Brain (1997), **120,** 1793–1803

Speechreading in the akinetopsic patient, L.M.

R. Campbell,[1] J. Zihl,[2] D. Massaro,[3] K. Munhall,[4] M. M. Cohen[3]

[1]*Department of Human Communication Science, University College London,UK,* [2]*Max Planck Institut für Psychiatrie, Klinisches Institut, München, Germany,* [3]*Program in Experimental Psychology, University of Southern California, Santa Cruz, California, USA, and* [4]*Psychology Department, Queen's University, Kingston, Ontario, Canada*

Correspondence to: Ruth Campbell, Department of Human Communication Studies, University College London, Chandler House, Wakefield Street, London WC1N 1PG, UK

Summary

Patient L.M. has a well-documented, long-standing and profound deficit in the perception of visual movement, following bilateral lesions of area V5 (visual movement cortex). Speechreading was explored in this patient in order to clarify the extent to which the extraction of dynamic information from facial actions is necessary for speechreading. Since L.M. is able to identify biological motion from point-light displays of whole-body forms and has some limited visual motion capabilities, we expected that some speechreading of faces in action would be possible in this patient. L.M.'s reading of natural speech was severely impaired, despite unimpaired ability to recognize speech-patterns from face photographs and reasonable identification of monosyllables produced in isolation. She was unable to track multisyllabic utterances reliably and was insensitive to vision when incongruent audiovisual speech syllables were shown. Point-light displays of speech were as poorly read as whole face displays. Rate of presentation was critical to her performance. With speech, as with other visual events, including tracking the direction of gaze and of hand-movement sequences, she could report actions that unfolded slowly (~one event per 2 s). In line with this, she was poor at reporting whether seen speech rate was normal, fast (double-speed) or slow (half-speed). L.M.'s debility is the converse of that reported for a patient with lesions primarily to V4 (H.J.A.), who is unable to speechread photographs of faces but can speechread moving faces. The visual analysis of both form and motion is required for speechreading; the neural systems that support these analyses are discussed.

Keywords: speechreading (lipreading); audiovisual fusion; biological motion; V5 lesions

Abbreviations: M-system = magnocellular system; P-system = parvocellular-system

Introduction

Speechreading, the extraction of speech information from the seen action of the lower face, especially the jaws, lips, tongue and teeth, is a natural skill in hearing people. It improves the understanding of speech in noise (Sumby and Pollack, 1954) and under clear hearing conditions (Reisberg *et al.*, 1987). When different auditory and visual streams are appropriately dubbed, perceptual illusions arise so that, while perceivers report 'what they hear' an influence of the seen mouth shapes and actions can be demonstrated (the McGurk fusion illusion) (*see* McGurk and MacDonald, 1976; Massaro and Cohen, 1983; Massaro, 1987).

Movements of the lower face during speech which facilitate speechreading might be regarded as a form of biological motion (Johansson, 1973; Oram and Perrett, 1994); this has been investigated using point-light displays. Such displays, where selected points on the surface of the moving body are illuminated, cannot be identified from a still-frame but are readily recognized when animated. The direction of body-movement, type of action, recognition of the number of actors, and even their identity, age or gender can all be identified from such whole body point-light displays (Jansson and Johansson, 1973; Cutting and Kozlowski, 1977; Kozlowski and Cutting, 1977; Cutting, 1978; Perrett *et al.*, 1990). Furthermore, movements of body-parts, including the hands and the face, can be similarly displayed, and facial expressions and sign-language movements can be perceived from them (Bassili,1978; Poizner *et al.*, 1981; Tartter and Knowlton, 1981).

Rosenblum *et al.* (1996) have found that a point-light display of illuminated dots over the cheeks, lips, teeth and tongue can be speechread; hearing subjects showed an improvement in shadowing speech in noise compared with a baseline condition where no visual display was seen. Such displays can also generate audiovisual fusion illusions. A point-light face display saying ‘ve’ dubbed to an audio ‘ba’ could generate the illusion that ‘va’ was spoken (Rosenblum and Saldaña, 1996). Such displays, which do not deliver the full visual form of the lips, mouth cavity, teeth, tongue and jaw, let alone the facial texture deformations associated with such movements, may nevertheless map effectively and directly onto the perceptual and representational mechanisms that support speechreading from natural visual sources. Visual movement, as indicated by such biological motion displays of a speaking face, may play a significant role in speechreading.

Since speaking is a natural dynamic event, it may be construed as (visual) biological motion and processed by mechanisms designed to perceive such actions. However, there may be reasons to consider it to be special. Facial speech actions have distinctive dynamic characteristics. The frequency content of speech movements is restricted to a small temporal range; 4 Hz is the modal syllable rate for reading aloud (Ohala, 1975) and this is reflected in the visual dynamic characteristics of face-surface actions. Movements of the lips can have higher frequency components; thus the opening movements for stop consonants such as /p/ can be rapid due to the combined aerodynamic and muscular forces at oral release. The net effect of a relatively low syllable rate and its modulation by some higher frequency components from consonant articulation produces a recognizable speech-specific temporal code for seen actions of the lower face (Vatikiotis-Bateson *et al*., 1996; Munhall and Vatikiotis-Bateson, 1997). As for image characteristics, speech-movements occur without translation of the moving parts across the field. But they have to be interpreted in the context of a moving head: a non-trivial computational problem (Black and Yacoob, 1995). The movements that cause the oral cavity to change in shape and size reflect the conjoint actions of sets of muscles primarily in one, frontoparallel plane. Because of its phonetic basis in speech, the order of face actions is crucial to its interpretation. For example, neither the first nor the last gesture in a sequence is privileged (‘pool’ versus ‘loop’). Finally, the natural context of speechreading is bimodal; we see and hear the speaker’s actions as a synchronized event. Such considerations suggest that speechreading may make use of rather different motion and form-from-motion mechanisms than those for other biological actions, including facial expressions or manual signs.

Occipital specialization for vision: effects of discrete lesions on visual tasks, including speechreading

Occipital areas are specialized for different aspects of visual perception. Areas V1 and V2 (striate cortex), the first cortical projection sites, map the full retinal field via projections from neural systems originating in the lateral geniculate nucleus. These distinct systems are functionally specialized for high luminance and colour vision (parvocellular, or P-system) and movement vision (low luminance, magnocellular or M-system) and are, generally, separately represented in V1 and V2 (there is some evidence that M- and P-systems can converge in V1; Sawatari and Callaway, 1996). Cortical regions V3–V5 (occipital prestriate regions) maintain the integrity of these neural systems to some extent, but also show further integration. V5 is said to be ‘movement cortex’, with cells specifically tuned to directional movement (largely M-system based). The perception of visual form depends on the integrity of areas V4 and V3 (and their higher projections in the temporal lobe). V4 may be organized to analyse form-from-colour (P-system based). V3 comprises mainly M-system cells and one hypothesis is that it may be specialized to analyse objects in/from motion, i.e. ‘dynamic form’ (Zeki, 1991, 1993). Recent studies (Beckers and Zeki, 1995; ffytche *et al*., 1995) suggest that V1 need not be the first cortical projection site for visual movement. Two parallel systems for movement processing may be engaged. There is fast direct activation of V5 from subcortical sites for high frequency movement components and serial activation of V5 via V1 and V3 for slower motion components ($<6°$/s).

While damage to V1 can lead to a lack of awareness of vision (blindsight) [though the overt perception of movement can be spared in such patients (Mestre *et al*., 1992; Barbur *et al*., 1993)], more specific functional deficits are observed when prestriate areas are damaged. In some of these patients biological motion may be relatively spared. Thus bilateral stroke patient H.J.A. (Humphreys and Riddoch, 1987; Humphreys *et al*., 1993) in whom areas V3 and V5 were relatively less damaged than V4, could identify moving forms that he could not see as still images. Whereas H.J.A. was unable to identify faces, facial expressions or mouth patterns accurately from still pictures or from stilled actions, he was able to identify all of these when the stimulus moved. Not only was H.J.A. able to identify natural seen speech events, he also showed essentially normal effects of vision on audition for audiovisual dubbed tokens (Campbell, 1992; Humphreys *et al*., 1993). However, the pattern of speech-reading impairment in two other patients with extensive bilateral posterior lesions and relative sparing of V5 does not suggest that functional V5 is sufficient for effective speechreading. Patient W.M. (Grüsser and Landis, 1991; Scheidler *et al*., 1992; Troscianko *et al*., 1996), with intact magnocellular function and undamaged V5, was completely insensitive to seen speech, whether still or moving. Similarly, patient D.F. (Milner *et al*., 1991; Humphrey *et al*., 1994), with spared ‘perception-through-action’ and limited visual movement perception, could not speechread faces in action or show any indication of influences of vision on reports of auditory speech tokens (Campbell, 1996).

Lesions of V5

Results of tests on patient L.M. may cast a clearer light on the involvement of cortical mechanisms of visual movement in speechreading. She has a well established and circumscribed lesion of lateral occipitotemporal regions. In particular, while V5 and some parts of surrounding posterior temporal lobe are damaged on the left and the right, areas V1 to V4 are essentially spared (Shipp *et al*., 1994). The extent of L.M.'s visual motion deficit and the patterns of sparing and impairment have been extensively investigated by Zihl *et al*. (1983, 1991) and are summarized in relevant detail below. She presents a unique opportunity to explore the extent to which visual motion perception, defined in terms of the function of visual movement cortex, may be integral to speechreading. In this paper we report how she performed a number of speechreading and associated tasks with a view to clarifying how the processing of visual movement might contribute to speechreading.

Case details

L.M. was 65 years old at the time of testing. At the age of 43 years she developed a sinus vein thrombosis which led to severe headache, vertigo and nausea, culminating in a state of stupor. Hospital examination (October 1978) revealed xanthochromic CSF, bilateral papilloedema and continuous delta–theta EEG activity. Cerebral arteriography showed occlusion of the parietal segment of the superior sagittal sinus and of cortical veins in the temporoparietal region; a number of abnormal 'corkscrew' veins were also evident. A subsequent (1980) CT scan revealed large bilateral lesions of occipitoparietal cortex. However, PET-MRI imaging (Shipp *et al*., 1994) showed grey-matter lesions to be confined to area V5 (occipitotemporal junction), bilaterally. The lesions were symmetrical, centred on Brodmann areas 37 and 19 in the lateral occipital gyri, extending ventrally to the position of the lateral occipital sulcus (obliterated), posteriorly to within 2 cm of the midline on each side at the borders of areas 18/19. Anteriorly, all of areas 19 and 37 were affected. The extension of the lesions was slightly different on the left and the right, with that on the left reaching the occipital continuation of the superior temporal sulcus, that on the right not extending quite so far. The right-sided lesion into occipitoparietal cortex was greater than that on the left, reaching dorsal parts of area 19 and possibly area 39. White-matter damage and ventricular enlargement was also evident on scan, and was more extensive on the right than the left side; on both sides much of the white-matter underlying lateral occipital cortex was destroyed, although medial cortex within calcarine and parieto-occipital sulci was intact.

Tests on L.M. since 1980 (Zihl *et al*., 1983, 1991; Hess *et al*., 1989; McLeod *et al*., 1989; Paulus and Zihl, 1989; Baker *et al*., 1991; Shipp *et al*., 1994) confirm that there is no visual field deficit when tested by static or dynamic stimuli and no indication of extinction on visual confrontation. Among tests reported in the normal range were the following: critical flicker fusion frequency; saccadic localization; tactile and acoustic motion tracking; distance perception in the frontoparallel plane and in depth; subjective estimation of horizontal and vertical; line bisection; line orientation matching; spatial position matching; matching of parts to the whole; face recognition and face constancy.

L.M. cannot perceive speeds in excess of 6–8°/s (Zihl *et al*., 1983, 1991). Psychophysical tests using drifting gratings showed a 20-fold increase in movement-detection thresholds compared with normal. In contrast, L.M.'s thresholds for contrast discrimination were only three times greater than normal. Since her basic temporal and contrast sensitivity functions were only modestly suppressed, Hess *et al.* (1989) suggested that signal processing up to the level of V1 was not grossly impaired, while Baker *et al.* (1991) showed that L.M.'s performance on a task of stochastic dot motion detection was similar to that of monkeys with selective lesions to V5 (Newsome and Paré, 1988).

Attested minimal motion perception abilities in L.M.

Despite her impairment, L.M. can detect the direction of cardinal motion in random-dot kinematograms. These are dot (visual noise) fields in which there is coherent displacement of the contrast values of the dots comprising the field, which is normally perceived as directional movement of the field. Shipp *et al.* (1994), from cortical imaging evidence, suggested that intact superior parietal and cuneus (within V3) regions are responsible for this ability in L.M., although they noted that her elevated motion thresholds do not correspond with motion-sensitivity capacities in V3. L.M.'s motion perception was very strictly circumscribed; she was unable to identify any directions other than up, down, left and right. When the dot-field included still random background dots or occasional dots moving in the opposite direction, L.M. was unable to perceive direction of movement. This has little effect on normal viewers. This factor, of the relative density of the coherently moving elements compared with other elements in the field, also has a marked effect on L.M.'s ability to detect shape through motion in random-dot kinematograms (Rizzo *et al*., 1995). While L.M. can perceive global coherent motion and direction of motion in such figures, her threshold for the detection of such figures (coherent movement of points describing a figure against a 'background' of non-coherent movement) is much higher than normal. L.M.'s inability to distinguish still from moving elements, or to establish direction of movement when some parts of the array are discrepant, characterizes other tests of her movement discrimination (Zihl *et al*., 1991; Rizzo *et al*., 1995). Thus she is unable to use the movement of object elements as an attentional grouping cue when still and moving parts are in the display (McLeod *et al*., 1989).

In summary, therefore, L.M.'s movement perception is extremely impaired when tested with overt tests requiring discrimination or reporting. She has very reduced ability to perceive directional coherent movement. Her movement thresholds (but not her temporal resolution thresholds) are higher than any yet described for patients with occipital lesions. Her processing of movement is especially compromised when some parts of the visual field are not moving or are moving in different directions. This also fits her phenomenological reports that she finds movement 'disturbing', that she sometimes sees moving objects as still ones (rather as if they were strobe-illuminated), and that both viewer-centred and object-centred movements are problematic.

For wholly coherent displays she can identify shape from motion for both 2-D and 3-D figures (Rizzo *et al*., 1995, McLeod *et al*., 1996). Perhaps surprisingly, she can distinguish a range of whole-person biological actions such as crouching, walking, dancing or jumping from dynamic point light displays alone (McLeod *et al*., 1996). In this task as in others, L.M. was more impaired when random still dots constituted the background, a manipulation that fails to affect normally sighted viewers. Such limited movement detection capacities may be sufficient to support speechreading, whether from full-forms or from point-light displays. Moreover, movements of the lower face during speech which facilitate speechreading could be considered to be a special form of biological motion, for their perception entails the identification of a dynamic human event, the identification of a spoken phrase or word, from sight alone. If L.M. shows reasonable speechreading it could indicate that there is a unitary class of biological actions which can be processed through routes that do not have to make use of V5.

Experimental tests

L.M. was tested over 3 days at the Max Plank Institut für Psychiatrie, Munich in 1995. All tasks were administered in German by J.Z. with R.C. in attendance. L.M. finds testing effortful and, as well as careful preparation for each test, numerous breaks were given throughout testing. Informed consent was obtained for testing from L.M. and from control subjects according to the Helsinki declaration (1991) and the study was approved by the Ethics Committee of the Max Plank Institut für Psychiatrie, Munich.

Speaking and gurning faces: photographs

Half-tone photographs of faces making speech sounds ('oo', 'ee', 'ah', 'sh', 'mm' and 'f', mixed with photographs of faces making non-speech gestures, including 'fish-faces' and tongue-protrusion (gurning faces) were shown to L.M. to sort into two piles: speech and non-speech. The faces were of five different individuals, of different size, three-quarter-view and full face, and included photographs of the lower-face alone. In total ~60 face pictures were shown. These pictures had previously been used to estimate the ability of patients T., D. and control subjects, who were also native German-speaking, to distinguish speaking from non-speaking faces (Campbell *et al*., 1986). L.M. was fast and accurate at this task, making no errors over the 10 min of testing.

Distinguishing vowel-shape from photographs

Following the sorting of face-images as speech or non-speech, L.M. was asked to identify the different vowel-types (~40 images). Once more she was accurate and fast, making no errors on this task.

Live-action tasks: syllables

L.M. was asked to report (aloud) simple silent speech actions produced by J.Z. and by R.C. ('Say what I say').

Vowel identification (/a:/, /u:/ and /i:/)

Thirty-two vowels were mouthed, one at a time, at a rate of 1/s from a resting open mouth shape (shwa). These were reported at maximum accuracy. These forms were identified slowly; L.M. sometimes touched her own mouth to confirm that the seen sound matched those which she reported.

Monosyllables

Interconsonantal vowels /ma:m/, /mu:m/ and /mi:m/. These monosyllables were produced at a speaking rate of 1/s from a resting closed mouth, with lengthy inter-presentation intervals (~6 s) for report. Two trials of 36 utterances, once with J.Z. speaking, once with R.C., generated accuracy of 27 out of 36 and 32 out of 36, respectively. L.M.'s errors were /mi:m/ for seen /ma:m/ or /mu:m/.

Bisyllables

Single bisyllables were presented for immediate verbal identification. These were all of the form /mumu/,/mama/, /mimi/, or any combination of these (e.g. /mima/, /mumi/ and /mamu/). Rate of utterance of bisyllables (silent mouthing) was 1/s. Two trials of 36 utterances, once spoken by J.Z., once by R.C. were presented. Accuracy was ~50% and was not affected by the speaker seen. All errors were of the form 'mimi', 'mama' or 'mumu' for presentations of 'mami', 'muma', 'mumi' etc. The position of the repeated syllable (first or second) was not systematically related to accuracy. L.M. was unimpaired at repeating varied bisyllables presented acoustically (whispered). Her misreports were specific to silent visual presentation.

Three native German control subjects of similar age were later tested in an identical fashion. Mean performance for the three tasks was 100% (task 1), 97% (task 2) and 97% (task 3). L.M.'s performance on tasks 2 and 3 was abnormal.

Lexical speechreading

Speechreading silently spoken numbers between 1 and 10 is readily accomplished by both English (Campbell and Dodd, 1980; Vitkovitch and Barber 1996) and German native speakers (Campbell *et al*., 1986, 1990). On first, live,

Table 1 *Number words: full-face and point light displays*

	L.M.		Control 1		Control 2	
	Full-face	Point light	Full-face	Point light	Full-face	Point light
Vision alone	3/19	2/19	63%	28.3%	56%	28%
Audiovisual (34 dB)	9/19	9/19	100%	100%	100%	100%
Audition alone	14/24		70%		67%	

presentation (J.Z.) of numbers mouthed at a rate of one per 1.5 s (slow rate) L.M. scored 28 out of 47 (~60%). At faster speech rates of one number per second she achieved a score of 26 out of 48 (54%). There was some patterning in her responses; she tended to confuse the numbers '3', '8' and '1'. Feedback was given after each response.

German number words: point-light and full-face versions of a synthetic speaking face

This test directly compared L.M.'s ability to read full-face displays and point-light displays. We reasoned that for L.M., a point-light display may be more readily perceived than a full-face display, which may be 'noisier' in terms of its component (moving and non-moving) parts. L.M. was able to identify such point-light displays of whole-body actions as long as the display contained no other elements. A wire-frame-based computer model of a speaking face, which uses parameters derived from the speech production of American English phonemes to control the face actions, has been widely used in experiments on audiovisual speech perception and formed the basis for the material used in testing (Cohen and Massaro, 1990,1993,1994; *see* below for details). This computer-generated display was programmed with the German number words 1 to 10 (/aintz/, /tzvai:/, /drai:/, /fi:r/ etc.), using the American-English phoneme repertoire (i.e. the synthetic face had a marked American accent). Six 48-word blocks were generated, with words in a randomized order, and stored to videotape, with a 3-s delay between each utterance. Synthetic point-light and full-face displays occurred unpredictably and in equal numbers within each block. The point-light displays comprised 28 illuminated points at specific face-feature landmarks on the lower half of the face of the wire-frame model. These followed precisely the same dynamic constraints as the full-face display. Point positions were as detailed in the experimental studies of Rosenblum and Saldaña (1996). It should be noted that the synthetic face maintains head position motionless in frame; this may simplify the resolution of speech movements which do not have to be parsed in relation to head movements.

This material was presented to L.M. under two conditions: as silent speech (vision alone) and as quiet ($<$35 dB) audiovisual material. In addition, a quiet audio-alone condition was run to check the comprehensibility of this American-English speech and to establish a sound level at which seeing the speaker should improve performance/report. L.M. was unable, because of fatigue, to view all of the trial series under all conditions; her performance shows actual numbers of trials performed (not including 10 practice trials) alongside those of control subjects who performed all six blocks of 48 numbers for the visual and audiovisual conditions, and one block of the audio-alone material.

Control subjects were females aged 59 and 61 years, of similar educational and social background to L.M. Both were native German speakers (*see* Table 1). The following points can be observed from the table. First, in these elderly German subjects, the silent synthetic face was 'speechreadable' as a full-face display, despite his American 'accent'. Secondly, point-light displays as silent visual displays were less speechreadable than full-face displays. Thirdly, control subjects showed a very marked gain when they saw and heard the face compared with hearing the (quiet) voice alone. That is, point light faces can effect an improvement in shadowing quiet speech, even when the speaker is unfamiliar and has a 'foreign' accent. L.M.'s performance under audition-alone was somewhat poorer than that of control subjects; that is, there is the potential for audiovisual performance to be enhanced relative to audition-alone. However, no such improvement was observed. It should be noted that she was at chance levels (i.e. ~10%) for the silent presentations. We conclude that L.M. is effectively unable to speechread the synthetic face in either full-face or point-light versions.

Audiovisual speech processing: discrete monosyllables

A videotape comprising auditory, visual and audiovisual (digitized) tokens of the monosyllables /ba/, /va/, /tha/ and /da/ was used for testing. Both a natural male speaker and the synthetic face were used, and each face appeared unpredictably within each trial run. A complete trial set comprised 36 dubbed syllables (four natural faces, four synthetic, all dubbed to the four speech sounds) and eight unimodal (four vision alone, four audition alone) tokens, all presented in random order. The videotape contained 15 such 44-trial sets, allowing reliable data to be obtained for individual cases. The auditory and the natural video tokens were taken from a single male speaker on the Bernstein and Eberhardt videodisk (1986) whose face was seen in the natural speech condition.

Synthetic visible speech

For the synthetic visible speech, as for the synthetic face speaking German numbers, a parametrically controlled

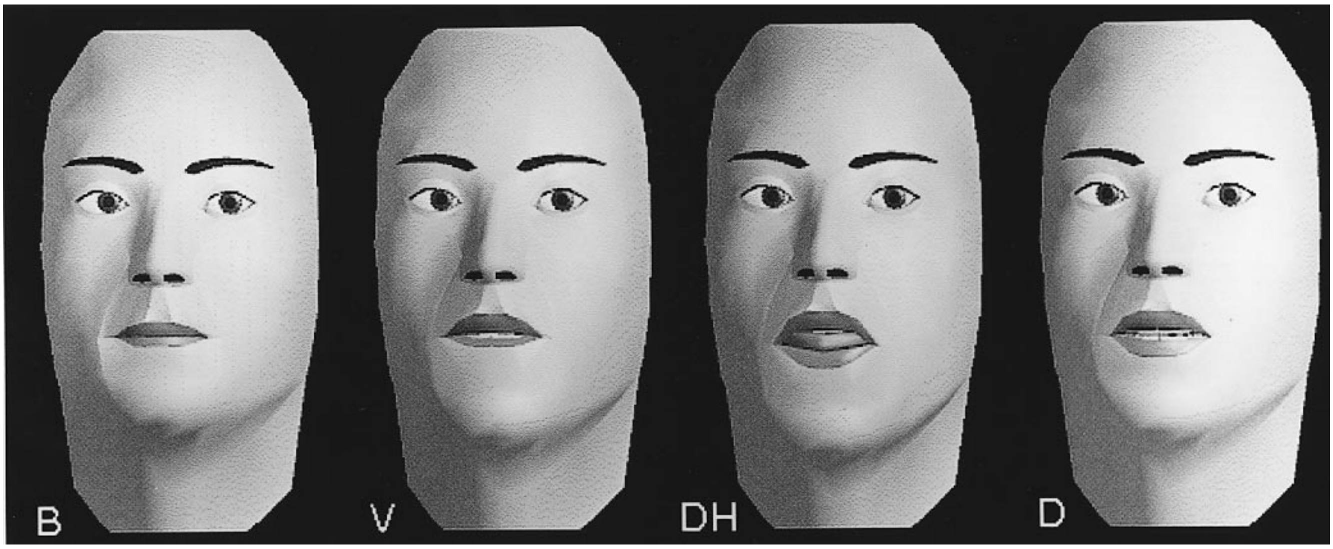

Fig. 1 The four consonant positions of the synthetic face used in testing. For point light displays, points were chosen that covered the lower part of the face, using the feature landmarks described by Rosenblum and Saldaña (1996). The syllable-reading material was in colour, the point light display was monochrome (light on dark).

polygon topology was used to generate the syllables (Cohen and Massaro, 1990, 1993, 1994). The animated display was created by modelling the facial surface as a polyhedral object composed of ~900 small surfaces arranged in 3-D, joined at the edges (Parke, 1975, 1982). The surface was smoothshaded using Gouraud's (1971) method. To achieve a more realistic synthesis, a tongue was added, with control parameters specifying its angle, length,width and thickness. The face was animated by altering the location of various points in the grid under the control of 50 parameters, 11 of which were used for speech animation. Each phoneme was defined in a table according to target values for the control parameters and their segment durations. Examples of the control parameters include jaw rotation, mouth *x*-scale, mouth *z*-offset, lipcorner *x*-width, lower lip 'f'-tuck and so forth. Parke's software, revised by Cohen and Massaro (1990, 1993) was implemented on a Silicon Graphics computer. The synthetic face was programmed to produce the syllables /ba/, /va/, /tha/ and /da/. Figure 1 shows the face at the onset of articulation of the four syllables.

Audiovisual speechreading: factorial combinations of seen and heard syllables

Audiovisual stimuli were created by computationally combining the auditory speech of the four syllables with the visual speech of each of these syllables. For both the natural visible and synthetic speech, the beginning of the auditory speech was synchronized with the consonantal release of the visual speech and the dynamic portion of the visual stimulus (from a resting face position) began before, and finished after, the corresponding auditory tokens. The durations of the visible speech were (approximately) 730 ms for /ba/, 730 ms for /va/, 900 ms for /tha/ and 667 ms for /da/. The corresponding durations for the four auditory tokens were 396, 470, 506 and 422 ms. A 100 ms, 1000 Hz warning tone was played 600 ms before each presentation, which occupied a 1300 ms interval.

Instructions to L.M. and to the control subjects were to watch and listen to the speaker and to identify the syllable by speaking it aloud; a full response choice was indicated, i.e. subjects were told that the syllable could be any of /ba/, /tha/, /da/, /va/, or a combination or blend of these (a consonant cluster). The image was shown on a large colour TV monitor and subjects were seated about 1 m from the screen. Each image subtended a visual angle of ~6° at this distance. Loudness level was set at 65 dB, and testing was in a quiet room.

L.M. was able to complete four trial blocks, after an initial practice session. The control subjects completed six trial blocks each. For each subject, the mean performance was calculated for each of the (44) presented tokens. These are summarized for L.M. and the two control subjects in Figs 2 and 3.

As these figures show, L.M. was very impaired at identifying these tokens by eye, but not by ear. Responses to /tha/ tokens, both auditory and visual, should not be expected to be correct, since this phoneme is absent in German. Despite this, for visual tokens without audition her responses were at chance levels, while there was no apparent influence of vision on audition in the combination responses. In contrast, both L.M.'s control subjects showed reliable identification of visual tokens /ba/, /va/ and /da/, and systematic influences of these forms in their responses to audiovisual tokens.

Visible speech rate

We have remarked that seen speech has a specific temporal code or pattern. One consequence is that people are as able

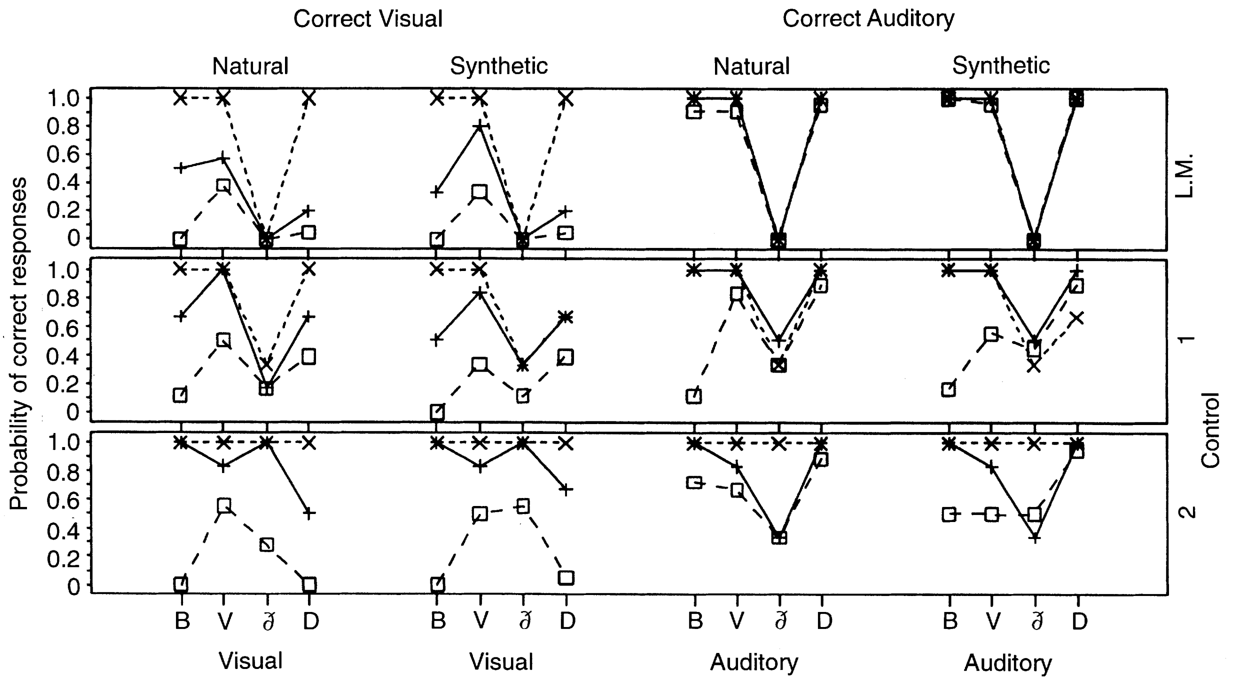

Fig. 2 Results from L.M. (top) and two control subjects. Average observed performance scored as accuracy/modality. Proportion correct on unimodal trials (solid lines and '+' signs) is shown for the visual syllable (*left* panel) condition. This panel also gives accuracy for the visual syllable for bimodal trials when the auditory information is consistent with the visual (small dashed line and '×' signs) and when the auditory information is inconsistent with the visual (large dashed line and open squares). Analogous measures are given for the auditory syllable on the *right* panel.

to distinguish fast and slow rates of speech by eye as by ear (Green, 1987). Furthermore, seen and heard speech interact in a rate-defined manner, so that seen speech-rate affects the phonetic categorization of heard speech tokens (Green and Miller, 1985). If the perception of dynamic facial actions is impaired, can changes in rate of speech be seen?

A videotape comprising excerpts from the Bernstein and Eberhardt (1986) videodisk database was the image source. This contains sentences and phrases selected from the CID Everyday Sentences set (Davis and Silverman, 1970). The excerpts used were of short English phrases spoken by a single male American speaker with normal intonation. They were re-recorded under three conditions: normal speed, half-speed and double-speed, and then assembled in random order. Each segment started and finished with a resting face of the speaker. L.M. was asked to judge whether the speech rate was slow, normal or fast for each token where the speech tokens were assembled in a random order. When the tape was running at normal speed she was correct for 90 out of 122 trials (~80%). This was a trivially easy task for control subjects (ceiling performance). When the videotape rate was set to double-speed, so that formerly slow events appeared at normal speed, normal ones fast, and fast ones at double speed, L.M.'s responses were always 'fast'; i.e. she could not detect 'normal' from 'fast' or 'very fast' items reliably (responses <20% correct). Once again, this task is trivially easy for control subjects. Although her verbal responses were inaccurate, she responded immediately and with some distress

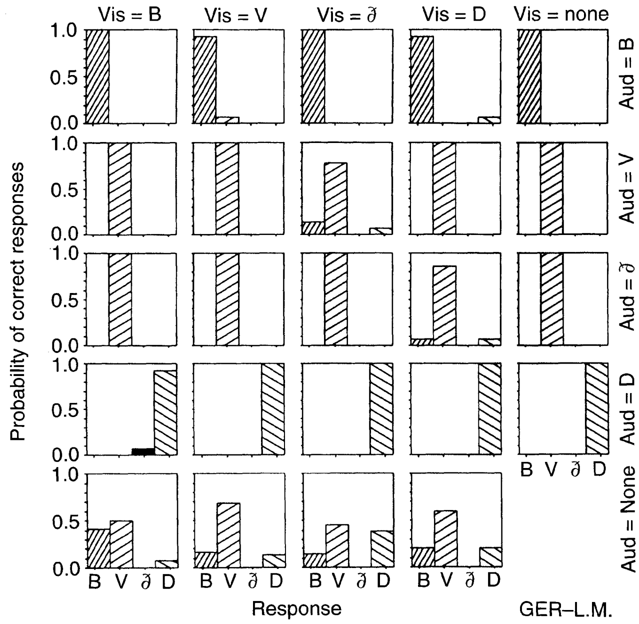

Fig. 3 L.M.'s average observed performance in terms of accuracy with respect to the modality of bimodal and unimodal tokens (visual horizontal, auditory vertical). Note (i) that L.M.'s identification of visual tokens was biased towards 'V' and was essentially within chance range and (ii) that the auditory-alone pattern (rightmost column) closely predicts her performance on all the bimodal events.

to the moving face, which she found very disturbing. It should be noted that it would be possible to assess rate of speech by estimating the duration of each utterance rather than its speed, and this may have been the basis for L.M.'s correct responses. However, even where perceived speech rate may have been confounded by other cues, L.M. was not able to detect rate of speech accurately under these easy conditions.

Other tests

One reason why L.M. was so impaired on tasks of natural speechreading could have been because these often involved complex relative motions of different parts of the display. Mouth rounding, occurs for example, on all (*x*, *y* and *z*) coordinates of a frontally viewed mouth for vowels (/u:/). Also, different parts of the face move at different rates; e.g. jaw drop actions are slower than many bilabial lip actions. Could L.M. perceive 'simpler' live actions performed at constant speed or using constrained single trajectories?

Numbers in the air

We were interested to find out to what extent L.M. would be able to identify known static forms by sight when these were displayed dynamically. To this end single digit forms (e.g. '2', '5') were drawn in the air by J.Z. with large continuous (right) hand movements moving, as far as possible, at constant speed. The numbers were drawn on an imaginary plane between the tester and L.M. (that is, they were drawn mirror-inverted) stretching from the head to the middle of the trunk. L.M. and the tester were about 1 m apart. At a drawing speed of about 2 s per number she reported 14 out of 24 number forms correctly. Her reports were extremely effortful. Occasionally she traced the form with her right hand on the table in front of her while watching, and when she did this she was generally correct. She consistently reported '9' as '0'. At faster rates (1/s) L.M. was unable to report any such number correctly.

Sequence of direction of gaze

L.M. could report cardinal (left–right–up–down) direction of movement under some conditions (Shipp *et al*., 1994). In this test she was asked to report sequences of eye movements performed by the tester who faced her at a distance of 1 m (J.Z.). Each sequence was reported after each three-movement action (e.g. 'left–down–right'). Every sequence started with eyes straight ahead (looking at L.M.). At rates of two movements per second, L.M. was unable to report any full sequence correctly (0 out of 10). Nevertheless she reported several of the individual items correctly, and these showed a marked serial position effect; eight out of 10 first positions, six out of 10 final positions and two out of 10 mid-positions were correctly reported. At slower rates (1/2 s) 20 out of 27 sequences were completely correct. Her errors showed no consistent mistakes in cardinal direction, but first and last position items were again more accurately reported than mid-positions.

Discussion

L.M. could identify still lip-pictures of speech and also some simple live-action monosyllables (/mu:/, /ma:/ and /mi:/). She was impaired at speechreading silent numbers, but this ability was not completely lost (performance was above chance). In contrast, she was unable to speechread even simple bisyllables presented live when the second syllable was different from the first, nor could she speechread rhyming monosyllables shown on video, where a single vowel followed different consonants (/ba/, /va/ and /da/). These debilities are not due to effects of head-movement interacting with speech-movements of the lower face, for they were as marked with the synthetic face which retained its position on the screen as with the 'natural' face. L.M. was unable to read moving speech actions under point-light conditions. She showed no discernible influence of vision on audition in a systematic test of audiovisual integration using /ba/, /va/ and /da/ tokens. She was poor at detecting whether a face was speaking quickly or slowly over the tested range. These impairments in perceiving human movements were not limited to seeing speech; she could only track direction of gaze sequences when they were produced slowly, with long resting states which she could identify and remember. She could identify number forms traced in the air, but again only when these were produced very slowly; this suggests an idiosyncratic strategy for identifying the event, possibly serial recall of the position of the hand. The speechreading tasks that she could achieve appear to rely on the analysis of mouth-shape, and this can be achieved from moving faces only at very slow rates of presentation and is lost when different mouth-shapes replace each other in sequence, leaving only the first (sometimes) and the last mouth shapes available for report.

Visible speech perception might be considered to be based entirely on the temporal properties of speech. That is, the sole input to the speechreading system could be either the coded movement trajectories of facial landmarks or the detection of direction of optical flow over the movement-deformed facial surface (Mase and Pentland, 1990; Rosenblum and Saldaña, 1996; Munhall and Vatikiotis-Bateson, 1997). If these dynamic processes were the basis for establishing long-term representations of seen speech, then L.M. might have been expected to have lost the ability to perform any speechreading task. Yet her ability to identify speech patterns from still images is unaffected by her lesion, even after 17 years, and she is also able to identify live speech actions to a limited degree when these are presented very slowly. To the extent that she can identify seen speech patterns, L.M. appears to read facial actions as if she were viewing still face shapes.

One possibility is that L.M. may have some residual movement vision allowing her to integrate stimulus movement over time, but only when the perceived events move extremely slowly (e.g. as for her correct reports of numbers drawn in the air and for sequences of direction of gaze). However, outside this 'speed window' she is unable to extract form from motion and must then use, as default states, the perceived static start and end positions. Images may occasionally be sampled from occasional intermediate positions where eye-movements might (accidentally) help to maintain a relatively sustained image for analysis.

Biological motion and speechreading: different substrates?

McLeod *et al.* (1996) reported that L.M. is able to categorize biological actions from moving pointlight displays. Yet she was unable to identify similar displays for speech (pointlight display of numbers). Seeing speech does not make use of the functional and neuroanatomical substrate that supports the identification of other body actions. Spared biological-motion perception, accompanying dense lesions affecting other aspects of visual processing, including other aspects of movement perception has been reported in other posterior-lesion patients (*see* e.g. Vaina *et al*., 1990). Biological motion perception through simple pointlight displays may be achieved by a variety of means and may make use of more extensively distributed neural substrates than V5. Howard *et al.* (1996) have shown that functional MRI activation for point-light and moving field displays extends to superior temporal areas, including regions activated by listening to speech. In the context of this particular finding, L.M.'s impaired speechreading is surprising; if any natural action could be construed as intrinsically polysensory and dependent on intact superior temporal areas, it is the perception of someone speaking. Indeed, functional MRI studies of speechreading in normal people implicate extensive bilateral activity in superior temporal areas, including those identified by Howard *et al.* (1996) (Calvert *et al*., 1997). However, it should be noted that L.M.'s lesions extend into superior temporal areas on the left and also that L.M.'s lesions would disconnect projections to these areas through V5.

Movement and form in natural speechreading

The dynamic signature of seen speech fits the demonstrated capacities of the 'slow' serial movement analysis system, which utilizes serial projections from V1 via V3 and V5, rather than the 'fast', direct projection to V5 (ffytche *et al*., 1995). Seeing natural speech probably utilizes a network involving the activation of visual areas V1–V5, serially, with some reciprocal activation between these areas (for example to take account of H.J.A.'s ability to perceive seen speech in motion, despite damage to V2 and V3 in addition to V4), and with projections to superior lateral temporal areas, especially those related to the perception of language. This network also allows for the ready integration of visual form and movement in the perception of natural speech. The different component speechreading abilities of L.M., who can perceive still-lipshapes but not moving ones, and of H.J.A. who shows the opposite pattern, can, in turn, be contrasted with the more complete speechreading failure of patient T. (Campbell *et al*., 1986), whose speechreading was impaired both for still and moving material, yet who had no low-level visual disturbance and no visual agnosia. This patient, with a left, medially placed, temporo-occipitoparietal lesion, was also densely alexic. Seeing speech, like reading text, may require the integrity of a left superior temporal site that maps the seen event to the language processing system. But the pattern of sparing, and impairment, in patients with more posteriorly placed lesions suggests that there are different visual components to speechreading and that the perception of both form and movement is required for natural speechreading to work.

Acknowledgements

We gratefully acknowledge comments on the manuscript from L. Rosenblum, L. Jakobson and two reviewers of an earlier version of this report. Particular thanks are due to L.M. for her patience and forbearance in testing.This research was supported in part by grants to D.M. and to M.M.C. from the Public Health Service (PHS R01 DC00236), the National Science Foundation, USA (BNS 8812728) and the University of California, Santa Cruz; to KM from NSERC and NIH grant DC-00594. A research fellowship (1994–1995) from the Leverhulme Trust to R.C. enabled testing to be done.

References

Baker CL Jr, Hess RF, Zihl J. Residual motion perception in a 'motion-blind' patient assessed with limited-lifetime random dot stimuli. J. Neurosci 1991; 11: 454–61.

Bassili J. Facial motion in the perception of facesand of emotional expression. J Exp Psychol Hum Percept Perform 1978; 4: 373–9.

Beckers G, Zeki S. The consequences of inactivating areas V1 and V5 on visual motion perception. Brain 1995; 118: 49–60.

Bernstein,LE, Eberhardt,SP. Johns Hopkins lipreading corpus [videodisk]. 1986.

Black MJ, Yacoob Y. Recognising facial expressions under rigid and nonrigid facial motions. In Proc Int Workshop Automat Face Gest Recn; 1995: 12–22.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK. et al. Activation of auditory cortex during silent lipreading. Science 1997; 276: 593–6.

Campbell R. The neuropsychology of lipreading. [Review]. Philos Trans R Soc Lond B Biol Sci 1992; 335: 39–45.

Campbell R. Seeing speech in space and time: neurological and psychological findings. In: Proc 4th Int Congress Speech Lang Process 1996: 1493–7.

Campbell R, Dodd B. Hearing by eye. Q J Exp Psychol 1980; 32: 85–99.

Campbell R, Landis T, Regard M. Face recognition and lipreading: a neurological dissociation. Brain 1986; 109: 509–21.

Campbell R, Garwood J, Franklin S, Howard D, Landis T, Regard M. Neuropsychological studies of auditory-visual fusion illusions. Neuropsychologia 1990; 28: 787–802.

Cohen MM, Massaro DW. Synthesis of visible speech. Behav Res Meth Instrum 1990; 22: 260–3.

Cohen MM, Massaro DW. Modelling coarticulation in synthetic visual speech. In: Thalmann NM, Thalmann D, editors. Models and techniques in computer animation. Tokyo: Springer-Verlag, 1993: 139–56.

Cohen MM, Massaro DW. Development and experimentation with synthetic visible speech. Behav Res Meth Instrum 1994; 26: 260–5.

Cutting JE.Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. Perception 1978; 7: 393–405.

Cutting JE, Kozlowski LT. Recognizing friends by their walk: gait perception without familiarity cues. Bull Psychon Soc 1977; 9: 353–6.

Davis H, Silverman SR. Hearing and deafness. 3rd ed. New York: Holt, Rinehart and Winston, 1970.

ffytche DH, Guy CN, Zeki S. The parallel visual motion inputs into areas V1 and V5 of human cerebral cortex. Brain 1995; 118: 1375–94.

Gouraud H. Continuous shading of curved surfaces. IEEE Trans Comput 1971; C-20: 623–8.

Green KP.The perception of speaking rate using visual information from a talker's face. Percept Psychophys 1987 587–93.

Green KP, Miller JL. On the role of visual rate information in phonetic perception. Percept Psychophys 1985; 38: 269–76.

Grüsser O-J, LandisT. Visual Agnosias and other disturbances of visual perception and cognition. In: Cronly-Dillon J, editor. Vision and visual dysfunction, Vol. 12. Basingstoke (UK): Macmillan, 1991: 394–5.

Hess RF, Baker CL Jr, Zihl J. The 'motion-blind' patient: low-level spatial and temporal filters. J Neurosci 1989; 9: 1628–40.

Howard RJ, Brammer M, Wright I, Woodruff PW, Bullmore ET, Zeki S. A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. Curr Biol 1996; 6: 1015–9.

Humphrey GK, Goodale MA, Jakobson LS, Servos P. The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. Perception 1994; 23: 1457–81.

Humphreys GW, Riddoch MJ. To see but not to see: a case study of visual agnosia. London: Lawrence Erlbaum, 1987.

Humphreys GW, Donnelly N, Riddoch J. Expression is computed separately from facial identity, and it is computed separately for moving and static faces. Neuropsychologia 1993; 31: 173–81.

Jansson G, Johansson G. Visual perception of bending motion. Perception 1973; 2: 321–6.

Johansson G. Visual perception of biological motion and a model for its analysis. Percept Psychophys 1973; 14: 201–11.

Kozlowski LT, Cutting JE. Recognizing the sex of a walker from a dynamic point-light display. Percept Pychophys 1977; 21: 575–80.

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976; 264: 746–8.

McLeod P, Heywood C, Driver J, Zihl J. Selective deficit of visual search in moving displays after extrastriate damage. Nature 1989; 339: 466–7.

McLeod P, Dittrich W, Driver J, Perrett D, Zihl J. Preserved and impaired detection of structure from motion by a 'motion-blind' patient. Visual Cognit 1996; 3: 363–91.

Mase K, Pentland A. Lipreading by optical flow. IEICE Transactions J73-D-II 1990; 6: 796–803.

Massaro DW. Speech perception by ear and eye: a paradigm for psychological inquiry. Hillsdale (NJ): Lawrence Erlbaum, 1987.

Massaro DW, Cohen MM. Evaluation and integration of visual and auditory information in speech perception. J Exp Psychol Hum Percept Perform 1983; 9: 753–71.

Milner AD, Perrett DI, Johnston RS, Benson PJ, Jordan TR, Heeley DW, et al. Perception and action in 'visual form agnosia'. Brain 1991; 114: 405–28.

Munhall K, Vatikiotis-Bateson E. The moving face in speech communication. In: Campbell R, Dodd B, Burnham D, editors. Hearing by eye II: the psychology of speechreading and audio-visual speech. Hove (UK): Psychology Press, 1997. In press.

Newsome WT, Paré EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J Neurosci 1988; 8: 2201–11.

Ohala JJ. The temporal regulation of speech. In: Fant G, Tatham MA, editors. Auditory analysis and perception of speech. London: Academic Press, 1975: 431–53.

Oram MW, Perrett DI. Responses of anterior superior temporal polysensory (STPa) neurons to 'biological motion' stimuli. J Cog Neurosci 1994; 6: 99–116.

Parke FI. A model for human faces that allows speech synchronized animation. Comput Graphics J 1975; 1: 1–14.

Parke FI. Parameterized models for facial animation. IEEE Comput Graphics 1982; 2: 61–8.

Paulus W, Zihl J.Visual stabilization of posture in a case with selective disturbance of movement vision after bilateral brain damage: real and apparent motion cues. Clin Vision Sci 1989; 4: 367–71.

Perrett DI, Harries MH, Benson PJ, Chitty AJ, Mistlin AJ. Retrieval of structure from rigid and biological motion: an analysis of the visual responses of neurones in the macaque temporal cortex. In: Blake A, Troscianko T, editors. AI and the eye. Chichester: John Wiley, 1990: 181–99.

Poizner H, Bellugi U, Lutes-Driscoll V. Perception of American Sign language in dynamic point-light displays. J Exp Psychol Hum Percept Perform 1981; 7: 430–40.

Reisberg D, McLean J, Goldfield A. Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. In Dodd B, Campbell R, editors. Hearing by eye: the psychology of lip-reading. London: Lawrence Erlbaum, 1987: 97–113.

Rizzo M, Nawrot M, Zihl J. Motion and shape perception in cerebral akinetopsia. Brain 1995; 118: 1105–27.

Rosenblum LD, Saldaña HM. An audiovisual test of kinematic primitives for visual speech perception. J Exp Psychol Hum Percept Perform 1996; 22: 318–31.

Rosenblum LD, Johnson JA, Saldaña HM. Point-light facial displays enhance comprehension of speech in noise. J Speech Hear Res 1997. 1996; 39: 1159–70.

Sawatari A, Callaway EM. Convergence of magno- and parvocellular pathways in layer 4B of macaque primary visual cortex. Nature 1996; 380: 442–6.

Scheidler W, Landis T, Rentschler I, Regard M, Baumgartner G. A pattern recognition approach to visual agnosia. Clin Vision Sci 1992; 7: 175–93.

Shipp S, de Jong BM, Zihl J, Frackowiak RSJ, Zeki S. The brain activity related to residual motion vision in a patient with bilateral lesions of V5. Brain 1994; 117: 1023–38.

Sumby WH, Pollack I. Visual Contributions to speech intelligibility in noise. JASA 1954; 26: 212–5.

Tartter VC, Knowlton KC. Perception of sign language from an array of 27 moving spots. Nature 1981; 289: 676–8.

Troscianko T, Davidoff J, Humphreys G, Landis T, Fahle M, Greenlee M, et al. Human colour discrimination based on a non-parvocellular pathway. Curr Biol 1996; 6: 200–10.

Vaina L.M., LeMay M, Bienfang DC, Choi AY, Nakayama K. Intact 'biological motion' and 'structure from motion' perception in a patient with impaired motion mechanisms: a case study. Vis Neurosci 1990; 5: 353–69.

Vatikiotis-Bateson E, Munhall KG, Hirayama M, LeeY, Terzopoulos D. Dynamics of facial motion in speech: kinematic and electromyographic studies of orofacial structures. In: Stork DG, Hennecke ME, editors. Speechreading by humans and machines: models, systems, and applications. NATO-ASI Series F, Vol. 150. Berlin: Springer-Verlag, 1996: 221–32.

Vitkovitch M, Barber P. Visible speech as function of image quality. Appl Cogn Psychol 1996; 10: 121–40.

Zeki S. Cerebral Akinetopsia (visual motion blindness). A review. [Review]. Brain 1991; 114: 811–24.

Zeki S. A vision of the brain. Oxford: Blackwell, 1993.

Zihl J, von Cramon D, Mai N. Selective disturbances of movement vision after bilateral brain damage. Brain 1983; 106: 313–40.

Zihl J, von Cramon D, Mai N, Schmid C. Disturbance of movement vision after bilateral posterior brain damage. Further evidence and follow up observations. Brain 1991; 114: 2235–52.

Received February 18, 1997. Accepted May 16, 1997