Giving Speech a Hand: Gesture Modulates Activity in Auditory Cortex During Speech Perception

Amy L. Hubbard,[1,2,3]* Stephen M. Wilson,[1] Daniel E. Callan,[3,4] and Mirella Dapretto[1,5]

[1]*Ahmanson-Lovelace Brain Mapping Center, University of California, Los Angeles, California*
[2]*Department of Applied Linguistics, University of California, Los Angeles, California*
[3]*Computational Neuroscience Laboratories, ATR Kyoto, Japan*
[4]*National Institute of Information and Communications Technology, ATR Kyoto, Japan*
[5]*Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, California*

Abstract: Viewing hand gestures during face-to-face communication affects speech perception and comprehension. Despite the visible role played by gesture in social interactions, relatively little is known about how the brain integrates hand gestures with co-occurring speech. Here we used functional magnetic resonance imaging (fMRI) and an ecologically valid paradigm to investigate how beat gesture—a fundamental type of hand gesture that marks speech prosody—might impact speech perception at the neural level. Subjects underwent fMRI while listening to spontaneously-produced speech accompanied by beat gesture, nonsense hand movement, or a still body; as additional control conditions, subjects also viewed beat gesture, nonsense hand movement, or a still body all presented without speech. Validating behavioral evidence that gesture affects speech perception, bilateral nonprimary auditory cortex showed greater activity when speech was accompanied by beat gesture than when speech was presented alone. Further, the left superior temporal gyrus/sulcus showed stronger activity when speech was accompanied by beat gesture than when speech was accompanied by nonsense hand movement. Finally, the right planum temporale was identified as a putative multisensory integration site for beat gesture and speech (i.e., here activity in response to speech accompanied by beat gesture was greater than the summed responses to speech alone and beat gesture alone), indicating that this area may be pivotally involved in synthesizing the rhythmic aspects of both speech and gesture. Taken together, these findings suggest a common neural substrate for processing speech and gesture, likely reflecting their joint communicative role in social interactions. *Hum Brain Mapp 30:1028–1037, 2009.* ©2008 Wiley-Liss, Inc.

Key words: gestures; speech perception; auditory cortex; magnetic resonance imaging; nonverbal communication

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsors: Foundation for Psychocultural Research-UCLA Center for Culture, Brain, and Development, ATR International, NSF EAPSI Program and an NRSA Predoctoral Fellowship (F31 DC008762-01A1) awarded to Amy L. Hubbard; Contract grant sponsors: Brain Mapping Medical Research Organization, Brain Mapping Support Foundation, Pierson-Lovelace Foundation, The Ahmanson Foundation, William M. and Linda R. Dietel Philanthropic Fund at the Northern Piedmont Community Foundation, Tamkin Foundation, Jennifer Jones-Simon Foundation, Capital Group Companies Charitable Foundation, Robson Family and Northstar Fund; Contract grant sponsors: National Center for Research Resources (NCRR), NIH (the contents of the project described are solely the responsibility of the authors and do not necessarily represent the official views of NCR or NIH); Contract grant numbers: RR12169, RR13642, RR00865.

*Correspondence to: Amy L. Hubbard, UCLA Department of Applied Linguistics, Ahmanson-Lovelace Brain Mapping Center, 660 Charles E. Young Drive South, Los Angeles, CA 90095-7085.
E-mail: ahubbard@humnet.ucla.edu

Received for publication 26 September 2007; Revised 20 January 2008; Accepted 15 February 2008
DOI: 10.1002/hbm.20565
Published online 15 April 2008 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Successful social communication involves the integration of simultaneous input from multiple sensory modalities. In addition to speech, features such as tone of voice, facial expression, body posture, and gesture all contribute to the perception of meaning in face-to-face interactions. Hand gestures, for example, can alter the interpretation of speech, disambiguate speech, increase comprehension and memory, and convey information not delivered by speech [e.g., Cook et al., 2007; Goldin-Meadow and Singer, 2003; Kelly et al., 1999; Kendon, 1972; McNeill et al., 1992, 1994]. Despite the visible role of gesture in everyday social communication, relatively little is known about how the brain processes natural speech accompanied by gesture.

Studies examining the neural correlates of co-occurring gesture and speech have focused almost entirely on iconic gestures (i.e., hand movements portraying an object or activity). For instance, Holle et al. [2007] showed that viewing iconic gestures (as compared to viewing self-grooming movements) led to increased activity in STS, inferior parietal lobule, and precentral sulcus. Willems et al. [2006] observed similar activity in Broca's area in response to both word–word and word–gesture mismatches. Using transcranial magnetic stimulation, Gentilucci et al. [2006] also demonstrated Broca's area involvement in iconic gesture processing. Further, studies using event related potentials suggest that iconic gestures engage semantic processes similar to those evoked by pictures and words [Wu and Coulson, 2005], that iconic gestures are integrated with speech during language processing [Kelly et al., 2004; Özyürek et al., 2007], and that integration of gesture and speech is impacted by the meaningfulness of gesture [Holle and Gunter, 2007].

The integration of auditory and visual cues during speech has been studied more extensively in the context of "visual speech" (i.e., the speech-producing movements of the lips, mouth, and tongue). Behavioral effects related to visual speech are similar to those observed for gesture, as concordant visual speech can aid speech perception [Sumby and Pollack, 1954] whereas discordant visual speech can alter auditory perception [McGurk and MacDonald, 1976]. Neuroimaging studies have shown that listening to speech accompanied by concordant visual speech yields greater hemodynamic activity in auditory cortices than listening to speech alone [e.g., Calvert et al., 1999, 2003]. In addition, multisensory integration of auditory and visual speech has been observed in the left planum temporale (PT), left superior temporal gyrus and sulcus (STG/S), and left middle temporal gyrus [MTG; Callan et al., 2001, 2003, 2004; Campbell et al., 2001; Calvert et al., 2000; Pekkola et al., 2006].

Here we used functional magnetic resonance imaging (fMRI) paired with an ecologically valid paradigm in order to investigate how speech perception might be affected by rhythmic gesture which accompanies speech. Descriptions of gestures which match the cadence of speech stem from as long ago as 60 A.D. [Quintilian, 1856]; these gestures have since been dubbed "batons," "beats," and "beat gesture" (i.e., rapid movements of the hands which provide "temporal highlighting" to the accompanying speech; McNeill, 1992]. Beat gesture has been shown to impact the perception and production of speech prosody [i.e., the rhythm and intonation of speech; Krahmer and Swerts, 2007], as well as to establish context in narrative discourse [McNeill, 1992]. To the extent that no prior study has focused on the neural correlates of beat gesture and that both visual speech and gesture (1) affect speech comprehension and (2) involve biological motion, we hypothesized that they might be subserved by overlapping neural substrates. Accordingly, we focused on superior temporal cortices as our a priori regions of interest.

MATERIALS AND METHODS

Subjects

Thirteen adult subjects (3 females; 27.51 $\pm$ 7.10 years of age) were recruited at Advanced Telecommunications Research Institute in Kyoto from a cohort of international visitors. All subjects were healthy, right-handed, native English speakers who neither spoke nor understood American Sign Language.

Stimulus Material

All video segments comprising the stimuli were culled from 2 h of spontaneous speech recorded in a naturalistic setting (i.e., the kitchen of a house). The recording featured a female native speaker of North American English who was naïve to the purpose of the recording. A set of questions relevant to the speaker's life and experiences was prepared prior to the recording. During the recording, the speaker was asked to stand in the kitchen and answer questions posed to her by the experimenter in the adjacent room. Great care was taken to remove speech articulators and other indices of fundamental frequency in an uncontrived, ecologically-valid manner. The illusion of a cupboard occluding the speaker's face was created by affixing a piece of plywood (stained to match the wood in the kitchen) to the wall above the stove. The recording was produced using a Sony DCR-HC21 Mini DV Handycam Camcorder secured on a tripod and tilted downward so that only the speaker's lower neck, torso area, and upper legs were visible. The speaker moved freely and expressed herself in a natural, conversational style throughout the recording. Importantly, although her head was behind the plywood board, her gaze was free to shift from the board directly in front of her to the observer sitting on the couch in the adjacent room. Following the spontaneous speech recording, 12 picture sequences were affixed to the plywood board in front of the speaker's face. The pictures depicted movements that represent words in American Sign Language (ASL) but which lack obvious iconic mean-

ing to nonsigners. The speaker, who neither spoke nor understood ASL, produced each set of movements one time. There were no words written on the pictures, and the speaker did not talk while producing the hand movements. Finally, the speaker was recorded as she stood motionless.

Videos were captured with a Sony Mini DV GV-D900 and imported using Macintosh OSX and iMovie. Final Cut Pro HD 4.5 was used to cut and export 24 18-s segments of speech with beat gesture to. avi movie files. Since the 24 segments were selected from 2 h of free-flowing speech with gesture, inclusion or exclusion of gesture type could be controlled by cropping. That is, it was possible to eliminate movements that communicated consistent semantic information in the absence of speech by beginning an 18-s segment after that gesture had occurred. As the benefits of segregating gesture into strict categories has recently come under scrutiny [McNeill, 2005], in order to maintain ecological validity, beat gesture (i.e., rhythmic gesture) was not limited to flicks of the hand for the purposes of this study. The stimuli segments contained both beat gesture (strictly defined) as well as rhythmic movements possessing minimal iconicity and metaphoricity. All three types of beat gesture described in Li et al. [2003]–beats with and without poststroke holds and movement to a different gesture space for subsequent beat gesture–occurred in our stimuli. Tuite [1993] and Kendon [1972] describe relationships between gestural and speech rhythm, but methods for studying this complex relationship remain elusive. Shattuck-Hufnagel et al. 2007 and Yasinnick [2004] are among the first to attempt to develop systematic, quantitative methods for investigating speech and gesture timing. Their ongoing work [Shattuck-Hufnagel et al., 2007] seeks to represent the relationship between pitch accents and corresponding gestural events.

In the absence of an established method for determining the direct relationship between speech and gesture timing in free-flowing speech, we attained 18-s segments of rhythmic gesture and speech by removing highly iconic gestures. A group of eight viewers (who were not subjects in the study) reported that semantic information could not be discerned by viewing the 24 video segments in the absence of speech. Additionally, one 18-s segment with a still frame of the speaker's body and 12 segments of ASL-based movements, consisting of 65 different signs, were selected. The selected ASL movements were noniconic, and a group of eight viewers (who did not participate in the study) confirmed that the movements did not elicit semantic information. The 24 segments of beat gesture and speech were used in the beat gesture with speech condition (as originally recorded) and in the beat gesture without speech condition (where the audio was removed; Fig. 1). The 12 ASL-based segments were used in the nonsense hand movement without speech condition (as originally recorded) and in the nonsense hand movement with speech condition (where they were paired with speech from the former 24 segments that were originally accompanied by beat gesture). Finally, the motionless recording of the speaker was used in the still frame without speech condition, used as baseline, and in the still frame with speech condition (where they were paired with speech from the 24 segments originally accompanied by beat gesture). One 18-s segment was shown per block, thus blocks were18 s long, with a 3-s cream-colored screen separating segments. Samples of these video clips are available online as supplemental materials (available at www.interscience.-wiley.com/jpages/1065-9471/suppmat.)

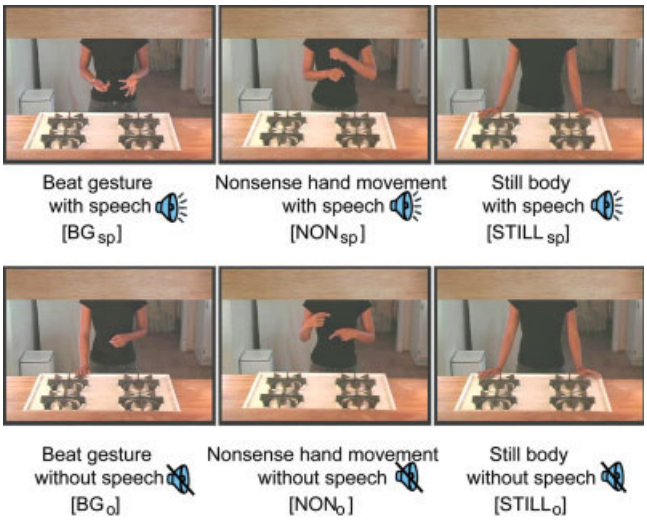

Figure 1.

Experimental paradigm. There were six conditions, obtained by crossing movement type (beat gesture, nonsense hand movement, still frame) by speech (present or absent). In the actual experiment, blocks were presented in pseudorandom orders counterbalanced across subjects.

The RMS energy of the audio segments was adjusted to be identical across stimuli. To prevent specific item effects (in terms of speech content), stimuli were counter-balanced across subjects such that one subject might hear and see segment no. 1 with the original beat gesture and speech, another subject might hear the speech of segment no. 1 while viewing one of the segments of nonsense hand movement, and yet another subject might hear the speech of segment no. 1 while viewing the still frame. For each subject, any part (speech and/or body movements) of the original 24 beat gesture segments and 12 nonsense hand movement segments occurred exactly one time over the two sessions. The order of presentation of the video segments was randomized subject to the constraints that there would be no serial occurrence of: (i) two identical conditions, (ii) three segments with speech, or (iii) three segments without speech. Each subject viewed a different randomization of the video sequences.

After the fMRI scan, subjects were given a short test with three multiple-choice questions based on the audio content of the final three audiovisual segments appearing

in the subject's fMRI session. Since stimuli were randomized and counter-balanced, each subject received a different set of test questions. This test was intended to promote attention during the passive fMRI task as subjects were informed about a post-scan test at the beginning of the fMRI session. The average accuracy for the 13 subjects was 90% correct.

Experimental Procedures

Prior to the fMRI scan, subjects received a short introduction to the task. They were shown a still picture of the video and told that the speaker, whose head was blocked by a cupboard in the kitchen, was talking to a person in the adjacent room. They were told to keep their eyes fixated on the speaker's torso at all times, even throughout the silent segments. Subjects were advised to pay attention during the entire scan because they would be given a post-scan test on what they saw and heard.

Subjects lay supine in the scanner bed while undergoing two consecutive fMRI scans; in each of these 6 min and 30 s scans, each condition occurred three times. Visual and auditory stimuli were presented to subjects using a magnet-compatible projection system and headphones under computer control. Subjects viewed visual stimuli via a Victor, Japan projector. The audiovisual stimuli were presented using full view in Real Player in order to ensure that subjects saw no words, numbers or time bars while viewing the stimuli. The images were projected first to a 150 × 200 mm screen placed just behind the subject's head. Subjects viewed a reflection of the 110 × 150 images (screen resolution 1024 × 768) via a small mirror adjusted to eye level. Auditory stimuli were presented using a Hitachi Advanced head set.

Images were acquired at Advanced Telecommunications Research Institute in Kyoto, Japan using a Shimadzu 1.5 whole body scanner. A 2D spin-echo image (TR = 300 ms, TE = 12.1 ms, matrix size 384 × 512, 5-mm thick, 5-mm gap) was acquired in the sagittal plane to allow prescription for the remaining scans. For each participant, a structural T2-weighted fast spin echo imaging volume (spin-echo, TR = 5468 ms, TE = 80 ms, matrix size 256 × 256, FOV = 224 cm, 30 slices, 0.875-mm in-plane resolution, 5-mm thick) was acquired coplanar with the functional scans to allow for spatial registration of each subject's data into a common space. The functional data were acquired during two whole-brain scans, each lasting 6 min and 30 s (264 images, EPI gradient-echo, TR = 3000 ms, TE = 49, flip angle = 90°, matrix size = 64 × 64, 3.5 mm in-plane resolution, 5 mm thick, 0 mm gap).

Data Analysis

Following image conversion, the functional data were analyzed using Statistical Parametric Mapping 2 (SPM2; http://www.fil.ion.ucl.ac.uk/spm/software/spm2/). Functional images for each participant were realigned to correct for head motion, normalized into MNI space [Collins et al., 1994; Mazziotta et al., 2001] and smoothed with a 7 mm Gaussian kernel. For each subject, condition effects were estimated according to the General Linear Model using a 6-s delay boxcar reference function. The still frame condition was implicitly modeled as baseline. The resulting contrast images were entered into second level analyses using random effect models to allow for inferences to be made at the population level [Friston et al., 1999]. Group activation maps were thresholded at $P < 0.01$ for magnitude, with whole-volume correction for multiple comparisons applied at the cluster level ($P < 0.05$). The SPM2 toolbox MarsBaR [Brett et al., 2002] was used to extract parameter estimates for each participant from regions of interest. Small volume correction was applied for selected contrasts of interest based upon previous research identifying superior temporal cortices (i.e., PT and STG/S) as areas of increased activity while viewing visual speech during speech perception and as putative sites of multisensory integration. For the contrast of speech with beat gesture versus speech with nonsense hand movement (Fig. 2b), small volume correction was based on a 14,000 mm^3 volume, a conservative estimate according to measurements of the auditory belt and parabelt regions reported in Sweet et al., 2005. This volume was defined by a sphere of 15 mm radius and centered at the functional maxima (x = −57, y = −12, z = 8). For the contrast of bimodal (beat gesture and speech) versus unimodal (still body and speech, beat gesture only) (Fig. 2c), small volume correction was based upon a 4,100 mm^3 volume defined by a sphere of 10 mm radius centered at the functional maxima (x = 57, y = −27, z = 8; identified as planum temporale per anatomical maps reported in a previous structural MRI study, Westbury et al., 1999)]. Cluster size and coordinates for peaks of activity for all contrasts of interest are presented in Supplementary Table I.

RESULTS

A direct comparison between speech accompanied by beat gesture versus speech accompanied by a still body (statistical activation map in Fig. 2a; Supplementary Table I) revealed greater activity in bilateral PT and posterior STG, two areas known to underlie both speech perception and the processing of biological motion. Greater activity for this contrast was also observed in visual cortices, associated with sensory processing, as well as in bilateral premotor and left parietal regions, perhaps reflecting "mirror neuron system" activity associated with the perception of meaningful actions [Rizzolatti and Craighero, 2004]. As compared to baseline (i.e., viewing a still body without speech), viewing speech accompanied by beat gesture (blue outlines in Figs. 2a and 3b; Supplementary Table I) led to increased activity in bilateral visual cortices (including visual motion area MT), primary auditory cortices, STG/S, MTG, inferior frontal gyrus, middle frontal

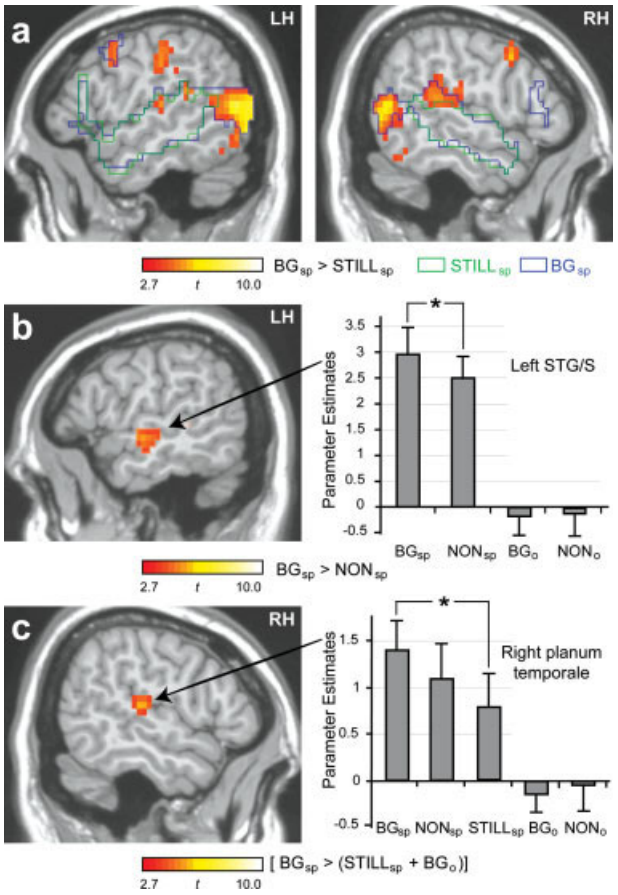

Figure 2.

Neural activity related to processing speech and speech accompanied by beat gesture. (**a**) Clusters depict areas of greater activity while listening to speech accompanied by beat gesture as compared to listening to speech accompanied by a still body. Areas of stronger activity for listening to speech while viewing a still body and for listening to speech while viewing beat gesture as compared to baseline (i.e., still body only) are shown in green and blue outlines, respectively. Specific contrasts are depicted using the abbreviated condition names defined in Figure 1. Group activation maps were thresholded at $P < 0.01$ for magnitude, with correction for multiple comparisons at the cluster level ($P < 0.05$). (**b**) Greater activity was observed in left STG/S while listening to speech accompanied by beat gesture as compared to speech accompanied by nonsense hand movement (maxima located at −57, −12, −8, MNI coordinates, $t = 4.23$, small volume corrected). Parameter estimates within this region for both conditions (relative to still frame, no speech) are shown in the accompanying graph. (**c**) Superadditive responses were observed in the right PT for the bimodal presentation of beat gesture and speech (maxima located at 57, −27, 8, MNI coordinates, $t = 4.68$, small volume corrected). Parameter estimates within this region for each condition (relative to still body, no speech) are shown in the accompanying graph. Error bars equal standard error of the mean. RH, right hemisphere; LH, left hemisphere.

gyrus, postcentral gyrus, and superior colliculi. Viewing speech accompanied by a still body (as compared to baseline) led to increased activity in several overlapping areas (green outlines in Fig. 2a; Supplementary Table I), such as bilateral STG/S, MTG, and IFG. Viewing beat gesture without speech as compared to baseline (Fig. 3a; Supplementary Table I) yielded significant increases in bilateral occipito-temporal areas including MT, right postcentral gyrus and intraparietal sulcus, and right posterior MTG and STG/S.

As compared to baseline, viewing speech accompanied by nonsense hand movement (Fig. 3d; Supplementary Table I) led to increased activity throughout visual and temporal cortex bilaterally as well as in bilateral postcentral gyrus, intraparietal sulcus, superior colliculi, and left middle frontal gyrus. Viewing nonsense hand movement without speech as compared to baseline (Fig. 3c; Supplementary Table I) yielded significant increases in bilateral occipito-temporal areas including MT, postcentral gyrus, intraparietal sulcus, superior and middle frontal gyrus as well as right posterior MTG and STS and right cerebellum.

To identify regions where increased activity might specifically reflect the integration of beat gesture and speech, we directly compared neural responses to speech accompanied by beat gesture versus speech accompanied by nonsense hand movement. Notably, this contrast revealed significant activity in left STG/S (Fig. 2b; Supplementary Table I), indicating that beat gesture, just as visual speech, modulates activity in left nonprimary auditory cortices during speech perception. For the inverse contrast—nonsense hand movement with speech versus beat gesture with speech—left cerebellum, postcentral gyrus, and intraparietal sulcus were significantly more active (Supplementary Table I).

To further examine regions where the presence of speech impacts gesture processing, we contrasted summed responses to unimodal conditions (still body with speech and beat gesture only) with responses to the bimodal condition (beat gesture with speech). Significantly greater responses to the bimodal presentation of beat gesture and speech (speech with beat gesture > speech with still body + beat gesture with no speech) were observed in right PT (Fig. 2c; Supplementary Table I). Parameter estimates for each condition in this contrast show that activity while silently viewing beat gesture was neither significantly below nor above baseline. Hence, right PT was recruited when beat gesture was presented in the context of speech, whereas in the absence of speech, gesture had no effect. No areas demonstrated superadditive properties for the combination of nonsense hand movement and speech (speech with nonsense hand movement > speech with still body + nonsense hand movement with no speech).

DISCUSSION

Few studies have attempted to characterize the brain's response to concurrently and spontaneously produced ges-

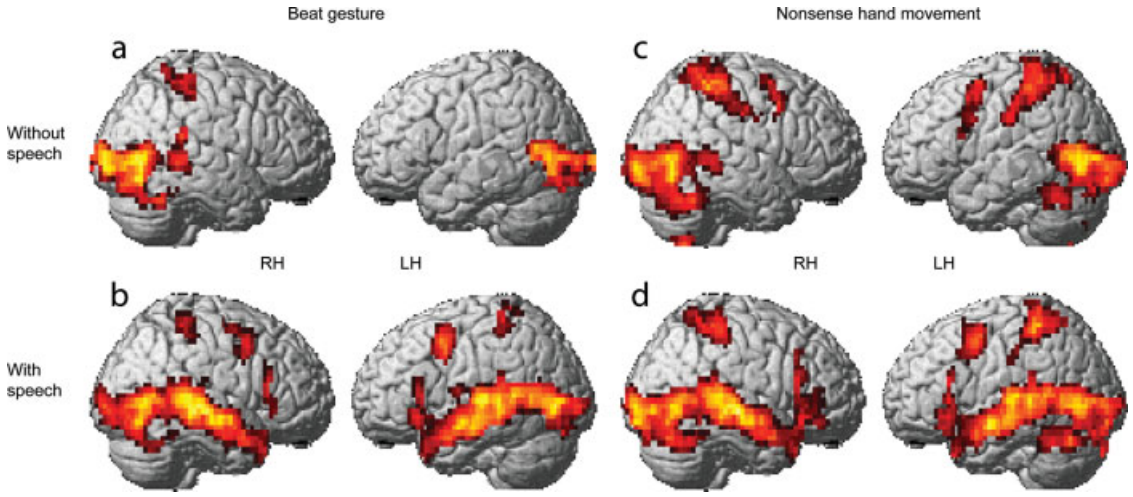

Figure 3.

Neural activity related to processing beat gesture and nonsense hand movements in the presence and absence of speech. Clusters depict areas of greater activity while (**a**) viewing beat gesture as compared to viewing a still body, (**b**) listening to speech accompanied by beat gesture as compared to listening to speech accompanied by a still body, (**c**) viewing nonsense hand movements as compared to viewing a still body, (**d**) listening to speech accompanied by nonsense hand movements as compared to listening to speech accompanied by a still body.

ture and speech. We hypothesized that neural responses to natural, rhythmic gesture accompanying speech would be observed not only in visual cortex but also in STG and STS, areas well-known for their role in speech processing. This hypothesis was guided by research on iconic gestures and deaf signers which indicate that STG/S plays a role in processing movement. Additional cues were provided by studies on visual speech showing STG/S to be crucially involved in audiovisual integration of speech with accompanying mouth movement. Supporting our hypothesis, bilateral posterior STG/S (including PT) responses were significantly greater when subjects listened to speech accompanied by beat gesture than when they listened to speech accompanied by a still body. Further, left anterior STG/S responses were significantly greater when listening to speech accompanied by beat gesture than when listening to speech accompanied by a control movement (i.e., nonsense hand movement). Finally, right posterior STG/S showed increased responses only to beat gesture presented in the context of speech, and not to beat gesture presented alone, suggesting a possible role in multisensory integration of gesture and speech. Related research in biological motion, deaf signers, visual speech, and iconic gesture highlight the importance of these current data.

As would be expected, canonical speech perception regions in STG/S showed increased bilateral activity while subjects heard speech accompanied by a still body or speech accompanied by beat gesture. Importantly, when directly comparing these two conditions, responses in the posterior portion of bilateral STG (including PT) were significantly greater when speech was accompanied by beat gesture. These data provide further support for STG/S as a polysensory area, as was originally suggested by studies in rhesus and macaque monkeys [Bruce et al., 1981; Desimone and Gross, 1979; Padberg et al., 2003]. Neuroimaging data has shown that STG/S–especially the posterior portion–is responsive to both visual and auditory input in humans. Studies in hearing and nonhearing signers strongly implicate the posterior temporal gyrus in language-related processing, regardless of whether the language input is auditory or visual in nature [MacSweeney et al., 2004, 2006]. Most recently, Holle et al. [2007] reported that the posterior portion of left STS showed increased activity for viewing iconic gestures as compared to viewing grooming-related hand movements. Wilson et al. [2007] found a greater degree of intersubject correlation in the right STS when subjects viewed an entire body (e.g., head, face, hands, and torso) producing natural speech as compared to when they heard speech alone. STG/S has also been shown to be more active while listening to speech accompanied by a moving mouth than while listening to speech accompanied by a still mouth [Calvert et al., 1997, 2003]. Interestingly, the stimuli in these studies may all be said to have communicative intent, suggesting that the degree of STG/S involvement may be mediated by the viewer's perception of the stimuli as potentially communicative. Such a characteristic of STG/S would be congruent with Kelly et al.'s [2006] finding that the central N400 effect (i.e., a response known to occur when incongruent stimuli are presented) can be eliminated when subjects know that gesture and speech are being produced by different speakers.

It is important to distinguish between the posterior portion of STG/S, and the STSp (posterior superior temporal sulcus). The latter is a much-discussed area in the study of biological motion [for review, see Blake and Shiffrar, 2006], as STSp has consistently shown increased activity for viewing point-light representations of biological motion [Grossman et al., 2004; Grossman and Blake, 2002]. Qualitative comparisons suggest that silently viewing beat gesture versus a still body, leads to increased activity in the vicinity of STSp (right hemisphere) as reported in Grossman et al. [2004]; Grossman and Blake [2002], and Bidet-Caulet et al. [2005]. Significant increases for speech-accompanied beat gesture over speech-accompanied still body, however, are anterodorsal to STSp. That is, speechless biological motion versus a still body yields significant increases in regions known to underlie processing of biological motion, but when accompanied by speech, biological motion versus a still body yields significant increases in an area more dorsal and anterior (to that identified by biological motion localizers). Once again, this suggests that the intent to participate in a communicative exchange (e.g., listening to speech) is a crucial determinant in how movement is processed. The idea that perception of gesture can be altered by the presence or absence of speech complements behavioral findings on gesture production, where it has been shown that the presence of speech impacts what is conveyed through gesture [So et al., 2005].

We would like to suggest that processing of movement may, in many cases, be context driven. Rather than processing speech-accompanying movement in canonical biological motion regions, perhaps movement is processed differently when it is interpreted (consciously or unconsciously) as having communicative intent. We are not the first to suggest that–in the case of language–the brain may not break down sensory input to its smallest tenets and then build meaning from those pieces. In a survey of speech perception studies, Indefrey and Cutler [2004] discovered that regions which are active while listening to single phonemes are not necessarily active while listening to a speech stream. Hence, it appears that the brain is not breaking the speech stream down into its component parts in order to extract meaning. Instead, the context in which the phonemes are received (e.g., words, sentences) determines neural activity. We are suggesting that this may be the case for biological motion as well—that biological motion with speech and without speech may be processed differently because of the contextualization afforded by speech.

Again when exploring activity within STG/S for the contrast of speech accompanied by beat gesture versus speech accompanied by a still body, it is notable that STG/S activity for this contrast includes PT bilaterally. Within this study, PT has emerged as a potentially critical site for the integration of beat gesture and speech. Contrasting responses to the co-presentation of speech and beat gesture with responses to unimodal presentation of speech (with a still body) and beat gesture (without speech), the right PT was identified as a putative site of gesture and speech integration (Fig. 2c).[1] In other words, in the right PT, beat gesture had no effect in the absence of speech. However, in the presence of speech, beat gesture resulted in a reliable signal increase in right PT.

Significant activity in bilateral PT (as well as inferior, middle, and superior temporal gyri) was observed by MacSweeney et al. [2004] while hearing nonsigners viewed blocks of British Sign Language and Tic Tac (a communicative code used by racecourse betters). We observed no activity in PT for either beat gesture or nonsense hand movements (which are based on ASL signs) when viewed without speech. MacSweeney et al. [2004], in addition to including a highly animated face in their stimuli, informed participants that the stimuli would be communicative and asked them to judge which strings of movements were incorrect. Thus, the participants had several cues indicating that they should search for meaning in the hand movements. In the current study, participants had no explicit instruction to assign meaning to the hand movement. Increased activity in PT was observed only when beat gesture was accompanied by speech and not when beat gesture was presented silently. Hence, it appears that PT activity, especially, is mediated by imbuing movement with the potential to convey meaning.

Considering what is known about PT activity, it is likely that beat gesture establishes meaning through its connection to speech prosody. PT has been shown to process meaningful prosodic and melodic input, as significantly greater activity has been observed in this area for producing or perceiving song melody versus speech [Callan et al., 2006; Saito et al., 2006] and for listening to speech with strong prosodic cues [Meyer et al., 2004]. Greater activity in PT has also been observed for listening to music with salient metrical rhythm [Chen et al., 2006], processing pitch modulations [Barrett and Hall, 2006; Warren et al., 2005], singing versus speaking, and synchronized production of song lyrics [Saito et al., 2006]. The observed right lateralization of multisensory responses for beat gesture and speech may be a further reflection of the link between speech prosody and beat gesture [Krahmer and Swerts, 2007]. Numerous fMRI, neurophysiological, and lesion studies have demonstrated a strong right hemisphere involvement in processing speech prosody [for review, see Kotz et al., 2006]. Along these lines, it has also been suggested that the right hemisphere is better suited for musical processing [Zatorre et al., 2002].

[1]The multisensory properties demonstrated by right PT were observed by utilizing a test for superadditivity. First described in single-cell studies, superadditivity is a property whereby neuronal responses to bimodal stimulus presentation are greater than the summed responses to unimodal stimulus presentations [Stein et al., 1993]. Although activity observed using the test for superadditivity may not reflect the same neuronal activity measured in the single multisensory integration cells which were originally identified with this approach [Stein et al., 1993; Laurienti et al., 2005], this test has been used by researchers of visual speech [Calvert et al., 2000; Campbell et al., 2001; Callan et al., 2003, 2004] to successfully identify areas involved in multisensory integration.

Our findings both confirm the role of PT in processing rhythmic aspects of speech and suggest that this region also plays a pivotal role in processing speech-accompanying gesture. This warrants future work to determine the degree to which PT responses may be modulated by temporal synchrony between beat gesture and speech. Additionally, further studies will be necessary to determine the impact of beat gesture in the presence of other speech-accompanying movement (e.g., head and mouth movement). In order to begin to investigate the neural correlates of beat gesture independently from other types of speech-accompanying movement, the current study recreates environmental conditions where gesture is the only speech-accompanying movement that can be perceived (e.g., viewing a speaker whose face is blocked by an environmental obstacle or viewing a speaker from the back of a large auditorium whose face is barely visible).

Whereas the contrast of beat gesture with speech versus still body with speech showed significant increases in bilateral posterior areas of STG/S, the contrast of beat gesture with speech versus nonsense hand movement with speech showed significant increases in left anterior areas of STG/S. In light of the role of left anterior STG/S in speech intelligibility [Scott et al., 2000; Davis and Johnsrude, 2003], these data suggest that natural beat gesture may impact speech processing at a number of stages. Humphries et al. [2005] found that the left posterior temporal lobe was most sensitive to speech prosody. It may be the case that beat gesture focuses viewers' attention on speech prosody which, in turn, leads to increased intelligibility and comprehension. Considering that responses to speech-accompanied beat gesture and nonsense hand movement are not significantly different within right PT, the synchronicity of beat gesture (or the asynchronicity of the random movements) may contribute to differential responses observed in anterior temporal cortex for listening to speech accompanied by these two types of movement.

Willems and Hagoort [2007] have suggested that the link between language and gesture stems from a more general interplay between language and action. Perhaps attesting to this interplay, no other regions besides the anterior STG/S were more active for speech with beat gesture compared to speech with nonsense hand movements. The stimuli and design of the present study were also significantly different from those of another recent study which showed increased responses in Broca's area for gesture-word mismatches [Willems et al., 2006]. Willems et al.'s findings are complimentary to those of the current study in that we investigated responses to gesture with very little semantic information, whereas Willems et al. examined the impact of semantic incongruency in gesture and speech.

Besides posterior temporal regions, we also observed greater activity for speech with beat gesture (as compared to speech with a still body) in bilateral premotor cortices and inferior parietal regions. This may reflect activation of the "mirror neuron system" [for review, see Rizzolatti and Craighero, 2004, and Iacoboni and Dapretto, 2006], whereby regions responsible for action execution (in this case, gesture production) are thought to likewise be involved in action observation. Wilson et al. [2007] also reported bilateral premotor activity for audiovisual speech (but not for audio-only speech), although this activity was ventral to that observed in the present study and did not reach significance when audiovisual and audio-only conditions were compared directly. This difference in localization might reflect the fact that, unlike the stimuli used in the current study, the speaker's head, face, and speech articulators were fully visible in the stimuli used by Wilson and colleagues (i.e., hand muscles are known to be represented dorsally to head and face muscles within the premotor cortex).

An important area for the processing of our ASL-derived nonsense hand movement was the parietal cortex. Parietal activity was consistently observed when beat gesture and nonsense hand movement (both with and without speech) were compared to baseline. In addition, parietal activity was significantly greater both when viewing nonsense hand movement accompanied by speech (as compared to viewing beat gesture accompanied by speech) and when viewing nonsense hand movement without speech (as compared to viewing beat gesture without speech). Interestingly, Emmorey et al. [2004, 2005, 2007] have identified parietal activity as being crucial to production of sign language. Considering that our subjects and the woman appearing in our stimuli neither spoke nor understood ASL, our data suggest that parietal regions may be optimized for perception of the types of movement used in ASL.

To conclude, our findings of increased activity in posterior STG/S (including PT) for beat gesture with speech indicate that canonical speech perception areas in temporal cortices may process and integrate not only auditory cues but also visual cues during speech perception. Additionally, our finding that activity in anterior STG/S is impacted by speech-accompanying beat gesture suggest differential but intertwined roles for anterior and posterior sections of the STG/S during speech perception, with anterior areas demonstrating increased effects for amplification of speech intelligibility and posterior areas demonstrating increased effects for the presence of multimodal input. In line with extensive research showing that speech-accompanied gesture impacts social communication [e.g., McNeill, 1992] and evidence of a close link between hand action and language [for review, see Willems and Hagoort, 2007], our findings highlight the important role of multiple sensory modalities in communicative contexts.

ACKNOWLEDGMENTS

The authors thank Olga Griswold for invaluable discussions throughout the length of this project and two anonymous reviewers for their helpful comments.

REFERENCES

Barrett DJ, Hall DA (2006): Response preferences for "what" and "where" in human non-primary auditory cortex. Neuroimage 32:968–977.

Bidet-Caulet A, Voisin J, Bertrand O, Fonlupt P (2005): Listening to a walking human activates the temporal biological motion area. Neuroimage 28:132–139.

Blake R, Shiffrar M (2007): Perception of human motion. Annu Rev Psychol 58:47–73.

Brett M, Anton JL, Valabregue R, Poline JB (2002): Region of interest analysis using an SPM toolbox. Neuroimage 16:S497 [abstract].

Bruce C, Desimone R, Gross CG (1981): Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384.

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001): Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. Brain Res Cogn Brain Res 10:349–353.

Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003): Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14:2213–2218.

Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004): Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J Cogn Neurosci 16:805–816.

Callan DE, Tsytsarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, Turner R (2006): Song and speech: Brain regions involved with perception and covert production. Neuroimage 31:1327–1342.

Calvert GA, Campbell R (2003): Reading speech from still and moving faces: The neural substrates of visible speech. J Cogn Neurosci 15:57–70.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. Science 276:593–596.

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999): Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport 10:2619–2623.

Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, Brammer MJ, David AS (2001): Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Brain Res Cogn Brain Res 12:233–243.

Chen JL, Zatorre RJ, Penhune VB (2006): Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage 32:1771–1781.

Collins DL, Neelin P, Peters TM, Evans AC (1994): Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192–205.

Cook SW, Mitchell Z, Goldin-Meadow S (2007): Gesturing makes learning last. Cognition doi: 10.1016. 6/j.cognition.2007.04.010.

Davis MH, Johnsrude IS (2003): Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Desimone R, Gross CG (1979): Visual areas in the temporal cortex of the macaque. Brain Res 178:363–380.

Emmorey K, Grabowski T, McCullough S, Damasio H, Ponto L, Hichwa R, Bellugi U (2004): Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. Brain Lang 89:27–37.

Emmorey K, Grabowski T, McCullough S, Ponto LL, Hichwa RD, Damasio H (2005): The neural correlates of spatial language in English and American sign language: A PET study with hearing bilinguals. Neuroimage 24:832–840.

Emmorey K, Mehta S, Grabowski TJ (2007): The neural correlates of sign versus word production. Neuroimage 36:202–208.

Friston KJ, Holmes AP, Price CJ, Buchel C, Worsley KJ (1999): Multisubject fMRI studies and conjunction analyses. Neuroimage 10:385–396.

Gentilucci M, Bernardis P, Crisi G, Dalla Volta R (2006): Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. J Cogn Neurosci 18:1059–1074.

Goldin-Meadow S, Singer MA (2003): From children's hands to adults' ears: Gesture's role in the learning process. Dev Psychol 39:509–520.

Grossman ED, Blake R (2002): Brain areas active during visual perception of biological motion. Neuron 35:1167–1175.

Grossman ED, Blake R, Kim CY (2004): Learning to see biological motion: Brain activity parallels behavior. J Cogn Neurosci 16: 1669–1679.

Holle H, Gunter TC (2007): The role of iconic gestures in speech disambiguation: ERP evidence. J Cogn Neurosci 19:1175–1192.

Holle H, Gunter TC, Ruschemeyer SA, Hennenlotter A, Iacoboni M (2007): Neural correlates of the processing of co-speech gestures. Neuroimage doi:10.1016/2007.10.055.

Humphries C, Love T, Swinney D, Hickok G (2005) Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Hum Brain Mapp 26:128–38.

Iacoboni M, Dapretto M (2006): The mirror neuron system and the consequences of its dysfunction. Nat Rev Neurosci 21:191–199.

Indefrey P, Cutler A (2004): Prelexical and lexical processing in listening. In: Gazzaniga M, editor. The Cognitive Neurosciences. Cambridge, Massachusetts: MIT Press.

Kelly SD, Barr D, Church RB, Lynch K (1999): Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. J Mem Lang 40:577–592.

Kelly SD, Kravitz C, Hopkins M (2004): Neural correlates of bimodal speech and gesture comprehension. Brain Lang 89:253–260.

Kelly SD, Ward S, Creigh P, Bartolotti J (2006): In intentional stance modulates the integration of gesture and speech during comprehension. Brain Lang 101:222–233.

Kendon A (1972): Some relationships between body motion and speech: An analysis of an example. In: Siegman AW, Pope B, editors. Studies in Dyadic Communication. Elmsford, New York: Pergamon. pp 177–216.

Kotz SA, Meyer M, Paulmann S (2006): Materalization of emotional prosody in the brain: An overview and synopsis on the impact of study design. Prog Brain Res 156:285–294.

Krahmer E, Swerts M (2007) Effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. Journal of Memory and Language 57:396–414.

Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005): In the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Exp Brain Res 166:289–297.

Li Q, Nakano Y, Nishida T (2003): Gestures realization for embodied conversational agents. Presented at The 17th Annual Conference of the Japanese Society for Artificial Intelligence, Niigata, Japan.

MacSweeney M, Campbell R, Woll B, Giampietro V, David AS, McGuire PK, Calvert GA, Brammer MJ (2004): Dissociating linguistic and nonlinguistic gestural communication in the brain. Neuroimage 22:1605–1618.

MacSweeney M, Campbell R, Woll B, Brammer MJ, Giampietro V, David AS, Calvert GA, McGuire PK (2006): Lexical and sentential processing in British sign language. Hum Brain Mapp 27:63–76.

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K., Woods R., Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B (2001): A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 356:1293-322.

McGurk H, MacDonald J (1976): Hearing lips and seeing voices. Nature 264:746–748.

McNeill D (1992): Hand and Mind: What Gestures Reveal About Thought. Chicago: University of Chicago Press.

McNeill D (2005): Gesture and Thought. Chicago: University of Chicago Press.

McNeill D, Cassell J, McCoullough KE (1994): Communicative effects of speech mismatched gestures. Res Lang Soc Interact 27:223–237.

Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004): Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89:277–289.

Özyürek A, Willems RM, Kita S, Hagoort P (2007): In-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. J Cogn Neurosci 19:605–616.

Padberg J, Seltzer B, Cusick CG (2003): Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: An analysis in the tangential plane. J Comp Neurol 467:418–434.

Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M (2006): Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. Hum Brain Mapp 27:471–477.

Quintilian (1856) Institutes of oratory. Honeycutt L, Ed. (Watson JS, Trans.).

Rizzolatti G, Craighero L (2004): The mirror-neuron system. Ann Rev Neurosci 27:169–192.

Saito Y, Ishii K, Yagi K, Tatsumi IF, Mizusawa H (2006): Cerebral networks for spontaneous and synchronized singing and speaking. Neuroreport 17:1893–1897.

Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Shattuck-Hufnagel S, Yasinnik Y, Veilleux N, Renwick M (2007): A method for studying the time alignment of gestures and prosody in American English: 'Hits' and pitch accents in academic-lecture-style speech. In: Esposito A, Bratanić M, Keller E, Marinaro M, editors. Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue. Amsterdam: IOS Press.

So C, Coppola M, Licciardello V, Goldin-Meadow S (2005): The seeds of spatial grammar in the manual modality. Cog Sci 29:1029–1043.

Stein BE, Meredith MA, Wallace MT (1993): The visually responsive neuron and beyond: Multisensory integration in cat and monkey. Prog Brain Res 95:79–90.

Sumby WH, Pollack I (1954): Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26:212–215.

Sweet RA, Dorph-Petersen KA, Lewis DA (2005): Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J Comp Neurol 491:270–289.

Tuite K (1993): The production of gesture. Semiotica 93:83–105.

Warren JD, Jennings AR, Griffiths TD (2005): Analysis of the spectral envelope of sounds by the human brain. Neuroimage 24: 1052–1057.

Westbury CF, Zatorre RJ, Evans AC (1999): Quantifying variability in the planum temporale: a probability map. Cereb Cortex 9: 392–405.

Willems RM, Hagoort P (2007): Neural evidence for the interplay between language, gesture, and action: A review. Brain Lang 101:278–289.

Willems RM, Ozyurek A, Hagoort P (2006): When language meets action: The neural integration of gesture and speech. Cereb Cortex doi:10.1093/cercor/bhl141.

Wilson SM, Molnar-Szakacs I, Iacoboni M (2007): Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. Cereb Cortex doi:10.1093/cercor/bhm049.

Wu YC, Coulson S (2005): Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology 42:654–667.

Yasinnik Y, Renwick M, Shattuck-Hufnagel S (2004): The timing of speech-accompanying gestures with respect to prosody. Proceedings of Sound to Sense, MIT.

Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: Music and speech. Trends Cog Neurosci 6:37–46.