ELSEVIER

# Impairment of speech-reading in prosopagnosia

## Beatrice de Gelder *, Jean Vroomen

*Department of Psychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

## Abstract

The face is a source of information processed by a complex system of partly independent subsystems. The extent of the independence of processing personal identity, facial expression and facial speech remains at present unclear. We investigated the speech-reading ability of a prosopagnosic patient, LH, who is severely impaired on recognition of personal identity and recognition of facial expressions. Previous reports of such cases raised the possibility that speech-reading might still be intact, even if almost all other aspects of face processing are lost. A series of speech-reading tasks were administered to LH including still photographs, video clips, short-term memory tasks for auditory and speech-read materials, and tasks aimed at assessing the impact of the visual input on auditory speech recognition. LH was severely impaired on these tasks. We conclude that in LH there is a strong association between severe face processing deficits and loss of speech-reading skills.   © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* Prosopagnosia; Speech-reading; Face processing

## 1. Introduction

The human face is a very rich source of information. Personal identity, age, gender, emotion, as well as speech can all be perceived from the face. The face is not the exclusive bearer of all these types of information. The voice, for example, can be equally informative about the gender, identity, or emotion of a speaker. Moreover, the information conveyed by the face is combined with that contributed by other sources. Speech is a particularly striking example of such multimodal information processing as it is conveyed by the voice as well as the face. The evidence about the combination of the two sources in a single percept is overwhelming. Seeing the face and watching the movements of the mouth are helpful for understanding speech, even in perfectly healthy individuals (e.g., Summerfield, 1991). The ability to speech-read is thus part of face processing skills, but its study belongs equally to the domain of speech processing as to that of inter-sensory integration. If so, an impairment in speech-reading ability may result either from a face processing deficit, a speech processing deficit, or a problem with inter-sensory integration.

The present report concerns the speech-reading skills of LH, a well-known prosopagnosic patient whose various face processing abilities have been documented by several researchers over the last two decades (e.g., Etcoff et al., 1991; Farah et al., 1995a; Levine and Calvanio, 1989). The main goal of our study is to investigate the extent to which LH's prosopagnosia has left intact his speech-reading ability.

* Corresponding author. Tel.: +31 13 466 2167; fax: +31 13 466 2370; e-mail: b.degelder@kub.nl.

## 1.1. Autonomy of different face processing abilities

Models of normal face processing such as the widely quoted model of Bruce and Young (1986) picture different kinds of facial information as so many separate processing routes, all taking off from the stage after which a face is recognized as such, sometimes called the structural face processing stage. The issue of the autonomy of these routes is not one that is well investigated. Recent evidence suggests that this autonomy might not be as radical as previously assumed. For example, a behavioral study by Walker et al. (1995) found that subjects who are familiar with a face are less susceptible to McGurk effects than subjects who are unfamiliar with it.

A strong impetus to the notion of autonomous subsystems for different face processing abilities came from the study of brain-damaged and other neurologically impaired patients, most importantly patients impaired in face processing (prosopagnosics). Such reports have raised the question whether all kinds of information carried by the face would be impaired in these cases (see (Damasio et al., 1990) for an overview). The currently available evidence points either way. Cases of dissociation between the various subcomponents of face recognition have been observed, most notably between personal identity and facial expression recognition. Other cases of prosopagnosia suggest, rather, an association of various face deficits and show that brain-damage affecting one component does not leave intact other face processing abilities.

## 1.2. Dissociation between face recognition and speech-reading

The issue of spared lipreading in prosopagnosics is particularly intriguing. Intuitively, it seems relatively straightforward to lump together various aspects of face processing that concern the major semantic components of information provided by the face such as personal identity, age or gender and contrast them all with speech-reading ability.

Over the last decade, the fate of speech-reading when a prosopagnosic disorder occurs, has been the topic of strong predictions. These were based on state of the art knowledge about the lateralization of face processes and of language processing skills. Given the dominance of the right hemisphere for the former and the left hemisphere for the latter, patients with impaired face processing skills were expected to have intact lipreading skills. This was indeed observed by Campbell et al. (1986) in a report about a double dissociation between lipreading and personal identity recognition in two brain-damaged patients.

The first report of just such a dissociation was offered by Campbell et al. (1986). Patient Mrs. D was highly agnosic with profound prosopagnosia, yet could sort pictures of faces according to speech sound and was sensitive to the effects of seeing the speaker in reporting heard speech (McGurk effects). She could speech-read silent spoken numbers as well as discriminate lipspoken vowels and consonants. By contrast, patient Mrs. T was unable to perform such tasks, although she had no difficulty recognizing faces or facial expressions or other visual objects, even though she was alexic. Mrs. T's lesion was unilateral and affected the left hemisphere, Mrs. D's only affected the right. However, more recently, a study of HJA (Campbell, 1992), who is a patient with prosopagnosia and visual agnosia with bilateral lesions of occipito-temporal areas, showed that he could not classify photographs of speaking faces. He was however completely normal with dynamic speech-reading stimuli. In bimodal speech tasks (in which visual and auditory input are provided simultaneously) he had normal audio-visual integration. The critical dissociation in this case thus seems not to be between speech versus non-speech aspects of face processing, but between recognizing information provided by still versus dynamic displays.

The importance of visual movement pathways for speech-reading is illustrated by patient LM (Campbell, 1996a). LM's lesion affected only the cortical visual movement areas, including area V5, and sparing areas V1-V4 which are all damaged in HJA. LM could only classify still photographs and did not show McGurk effects. This dissociation between static and dynamic inputs to speech-reading would imply that at least in some basic

sense, perception of static forms and that of movement patterns can each independently access speech representations.

An extensive study of speech-reading ability in a visual agnosic patient was presented by de Gelder et al. (1998b). Their patient AD was impaired in all aspects of face processing, except structural face recognition. Auditory language processing was normal, but speech-reading from still faces was entirely lost. There was no indication of a movement perception disorder in this patient. With dynamic displays of talking faces she yielded a slightly better speech-reading score than with still faces. Surprisingly, the movement of the lips was perceived in a systematic fashion, but the link with the correct phoneme was not made. Since some dynamic speech-reading ability was preserved, we asked whether there would at least be some evidence of audio-visual bias (an impact of the visual modality on the to be evaluated auditory input) and audio-visual integration. Instead, we found no impact of her partly preserved dynamic speech-reading ability in bimodal situations. This suggests that BC has a route to speech-reading, but its output does not merge with auditory speech, either because it is too weak or too coarse, or because her partly preserved speech-reading is qualitatively different from that of normals. The finding of a spared ability for speech-reading from motion observed in HJA clearly does not generalize to AD, although there are many similarities between the two patients in the site of the lesions as well as in object and face recognition impairments.

## 1.3. Case presentation

LH is a 44 yr-old man who is prosopagnosic as a consequence of an automobile accident 24 years earlier. Brain-damage from the accident and subsequent surgery consists of bilateral occipitotemporal lesions as well as right frontal and anterior temporal lesions. Additional neuropsychological information is found in (Levine et al., 1980). His verbal IQ was 132 and his performance IQ was 92 as shown in tests administered during his recovery. He is profoundly prosopagnosic and unable to recognize even his wife or children. His face pro-

cessing impairments have been reported in detail, most recently by Etcoff et al. (1991), Farah et al. (1995b), de Gelder and Etcoff (1997) and de Gelder (submitted). LH has intact abilities in the domain of language. We have no way of appreciating LH's speech-reading ability before his accident. When questioned about this informally, he seemed convinced that before as well as after his accident he could deal perfectly normally with information provided by the movements of the speaker's lips which he claimed to be aware of and paying attention to. Neurologists and neuropsychologists have never reported any problems in the area of movement perception, although this has not been tested formally. However, a comparison of the neuroanatomical loci of the lesions of LH and of patient LM suffering from motion perception disorder (Zihl et al., 1983) makes it seem unlikely that there would be a deficit in motion perception with a neuroanatomical basis in LH.

## 2. Method and results

We tested LH with a task of speech-reading from still faces, with a single digit speech-reading task with still and moving faces, with an audio-visual memory task as well as with a set of bimodal tasks.

## 2.1. Recognition of speechsounds from still photographs

LH was presented with a series of 16 black and white photographs. They represented four different actors each with four different mouth positions (saying /a/, /i/, /o/ or making a grimace). He was given the pictures one by one and told about the four response choices which were written down on cards in front of him. He was asked to put each photograph down next to what he deemed to be the correct response. LH was confident about his answers, but the results showed that he performed at chance level, $Z = 0.0$, NS, since only 4 out of the 16 trials were correctly recognized (2 grimaces, and 2 /i/'s). Most noteworthy is that LH did not distinguish between a mouth position that corre-

sponded to a speech sound and a mere grimace. Since LH is unable to recognize reliably any facial expression from still photographs (except happiness), his poor result with still faces is not surprising. On the other hand, if we expect speech-reading ability to stand apart from other face processes, he might have some preserved speech-reading. Of course, recognizing speech sounds from still photographs is an unnatural task and it has often been argued that it does not offer good evidence about speech-reading abilities (but see (Campbell et al., 1996)).

The total inability to process facial expressions, including those corresponding to speech sounds, has been observed in other cases of prosopagnosia. Will LH, like HJA and BC, perform better when he has information provided by short video clips?

## 2.2. Recognizing spoken digits

LH was shown short video clips of a female speaker articulating one by one the digits 1, 2, 4, 5, 6, 7, 8 or 9 in random order with each digit presented twice. Tests with normal controls had shown that these digits were clearly speech-readable. LH's performance was surprisingly poor. He recognized only 2 out of 16 digits, which is again at chance level, $Z = 0.0$, NS. Thus, with moving stimuli, he did not improve when compared to the result obtained with the still photographs.

## 2.3. Serial recall of audio-visual, auditory and speech-read digits

Given the previous result, it was of interest to us whether there would be any difference between the performance of LH on audio-only trials as contrasted with audio-visual presentation. If performance would be poorer in the latter than in the former case, it would suggest an interference from intact face processing, at least from intact structural recognition of a face. Such an effect was reported previously by de Gelder et al. (1998a).

A videotape was constructed of the same female speaker used in the single digit task, but this time she pronounced digit lists. The video was then edited so that the digit lists were presented in an audio-visual mode, an audio-only mode in which the sound was heard, but with a blank screen, and a speech-reading mode in which the speaker was visible, but without sound. Each list was composed by drawing without replacement from the numbers 1, 2, 4, 5, 6, 7, 8 and 9 in pseudo-random order. A list consisted of eight digits. Each presentation mode consisted of one block of eight digit lists. The audio-visual block was presented first, followed by the audio-only block, and then the speech-read block. Instructions specified to attend to the auditory, the audio-visual and the visual input respectively, and to report the digits in the order as presented.

Digits were scored as correct if reported in the correct order. LH's performance in the audio-visual and audio-only mode was almost flawless: 97% (62 out of 64) and 100% (64 out of 64) correct, respectively, $\chi^2_{(1)} = 2.03$, NS. However, when the digit-lists had to be speech-read, performance dropped to only 3% (2 out of 64) correct, $\chi^2_{(1)} = 108.8$, $p < 0.001$. For comparison, a control subject of similar age and sex obtained a score of 94% (60 out of 64) correct with audio-visual presentation, 100% (64 out of 64) correct with audio-only presentation, and 96% (61 out of 64) correct with the speech-read lists. So LH only had problems when the lists had to be speech-read, but his score was equally high for auditory-only and audio-visual presentation. Thus, the presence of the speech-read information embedded in a visual stimulus he cannot deal with (i.e., the face) does not hinder his performance, although a direct comparison between the auditory and the audio-visual performance is complicated, because of ceiling effects.

Our second set of tasks was intended to assess speech-reading ability in bimodal situations under conditions where either the visual input was not explicitly attended to or had to be reported as was the case in some of the previous tasks. We considered two ways in which some residual speech-reading ability might manifest itself. One was through an impact on the auditory perception of natural speech sounds as in the McGurk effect, another was through a systematic shift of an ambiguous synthesized /ba-da/ continuum.

### 2.4. Auditory processing, speech-reading and audio-visual conflict

We used a video recording of a female speaker pronouncing a series of VCV sequences (de Gelder et al., 1991, 1998b). Each sequence consisted of one of the four plosive stops /p, b, t, d/ or a nasal /m, n/ in between the vowel /a/ (e.g., /aba/ or /ana/). There were three presentation conditions: an audio-visual, an audio-only and a visual-only presentation. In the audio-visual presentation, dubbing operations were performed on the recordings so as to produce a new video-film comprising six different audio-visual combinations: auditory /p, b, t, d, m, n/ were combined with visual /t, d, p, b, n, m/, respectively. The visual place of the articulation feature thus never matched the auditory place feature. Appropriate dubbing ensured that there was audio-visual coincidence of the release of the consonant in each utterance. In addition, unimodal presentation conditions were produced. For the audio-only condition, the original auditory signal was dubbed on a blank screen. For the visual-only condition, the auditory channel was deleted from the recording, so the subject had to rely entirely on speech-reading. Each presentation condition comprised three replications of the six possible stimuli. LH was instructed to watch the speaker and repeat what she said.

In the audio-visual conflict condition, there were only two fusions (i.e. a combination of a heard and a seen input leading to a new percept) out of 18 trials (11%) while normal performance is about 51%, $\chi^2_{(1)} = 6.41$, $p < 0.05$ (see (de Gelder et al., 1991)). In all other trials he reported the audio-part of the audio-visual stimulus. In the audio-only trials he always reported the correct phoneme. For the visual-only trials, two response categories were made, based on two broad viseme classes: lingual (d, t or n) or bilabial (b, p, m). Performance was in this case only 22% correct (4 out of 18 trials), whereas in normal subjects, it is about 84%, $\chi^2_{(1)} = 13.79$, $p < 0.001$ (see (de Gelder et al., 1991)). On the nine bilabial trials, he reported two times a bilabial phoneme, the other times a lingual phoneme. On the nine lingual trials, he reported two times a lingual phoneme, and seven times a bilabial phoneme.

LH is thus normal at processing the auditory input presented in the absence of a face. But he performs poorly when having to report what is said by a face in the absence of any auditory input. Therefore it is not surprising that he does not show fusions or blends and only tends to report the audio part of a bimodal stimulus when there is a conflict between the information in audition and vision. Such a result looks straightforward and implies that his prosopagnosia strongly affects his speech-reading ability and that LH cannot process visual speech.

### 2.5. Recognition of speech from a synthetic face and voice

The next task focuses on audio-visual bias with synthetic ambiguous stimuli that may offer a more fine-grained appraisal of LH's speech processing abilities. The task is a variant of the well-known categorical perception paradigm, and requires the use of synthetic speech as well as a synthetic face. As in the previous task, the materials consisted of bimodal as well as unimodal trials. An important difference from the previous task is that this time the unimodal auditory trials always consisted of a speech stimulus combined with a static face. This allows us to appreciate whether the auditory speech channel is autonomous and still robust in the presence of a face.

The task consisted of a tape showing an artificially created synthesized face (Massaro and Cohen, 1990). The synthetic face is controlled by 11 display parameters which determine jaw rotation, lip protrusion, upper lip raise, etc. By varying these parameters, a dynamic face is created that articulates 'ba', 'da' or any intermediate position between these two syllables. In the test, five levels of audio speech varying between 'ba' and 'da' were crossed with five levels of visible speech varying between 'ba' and 'da'. These 25 stimuli comprise the audio-visual condition. The auditory stimuli were also presented with a still face, and the visual stimuli were also presented alone, so that there was a total of $25 + 5 + 5 = 35$ independent stimulus events. The whole test consisted of 6 blocks of these 35 trials in randomized order for a total of 210 trials.

The performance of LH was compared with that of four control subjects of similar age to LH. All participants were instructed to listen and to watch the video and to identify each token as 'ba', 'da', 'bda', 'dba', 'va', 'tha', 'ga' or 'other'. There were thus 8 response possibilities × 35 trial types = 280 categories. In order to decrease this number, we scored the number of 'ba' and 'bda' responses as one category, and 'da' and 'tha' responses as another category, because these categories are visually very similar and they accounted for 80% of LH's judgements. We then computed four different performance measures: the visual and auditory influence in the bimodal condition, and the percentage correct in visual-only and audio-only trials. For the visual influence in bimodal trials, a visual 'ba' (i.e., the first two levels of the visual 'ba-da' continuum) should – compared with visual 'da' – increase the number of 'ba' and 'bda' responses, and a visual 'da' (i.e., the final two levels of the visual 'ba-da' continuum) should – compared with visual 'ba' – increase the number of 'da' or 'tha' responses. The bigger these differences, the bigger the visual influence in audio-visual trials. The same logic was applied to the auditory influence in bimodal trials. An auditory 'ba' (i.e. the first two stimuli of the auditory 'ba-da' continuum) should, compared to auditory 'da' (i.e. the final two stimuli of the auditory 'ba-da' continuum), increase the number of 'ba' and 'bda' responses, and an auditory da should, compared to auditory 'ba' increase the number of 'da' responses. The difference should give an indication of the auditory influence in audio-visual trials. For the visual-only trials, we computed the number of correct identifications. That is, the number of 'ba' or 'bda' responses when the first two stimuli of the visual 'ba-da' continuum were presented, and the number of 'da' or 'tha' responses when the final two stimuli of the visual 'ba-da' continuum were presented. For the audio-only trials, we computed the number of 'ba' and 'bda' responses when auditory 'ba' was presented, and the number of 'da' and 'tha' responses when auditory 'da' was presented.

LH had a negative visual influence in bimodal trials, i.e. −14%, compared to 26% (range 8–43%) for control subjects, $Z = 2.28$, $p < 0.01$. Thus, the combination of an auditory stimulus with a visual 'ba' increased the number of 'da' responses, and the combination of an auditory stimulus with a visual 'da' increased the number of 'ba' responses. His auditory influence in bimodal trials was normal: 37% for LH versus 26% (range 8–43%) for control subjects, $Z = 0.21$, NS. This is also consistent with the fact that there is no auditory processing difficulty. On the other hand, on audio-only trials he performed poorly: 17% correct for LH versus 64% (range 50–79%) for control subjects, $Z = 3.35$, $p < 0.01$. On 53% of the audio-only trials, he responded with phonemes like 'fa', 'ta' or 'ma'. There is no obvious explanation for this, except that it may be due to the fact that synthetic speech was used.

The most surprising result was that in the visual-only trials LH performed within the range of control subjects: 58% correct versus 55% (range 46–67%) for the controls, $Z = 0.33$, NS. His speech-reading performance with the artificial visual stimuli was thus superior to that observed in our previous tasks. We observed a similar improvement with BC (de Gelder et al., 1998b) who could not speech-read, but who nevertheless could reliably discriminate between a bilabial/non-bilabial stimulus. However, on some occasions she would classify a bilabial stimulus as /ba/ (a correct response), but a week later she labeled the same stimulus consistently as /da/. This raises the possibility that LH is able to discriminate bilabial from non-bilabial face movements, but as such, it is not clear whether it also leads to a speech percept that is comparable to that of normals.

## 3. Discussion

Evidence from normal studies has established very convincingly that speech-reading is a skill grounded in our linguistic abilities to the same extent and for the same reasons that understanding of auditory speech is. Starting from that evidence, and combining it with findings about the selective lateralization of speech versus face processing provided the theoretical motivation for previous reports about a dissociation between the two. In LH, however, speech-reading is almost entirely wiped out as a consequence of his

prosopagnosia. The close link between auditory and visual input for speech in normals does not safeguard speech-reading ability in the case of a generalized face recognition disorder. The results of LH raise several issues that need to be pursued and they point to several aspects of the speech-reading impairment in prosopagnosics that require attention.

A striking finding given that LH has no problem perceiving movement is the absence of an improvement with dynamic stimuli. We must therefore conclude that the perception of movement per se is not sufficient to sustain lipreading. The latter does seem to require skills related to face perception above and beyond a generic ability to process movement and an intact auditory speech ability. In this respect, our study of LH replicated the findings of patient QL (see discussion in (Campbell, 1996b)) who was likewise unable to speech-read from dynamic stimuli in the absence of a disorder of motion perception.

Only the data from the task where a synthetic face is used seems to represent an exception to this picture of a total loss of speech-reading. The synthetic face is a stimulus where all detail is left out. No research is available that contrasts synthetic versus naturally presented visual speech that would allow us to appreciate whether the former is easier or more difficult than the latter. One might argue that speech from a synthetic talking face presents a simplified stimulus which is therefore easier to perceive. So at present, we cannot entirely exclude that the synthetic face display improves perception of the lipmovement which then might explain the good performance for speech-reading in this task. But an alternative explanation is probably called for. There is reason to wonder whether in the task where a synthetic face is used, LH actually speech-reads rather than systematically reacting to the perceived movement, as is easy to do given the pair of syllables used. If we take the good performance in the synthetic face task to reflect systematic movement perception, but not speech-reading, then the absence of any visual bias in bimodal trials can be explained. Moreover, it underscores the dissociation between the ability to perceive movement and that of being able to speech-read.

Finally, we note the intriguing fact that in one task auditory recognition was severely handicapped by the presence of a still face. The same phenomenon was observed in AD. Further research is needed to decide whether this might be a consequence of an expectancy bias, or more interestingly, an interference originating in preserved face perception abilities. Both LH (de Gelder and Etcoff, 1997) and AD (de Gelder et al., 1998a) have impaired face recognition but show evidence for spared structural encoding of faces and continue to perceive faces as faces. The interference from preserved structural encoding of faces observed previously in our face matching and recognition tasks might thus extend to the processing of visual speech.

## Acknowledgements

## References

Bruce, V., Young, A., 1986. Understanding face recognition. Brit. J. Psych. 77, 305–327.

Campbell, R., 1992. The neuropsychology of lipreading. Phil. Trans. Roy. Soc. London, B 335, 39–45.

Campbell, R., 1996a. Seeing brains reading speech: A review and speculations. In: Stork, D.G., Hennecke, M.E. (Eds.), Speechreading by Humans and Machines: Models, Systems, and Applications. Springer, Berlin, pp. 115–134.

Campbell, R., 1996b. Seeing speech in space and time. In: Proc. 4th Internat. Conf. Spoken Language Processing. Philadelphia, PA, October 1996.

Campbell, R., Landis, T., Regard, M., 1986. Face recognition and lipreading: A neurological dissociation. Brain 109, 509–521.

Campbell, R., de Gelder, B., de Haan, E., 1996. The lateralization of lipreading: A second look. Neuropsychologia 34, 1235–1240.

Damasio, A.R., Tranel, D., Damasio, H., 1990. Face agnosia and the neural substrates of memory. Ann. Rev. Neurosci. 13, 89–109.

Etcoff, N.L., Freeman, R., Cave, K.R., 1991. Can we lose memories of faces? Content specificity and awareness in a prosopagnosic. J. Cogn. Neurosci. 3, 25–41.

Farah, M., Levinson, K., Klein, K., 1995a. Face perception and within-category discrimination in prosopagnosia. Neuropsychologia 33, 661–674.

Farah, M.J., Wilson, K.D., Drain, M.H., Tanaka, J.R., 1995b. The inverted face inversion effect in prosopagnosia: evidence for mandatory, face-specific perceptual mechanisms. Vision Res. 35, 2089–2093.

de Gelder, B., submitted. Inversion superiority in prosopagnosia extends to objects.

de Gelder, B., Etcoff, N., 1997. Component-based strategies do not compensate for loss of face and object recognition. In: Meeting of the Psychonomic Society, Philadelphia, PA, 22–23 November 1997.

de Gelder, B., Bachoud-Lévi, A.-C., Degos, J.D., 1998a. Inversion superiority in visual agnosia may be common to a variety of orientation polarised objects besides faces. Vision Res.

de Gelder, B., Vroomen, J., Bachoud-Levi, A.-C., 1998b. Impaired speechreading and audio-visual speech integration in prosopagnosia. In: Campbell, R., Dodd, B., Burnham, D. (Eds.), Hearing by Eye II, Advances in the Psychology of Speechreading and Auditory-Visual Speech. Psychology Press, Hove, pp. 195–207.

de Gelder, B., Vroomen, J., van der Heide, L., 1991. Face recognition and lipreading in autism. Eur. J. Cogn. Psych. 3, 69–86.

Levine, D.N., Calvanio, R., 1989. Prosopagnosia: A defect in visual configural processing. Brain and Cogn. 10, 149–170.

Levine, D.N., Calvanio, R., Wolf, E., 1980. Disorders of visual behavior following bilateral posterior cerebral lesions. Psych. Research 41, 217–234.

Massaro, D.W., Cohen, M.M., 1990. Perception of synthesized audible and visible speech. Psych. Sci. 1, 1–9.

Summerfield, Q., 1991. Visual perception of phonetic gestures. In: Mattingly, I.G., Studdert-Kennedy, M. (Eds.), Modularity and the Motor Theory of Speech Perception. LEA, Hillsdale, NJ, pp. 117–137.

Walker, S., Bruce, V., O'Malley, C., 1995. Facial identity and facial speech processing: familiar faces and voices in the McGurk effect. P&P 57, 1124–1133.

Zihl, J., von Cramon, D., Mai, D., 1983. Selective disturbances of movement vision after bilateral brain damage. Brain 106, 313–340.