashion, then we do not expect to observe a syllabic interference effect in this experiment. That is, the orthogonal condition should behave as the homogenous condition. However, if audiovisual integration does not occur automatically, then we should observe syllabic interference.

5.1. Results and discussion

We conducted an ANOVA on correct RTs (see Fig. 1C) as in the previous experiments, including the factors stimulus type (matching "di" or McGurk "di") and condition. The effect of stimulus type was marginally significant [$F(1,39) = 3.5$, $P = 0.067$] because stimuli containing a McGurk "di" syllable were responded to 9 ms slower than stimuli containing a matching audiovisual "di". Critically, however, there was no effect of condition (2 ms difference) nor any interaction between condition and type of stimulus [both $F < 1$]. The effect of condition was not significant for any of the two types of stimulus individually ($t < 1$ for McGurk "di", and $t(39) = 1.4$, $P = 0.156$ for matching "di"). Accuracy was, again, high (98% overall), and none of the two main terms in the analysis [both $F < 1$] nor the interaction [$F(1,39) = 2.7$, $P = 0.104$] reached significance. Further analyses pooling the RT data of Experiments 2 and 3 revealed that there was a significant interaction between experiment and condition [$F(1,78) = 4.8$, $P < 0.05$], while no other terms in the analysis resulted in a significant difference. Specifically, overall RTs were not different across experiments ($F < 1$).

The results of Experiment 3 provide a critical test for the hypothesis that audiovisual integration occurs automatically. Here, the results observed can not be predicted on the basis of neither visual nor auditory information alone, but only as a result of the integrated percept. As in Experiment 2, the McGurk event occurred on the task-irrelevant second syllable of the utterance that people had to classify, and yet it proved effective enough to cancel out the interference predicted on the basis of auditory information (that varied between /di/ and /bi/) and on the basis of visual information (that varied between [di] and [gi]).

6. General discussion

The present data strongly converge on the idea that audiovisual integration of speech information proceeds in an automatic manner. The present demonstration has the advantage that selective attention was measured in an indirect way (participants were never asked about the integrated percept), and the response set ("ta" vs. "to") was orthogonal to the factor of interest (the McGurk illusion). Potential biases introduced by the instructions or by the presence of mismatching audiovisual events are therefore unlikely to account for the present data. It is interesting to note that, while in Experiment 2

participants were unable to focus on the auditory component of the stimuli in order to avoid the interference produced by the McGurk syllables, in Experiment 3 the McGurk illusion released participants from the interference predicted on the basis of any input modality when considered in isolation.

Speeded classification tasks have been used with bimodal stimuli before, but usually their application has been mostly restricted to non-linguistic materials using dimensions that are arbitrarily paired (such as pitch and color; see Marks, in press, for a review). There is one study where the speeded discrimination task has been used with bimodal speech stimuli (Green & Kuhl, 1991). Green and Kuhl (Experiment 2) found that place of articulation and voicing are integrally processed even if specified in different sensory modalities. However, in contrast with the present study, Green and Kuhl's participants made phonetic judgments directly on the McGurk stimulus, and their orthogonal lists always contained variation in the visual component. Therefore, their results are inconclusive with respect to the automaticity issue.

The present study provides a strong demonstration of automaticity during audiovisual integration of the auditory and visual components of a speech stimulus. This result not only coincides with the assumption that audiovisual integration is an obligatory process, but is also in keeping with recent neuroimaging findings where audiovisual integration results in enhanced activation of brain areas related to the initial stages of speech perception (Calvert et al., 1997). Our results challenge the claim that cross-modal binding necessitates attention (Treisman & Gelade, 1980), as well as the prediction that selective attention will proceed independently for each sensory modality (e.g. Wickens, 1984), and conforms to models predicting attention-free audiovisual integration (i.e. Massaro, 1998). The current demonstration of automaticity in audiovisual speech integration is also in agreement with theories of speech perception that argue for abstract motor-based representations (e.g. Liberman & Mattingly, 1985). According to this theory, auditory and visual speech cues would map directly onto a common representational code and therefore integration would occur at early processing stages as a natural consequence of this mapping.

Finally, the present findings open up an interesting question for future research. As several recent studies show, certain classical manifestations of automatic processes can break down under heavy attentional demands. These include pop-out or parallel search, word reading, and even visual motion after-effects (Joseph, Chun, & Nakayama, 1997; Rees, Frith, & Lavie, 1997; Rees, Russell, Frith, & Driver, 1999; see also Sinnett, Soto-Faraco, & Costa, 2003). It will be interesting to address whether this is also the case for audiovisual integration of speech (see Soto-Faraco, Navarra, Alsius, & Campbell, 2003, for an initial attempt).

Acknowledgements

This work has been supported by grants from the DGCYT BSO2002-01545 (Ministerio de Ciencia y Tecnología) and the James S. McDonnell Foundation JSMF-20002079.

References

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*, 593–596.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*, 495–506.

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*, 460–467.

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*, 808–810.

Easton, R. D., & Basala, M. (1982). Perceptual dominance during lipreading. *Perception and Psychophysics*, *32*, 562–570.

Garner, W. R. (1974). The processing of information and structure. Hillsdale, NJ: Erlbaum.

Green, K. P., & Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 278–288.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature*, *387*, 805–807.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). *Scientific psychology series*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Marks, L. E (in press). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Massaro, D. W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry* (pp. 66–74). Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1998). Perceiving talking faces: from speech perception to a behavioral principle. Cambridge, MA: MIT Press.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *265*, 746–748.

Pallier, C (1994). *Rôle de la syllabe dans la perception de la parole: études attentionelles*. Doctoral dissertation presented at the École des Hautes Études en Sciences Sociales, Paris (available online at: http://www.pallier.org/papers/thesis/thsplra4.pdf)

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, *278*, 1616–1619.

Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattentional blindness versus inattentional amnesia for fixated but ignored words. *Science*, *286*, 2504–2507.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In B. Dodd, & R. Campbell (Eds.), *Hearing by eye: the psychology of lip-reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosemblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 318–331.

Sinnett, S. C., Soto-Faraco, S., & Costa, A (2003). *Inattentional blindness for words*. Poster presented at the XIIIth congress of the European Society for Cognitive Psychology, Granada, Spain.

Soto-Faraco, S., Alsius, A., Navarra, J., & Campbell, R (2003). *Assessing automaticity in audiovisual integration of speech*. Poster presented at the 4th International Multisensory Research Forum, Toronto, Canada.

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

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audiovisual incompatibility perception in vowels. *Quarterly Journal of Experimental Psychology*, *36A*, 51–74.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, *81*, 755–764.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). San Diego, CA: Academic Press.