International Journal of Language & Communication Disorders
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393930

Gesture and speech integration: an exploratory study of a man with aphasia

Naomi Cocks [a]; Laetitia Sautin [a]; Sotaro Kita [b]; Gary Morgan [a]; Sally Zlotowitz [c]
[a] Language and Communication Science, City University, London, UK [b] University of Birmingham, Birmingham, UK [c] University College London, London, UK

First Published on: 25 September 2008

INT. J. LANG. COMM. DIS.
2008, iFirst Article, 1–10

Gesture and speech integration: an exploratory study of a man with aphasia

Naomi Cocks†, Laetitia Sautin†, Sotaro Kita‡, Gary Morgan† and Sally Zlotowitz§

†Language and Communication Science, City University, London, UK
‡University of Birmingham, Birmingham, UK
§University College London, London, UK

(*Received 30 October 2007; accepted 6 June 2008*)

Abstract

Background: In order to comprehend fully a speaker's intention in everyday communication, information is integrated from multiple sources, including gesture and speech. There are no published studies that have explored the impact of aphasia on iconic co-speech gesture and speech integration.
Aims: To explore the impact of aphasia on co-speech gesture and speech integration in one participant with aphasia and 20 age-matched control participants.
Methods & Procedures: The participant with aphasia and 20 control participants watched video vignettes of people producing 21 verb phrases in three different conditions, verbal only (V), gesture only (G), and verbal gesture combined (VG). Participants were required to select a corresponding picture from one of four alternatives: integration target, a verbal-only match, a gesture-only match, and an unrelated foil. The probability of choosing the integration target in the VG that goes beyond what is expected from the probabilities of choosing the integration target in V and G was referred to as multi-modal gain (MMG).
Outcomes & Results: The participant with aphasia obtained a significantly lower multi-modal gain score than the control participants ($p<0.05$). Error analysis indicated that in speech and gesture integration tasks, the participant with aphasia relied on gesture in order to decode the message, whereas the control participants relied on speech in order to decode the message. Further analysis of the speech-only and gesture-only tasks indicated that the participant with aphasia had intact gesture comprehension but impaired spoken word comprehension.
Conclusions & Implications: The results confirm findings by Records (1994) that reported that impaired verbal comprehension leads to a greater reliance on gesture to decode messages. Moreover, multi-modal integration of information from speech and iconic gesture can be impaired in aphasia. The findings

Address correspondence to: Naomi Cocks, Language and Communication Science, City University, Northampton Square, London, EC1V 0HB, UK; e-mail: naomi.cocks.1@city.ac.uk

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online
http://www.informahealthcare.com
DOI: 10.1080/13682820802256965

highlight the need for further exploration of the impact of aphasia on gesture and speech integration.

Keywords: Gesture, aphasia, multi-modal integration, comprehension.

What this paper adds

What is already known on this subject

Very little is known about the impact of aphasia on gesture and speech integration in language comprehension tasks. Only one published study has investigated speech and gesture integration and this study was limited to the integration of pointing gestures and speech (Records 1994).

What this study adds

This study explored the extent to which participants used co-speech gesture in a sentence–picture-matching task. It also outlined a new methodology for determining speech and gesture integration. Results are reported on this task for a man with Broca's aphasia and 20 age-matched control participants. The findings suggest that the participant with aphasia relied on gesture when he found the integration task difficult, whereas the control participants relied on speech. Furthermore, the participant with aphasia was impaired in integration of information from speech and iconic gestures.

Introduction

Communication involves both verbal and non-verbal information exchange. Co-speech iconic gestures are movements of the upper limbs which depict directly the attributes or actions associated with a particular object or event, e.g. moving curved hands in concentric arcs to represent 'a ball', or moving a hand from side to side to indicate 'writing'. These gestures co-occur with speech during everyday interaction (McNeill 2000). Iconic gestures have been shown to make a significant contribution to our comprehension of speakers' intentions (e.g., Beattie and Shovelton 1999). Furthermore, in natural conversations, iconic co-speech gestures are often used to convey information that may not be overtly conveyed verbally such as object size, object location, manner of movement, spatial relationships and an object's path of movement (Kita and Özyürek 2003). In order to understand fully the speakers' intention, the addressee is required to comprehend both the speech and the gesture and then integrate the information gained from the two modalities.

There have been a number of studies that have investigated the impact of aphasia on pantomime gesture comprehension. Pantomime gestures are produced in the absence of speech. Tasks usually involve the standard usage of an object being gestured and an individual indicates the object that the gesture refers to (Duffy and Duffy 1981, Daniloff *et al.* 1982, 1986, Varney 1982, Lambier and Bradley 1991, Thorburn *et al.* 1995). The findings of these studies have indicated that impaired comprehension of pantomime is unrelated to severity of aphasia (Daniloff *et al.* 1982). It has, however, been found to be more frequent in participants with posterior lesions than participants with anterior lesions (Varney 1982, Daniloff *et al.* 1986, Lambier and Bradley 1991). However, pantomime is produced in the absence of speech and thus comprehension of pantomime requires only the comprehension of one modality and does not require integration.

Similarly, redundant gesture comprehension tasks do not require integration. Redundant gesture tasks are where the same meaning is portrayed in both speech and gesture and thus gesture and speech integration is not required to determine the full meaning of the message, e.g. 'brush your teeth' said verbally combined with stereotypical tooth brushing gesture. Research which has investigated redundant gesture comprehension with participants with aphasia has found that the addition of redundant gesture increases the accuracy of comprehension (Yorkston *et al.* 1979).

One study that has investigated the impact of aphasia on gesture and speech integration is that by Records (1994). This study investigated whether the reliance on pointing gestures increased with verbal message ambiguity. The findings suggested that when verbal information is ambiguous, individuals with aphasia become more reliant on co-speech pointing gestures to determine the speaker's intention. Records (1994) makes an important contribution to our understanding of the impact of aphasia on gesture and speech integration. However, the findings are limited to pointing gestures. Investigating the use of iconic gestures by individuals with aphasia allows for assessment of more complex meanings, such as those communicated in verb phrases.

Imaging techniques such as functional magnetic resonance imaging (fMRI) have been used to investigate the neurological basis of iconic co-speech gesture and speech integration and implicate an important role for Broca's area (Willems *et al.* 2007). This would suggest that if Broca's area is damaged, as is the case in some aphasias, an individual may have difficulty with iconic co-speech gesture and speech integration.

Integration is more than the sum of the two parts. When integration occurs, the certainty in decoding the message from multimodal input is higher than the certainty derived from separate considerations of each modality. We refer to such an increase as 'multimodal gain'. Such a gain occurs when two modalities mutually enhance their informativeness, in other words when there is a synergy effect of considering two modalities together while decoding (Kelly *et al.* 1999). For a more detailed explanation of the calculation of multi-modal gain, see the data analysis section below.

The current study explored co-speech iconic gesture comprehension in a novel methodology, in one participant with Broca's aphasia with impaired comprehension (SR) and 20 control participants. The assessment tool developed for this project was used to determine the success of the participant at iconic co-speech gesture and speech integration and their gesture and speech comprehension independently. An error analysis was used to indicate whether participants relied on either gesture or speech in the integration condition.

Methods

Case information

The patient was SR, a right-handed English-speaking male aged 75 years who presented with a dense right-sided paresis of both the upper and lower limbs. He experienced a left middle cerebral artery cerebral vascular accident (CVA) when he was 69 years of age. Unfortunately, computed tomographic (CT) scan results indicating exact location of damage were not available. Before his CVA, SR worked as an electrical engineer.

SR was assessed on the Western Aphasia Battery (WAB) (Kertesz 1982) and obtained the following scores: aphasia quotient=24.9; fluency=1.00; comprehension=5.85; repetition=2.4; and naming=1.2. These scores indicate a classification of severe Broca's aphasia with impaired expression and verbal comprehension (Kertesz 1982).

Control participants

SR's performance was compared with 20 English-speaking control participants (eleven female, 17 right handed) aged 60–79 years (mean=68.6 years, standard deviation (SD)=5.71 years). Control participants were recruited from a range of community groups, e.g. churches. Control participants had no history of severe head trauma, stroke or progressive neurological disease. Two participants wore glasses and reported that they were able to see the screen clearly with their glasses. They wore their glasses throughout the experiment. Four participants reported very mild hearing loss but indicated in the trial items that they were able to hear the verbal stimulus clearly.

Creation of stimuli

An actor, whose face was covered to conceal the lip movements, produced 21 combinations of an iconic gesture and a short sentence. They expressed common everyday actions (e.g., writing, driving, cutting). From the recording of each speech–gesture combination, three versions of vignettes (total 63) were created by video editing software: verbal plus gesture (VG) (the original video recording), gesture only (G) (speech muted), and verbal only (V) (video replaced by a still picture of the actor).

Each speech–gesture combination (e.g., 'they paid' with a gesture depicting somebody writing) had corresponding four colour photographs as choices in the response booklet: (1) integration target (paying with a cheque), (2) verbal-only match (paying with cash), (3) gesture-only match (writing a letter), and (4) unrelated foil (reading a book). Both the integration target and the verbal-only match were semantically congruent with the speech and therefore were both equally likely to be selected by the control participants in the V condition. This meant that in the V condition only, both the integration target and the verbal-only match were correct. In the other two conditions (VG and G) the verbal-only match was incorrect. Similarly in the G condition, both the integration target and the gesture-only match were semantically congruent with the gesture and therefore were both equally likely to be selected by the control participants. In the other conditions (V and VG) the gesture-only match was incorrect. However, it was the VG condition that was of greatest interest, because in this condition if participants integrated the information from the speech and the gesture, the integration target was the only congruent choice. It is the gain in integration target choice between the unimodal tasks and VG that was of most interest. The unrelated foil was created by combining semantic associates of elements of the gesture-only match (e.g., reading for writing, a book for a letter), but it was not congruent with the speech or the gesture. The photographs were arranged in A4 response booklets in such a way that four choice photographs could be seen simultaneously and that the positions of the four choices on the page varied. All photographs contained an individual or a relevant body part, e.g. a hand carrying out an action.

Stimulus presentation

Each participant saw the 21 speech–gesture combinations in all three conditions (total 63 stimuli) in a semi-randomized order.[1] The 21 speech–gesture combinations were split into three groups (seven in each). Each group had a different presentation order. These were as follows: V-VG-G, G-V-VG, and VG-G-V. Three counterbalancing sets were created so that each speech–gesture combination was presented in all three condition orders across the sets. Three response booklets corresponding to the three counterbalancing orders were created. All participants used all three response booklets resulting in a total of 63 trials.

Procedure

The 63 video vignettes embedded in a Power Point presentation were presented on a laptop with a 15.4-inch screen. The average duration that each vignette was shown for was 5 seconds. After each vignette, participants pointed to one of four colour photos in the response booklet in front of them that 'best matched the message portrayed in the video'. Participants were not instructed to attend to either gesture or speech or both, simply instructed to point to the picture that best matched the message. All participants were required to respond within 3 minutes of the presentation of the vignette. As there were two correct responses in the V and G conditions, some participants required this amount of time to determine which response they considered most correct. All participants had three practice trials. All responses and all errors were recorded.

Data analysis

It is possible to get the integration target in the VG condition by just understanding gesture or just understanding speech and not integrating (Kelly *et al.* 1999). In order to determine the probability of getting the correct answer without integrating, it is necessary to estimate the relative contributions of speech and gesture in the participants' decision in the VG condition. We assume that the participants rely more on the stronger (more intact) modality in the decision. Thus, we calculate the relative strengths of the modalities as the relative proportions of the probabilities of getting the matched choice (i.e. the integration target or the matching response) in the V and the G conditions. The relative strengths are our estimates of percentage contributions of speech and gesture to the participants' decision in the VG condition, as in the following formulae:

Percentage contribution of speech (PCS) = [%_matched_choice_in_V/(%_matched_choice_in_V+%_matched_choice_in_G)]

Percentage contribution of gesture (PCG) = [%_matched_choice_in_G/(%_matched_choice_in_V+%_matched_choice_in_G)]

Therefore, if an individual is estimated to be more reliant on one modality (e.g., gesture, as may be the case in aphasia), given the relative strengths of the modalities, then this modality will be given a higher percentage contribution. In order to determine the probability of getting the integration target without integration we used the percentage contributions in the following formula:

> Probability of getting the integration target in VG without_integration = [(PCS * percentage integration target_choice_in_V) + (PCG * percentage integration target_choice_in_G)]

However, it is the integration score that was of most interest. If the percentage of integration targets chosen is higher than the probability of choosing the integration target without integration then multi-modal gain (MMG) has occurred. The gain stems from the fact that two modalities can mutually enhance their informativeness in the decoding process. The multi-modal gain score indicates how much gain the individual obtained by integrating the information from gesture and from speech. The formula we used was as follows:

> MMG = percentage integration target_choice_in_VG – probability of getting the integration target in VG without_integration

Results

Only one control participant did not choose the integration target in the verbal gesture (VG) condition more than the other conditions. This participant only chose the integration target in VG 38% of occasions. This was more than 2 SDs below the mean and was therefore considered an outlier. This participant was removed from the data for all further analyses. With the removal of this participant, the control participants chose significantly more integration targets in the VG condition than the other two conditions ($t(36) = 11.25$, $p < 0.05$; $t(36) = 6.57$, $p < 0.05$) (figure 1). We then assessed whether SR integrated information from speech and gesture to the

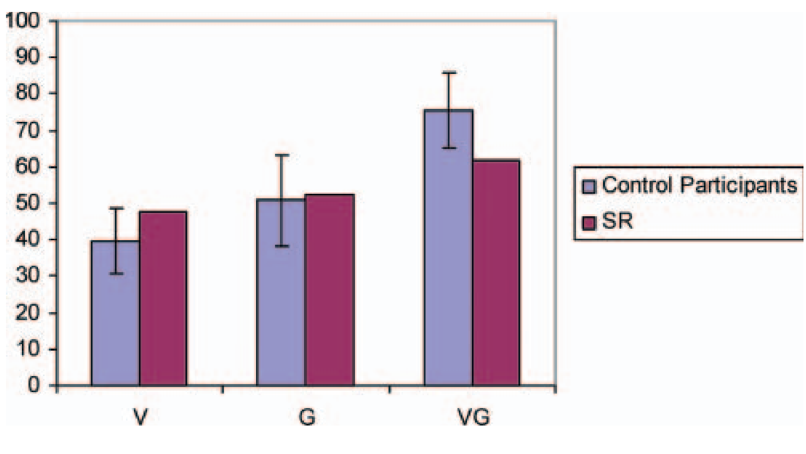

Figure 1. Mean percentage of integration targets chosen in each of the three conditions (V = verbal only, G = gesture only, and VG = verbal and gesture). Note that in VG the integration targets are the only correct choice. In G, both the integration targets and the gesture-only match were correct. In V, both the integration targets and verbal-only match are correct.

same degree to the control participants. To this end, we calculated multi-modal gain (MMG) scores for each participant, according to the formulae described in the method section. SR's MMG score (11.6%) was significantly lower than the control group (mean=30%, SD=11.13) (z=–1.66, $p<0.05$).

Error analysis

Verbal-only condition (V)

A more detailed error analysis of V confirmed initial findings on the WAB, that SR had difficulty comprehending the verbal message. In the verbal only condition, both the integration targets and the verbal-only matches are correctly matched choices (i.e., the sentence 'he paid' matches both paying by cheque and paying by cash, but it does not match writing a letter). The mean percentage of correctly matched choices and the incorrectly matched choices (chosen by SR and the control participants is presented in figure 2. As predicted, the majority of the control participants selected the correctly matched choices, either the verbal-only match or the integration target. SR however, selected the incorrectly matched choices (gesture-only matches (GM) or unrelated foils (UF)) on significantly more occasions (28.5%) than the control participants (mean=2.25%, SD=3.32) (z=7.91, $p<0.05$) further confirming he had difficulty with understanding the verbal message.

Gesture-only condition (G)

Analysis of the G condition suggested that SR had preserved gesture comprehension. The mean percentage of integration targets (IT), verbal-only matches (VM), gesture-only matches (GM) and unrelated foils (UF) chosen by the control participants and SR is presented in figure 3. In this condition the correctly matched choices were the integration target and the gesture-only match. As predicted, the control participants and SR selected either the integration target or the gesture-only

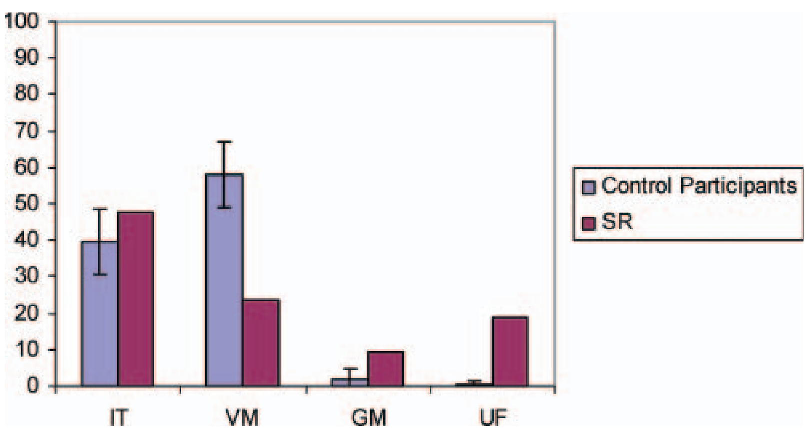

Figure 2. Mean percentage of correctly matched responses (IT=integration target, VM=verbal-only match), and the incorrectly matched responses (GM=gesture-only match) and UF (unrelated foil) chosen by the control participants and SR in V (verbal-only condition). Both the integration target and the verbal-only match were congruent with the stimulus.

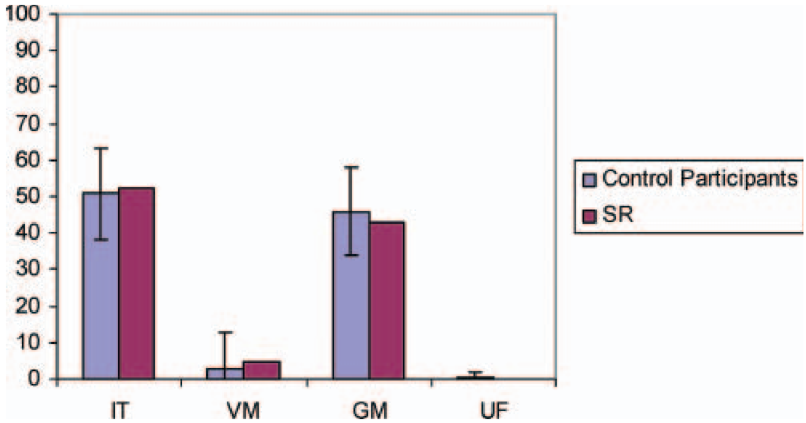

Figure 3. Mean percentage of IT (integration target), VM (verbal-only match), GM (gesture-only match) and UF (unrelated foil) chosen by the control participants and SR in G (gesture only condition). Both the integration target and the gesture-only match were congruent with the stimulus.

match on nearly all occasions. There was no significant difference between the percentage of times that SR selected the UF or the VM (4.76%) compared with the control participants (mean=3.5%, SD=4.15) (z=0.304, p>0.05).

Verbal and gesture condition (VG)

Similar to V, error analysis of VG also indicated differences between the control participants and SR. The mean percentage of integration targets (IT), verbal-only matches (VM), gesture-only matches (GM) and unrelated foils (UF) chosen by SR and the control participants is presented in figure 4. It is important to note that in this condition, only the integration target counts as a correct response. The errors

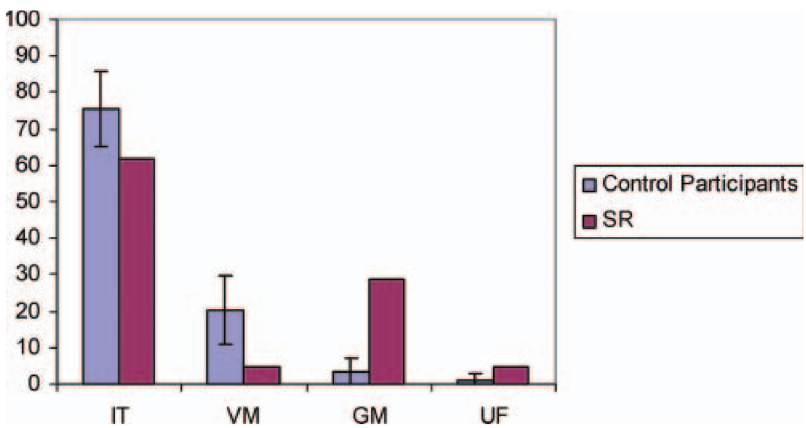

Figure 4. Mean percentage of integration targets (IT), verbal-only matches (VM), gesture-only matches (GM), and unrelated foils (UF) chosen by the control participants and SR in VG (the verbal and gesture condition). Note that in this condition only the integration targets are the correctly matching choice.

made by SR were mostly selection of the GM, whereas the errors made by the control participants were mostly VM. The percentage of times that GM was chosen by SR (28.6%) was significantly higher than the control participants (mean=3.5%, SD=3.5%) (z=7.17, p<0.05). The difference between the percentage of times that VM was chosen by SR (4.76%) and VM was chosen by the control participants (mean=20.3%, SD=9.5) (z=−1.64, p=0.05), was very close to significant. This indicates the high dependence of SR on gesture to decode messages. This differs to the control participants who relied more on speech.

Discussion

This study investigated iconic gesture and speech integration in one participant with Broca's aphasia (SR) and 20 control participants. One control participant was removed from the sample as they chose the integration target on only 38% of occasions.[2] SR obtained a significantly lower multi-modal gain score than the control participants, indicating that SR had an impaired ability to integrate information from iconic gesture and speech. A more detailed error analysis indicated that SR processed the information in the trials in a different way to the control participants. SR relied more on gesture when gesture and speech integration was required, whereas the control participants relied more on verbal input. Further error analysis of V and G conditions suggested that this may be because SR had impaired verbal comprehension but intact gesture comprehension.

The current study adds support to the findings of Records (1994) that aphasia can impact on gesture and speech integration. It is also adds further support to the finding that individuals with low comprehension abilities due to Broca's aphasia may rely more heavily on gesture to decode messages when gesture and speech are combined. Individuals with intact comprehension however, rely more heavily on verbal information. While Records (1994) found that this was the case for pointing gestures, the current study provides evidence that this is also the case for co-speech iconic gesture. Furthermore, the finding from the current study may provide support to the findings of Willems *et al.* 's (2007), that Broca's area is implicated in speech–gesture integration. However, it should be noted that the lack of CT scan limits us to draw a firm conclusion about functional localization, though the profile of SR's deficits strongly suggests a lesion in Broca's area.

While the findings of this study are limited, as they are based on just one participant with aphasia and a methodology which allowed for analysis of just one type of linguistic phrase and gesture type, the results imply that there is a need for further research in this area. Our future work will study the impact of aphasia on gesture and speech integration in a larger group of individuals to ascertain whether the findings of this study can be generalized. Furthermore, we expect that this research will contribute to the development of assessments that can be used by speech and language therapists to determine whether gestures will facilitate or hinder an individual client's comprehension.

Acknowledgements

The authors would like to thank Anna Caute and Lucy Dipper for helpful comments on an earlier draft of this paper, to Razvanah Shah for helping with some of the data collection, and all the participants who took part in the study. This study was funded by a grant from the City University Pump-Priming Scheme. ***Declaration of interest:*** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

1. More information about the exact stimulus can be obtained by contacting the first author on request.
2. It is not clear why this participant could not integrate, but given the participant's age, it could be the impact of an undiagnosed neurological condition.

References

Beattie, G. W. and Shovelton, H., 1999, Mapping the range of information contained in the iconic gestures that accompany speech. *Journal of Language and Social Psychology*, **18**, 438–462.

Daniloff, J. K., Fritelli, G., Buckingham, H. W., Hoffman, P. R. and Danilogg, R. G., 1986, Amer-Ind versus ASL: recognition and imitation in aphasic subjects. *Brain and Language*, **28**, 95–113.

Daniloff, J. K., Noll, J. D., Fristoe, M. and Lloyd, L. L., 1982, Gesture recognition in patients with aphasia. *Journal of Speech and Hearing Disorders*, **47**, 43–49.

Duffy, R. J. and Duffy, J. R., 1981, Three studies of deficits in pantomimic expression and pantomimic recognition in aphasia. *Journal of Speech and Hearing Research*, **46**, 70–84.

Kelly, S. D., Barr, D. J., Church, R. B. and Lynch, K., 1999, Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, **40**, 577–592.

Kertesz, A., 1982, *Western Aphasia Battery* (New York, NY: Grune & Stratton).

Kita, S. and Özyürek, A., 2003, What does cross-linguistic variation in semantic co-ordination of speech and gesture reveal? Evidence of an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, **48**, 16–32.

Lambier, J. D. and Bradley, D., 1991, The effect of physical similarity on pantomime recognition in aphasia. *Aphasiology*, **5**, 23–37.

McNeill, D., 2000, *Language and Gesture* (Cambridge: Cambridge University Press).

Records, N., 1994, A measure of the contribution of a gesture to the perception of speech in listeners with aphasia. *Journal of Speech and Hearing Research*, **37**, 1086–1099.

Thorburn, L., Newhoff, M. and Rubin, S. S., 1995, Ability of subjects with aphasia to visually analyse written language, pantomime, and iconographic symbols. *American Journal of Speech–Language Pathology*, **4**, 174–179.

Varney, N. R., 1982, Qualitative aspects of pantomime recognition defect in aphasia. *Brain and Cognition*, **1**, 132–139.

Willems, R. M., Özyürek, A. and Hagoort, P., 2007, When language meets action: the neural integration of gesture and speech. *Cerebral Cortex*, **17**, 2322–2333.

Yorkston, K. M., Beukelman, D. R. and Waugh, P. F., 1979, A comparison of verbal, pantomime and combined instruction modes with severely aphasic individuals. Paper presented at the Clinical Aphasiology Conference.