



Review

# The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism

Elizabeth Redcay\*

*Department of Psychology, University of California, San Diego, 8110 La Jolla Shores Dr., Suite 201, La Jolla, CA 92037, USA*

Received 26 October 2006; received in revised form 8 March 2007; accepted 12 June 2007

## Abstract

Within the cognitive neuroscience literature, discussion of the functional role of the superior temporal sulcus (STS) has traditionally been divided into two domains; one focuses on its activity during language processing while the other emphasizes its role in biological motion and social attention, such as eye gaze processing. I will argue that a common process underlying both of these functional domains is performed by the STS, namely analyzing changing sequences of input, either in the auditory or visual domain, and interpreting the communicative significance of those inputs. From a developmental perspective, the fact that these two domains share an anatomical substrate suggests the acquisition of social and speech perception may be linked. In addition, I will argue that because of the STS' role in interpreting social and speech input, impairments in STS function may underlie many of the social and language abnormalities seen in autism.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* fMRI; Development; Language; Eye gaze; Biological motion

## Contents

1. Introduction . . . . .	124
2. A common function is performed by the STS in both the language and social domains . . . . .	124
2.1. The STS and language functions . . . . .	124
2.2. The STS and social functions . . . . .	126
2.3. The role of STS connections with other brain regions . . . . .	130
2.4. Hemispheric biases are seen in the STS . . . . .	130
3. The STS is important to receptive language acquisition . . . . .	130
3.1. Behaviors important to language acquisition are mediated by the STS . . . . .	130
3.2. The STS becomes tuned to respond to meaningful input . . . . .	133
3.3. Other regions involved in language learning . . . . .	133
4. Given the above, abnormalities in this region may be critical to the emergence of autism . . . . .	134
4.1. Anatomical evidence of STS abnormalities in autism . . . . .	134
4.2. Functional evidence of STS abnormalities in autism . . . . .	135
4.3. Role of connectivity in STS impairments in autism . . . . .	137
5. Conclusion . . . . .	137
Acknowledgments . . . . .	138
References . . . . .	138

\*Tel.: +1 858 551 7925x231; fax: +1 858 551 7931.

E-mail address: [eredcay@ucsd.edu](mailto:eredcay@ucsd.edu)

## 1. Introduction

Fundamental questions in cognitive science, psychology, and cognitive neuroscience include how do we process our world, how is the brain organized, and in developmental science, how do we acquire these processes. Neuroimaging, particularly functional magnetic resonance imaging (fMRI), gives the advantage of identifying the neural substrates for cognitive processing in healthy participants. With the fMRI method, one typically chooses a cognitive faculty and then identifies the structures underlying the behavior associated with this faculty. Thus, because this technique is only able to answer the question of “where” cognitive processing is taking place, the advent of fMRI has brought with it increased claims of modularity. For example, when presented with auditory language stimuli, lateral superior temporal regions will show increases in blood oxygenated level dependent (BOLD) signal as compared to a baseline leading researchers to suggest (or ‘confirm’ previous neuropsychological data) that these regions within superior temporal cortex are language-specific. However, if one examines a number of studies spanning multiple cognitive domains, it is apparent that the same structures show signal changes to a diverse number of tasks. One such example is Broca’s area, located in the inferior frontal gyrus. While this region was traditionally thought to be specific to speech production based on patients with lesions to this area, studies have now revealed its role in visuo-motor integration, lexico-semantic decision-making, and tone discrimination (Muller et al., 2001). Thus, this piece of cortex likely performs a domain-general function that underlies these diverse processes.

A greater question, then, for cognitive neuroscience may be not where in the brain is a specific task performed, but rather, how? What is the underlying mechanism? ‘How’ questions are inherently more theoretical and difficult to test than ‘where’ but may be essential to advancing our understanding of how we process our world. In this review, I will focus on the superior temporal sulcus (STS). This region is involved in different aspects of social cognition and language processing. I propose that a common function it serves in both of these domains is the parsing of sequences of inputs into discrete units and the extraction of meaning from these units. The STS shows the greatest response to meaningful stimuli of communicative significance. I argue that in development this mechanism of STS function is critical to early receptive language development as language is learned through the joint interaction of child and caregiver through non-verbal, social means. Additionally, I provide evidence for STS abnormalities in autism and propose these may relate to the pervasive social and linguistic deficits seen in this disorder.

## 2. A common function is performed by the STS in both the language and social domains

### 2.1. The STS and language functions

First, I will present evidence for the role of the STS in adult language processing: both auditory and visual. In the

auditory domain, human speech is “a continuously varying stream of sound containing multiple frequencies (formants) changing over time.” (Rimol et al., 2005). As such the ability to parse information over time is inherent to speech, and auditory, perception. To comprehend speech, one must first be able to parse this stream of changing frequencies into meaningful discrete elements, such as phonemes, syllables, words, and phrases, and determine the meaning from these elements. Evidence reviewed below suggests this very mechanism is performed by the STS.

The STS is activated by these meaningful discrete elements ranging in complexity and duration from tones to narratives. However, the degree and location of activation in the STS varies by stimulus complexity and ‘meaningfulness’. For example, greater activation is seen in the STS to narratives, in which the meaning of the entire passage must be tracked, than to sentences (Xu et al., 2005) (Fig. 1A) or reversed stories (Crinion et al., 2003). However, the response to sentences is greater than to meaningless pseudo-word sentences (Roder et al., 2002), a finding for which the authors suggested a specific role of the STS in semantic processing. Other researchers finding similarly greater activity to sentences than non-speech have interpreted the STS, specifically anterior STS (aSTS), as important to intelligible speech processing (Scott et al., 2000, 2006). In one study (Scott et al., 2000), two conditions of intelligible speech (easy sentences and noise-vocoded speech) and two conditions of unintelligible speech (reversed speech and reversed noise-vocoded speech) were presented. Noise-vocoded speech is created by manipulating the periodicity and pitch information extracted from speech excerpts and the resulting sound is like a harsh whisper. This speech is intelligible after a brief training session, although difficult to understand. aSTS activity was greater to the intelligible speech (both easy and noise-vocoded speech) than to the unintelligible reversed speech and noise-vocoded speech. Greater STS activation is also seen at the single word level, as shown by greater activation to words when compared to signal correlated noise (SCN) (Wise et al., 2001). SCN is created by adding white noise to the speech envelope thus producing a stimulus with similar acoustic properties but no phonetic, lexical, or semantic information. Wise et al. (2001) also report STS activation in this same posterior STS (pSTS) site during retrieval of single words from within a specified category (i.e. fish: cod, salmon, trout). Thus, summary of these studies suggest the STS plays a specific role in higher-linguistic processing of meaningful (i.e. intelligible, semantic) speech.

Interestingly, studies examining complex, meaningful (i.e. semantic, familiar) but non-linguistic sounds such as environmental sounds, tool sounds, animal sounds, or human non-speech vocal sounds also elicit STS activation. These studies suggest the STS is not responding to the linguistic aspects of speech specifically. Meaningful sounds (i.e. tools, animals, liquids, dropped objects) and non-meaningful sounds (such as reversed sounds) both activate

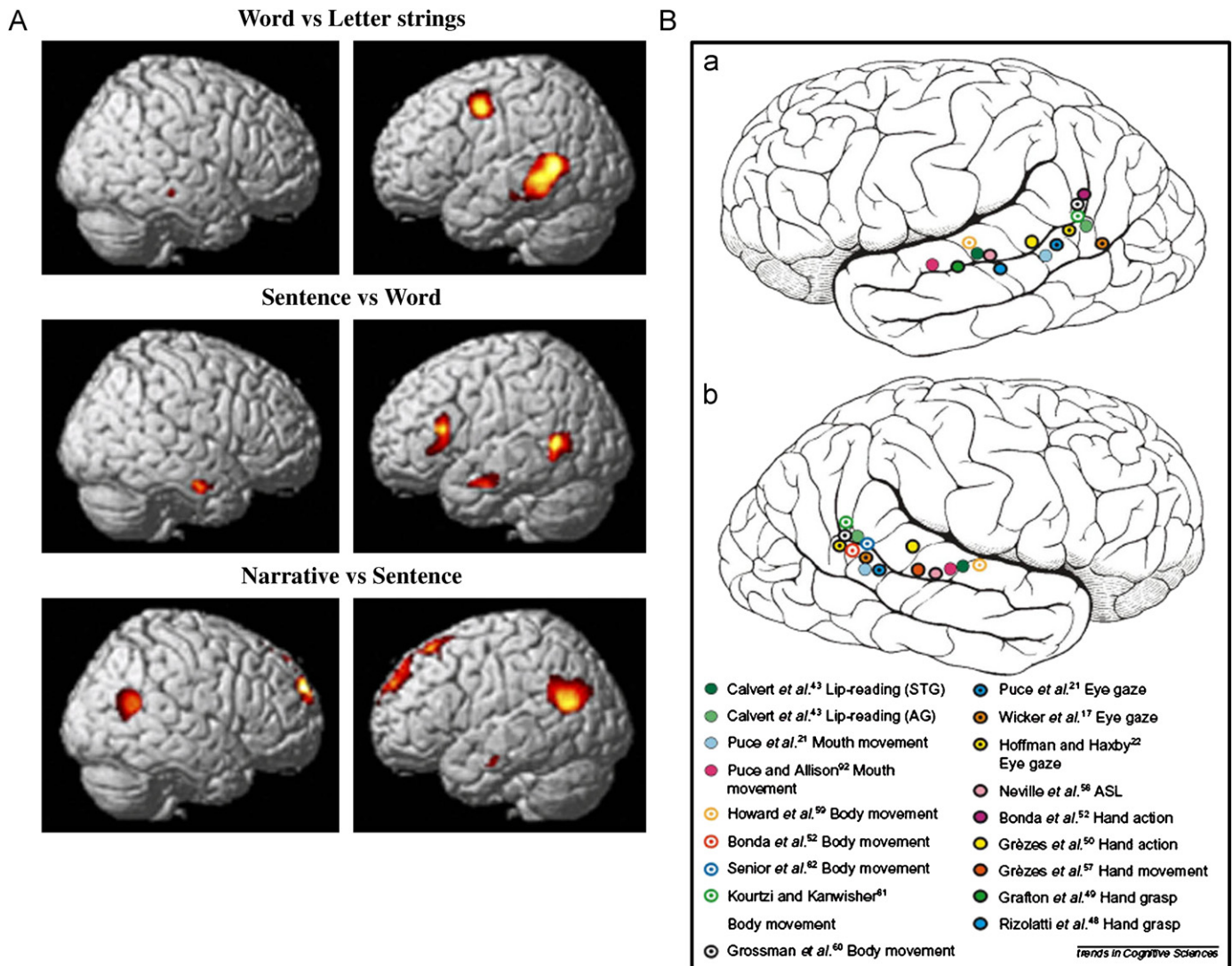


Fig. 1. Activation of the superior temporal sulcus by language (A) and social (B) stimuli. (A) This image reveals variation in both the degree and location of activation for language stimuli. Each component of a narrative (e.g. letter strings, words, and sentences) activates the STS but the greatest activation is seen to the most complex stimulus, a narrative. This hierarchical response of the STS is seen in social perception as well as speech perception. (B) This image, taken from Allison *et al.* (2000), plots the point of peak activation to various forms of biological motion perception in the left hemisphere (a) and the right (b). The numbers in superscript refer to references found in Allison *et al.* (2000). Note for both (A) and (B) the distributions for activation to both biological motion and speech perception are bilateral and along the anterior to posterior axis. Panel (A) was reprinted from *Neuroimage*, 25, Xu, J., Kemeny, S., Park, G., Frattali, C., Braun, A. "Emergent features of word, sentence, and narrative comprehension." 1002–1015. Copyright (2005) with permission from Elsevier. Panel (B) was reprinted from *Trends in Cognitive Sciences*, 4, Allison, T., Puce, A., McCarthy, G. "Social perception from visual cues: Role of the STS region", 267–278, Copyright (2000), with permission from Elsevier.

bilateral posterior and middle STS (pSTS and mSTS) (Lewis *et al.*, 2004, 2005). However, meaningful sounds show a left lateralization when compared to non-meaningful sounds (Engelien *et al.*, 2006; Lewis *et al.*, 2004). Within the category of meaningful sounds, those that are human voice-specific (i.e. laughs, cries) elicit greater STS activation than those that are non-human environmental or mechanical sounds (Belin *et al.*, 2000). Animal vocal sounds also elicit a greater degree of left STS activation than non-vocal sounds, but this difference is smaller than that of human vocal sounds (Belin *et al.*, 2004; Fecteau *et al.*, 2004). In sum, the STS is activated by complex, meaningful sounds that are both linguistic and non-

linguistic and preferential activation is seen to human produced sounds.

In addition, studies with stimuli containing no lexical or semantic information also show activation within STS, albeit not as strong as auditory stimuli with linguistic-communicative significance such as speech or familiar sounds. Hearing pseudo-words elicits greater STS activation than tones (Binder *et al.*, 2000). These authors interpret this greater activation to complex speech and non-speech as evidence for the STS' role in the analysis of complex acoustic features rather than lexical-semantic processing described in the above paragraph. These complex acoustic features would not necessarily have to

connote meaning to elicit STS activity as seen in another study in which monosyllabic nonsense words showed a greater STS response than sine wave analogs (Vouloumanos et al., 2001). Sinewave analogs of speech maintain the same pitch countour, amplitude envelope, relative formant amplitude, and relative intensity of their speech counterparts (Vouloumanos et al., 2001), but the frequencies are altered. Thus, when first heard these sinewave analogs sound like an “electronic glissado”; however, with training, subjects report hearing a syllable (Dehaene-Lambertz et al., 2005). Comparison of perception of the sinewave analog as a syllable versus perception as noise results in greater STS activation (Dehaene-Lambertz et al., 2005; Mottonen et al., 2006). Phonemes show greater mSTS activation than non-phonemic auditory patterns (Liebenthal et al., 2005). At an even simpler level, single consonants repeated in a random order produce greater activity in pSTS than noise (Rimol et al., 2005). Finally, even perception of sounds with no phonetic components, such as sine wave analogs not perceived as speech, show a greater response in the STS than perception of simple tones (Vouloumanos et al., 2001). The common denominator of activation in the STS in these studies appears to not be meaningfulness but rather perception of units of sound, such as sentences, words, laughter, phonemes, and even consonants. However, with increasing complexity and linguistic or communicative relevance, activation within the STS increases.

Specht and Reul (2003) examined this hierarchical nature of STS activation by examining the response to words, sounds, and tones in the STS in the same experiment. These authors found an interaction by hemisphere in speech and non-speech processing. Specifically, hearing words elicited greater activity than tones in both left and right hemispheres but presentation of words showed greater activity than sounds in left STS (lSTS) while presentation of sounds showed greater activity than words in right STS (rSTS) (Specht and Reul, 2003). In a separate study of words, syllables, and familiar sounds (i.e. animal and mechanical), functional variation was seen along the anterior to posterior axis. aSTS and pSTS were specific to speech while mSTS responded to both sounds and speech (Giraud and Price, 2001).

Interestingly, the rate of presentation of auditory stimuli affects STS response. In one study, rate of word presentation varied from 1 stimulus per minute to 75 per minute, with 4 rates in between. When the linear response to rate of word presentation was modeled, activation was found along the anterior to posterior extent of the STS, with the greatest response to rapid acoustic presentation rates. Wise et al. (2001) interpret these findings, in addition to their findings of greater STS activation to word recall and verb generation, as evidence for the STS involvement in “transiently representing phonetic sequences, whether heard or internally generated and rehearsed.”

In sum, the common underlying function of the STS in the auditory language domain is parsing of this rapid stream of input into discrete meaningful elements. Greater

STS activation is seen with both increasing degrees of significance (i.e. semantic or paralinguistic information) and complexity in the auditory input, with the weakest activation seen to the simplest, non-meaningful, non-social sounds: tones. Because the STS is activated by auditory sounds of almost all levels of complexity and meaningfulness, to differing degrees, an important question is do complex stimuli that engage similar semantic and linguistic analysis but in the visual domain elicit STS activity? In other words, is the STS specific to auditory processing or is it responsive to the patterns associated with linguistic processing. The literature on the brain response to sign language can address this question.

Studies of the neural bases of sign language (SL) typically measure the neural response to SL in deaf native signers, hearing signers, and hearing non-signers. Evidence from these studies suggests auditory and visual processing of linguistic stimuli, including signed sentences, words, and sublexical phonetic units, utilize very similar neural substrates (Neville et al., 1998; Petitto et al., 2000; Sakai et al., 2005), although some have found a greater extent of STS activation in the right hemisphere in signers (Neville et al., 1998). Another study found a left-lateralized response in deaf native signers to linguistic facial expressions with a verb sign (McCullough et al., 2005). When presented with SL alone, non-signers may activate the STS but to a lesser degree than signers because these gestures hold no meaning to non-signers (Levanen et al., 2001; MacSweeney et al., 2004; Sadato et al., 2004). Further, this STS activation in non-signers tends to be more posterior (MacSweeney et al., 2004). Additionally, studies of speech reading from the face find a greater response in aSTS to speech reading as compared to meaningless face motions (Campbell et al., 2001).

As in the auditory domain, the STS is also activated by non-linguistic, but meaningful, inputs. In the visual domain, these “inputs” are communicative gestures. Studies suggest processing of these communicative gestures is supported by STS regions (Gallagher and Frith, 2004; MacSweeney et al., 2004; Nakamura et al., 2004), but the activation is greater to sign language than to gestures only to native signers (both deaf and hearing) (MacSweeney et al., 2004). This parallels findings in the auditory domain in which activation is greater to linguistic input than to semantic non-linguistic input such as environmental sounds or laughter. Overall, like spoken language, sign language appears to be processed by a network of regions including STS, and the degree of activation is related to the meaningfulness of the information. Thus, linguistic activation of STS is not specific to the auditory modality but rather is involved in parsing auditory or visual sequences of input into discrete units that can convey meaningful, communicative information.

## 2.2. *The STS and social functions*

A separate literature reveals the role of the STS in social perception. Social perception refers to the initial stages of

evaluating the social communicative intentions of others by analysis of eye-gaze direction, facial expressions, body movements, and other types of biological motion (Allison et al., 2000) (Fig. 1B). Like in the language domain, an important component to STS activation in social perception is temporal processing. Hearing a sentence requires representing each linguistic unit heard over time and then extracting the intended meaning. Viewing a facial expression requires representing each facial cue over time and assigning meaning to the specific order of facial cues. Importantly, the greatest response within the STS is seen to dynamically changing, social stimuli, although a response is also seen to static biologically relevant stimuli, such as faces (LaBar et al., 2003), or stimuli with implied motion (Jellema and Perrett 2003a). Further, perception of this motion, or implied motion, recruits a large response from cells within the STS only if the visual actions follow a specific sequence (Jellema and Perrett, 2003b). Thus, while speech and social perception appear to be separate behavioral domains, they both share a common function of first parsing sequences of input into units and second interpreting meaning from those units in the STS.

Social stimuli which elicit activation within the STS fit broadly into the category of biological motion, or the “visual perception of a biological entity engaged in a recognizable activity” (Pelphrey and Morris, 2006). The perception of biological motion engages a broad network of brain regions including superior temporal areas, and visual regions within both ventral (object) and dorsal (motion) regions (Vaina et al., 2001). The STS receives inputs on both form and motion and integrates these to identify a moving form and to extract social significance from this form. This function is seen in a diverse number of biological motion tasks ranging from eye gaze perception to perception of a social, moving form from point-light displays. While motion, or implied motion, must be present to activate the STS, the degree and location of activation appears to vary by degree of complexity, type of motion, and biological relevance. Specifically, like in the language domain, a hierarchy of STS activation is seen with the greatest degree of activation to dynamic, complex, and socially meaningful stimuli (e.g. emotional facial expressions) and the lowest to non-meaningful, non-social motion (i.e. random motion), with no activation seen to static, non-socially relevant stimuli. These stimuli will be reviewed in turn below.

Facial expressions are a complex form of biological motion in which a number of facial muscles change over time in a specific sequence to convey a particular emotion. It has been proposed that the role of the STS in face perception is to process changeable aspects of faces—perception of eye gaze, expression, and lip movement (Haxby et al., 2000, 2002). A separate region, the fusiform gyrus (FG) is involved in invariant aspects of faces—perception of unique identity (Haxby et al., 2000, 2002), highlighting the dissociation between recognizing who a person is and what a person is trying to communicate.

LaBar et al. (2003) presented subjects with identity morphs in which one neutral face morphed into a different person’s neutral face and emotion morphs in which the same face morphed from a neutral to angry or fearful expression. In comparison to the identity morphs, emotion morphs elicited greater responses in aSTS and pSTS, in addition to ventral anterior cingulate gyrus, ventromedial prefrontal cortex, middle frontal gyrus, and medial FG (LaBar et al., 2003). The authors suggest the STS is responsive to changes across emotions, but not identities, because it is sensitive to motion that is biologically plausible, such as changes in facial muscles on the same face.

One type of biological motion involving the face is a change in eye gaze direction. This powerful social cue can convey an enormous range of social and communicative signals including boredom, envy, disgust, fear, interest in another object, or sharing of interest in that object, for example. When subjects were asked to selectively attend to eye gaze direction, more activity was seen in the left STS and intraparietal sulcus (IPS) than when they were asked to attend to person identity; however when attending to identity versus gaze, greater activation was seen in right lateral FG and inferior occipital gyrus (IOG) (Hoffman and Haxby, 2000). Furthermore, viewing an averted as opposed to direct gaze activates STS and IOG to a greater extent (Hoffman and Haxby, 2000). A series of experiments with a brain-damaged patient who had a lesion of the right superior temporal gyrus (STG) revealed a selective deficit in using eyes as a cue to shift attention (Akiyama et al., 2006a, b), suggesting this region is critical to perceiving eye gaze shifts as meaningful. A similar conclusion can be drawn from an fMRI study in which subjects were shown faces with eye gaze cues, faces with direct gaze and an arrow drawn on top of the face, faces with non-meaningful eye gaze (i.e. cross-eyes), and arrows alone (Hooker et al., 2003). Importantly, images were presented in a way that implied eye motion by presenting a series of pictures with eye gaze shifts in 10 different locations. Bilateral STS activation was found when subjects viewed a face with averted gaze as compared to three control tasks (an arrow pointing the same direction, an arrow pointing on a face with direct eye gaze, and a face with eye motion with no directional information). One could argue that an arrow is not as complex as a face and perhaps this is why greater activation was seen to a gaze shift than to an arrow pointing. However, in a separate fMRI study in which complexity is controlled, eyes still elicit greater STS activation (Kingstone et al., 2004). In this study ambiguous cartoon stimuli were presented to participants and in one condition were described as eyes with a hat while in another condition were described as a car with headlights. In both cases, movement of the “eyes” or “headlights” triggered an attentional shift to an asterisk. However, STS activation was only seen in the condition in which subjects were told the cartoon was eyes and a hat. Thus, activation is greatest to a social stimulus (i.e. a face) that is conveying information to the subject (averted gaze in this study).

Further, from the Hooker et al. study, one might suggest that rather than social relevance, the STS may simply be more responsive to averted gaze than to direct. However, an fMRI study (Pelphrey et al., 2004) examined the STS response to averted vs. direct gaze and found greater STS activation to direct gaze. In this case, a virtual person appeared to be walking towards the subject with averted gaze and as he approached he shifted his gaze to direct. In this case, the shift to direct gaze is more socially relevant and therefore meaningful than the averted gaze. Thus, a socially significant gaze shift (or implied shift) seems to be the specific stimulus that elicits STS response in studies of eye gaze processing.

The above studies reveal that the role of the STS in face processing is in identifying the changeable aspects of faces, such as eye gaze, head orientation, lip-reading, and facial expressions (particularly when conveying an emotion). Each of these “changeable aspects” convey social communicative significance through specific, sequenced movements. For example, a facial expression can be thought of as a series of visual-spatial components (of a face) that are ordered in a specific temporal manner to convey a particular emotion. This is very similar to a sentence that is formed by connecting discrete elements (from the level of consonant to word) in a specific, meaningful order to convey a thought. Ekman and Friesen (1978) have identified 46 discrete muscle movements of the face that they term action units (AU). Combinations of these AUs code for different facial expressions. In the auditory domain, an AU can be thought of as a single phoneme, or even letter of the alphabet, combinations of which lead to an infinite set of possible expressions, although only a handful do we recognize as meaningful (Fig. 2). I propose the function of the STS is to perceive, or extract, a

meaningful unit created from a sequence of input units (i.e. AUs or phonemes) presented over time.

Motion in other parts of the head and body can also elicit strong STS activity. For example, when a subject approaches a person (through virtual reality in the scanner) STS activation is seen when the person is making a gesture but not when he is standing still (Morris et al., 2005). In an fMRI adaptation experiment, Andrews and Ewbank (2004) found greater activation in STS when head orientations (viewpoints) were varied but identity was kept constant than when identities varied but viewpoint was kept constant (Andrews and Ewbank, 2004). STS activation is seen to movement of different regions of the face and body separately (Pelphrey et al., 2005b; Puce et al., 1998). Finally, observation of simple finger tapping can also elicit STS activation (Iacoboni et al., 2001).

In addition to recognizing human activity, an important component of STS function is the ability to extract meaning from that activity. Greater activation in the STS is seen when the subject infers a goal or intention of the other person as compared to simple perception of biological motion. In fact, studies of biological motion perception often imply goals or intentions of an observer. A shift in eye gaze, for example, is a movement that conveys social meaning but it also can elicit conjecture on what the person who is doing the gaze shift is thinking. This function may explain why STS activation appears in studies of theory of mind (TOM) perception, or having an understanding of the goals, intentions, or beliefs of another person (Frith and Frith, 2003). Evidence from lesion and fMRI studies in humans and single cell recording studies in macaque monkeys suggests that when a person’s (or object’s) actions implies goal-directed behavior greater STS activation is seen. Patients with lesions to the

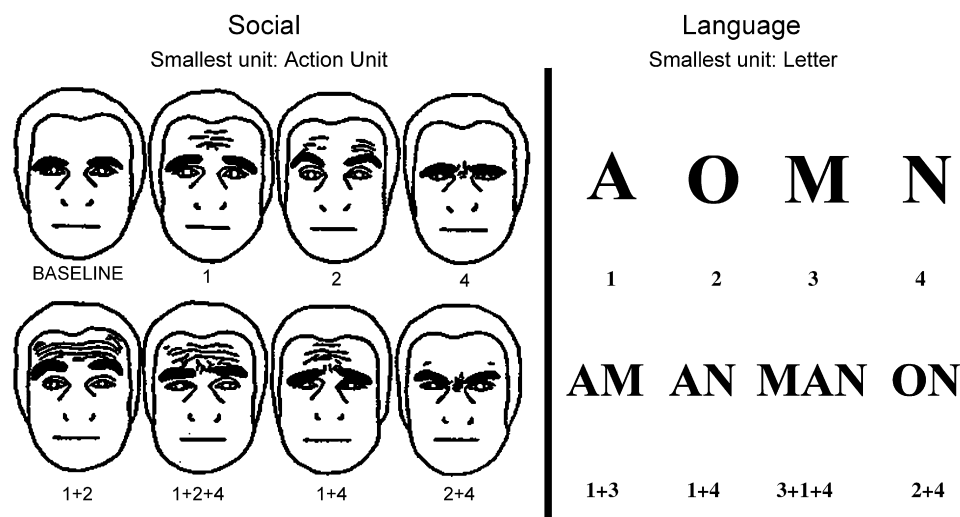


Fig. 2. A shared mechanism for processing facial expressions and words. The figure on the left depicts three facial action units and the resulting expressions from combination of these three (taken from Hager, J.C., Ekman, P. The inner and outer meanings of facial expressions. In: Cacioppo, J.T., Petty, R.E. (Eds.), *Social Psychophysiology*, Guilford Press, New York, p. 291, Copyright (1983), with permission from Guilford Press). The figure on the right describes an analogous process in the language domain of 4 letters arranged in such a way as to form a meaningful word. The smallest unit in the social domain of facial expression processing is a facial action unit. The smallest unit in the language domain is a letter. These units are combined to create a form of communicative significance (i.e. a facial expression or word).

temporoparietal junction, (which includes the STS region) were impaired on questions which required reasoning about another's belief but not on questions requiring memory or factual reasoning (Samson et al., 2004). An fMRI study of healthy adults showed that a person walking behind a bookcase and pausing before stepping out from the bookcase led to greater STS activation than someone not stopping behind the bookcase (Saxe et al., 2004). The authors suggest the subject is detecting goal-directed behavior by the person stopping behind the bookcase. A second fMRI study presented simple moving geometric shapes and similarly found the greatest STS response when the shape appeared to perform a goal-directed behavior, such as anticipating the path of a second shape (Schultz et al., 2004). A third study employed a virtual reality game in which participants became taxi cab drivers (Spiers and Maguire, 2006). Subjects reviewed the game they played outside of the scanner and reported their spontaneous mental thoughts during different events. These thoughts were divided into either attributing intentions to agents in the environment of the game (e.g. pedestrian crossing the street) or thinking about their own thoughts with regard to their customers or the experimenters. The control was "coasting" in which they were not thinking about anything in particular. Only a region within the right pSTS was more active in response to all mentalizing thoughts than to coasting. Thus, the STS region is involved in mentalizing, especially during attributing an intention or goal-directed behavior towards something (usually someone) in the environment.

Single cell recording studies of macaque monkeys can shed light on how the function of the STS region is to integrate across a number of different types of moving forms and extract meaning from them. These studies have revealed that the STS contains neurons selectively responsive to a variety of different parts of the body and face, and also specific head, body, and eye orientations (reviewed in Jellema et al., 2000). Additionally, certain neurons will fire in response to specific actions and in particular when these actions are performed with an object. An impressive finding from single cell recording is that other neurons in the STS will fire when a person reaches for an object only if the gaze of the other person matches the object. This finding provides strong evidence that the STS is involved in attributing intentions or goals to another person. The close proximity of these cells which show responses to different parts of the body and to specific movements and sequences of action makes the STS well suited to interpreting goal-directed behavior of another person based on the other person's actions (Jellema et al., 2000).

The STS also contains neurons sensitive to both auditory and visual stimuli (Bruce et al., 1981; Benevento et al., 1977; Barraclough et al., 2005). This integration of auditory and visual stimuli within individual neurons in the STS may contribute to its performance of a similar function in both the auditory and visual domains. Social cognition in the auditory domain includes perception of

emotional prosody, vocal sounds, and the sound of human action, to name a few. For example, identification of emotional intonation as compared to vowel identification revealed a region in the right STS (Wildgruber et al., 2005). Greater bilateral mSTS activation is seen to angry word-like utterances as opposed to neutral utterances (Grandjean et al., 2005). Further, the STS has been called the 'auditory face' region because it responds maximally to human vocal as opposed to non-human vocal or non-vocal sounds (Belin et al., 2004). Finally, pSTS activation is seen to the sound of footsteps which overlaps with the same region which is responsive to the visual display of a person walking (Bidet-Caulet et al., 2005). While auditory stimuli are generally not described in terms of 'motion', auditory stimuli inherently contain a temporal component in that units of information must be presented over time and the sequence of temporal presentation is critical. Thus, perception of these auditory social stimuli would necessitate similar processing as that of a social, visual stimulus that changes over time through motion. In sum, an analogous function can be seen in both auditory and visual social perception in the STS.

As mentioned previously, a hierarchical response is seen within the STS such that the greatest response is seen during meaningful biological motion perception (either visual or auditory as reviewed above) as compared to non-biological motion perception. However, non-biological motion perception still elicits greater STS activation than static, non-social images. In a study aimed at elucidating the hierarchical nature of responses in the STS, subjects were shown four stimuli: two types of biological motion stimuli (human and robot) and two types of mechanical motion stimuli (grandfather clock and random mechanical motion) (Pelphrey et al., 2003). Both types of stimuli elicited activation within the right STS; however biological motion stimuli elicited greater right STS activation than non-biological motion. In general, when presented with non-biological motion stimuli, the degree of activation within the STS is greatest if the motion is articulated (like that of typical human motion) as opposed to simple translational or rotational movement. For example, Beauchamp et al. (2002) presented subjects with human and tool stimuli that were static, showed articulated motion (i.e. jumping jacks) or showed unarticulated motion (i.e. rotation or translation of body or tool). Activity in the STS was elicited by all stimuli however a greater response was seen to biological than non-biological and to articulated as compared to unarticulated motion. Static images also activated the STS (to a lesser degree than moving) but these static images did have implied articulated motion. For example, the static image of a person had arms and legs extended as if in the middle of a jumping jack. In a separate series of experiments in which point-light displays, silhouettes, and words were presented, Peuskens et al. (2005) identified a similar response hierarchy within the STS and conclude that the STS is maximally responsive to human action. Activation in the

right STS was greatest to the point-light display of biological motion, which contains both the form of a figure and articulated motion, than in any other condition.

The STS is also responsive to motion that is non-social, non-articulated, but is still complex. Specifically, rSTS showed greater activation to second order as compared to first-order motion (Noguchi et al., 2005). Second-order motion was created by selecting dots in a rectangle area to move 3 times as fast as the background dots. This is contrasted with first-order motion in which the dots within the rectangle area simply change luminance. Detecting motion cues from a motion background is more complex than detecting motion without background noise, and presumably this complexity elicits greater STS activation.

From the above studies it is clear that the STS is central to the perception of meaningful and complex motion. Specifically, it is responsive to motion that is articulated and produced by a form which conveys significance (i.e. person, tool, or robot). Static images that imply motion or coherent patterns of motion also activate the STS but to a lesser degree. Importantly, the greatest degree of STS activation is seen to dynamic stimuli which elicit mentalizing, or attributing intentions or goals to the social form that is producing the motion. In sum, these studies suggest that the STS is involved in biological motion perception because it responds to motion and parses that motion into meaningful, coherent units. An analogous function in the auditory domain is seen in its role in prosody, emotion perception, and semantic, linguistic analysis. What is the common denominator? Speech is a sequence of auditory units of different frequencies presented over time. Biological motion is a sequence of changes in spatial location over time. The STS appears to parse these units and integrate them into a meaningful whole.

### 2.3. *The role of STS connections with other brain regions*

Research into the anatomical connections of the STS from the non-human primate literature gives evidence for why the STS is so well suited for its ability to parse streams of information from both auditory and visual domains and also extract meaning from these inputs. The STS has reciprocal connections to visual, auditory, and even somatosensory cortices (Barnes and Pandya, 1992). In addition, higher-order association cortices project to the STS, including prefrontal and parietal areas (Ban et al., 1991; Cusick et al., 1995; Seltzer et al., 1996; Seltzer and Pandya, 1989). This convergence of overlapping inputs from multiple areas is particularly apparent in area TPO, a polysensory association area located in the pSTS (Seltzer et al., 1996). Different regions of TPO have reciprocal projections to different areas of prefrontal cortex. Specifically, rostral TPO projects to ventral (BA 13, 12, 11), medial (BA 24, 32, 14, 9) and lateral (BA 10, 12, 46) regions of prefrontal cortex while the mid-portion of TPO projects to dorsal areas (BA 46, 9, 10) and the caudal portion projects to caudal areas (BA 46, 8, 6) (Seltzer and Pandya,

1989). The polysensory neurons and the overlapping connections from primary cortices and higher-order cortices make the STS particularly suited for parsing streams of input across modalities and extracting meaning from them.

### 2.4. *Hemispheric biases are seen in the STS*

The review of both domains suggests a preference of the left STS in processing language stimuli and the right STS in processing social and emotional stimuli. However, most studies reveal STS activation within both hemispheres, with one side showing a greater extent or intensity of activation (Fig. 1). Some have suggested that the reason for lateralization effects is that the left hemisphere is sensitive to temporal qualities of sound while the right is sensitive to spectral, such as pitch (Zatorre and Belin, 2001). Recent evidence suggests that, in the STS, both hemispheres are sensitive to temporal structure (Boemio et al., 2005). Specifically, for sounds with and without spectral variation, the lSTS showed a bias for sound segment lengths of 25–30 ms while rSTS showed a bias for sound segment lengths of 200–300 ms. Thus, stimuli of shorter processing duration, such as rapidly changing auditory, linguistic stimuli may show a left-bias in STS while stimuli of longer processing duration, such as changes in facial expression, body movement, or emotional vocal expressions may show a right-bias.

When do these “specializations” emerge? If the mechanism of STS function is similar across both social and speech perception, are these behavioral domains separate in early childhood? Behavioral evidence suggests not. Many of the behaviors described above and additional behaviors that are supported by STS are critical to language acquisition. These will be reviewed in the discussion below.

## 3. **The STS is important to receptive language acquisition**

### 3.1. *Behaviors important to language acquisition are mediated by the STS*

While there is evidence for speech perception ability (e.g. phonetic discriminations) in early infancy, receptive language acquisition (including word and sentence comprehension) does not occur until the end of the first year and into the second year of life. Typically, a child’s first comprehended word occurs around 8 months of age (Bates et al., 2003; Fenson et al., 1994). Before language is acquired, children must utilize non-verbal cues to both make sense of their world and to communicate. Their first means of communicative input is often from their mother’s and other’s facial expressions. As children develop, more complex non-verbal skills are acquired such as initiating and responding to joint attention (RJA) and interpreting and making gestures. Researchers have proposed that these joint engagement interactions are “based on social processes more basic than language and that they therefore



make language acquisition possible by creating a shared referential framework within which the child may experientially ground the language used by adults.” (Carpenter et al., 1998, p. 24) In fact, time spent in joint engagement at 10 months correlates significantly with language comprehension between 10 and 15 months (Carpenter et al., 1998).

In the above section, I proposed a common neural substrate underlying social and speech perception, namely the STS. This supports the idea that a child’s ability to interpret social communicative cues is necessary for the acquisition of receptive language. It follows then that the STS’ role in interpreting social and non-verbal, prelinguistic cues early in development could be the same mechanism for interpreting linguistic cues once language is acquired. In other words, cortex specialized for auditory language reception later in life may have been responsible for interpreting non-verbal communicative cues early in life via its ability to parse sequences of auditory or visual inputs and extract communicative meaning. This same ability then becomes critical to perceiving streams of speech. In fact, Carpenter et al. (1998) suggests that language should be thought of as another social skill and as such it fits into the social communicative behaviors of following gaze, pointing, and gesture. Interestingly, all of these functions are mediated by the STS at least in the adult, lending a neural basis to this hypothesis.

In this section, evidence for the role of non-verbal behaviors (such as joint attention, gesture, and temporal processing) in receptive language learning will be reviewed. Evidence from the adult literature suggests these skills are mediated by the STS. An important caveat to keep in mind is that the neural substrates underlying behaviors in the adult may not be the same as those in the child. Thus, evidence for STS involvement in these behaviors in infants and children needs to be empirically tested before strong assertions can be made.

One behavior that has consistently been shown to be predictive of later language skills is a child’s ability to engage in joint attention. Joint attention can be divided into two separate behaviors: initiating joint attention (IJA) and RJA. RJA is the ability to follow gaze or point towards an object and to infer that the person is requesting shared attention of the joint attention initiator and the object. Coordination of joint attention is a platform by which caregivers can label words. Tomasello and Todd (1983) were the first to provide direct evidence for the critical role of joint attention in language development. They showed that the amount of time spent engaging in joint attention between mother and child dyads during a 6 month period was predictive of vocabulary development at the end of this period. Other subsequent studies have supported this finding through significant correlations between RJA bids in the second year of life and later vocabulary development (Morales et al., 2000; Mundy et al., 2003; Mundy and Gomes, 1998). In the joint attention situation, word learning is dependent on the ability for a child to follow attention to the object being

labeled, to integrate the auditory label with the visual input of the object, and to infer the speaker’s intention in labeling that object.

Evidence from the adult fMRI literature suggests the STS is involved in the three components listed above as critical to learning language in a joint attention context: following a speaker’s attention, auditory–visual integration, and inferring a speaker’s intention. First, I will present behavioral evidence that suggests a link between these separate components of joint attention and language learning. Second, I will provide evidence suggesting that these components of joint attention may be mediated by the STS.

Following a speaker’s attention involves the ability to follow eye gaze shifts, head turns, and points to a target either within or beyond a child’s visual field. Following attention (either through gaze or point following) at 9–10 months is shown to correlate significantly with referential language at 12 months of age (Carpenter et al., 1998). An infant’s ability to match his mother’s direction of gaze at 6 months significantly correlates with his receptive language vocabulary at 12 months and expressive vocabulary at 18 months (Morales et al., 1998). The ability to respond to joint attention through gaze and point cues at 14–17 months is predictive of vocabulary development 4 months later (Mundy and Gomes, 1998). Additionally, an infant’s ability to locate targets outside of his visual field at 15 months predicts expressive language ability at 24 months (Delgado et al., 2002). A combination of these cues, such as head turn or pointing, from the adult may help the child follow attention to more difficult objects such as those behind the child’s head (Deak et al., 2000). Studies reviewed above suggest that in the adult, the ability to follow eye gaze, head movement, and pointing involves the STS. These abilities may be mediated by the STS in infants and children as well; however, as noted previously, future studies are needed to verify this as often brain regions utilized for a particular task in an adult or not the same as those utilized early in development.

Auditory–visual integration is necessary for being able to understand that the object one sees “matches” the word one hears in a joint attention context. This skill is particularly important for successful mapping of a referent with its label during the joint attention context. Molfese (1990) recorded brain auditory evoked responses (AER) to novel words that were correctly matched with the object the infant was looking at and words that were incorrectly matched (Molfese, 1990). A “correct match” refers to a novel object being paired with the label learned several days prior to AER testing. A difference was seen in the brain response to auditory and visual matched pairings as compared to unmatched word-object pairings in 14-month-old infants over bilateral frontal electrodes soon after the stimulus is presented and over left frontal, temporal, and parietal electrode sites later. Spatial localization is not a strength of the ERP method and thus, it remains unknown whether the STS is involved in auditory–visual integration

in infants. In adults, however, the integration of auditory and visual inputs involves the STS, suggesting this might be an important region for auditory–visual integration in infants and children. Auditory combined with visual presentation of congruent words resulted in greater activity within the STS than to either the written or heard word alone in an fMRI study (Wright et al., 2003). Auditory and visual matching for objects is also seen in the STS. Beauchamp et al. (2004) examined the neural response to pictures of objects, sounds they made, and the integration of the two. They found that the only regions which showed equal response to both auditory and visual conditions alone and enhanced response when auditory and visual conditions were presented together was the pSTS and MTG (Beauchamp et al., 2004). Single cell recording from monkeys revealed that single neurons within STS show multimodal responses, in that firing is elicited by auditory and visual stimuli alone but is greatest when these two inputs are paired, given that they are congruent (i.e. sound of foot tapping paired with image of foot tapping) (Barracough et al., 2005). Sestieri and colleagues suggest the STS serves to bind “semantic” audiovisual information as compared to “spatial” audiovisual information (Sestieri et al., 2006).

In addition to following attention, and correctly integrating auditory and visual input, a child must be able to understand that the adult is intending to label a specific object. Two-year-olds presented with a series of labeled objects can use cues to determine which object the speaker is intending to label. In one example, an adult is searching for an object and rejects certain other objects in doing so. Two-year-olds will learn the label for the object that is searched for but not the one that is rejected (Tomasello and Barton, 1994). Two-year-olds also understand that an adult is more likely to label an object that is novel to that adult (Akhtar and Tomasello, 1996). In this study, 2 experimenters, an adult, and a child play with 3 toys. The adult leaves and the 2 experimenters and the child play with the fourth toy. When the adult returns he looks into the bag and proclaims, “Look, that’s a gazzer.” The child is more likely to guess the 4th toy is a gazzer than any of the other toys, suggesting they understand that the adult labels the toy that is novel to the adult. Diesendruck et al. (2004) demonstrated that learning the label for this 4th toy was not due to contextual changes alone. In their study, the 4th toy was labeled in a different context (e.g. at a separate table) either due to intentionally moving to a new spot or accidentally (e.g. dropping the toy) (Diesendruck et al., 2004). Two-year-olds associated the experimenters label with the toy presented in a novel context only when the experimenter intentionally manipulated the novel context, suggesting the 2-year-olds understood that this toy was intentionally different from the other three. While these studies were conducted on 2-year-olds, younger children also show evidence for understanding other’s intention as being critical to word learning. Eighteen-month-old infants successfully learned a word label if the verbal label was said

by a person looking at the toy the infant was holding but did not learn the label if the word was said through a speaker (Baldwin et al., 1996). At 16 months of age, infants successfully learned the label for an object either when an adult labeled the object they were playing with or when an adult looked at another object and labeled it (Baldwin, 1991). Thus, the infant is not just performing a simple auditory–visual mapping. Rather, the infant can understand that the speaker is intentionally labeling a specific object. Understanding the intentionality of pointing and labeling a specific object is critical to being able to learn new words through joint attention. In the previous section, evidence suggested that in adults interpreting the intentions or goals of another person based on their actions involves the STS. In sum, three components important to RJA bids are mediated by the STS in adults, suggesting RJA is also mediated by the STS but this study has not yet been published.

In addition to joint attention ability, gesture has been shown to be important to language learning. Goldin-Meadow (2000) has suggested that an important behavior in cognitive development is the use and interpretation of gesture. Children who are about to progress in a cognitive area often show a mismatch between what they gesture and what they say. Auditory language comprehension has been shown to correlate with both comprehension and production of meaningful gestures in normal and aphasic adults (Wang and Goodglass, 1992). Gesture may aid in language production as well as comprehension. Congenitally blind children gesture while they speak even if speaking to other blind children (Iverson and Goldin-Meadow, 1997). This suggests that gesture is a tool to aid in the speaker’s, as well as the listener’s, comprehension. Furthermore, Bates and Dick (2002) review over 15 references marking language milestones and their gestural correlates between ages 6 and 30 months. Briefly, at age 6–8 months, canonical babbling occurs as well as rhythmic hand movements. During the 8–10-month period, word comprehension increases as do deictic gestures, such as showing and pointing, and gestural routines, such as waving goodbye. As word production begins, often between 11 and 13 months, recognitory gestures, such as putting a brush to the hair or a cup to the lip, are seen. Finally, at 18–20 months, when children typically begin using word combinations, gesture–word and gesture–gesture combinations are seen. These temporal correlations between gesture and language development suggests gesture may be tightly linked with language development. In fact, Goldin-Meadow (1998) suggests that “gesture and speech share a common cognitive representation; that is, before the communication unfolds, gesture and speech are part of a single idea.” As discussed above, the STS fires in response to meaningful gestures and speech in adults. Thus, the STS may serve as the neural link between gesture and speech, which after all “share a common cognitive representation.” Furthermore, research on native ASL signers reading ASL and English speakers reading English has shown that both utilize the STS,

although signers recruit greater right hemisphere STS than English speakers (Neville et al., 1998). This finding suggests that the STS is activated by ‘gesture speech’ or written speech. It would therefore make sense that a similar neural substrate for speech comprehension would be involved in the acquisition of ‘visual speech’, or gesture, comprehension.

A final behavior important to language learning discussed here is a child’s ability to temporally sequence and parse language. As a child’s production increases, so does the rate at which a word is anticipated in a sentence. Whereas 12-month-old infants may process the sentence “Where’s the Baby?” by the time the speaker got to “.by”, 24-month-old infants would respond by “Ba..” (Fernald et al., 2001). The STS’ role in learning temporal and acoustic sequences specific to speech stimuli may contribute to a faster processing rate as speech comprehension and production increase. Parsing in the visual domain appears much earlier in infants. Baldwin et al. (2001) show that 10–11-month-old infants are able to parse actions into discrete units. A study using fMRI with adults (Zacks et al., 2001) shows that the STS is involved in parsing events at least in the visual domain. In this study, subjects viewed movies of everyday events, such as ironing or making the bed, and were asked to segment the activity into natural and meaningful events (Zacks et al., 2001). A network of regions, including STS (Zacks et al., 2006), was found to be more active during the trials in which subjects marked a natural boundary. Activation was even greater when this boundary marked a large event than when it marked a small event. An analogous function in the language domain would be parsing natural boundaries for words or sentences.

While these behaviors do not account for all necessary components to language processing, they are nonetheless significant. A child’s ability to follow eye gaze and integrate auditory and visual inputs is important for the development of joint attention, which has been shown to be necessary for word learning (Moore et al., 1999). Furthermore, interpretation and performance of gestures are important to a child’s verbal and non-verbal communication skills and cognitive development. The common denominator across all of these functions sub-served by the STS is the parsing of a sequence of inputs into a meaningful, coherent whole.

### 3.2. *The STS becomes tuned to respond to meaningful input*

I propose that cells within the STS become tuned to respond to meaningful auditory input, such as words, and visual input, such as shifts in eye gaze or pointing, throughout development. STS neurons have been found in adult non-human primates that are responsive to specific sequences of actions that convey meaning. Jellema and Perrett (2003b) presented pictures of sequences of body postures to monkeys while recording from cells within the STS. If a monkey learned that posture ‘X’ following

posture ‘B’ is meaningful, cells would fire to ‘X’ if it followed ‘B’ but not if it followed posture ‘A’. Previous studies reveal involvement of the STS in perceiving point-light displays when they appear to move in a biological way (Grossman and Blake, 2002), however, when the dots forming the perception of biological motion are overlaid on noise, this perception becomes more difficult. When subjects were trained to recognize the moving figure form from the noise, activation was seen in right STS (Grossman et al., 2004). These findings suggest the STS is maximally responsive to coherent sequences of input and that perhaps with repetition, the STS can be tuned to decipher socially or linguistically meaningful from non-meaningful sequences of input. Early on, if the child learns that following gaze or gestures of the caretaker often results in finding a new or interesting object, and perhaps more importantly in jointly sharing in a social venture, these cues may become more salient to him or her. With repeated occurrences, the cells within the STS may receive feedback that this behavior is socially meaningful, and thus may become tuned to respond preferentially to specific eye gaze or gesture acts. The child may then begin to substitute words for gestures and requests for points.

By adulthood, the neural response to these words and requests is left-biased while the response to gestures and points is right-biased. However, early on this lateralization may not be so evident. Interestingly, language directed at infants tends to be rich in prosody, with wide pitch contours. The brain response to this type of prosodic speech is typically seen in right as opposed to left STS in adults. Perhaps the right STS is particularly critical in early language learning because the functions of the right STS (detecting prosody, biological motion, gestures, and interpreting intentions) are the communicative skills available to the pre- or peri-linguistic child. In fact, studies that have examined language skill in young children with early focal lesions suggests the right hemisphere, not left, is particularly important in word comprehension (review: Bates, 1997). Furthermore, a PET study of children 18 days to 19 years provides some evidence for a shift in metabolic activity from the right to left hemisphere at around the fourth year of life (Chiron et al., 1997). These findings suggest an early developmental importance of the right STS in particular for social processing and initial language learning. Perhaps, as speech perception requires increasingly rapid parsing, the left hemisphere, which is anatomically better suited for rapid auditory processing (Hutsler, 2003) may take over.

### 3.3. *Other regions involved in language learning*

As discussed above, this review is not intended to argue that behaviors mediated by the STS subserve all of language acquisition. Rather the function of the STS appears to be specific to perception, rather than production. For example, STS activation is elicited by hearing words, seeing biological motion and interpreting intentions, to

name a few. A child's productive activity through interaction with his or her own world is also of fundamental importance to acquiring language. Correlations are seen between an infant's initiating of joint attention or social interactions with the mother and language ability (Goldstein et al., 2003; Mundy et al., 2003). Studies have linked IJA as well as other social behaviors with prefrontal brain regions including medial frontal cortex (Mundy, 2003; Mundy et al., 2000, 2003).

Another region in prefrontal cortex, inferior frontal gyrus (IFG), has been linked to early language learning not only because of its role in adult language production and acquisition of a second language (Sakai, 2005) but also through behaviors that are important for language learning such as imitation. Mirror neurons, which are activated when an animal, or human, is planning a specific arm movement or gesture and when an animal observes someone else doing that same movement are found in area F5 (for which the human homolog is the inferior frontal gyrus and adjacent ventral premotor cortex) and rostral inferior parietal lobule (reviewed in Rizzolatti et al., 2001; Iacoboni, 2005). Thus, some researchers have focused on the role of the inferior frontal regions in the link between gesture, imitation, and language learning (e.g. Rizzolatti and Arbib 1998; Bates and Dick, 2002). Interestingly, fMRI activity that reflects a pattern congruent with mirror neuron activity is also observed in the STS (Iacoboni et al., 2001). While imitation of a gesture or word is important to word learning and production, the ability to signal meaning from a sequence of gestures, such as eye gaze or pointing, and from a sequence of auditory inputs is also crucial to receptive language acquisition.

Thus, a network of regions is clearly important in the acquisition of both expressive and receptive language. The evidence reviewed above on the role of the STS in the perception of speech, gesture, and in skills critical to early social communication and language learning, provides strong evidence for the STS as a neural component to the acquisition of receptive language.

#### **4. Given the above, abnormalities in this region may be critical to the emergence of autism**

Understanding the role of the STS in normal language acquisition could have a significant impact on our understanding of autism: a disorder characterized by abnormal language and social cognition. The above evidence suggests these two domains may share a common substrate of the STS and may be linked particularly early in development. Thus abnormalities in this region may be critical to the behavioral phenotype seen in autism. The STS has previously been proposed to be a critical component of the abnormal neural circuitry underlying deficits in social perception (Pelphrey et al., 2004; Zilbovicius et al., 2006) and basic visual perceptual abnormalities, such as motion perception (Dakin and Frith, 2005) in autism. I extend these proposals by suggesting that disruption to the STS,

particularly early in development, may contribute to the perceptual, social, and language deficits seen in autism.

#### *4.1. Anatomical evidence of STS abnormalities in autism*

A large body of neuroanatomical and neurofunctional evidence suggests the temporal lobes, and particularly the STS/STG, are abnormal in autism. Rapid brain growth is seen in autism between birth to 6–14 months of age (Courchesne et al., 2003). This rapid rate of brain growth continues into early childhood, after which time brain size is not significantly different from normal (Courchesne et al., 2001; Redcay and Courchesne, 2005). Evidence suggests this rapid growth occurs in an anterior to posterior gradient with frontal and temporal lobes most affected (Courchesne et al., 2004). Specifically, MRI volumetric studies reveal significant enlargement of gray matter in the frontal and temporal lobes in 2–4-year-old children with autism as compared to controls (Carper et al., 2002). A voxel-based morphometry study found decreased concentrations of gray matter in autism localized bilaterally to the superior temporal sulci (Boddaert et al., 2004b). Analysis of cortical sulcal maps in autistic children revealed bilateral anterior and superior shifting of the superior frontal sulci and right anterior shifting of the STS, sylvian fissure, and left inferior frontal sulcus (Levitt et al., 2003). The authors interpret anterior shifting of sulci as indicating delayed or incomplete sulcal development based on developmental sulcal mapping studies of typical children. Furthermore, cortical thickness has been shown to be reduced in a number of brain regions in autism and this reduction correlates significantly with autism symptoms in the STS and inferior frontal gyrus in addition to a number of parietal regions (inferior parietal lobule, supramarginal gyrus, and superior parietal lobule) and one occipital region (IOG). Reduced cortical thickness suggests either early primary neural abnormalities in these regions or secondary abnormalities resulting from aberrant cortical connectivity (Hadjikhani et al., 2006).

Temporal lobe abnormalities have also been found in post-mortem studies. Four of the six autism brains that Bailey et al. (1998) examined contained temporal lobe abnormalities. These abnormalities include mild widening of temporal sulci in one case (Case 1), abnormal temporal laminar patterns in two cases (Cases 1 and 4), thickening of the STG in two cases (Cases 2 and 4), and scattered mature neurons within white matter in the STG in one case (Case 5). Neuronal integrity, as measured by levels of NAA recorded during MRS Spectroscopy, is impaired in the lateral temporal regions in one study (Hisaoka et al., 2001). The primate anatomical literature has shown that the STS has rich projections to the frontal lobes (Seltzer and Pandya, 1989) and the cerebellum (Schmahmann and Pandya, 1991); areas which have been consistently implicated as abnormal or impaired in autism (Carper and Courchesne, 2000). Finally, a study of cerebral blood flow during sedation in children with autism revealed

hypoperfusion of bilateral STG/STS (Ohnishi et al., 2000; Zilbovicius et al., 2000) and in fact this hypoperfusion in the left STG/S correlated with degree of autism severity (Gendry Meresse et al., 2005). Thus, the above evidence suggests that dysfunction of the temporal lobes may play a significant role in the development of autism.

#### 4.2. Functional evidence of STS abnormalities in autism

Behavioral and neurofunctional data also suggest impairments of the STS in autism. As described above, a primary social cognitive function that recruits the STS is biological motion perception, or the identification of a social form from a moving entity, and attributing intentions or goals to that entity. In the language domain, the STS is involved in identifying linguistic units from a stream of auditory information and extracting the communicative significance of these units. Thus, behaviors that engage the STS include attributing intentions to others, perception of a social form from sparse, moving information, perception of the changeable aspects of faces such as eye gaze and expressions, complex motion perception, prosody perception, and narrative comprehension, to name a few. Many of these behaviors are impaired in autism and elicit abnormal STS activation. In the following section, behavioral and neurofunctional studies of these behaviors in autism will be reviewed.

Much behavioral work in autism reveals deficits in understanding that another person has a mental state that is different from one's own (or ToM). As reviewed above, the role of the STS in 'mentalizing' is primarily in determining the goals or intentions of a social being (or object) based on its actions. One such example is in determining the intention of another person based on eye gaze direction. Behavioral studies of eye gaze perception in autism suggest that individuals with autism can be cued by eye gaze direction or shifts in direction but that extracting meaning from eye gaze direction is not an automatic process and may be completely impaired. In one study, children were presented with pictures of a face below two candy bars; one to the left and one to the right (Baron-Cohen et al., 1995). The children were presented with images of the face looking towards one of two candy bars or looking forward (i.e. at the participant). They were asked to determine what candy bar the face in the picture wanted. Children with autism could detect the gaze shift but they were not able to answer which candy bar the face would want based on eye gaze direction. This suggests that while they processed a gaze shift, they were not able to extract meaning from the gaze direction. A second behavioral study found that children with autism were most impaired relative to their verbal mental age- and chronological age-matched controls at matching gaze direction and facial expressions and relatively less impaired at matching identity, gender, and lip-reading (Gepner et al., 2001). Thus, while perceiving and interpreting gaze directions and expressions may be a natural function for

typical individuals, the interpretation of gaze does not seem to be automatic in the case of autism.

An fMRI study by Pelphrey and colleagues (Pelphrey et al., 2005a) provides a neural correlate for these behavioral findings. Autism and control participants were presented with a face that made eye gaze shifts after a flickering checkerboard appeared in one of the four corners of the screen. In congruent trials, the eye gaze shifts were in the direction of the checkerboard, while in incongruent trials, the shifts were made to a corner of the screen that did not contain the checkerboard. STS activity was seen in both groups to shifts of eye gaze; however, while controls showed greater right STS activation to incongruent than to congruent gaze shifts, individuals with autism did not show differential STS activation between the conditions. The authors suggest the greater STS activation to incongruent than to congruent gaze shifts in controls reflects the greater processing demand placed on the STS when the action of the face violates the expectation of what the subject thought the face would do. In other words, greater STS activity may index mentalizing about the goals and intentions of the face on the part of the subject. Thus, the finding of no difference in STS activation between conditions suggests that in autism, the STS can be recruited to detect shifts in eye gaze; however, unlike controls, no greater processing demand is placed on the STS when the face acts in a way that is not congruent with the subjects' expectations of the actions or intentions of the face. Presumably, this is because expectations of another's intentions are not automatically formed in autism based on eye gaze shifts.

Intentions can be attributed based on actions even when the moving entity is not a human form. In one study, participants were presented with moving shapes, some of which appeared to be performing actions, such as coaxing or tickling (Castelli et al., 2002). Autistic subjects were less likely to attribute mental states to the action of the shapes. When viewing the actions that elicited mentalizing in controls, reduced functional activity was seen in the autism group in the medial prefrontal cortex, STS, and temporal-parietal junction. Interestingly, the difference between conditions is simply whether an intention, or mental state, can be attributed to the movement of geometric figures. In autism, STS activity did not reflect this distinction.

Social stimuli that change over time such as facial expressions and vocal sounds elicit greater STS activity in controls than in autism. Given the above studies, the reduced STS activity in autism may be due to an absence of extracting social communicative meaning of the gaze directions, expressions, or vocal sounds. In a study of face processing in autism, Pierce et al. (2001) presented a series of static faces that contained neutral expressions. The autism group showed reduced STS activity to faces (as well as reduced FG and amygdala activation) as compared to controls. In the auditory domain, additional STS processing is seen in controls, but not in autism, to the presentation of vocal as compared to non-vocal sounds

(Gervais et al., 2004). As described above, vocal sounds are stimuli of social relevance and have been described as auditory ‘faces’. Unlike static faces, vocal sounds contain a temporal component. Presumably the lack of additional STS processing in autism to vocal sounds reflects an absence of implicit social processing similar to that seen in the eye gaze and face tasks of Pelphrey et al. (2005a) and Pierce et al. (2001), respectively. The perception of a human activity from point-light animations is also impaired in children with autism (Blake et al., 2003) but surprisingly this has not yet been tested with fMRI. Presumably, activation would be seen in the STS to both conditions in both groups but controls would show greater activation to the point-light displays that were perceived as biological motion.

Based on findings from the studies above, it appears that while controls recruit additional activation within the STS to extract an agent’s motive, goal, or intention, subjects with autism do not. This would imply the STS abnormalities in autism are restricted to higher-order socio-emotional processing in autism. However, some evidence suggests that even more basic motion perception might be impaired in autism as well (reviewed in Dakin and Frith, 2005; Milne et al., 2005). When viewing a large radial flow, typical children make postural adjustments; however, this is not the case in autism. In fact, the degree of postural hypo-reactivity positively correlated with autism severity (Gepner and Mestre, 2002). Additionally, when viewing dots in motion, children with autism require a greater number of dots moving coherently to be able to detect their direction (Milne et al., 2002; Pellicano et al., 2005; Spencer et al., 2000). Some (Spencer et al., 2000) have attributed this to a dorsal stream deficit, while others (Pellicano et al., 2005) to a global integration deficit, and yet others (Bertone et al., 2003) to a difficulty in perceptual processing of complex stimuli. It has also been suggested that these more basic deficits may contribute in some way to the higher-order deficits seen in autism (Rivera and Koldewyn, 2005). It is unclear whether these behavioral abnormalities in motion perception in autism would be reflected in the STS as these tasks have not yet been performed with fMRI in autism to the best of my knowledge. The few extant imaging studies of motion processing, albeit higher-order motion processing, reviewed here suggest the STS is recruited in autism for some types of motion processing, but not to use motion cues to make an inference of another’s intentions or goals. Further imaging work of basic motion and biological motion perception in autism is needed to disentangle this apparent discrepancy.

An important observation regarding motion perception in autism is made by Gepner and Mestre (2002). They found autistic children were better at processing small squares moving if they were moving slowly (5–15 mm/s) than if they were moving quickly (30–50 mm/s). They suggest many of the social-emotional impairments, such as face and emotion processing can be explained by an underlying deficit in rapid motion processing. While they

propose that the abnormalities in the cerebellum and magnocellular pathways may account for this deficit, the STS may likely be a part of this circuit as well. As reviewed above, the STS is activated by stimuli of complex motion, or implied motion, and also shows a hierarchical response in the degree of activation with both complexity and socio-emotional significance of the stimuli. Perhaps, the lack of STS recruitment in higher-order tasks of social cognition involving emotion could be ameliorated with slowed presentation of these stimuli. In fact, slowed presentation of videos of actors performing an emotion did lead to emotion recognition in a group of autistic children (Gepner et al., 2001). However, this group of autistic children also was able to recognize emotions on still face actors, a finding contrary to previous studies (Gepner et al., 1994). Clearly this is a topic that needs to be re-addressed. Interestingly, anecdotal reports offer some support for the hypothesis that rapid temporal processing may underlie some of the communication deficits seen in autism. For example, an article by Noens and vanBerckelaer-Onnes (2005) describes an engineer with high functioning autism who “needs time to process incoming stimuli step-by-step. Since natural communication modes are so fast and transient, he frequently fails to understand the most essential information” (Noens and vanBerckelaer-Onnes, 2005, p. 143). Someone unable to parse rapidly changing input would likely have difficulties following and interpreting meaningful facial cues. Additionally, difficulties in processing rapid input over time could lead to difficulties in speech perception.

One diagnostic indicator of autism disorder is delayed or absent language. Children with autism are significantly impaired on both language comprehension and production. In a study utilizing the MacArthur Communicative Development Inventories (CDI), sentence and phrase *comprehension* were more delayed than word production relative to a normative sample. The authors suggest this may be due to a greater deficit in processing sequences of words, rather than words alone (Charman et al., 2003). Often it is noted that the typical sequence of word production following gains in comprehension, is not followed in autism, resulting in inappropriate word production. These findings suggest deficits in speech comprehension may be particularly detrimental to language development overall. The profile of language abnormalities in autism is different from language delay or specific language impairment (Fisher et al., 2005). Children with autism show the most consistent and greatest difference from typical children on language tasks that require an understanding of an other’s communicative intentions (reviewed in Sabbagh, 1999), such as prosody, metaphor, extracting themes from a narrative, or discourse (reviews, Lord and Paul, 1997; Wetherby et al., 2000). They perform best at responding to direct requests as opposed to indirect requests (Paul and Cohen, 1985) suggesting implicit processing of a speaker’s intent in language is impaired. Some linguistic structural abnormalities are

present, such as deficits in vocabulary, syntax, grammar, and morphology but these are more variable than pragmatic deficits (review, Tager-Flusberg and Joseph, 2003).

Only a handful of studies have examined the neural bases of language processing in subjects with autism and have found mixed results with regard to STS activation. Two studies of passive listening to speech-like sounds (one study was performed under sedation) revealed significantly reduced temporal lobe activity (STG and MTG) in autistic patients as compared to controls (Boddaert et al., 2003, 2004a). Interestingly, three fMRI studies of high-functioning adults with autism, one of irony comprehension (Wang et al., 2006), one of semantic vs. perceptual processing (Harris et al., 2006), and one of explicit sentence comprehension (Just et al., 2004), found greater activity in STG/S in participants with autism than in IQ matched controls. Two of these (Harris et al., 2006; Wang et al., 2006) contained a number of participants with Asperger's disorder. Asperger's disorder is characterized by deficits in social behavior but not by language impairments and thus studying language with a mix of Asperger's and autism participants could possibly confound group results. Nonetheless, these studies do offer valuable information that suggests that high-functioning adults with autism spectrum disorders (ASD) can engage STG/S for language processing. Given the pattern of behavioral abnormalities, it is possible that high-functioning ASD individuals are able to perform a language task (although often with more effort and poorer performance), but, unlike controls, without explicit instruction or other motivation they are not compelled to do so. This is analogous to evidence from the social perception literature as discussed above. Behavioral evidence does show that autistic individuals are worse at extracting speech signal from noise (Alcantara et al., 2004) and show reduced orienting to speech as compared to non-speech sounds (Klin, 1991). Together these findings suggest speech perception is not an automatic process in autism and this lack of automatic orienting, perception, and interpretation of speech may be a critical deficit. Perhaps, as discussed above, the lack of automatic interpretation of speech is due to a deficit in rapid "auditory motion" integration. Speech requires a very quick integration of multiple words, stress patterns, and intonational cues. High-functioning autistic individuals may be able to parse the simple word content but integrating other contextual cues creates a task that is too complex.

#### 4.3. Role of connectivity in STS impairments in autism

A large body of neuroanatomical, neurofunctional, behavioral, and anecdotal evidence (reviewed above) suggests the temporal lobes and particularly the STS/STG, are impaired in autism. While more research is required, an emerging picture of STS dysfunction in autism is not specifically in the ability to parse streams of

information, but rather the ability to extract meaning from them automatically. However, some evidence suggests there is a deficit in integration of rapidly presented information (Gepner and Mestre, 2002; Noens and vanBerckelaer-Onnes, 2005) or complex motion processing (Bertone et al., 2003). The ability to extract meaning from inputs may require not only intact STS but also intact input from other higher-order association regions such as frontal cortex. Much current research has identified abnormalities in both structural connectivity (which refers to the anatomical connections between regions as measured through integrity of white matter tracts or volume of white matter) and functional connectivity (which refers to the coherence in activity patterns between regions) in autism (Belmonte et al., 2004; Courchesne and Pierce, 2005). For example, white matter tracts (measured through DTI) are abnormal (Barnea-Goraly et al., 2004). White matter volume is increased at young ages in autism (Courchesne et al., 2001; Herbert et al., 2004) and higher-order association cortices are relatively more impaired structurally (Carper and Courchesne, 2005; Carper et al., 2002). Studies of functional connectivity have found reduced connectivity in autism, particularly within and between frontal and parietal (or occipital) regions, during sentence comprehension judgments (Just et al., 2004), executive function (Just et al., 2006), working memory (Koshino et al., 2005), and visuomotor performance (Villalobos et al., 2005). The only study to examine connectivity during a task that produced reduced STS activity in autism (Castelli et al., 2002) found reduced connectivity between STS and extrastriate cortex. In this study, the connectivity was assessed relative to a region in extrastriate cortex, not STS. It would be interesting to know if reduced connectivity was present between the STS and medial prefrontal cortex in this study and whether that reduced prefrontal input contributed to the inability to attribute intentions to the shapes. At the present time, however, it is not clear how much prefrontal cortical input contributes to TOM processes in the STS.

## 5. Conclusion

I argue that a common mechanism for both social and speech perception is performed by the STS, namely parsing rapidly changing auditory and visual input and extracting meaning from this input. The ability to integrate changing auditory and visual cues (both within and across modalities) and extract social or communicative significance of these cues may be critical to word learning. Based on the behavioral and functional literature currently available, the particular deficit in STS function in autism may be the automatic extraction of meaning from rapidly presented visual or auditory cues. However, some deficits in simple parsing of these rapidly presented cues may also exist. This inability to assign meaning to cues may be an underlying cause for the impaired language learning, particularly in pragmatic aspects of language, which are characteristic of

autism. Further research is needed to determine the role of the STS and its connectivity with other regions in the emergence of the pervasive communicative deficits seen in autism.

A complex neurobehavioral disorder such as autism cannot be explained by abnormalities in one brain area, nor is language acquisition dependent on the development of just one area. Development progresses through a complex interaction of many brain and behavioral changes and experiences. The current review aimed to focus on the role of an adult ‘language comprehension region’ from a developmental ontogenetic perspective based on evidence of its role in both non-verbal behaviors important to language acquisition and speech perception itself. Evidence that this area may be abnormal in a disorder characterized by language and social communication adds additional support to this hypothesis and suggests STS dysfunction may be a critical contributing factor in the emergence of the autistic phenotype.

### Acknowledgments

I am grateful to Dr. Eric Courchesne, Dr. Laura Schreibman, Dr. Leslie Carver, Dr. Victor Ferreira, and Dan Kennedy for insightful comments and discussion on the manuscript.

### References

- Akhtar, N., Tomasello, M., 1996. Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology* 14, 79–93.
- Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Nakachi, R., Kashima, H., 2006a. A deficit in discriminating gaze direction in a case with right superior temporal gyrus lesion. *Neuropsychologia* 44, 161–170.
- Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Umeda, S., Kashima, H., 2006b. Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia* 44, 1804–1810.
- Alcantara, J.I., Weisblatt, E.J.L., Moore, B.C.J., Bolton, P.F., 2004. Speech-in-noise perception in high-functioning individuals with autism or Asperger’s syndrome. *Journal of Child Psychology and Psychiatry* 45, 1107–1114.
- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. *Trends in Cognitive Science* 4, 267–278.
- Andrews, T.J., Ewbank, M.P., 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage* 23, 905–913.
- Bailey, A., Luthert, P., Dean, A., Harding, B., Janota, I., Montgomery, M., et al., 1998. A clinicopathological study of autism. *Brain* 121, 889–905.
- Baldwin, D.A., 1991. Infants’ contribution to the achievement of joint reference. *Child Development* 62, 875–890.
- Baldwin, D.A., Baird, J.A., Saylor, M.M., Clark, M.A., 2001. Infants parse dynamic action. *Child Development* 72, 708–717.
- Baldwin, D.A., Markman, E.M., Bill, B., Desjardins, R.N., Irwin, J.M., Tidball, G., 1996. Infants’ reliance on a social criterion for establishing word–object relations. *Child Development* 67, 3135–3153.
- Ban, T., Shiwa, T., Kawamura, K., 1991. Cortico-cortical projections from the prefrontal cortex to the superior temporal sulcal area (STS) in the monkey studied by means of HRP method. *Archives Italiennes de Biologie* 129, 259–272.
- Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., Reiss, A.L., 2004. White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biological Psychiatry* 55, 323–326.
- Barnes, C.L., Pandya, D.N., 1992. Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *Journal of Comparative Neurology* 318, 222–244.
- Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J., Walker, J., 1995. Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology* 13.
- Barracough, N.E., Xiao, D., Baker, C.I., Oram, M.W., Perrett, D.I., 2005. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience* 17, 377–391.
- Bates, E., 1997. Origins of language disorders: a comparative approach. *Developmental Neuropsychology* 13, 447–476.
- Bates, E., Dick, F., 2002. Language, gesture, and the developing brain. *Developmental Psychobiology* 40, 293–310.
- Bates, E., Thal, D., Finlay, B., Clancy, B., 2003. Early language development and its neural correlates. *Child Neuropsychology* 8 (Part II), 525–592.
- Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159.
- Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience* 7, 1190–1192.
- Belin, P., Fecteau, S., Bedard, C., 2004. Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences* 8, 129–135.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. *Nature* 403, 309–312.
- Belmonte, M., Cook Jr., E.H., Anderson, G., Rubenstein, J., Greenough, W., Beckel-Mitchener, A., et al., 2004. Autism as a disorder of neural information processing: directions for research and targets for therapy. *Molecular Psychiatry* 9, 646–663.
- Bertone, A., Mottron, L., Jelenic, P., Faubert, J., 2003. Motion perception in autism: a “complex” issue. *Journal of Cognitive Neuroscience* 15, 218–225.
- Benevento, L.A., Fallon, J., Davis, B.J., Rezak, M., 1977. Auditory–visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental Neurology* 57, 849–872.
- 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.
- Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., et al., 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex* 10, 512–528.
- Blake, R., Turner, L.M., Smoski, M.J., Pozdol, S.L., Stone, W.L., 2003. Visual recognition of biological motion is impaired in children with autism. *Psychological Science* 14, 151–157.
- Boddaert, N., Belin, P., Chabane, N., Poline, J.B., Barthelemy, C., Mouren-Simeoni, M.C., et al., 2003. Perception of complex sounds: abnormal pattern of cortical activation in autism. *American Journal of Psychiatry* 160, 2057–2060.
- Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., et al., 2004a. Perception of complex sounds in autism: abnormal auditory cortical processing in children. *American Journal of Psychiatry* 161, 2117–2120.
- Boddaert, N., Chabane, N., Gervais, H., Good, C.D., Bourgeois, M., Plumet, M.H., et al., 2004b. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *NeuroImage* 23, 364–369.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience* 8, 389–395.
- Bruce, C., Desimone, R., Gross, C.G., 1981. Visual properties of neurons in polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology* 46, 369–384.



- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., et al., 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). *Cognitive Brain Research* 12, 233–243.
- Carpenter, M., Nagell, K., Tomasello, M., 1998. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society of Research in Child Development* 63 (i–vi), 1–143.
- Carper, R.A., Courchesne, E., 2000. Inverse correlation between frontal lobe and cerebellum sizes in children with autism. *Brain* 123, 836–844.
- Carper, R.A., Courchesne, E., 2005. Localized enlargement of the frontal cortex in early autism. *Biological Psychiatry* 57, 126–133.
- Carper, R.A., Moses, P., Tigue, Z.D., Courchesne, E., 2002. Cerebral lobes in autism: early hyperplasia and abnormal age effects. *Neuroimage* 16, 1038–1051.
- Castelli, F., Frith, C., Happe, F., Frith, U., 2002. Autism, asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125, 1839–1849.
- Charman, T., Drew, A., Baird, C., Baird, G., 2003. Measuring early language development in preschool children with autism spectrum disorder using the macarthur communicative development inventory (infant form). *Journal of Child Language* 30, 213–236.
- Chiron, C., Jambaqué, I., Nabbout, R., Lounes, R., Syrota, A., Dulac, O., 1997. The right brain hemisphere is dominant in human infants. *Brain* 120, 1057–1065.
- Courchesne, E., Pierce, K., 2005. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current Opinion in Neurobiology* 15, 225–230.
- Courchesne, E., Karns, C., Davis, H.R., Ziccardi, R., Carper, R., Tigue, Z., et al., 2001. Unusual brain growth patterns in early life in patients with autistic disorder: an MRI study. *Neurology* 57, 245–254.
- Courchesne, E., Carper, R., Akshoomoff, N., 2003. Evidence of brain overgrowth in the first year of life in autism. *Journal of the American Medical Association* 290, 337–344.
- Courchesne, E., Redcay, E., Kennedy, D.P., 2004. The autistic brain: birth through adulthood. *Current Opinion in Neurology* 17, 489–496.
- Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.J., 2003. Temporal lobe regions engaged during normal speech comprehension. *Brain* 126, 1193–1201.
- Cusick, C.G., Seltzer, B., Cola, M., Griggs, E., 1995. Chemoarchitectonics and corticocortical terminations within the superior temporal sulcus of the rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex. *Journal of Comparative Neurology* 360, 513–535.
- Dakin, S., Frith, U., 2005. Vagaries of visual perception in autism. *Neuron* 48, 497–507.
- Deak, G.O., Flom, R.A., Pick, A.D., 2000. Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology* 36, 511–523.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., Dehaene, S., 2005. Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33.
- Delgado, C.E., Peter, M., Crowson, M., Markus, J., Yale, M., Schwartz, H., 2002. Responding to joint attention and language development: a comparison of target locations. *Journal of Speech, Language, and Hearing Research* 45, 715–719.
- Diesendruck, G., Markson, L., Akhtar, N., Reudor, A., 2004. Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science* 7, 33–41.
- Ekman, P., Friesen, W.V., 1978. Facial action coding system: A technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press.
- Engelien, A., Tuscher, O., Hermans, W., Isenberg, N., Eidelberg, D., Frith, C., et al., 2006. Functional neuroanatomy of non-verbal semantic sound processing in humans. *Journal of Neural Transmission* 113, 599–608.
- Fecteau, S., Armony, J.L., Joannette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23, 840–848.
- Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., Pethick, S.J., 1994. Variability in early communicative development. *Monographs of the Society of Research on Child Development* 59, 1–173 (discussion 174–85).
- Fernald, A., Pinto, J.P., Swingle, D., Weinberg, A., McRoberts, G.W., 2001. Rapid gains in speech of verbal processing by infants in the 2nd year. Language development: the essential readings. *Essential readings in developmental psychology* vii, 375.
- Fisher, N., Happe, F., Dunn, J., 2005. The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry* 46, 409–419.
- Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* 358, 459–473.
- Gallagher, H.L., Frith, C.D., 2004. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia* 42, 1725–1736.
- Gendry Meresse, I., Zilbovicius, M., Boddaert, N., Robel, L., Philippe, A., Sfaello, I., et al., 2005. Autism severity and temporal lobe functional abnormalities. *Annals of Neurology* 58, 466–469.
- Gepner, B., Mestre, D., 2002. Rapid visual–motion integration deficit in autism. *Trends in Cognitive Sciences* 6, 455.
- Gepner, B., De Schonen, S., Buttin, C., 1994. Face processing in young autistic children. *Infant Behavior and Development* 17, 661.
- Gepner, B., Deruelle, C., Grynfeldt, S., 2001. Motion and emotion: a novel approach to the study of face processing by young autistic children. *Journal of Autism and Developmental Disorders* 31, 37–45.
- Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., et al., 2004. Abnormal cortical voice processing in autism. *Nature Neuroscience* 7, 801–802.
- Giraud, A.L., Price, C.J., 2001. The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience* 13, 754–765.
- Goldin-Meadow, S., 1998. The development of gesture and speech as an integrated system. *New Directions for Child Development* 79.
- Goldin-Meadow, S., 2000. Beyond words: the importance of gesture to researchers and learners. *Child Development* 71, 231–239.
- Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences* 100, 8030–8035.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., et al., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience* 8, 145–146.
- Grossman, E.D., Blake, R., 2002. Brain areas active during visual perception of biological motion. *Neuron* 35, 1167–1175.
- Grossman, E.D., Blake, R., Kim, C.-Y., 2004. Learning to see biological motion: brain activity parallels behavior. *Journal of Cognitive Neuroscience* 16, 1669–1679.
- Hadjikhani, N., Joseph, R.M., Snyder, J., Tager-Flusberg, H., 2006. Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex* 16, 1276–1282.
- Harris, G.J., Chabris, C.F., Clark, J., Urban, T., Aharon, I., Steele, S., et al., 2006. Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain and Cognition* 61, 54–68.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends in Cognitive Sciences* 4, 223–233.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2002. Human neural systems for face recognition and social communication. *Biological Psychiatry* 51, 59–67.
- Herbert, M.R., Ziegler, D.A., Makris, N., Filipek, P.A., Kemper, T.L., Normandin, J.J., et al., 2004. Localization of white matter volume

- increase in autism and developmental language disorder. *Annals Neurology* 55, 530–540.
- Hisaoka, S., Harada, M., Nishitani, H., Mori, K., 2001. Regional magnetic resonance spectroscopy of the brain in autistic individuals. *Neuroradiology* 43, 496–498.
- Hoffman, E.A., Haxby, J.V., 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience* 3, 80–84.
- Hooker, C.I., Paller, K.A., Gitelman, D.R., Parrish, T.B., Mesulam, M.-M., Reber, P.J., 2003. Brain networks for analyzing eye gaze. *Cognitive Brain Research* 17, 406–418.
- Hutsler, J.J., 2003. The specialized structure of human language cortex: pyramidal cell size asymmetries within auditory and language-associated regions of the temporal lobes. *Brain and Language* 86, 226–242.
- Iacoboni, M., 2005. Neural mechanisms of imitation. *Current Opinion in Neurobiology* 15, 1–6.
- Iacoboni, M., Koski, L.M., Brass, M., Harold, B., Woods, R., Dubeau, M.-C. et al., 2001. Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences* 98, 13995–13999.
- Iverson, J.M., Goldin-Meadow, S., 1997. What's communication got to do with it? Gesture in children *blind* from birth. *Developmental Psychology* 33, 453–467.
- Jellema, T., Perrett, D.I., 2003a. Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia* 41, 1728–1737.
- Jellema, T., Perrett, D.I., 2003b. Perceptual history influences neural responses to face and body postures. *Journal of Cognitive Neuroscience* 15, 961–971.
- Jellema, T., Baker, C.I., Wicker, B., Perrett, D.I., 2000. Neural representation for the perception of the intentionality of actions. *Brain and Cognition* 44, 280–302.
- Just, M.A., Cherkassky, V.L., Keller, T.A., Minshew, N.J., 2004. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain* 127, 1811–1821.
- Just, M.A., Cherkassky, V.L., Keller, T.A., Kana, R.K., Minshew, N.J., 2006. Functional and anatomical cortical underconnectivity in autism: evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*.
- Kingstone, A., Tipper, C., Ristic, J., Ngan, E., 2004. The eyes have it!: an fMRI investigation. *Brain and Cognition* 55, 269–271.
- Klin, A., 1991. Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders* 21, 29–42.
- Koshino, H., Carpenter, P.A., Minshew, N.J., Cherkassky, V.L., Keller, T.A., Just, M.A., 2005. Functional connectivity in an fMRI working memory task in high-functioning autism. *Neuroimage* 24, 810–821.
- LaBar, K.S., Crupain, M.J., Voyvodic, J.T., McCarthy, G., 2003. Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex* 13, 1023–1033.
- Levanen, S., Uutela, K., Salenius, S., Hari, R., 2001. Cortical representation of sign language: comparison of deaf signers and hearing nonsigners. *Cerebral Cortex* 11, 506–512.
- Levitt, J.G., Blanton, R.E., Smalley, S., Thompson, P.M., Guthrie, D., McCracken, J.T., et al., 2003. Cortical sulcal maps in autism. *Cerebral Cortex* 13, 728–735.
- Lewis, J.W., Wightman, F.L., Brefczynski, J.A., Phinney, R.E., Binder, J.R., DeYoe, E.A., 2004. Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex* 14, 1008–1021.
- Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Janik, J.J., DeYoe, E.A., 2005. Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience* 25, 5148–5158.
- Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. *Cerebral Cortex* 15, 1621–1631.
- Lord, C., Paul, R., 1997. Language and communication in autism. *Handbook of autism spectrum and pervasive developmental disorders*, 195–225.
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A.S., McGuire, P.K., et al., 2004. Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage* 22, 1605–1618.
- McCullough, S., Emmorey, K., Sereno, M., 2005. Neural organization for recognition of grammatical and emotional facial expressions in deaf asl signers and hearing nonsigners. *Brain Research. Cognitive Brain Research* 22, 193–203.
- Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., Plaisted, K., 2002. High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 43, 255–263.
- Milne, E., Swettenham, J., Campbell, R., 2005. Motion perception and autistic spectrum disorder. A review. *Current Psychology of Cognition*. 23, 3–34.
- Molfese, D.L., 1990. Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain and Language* 38, 345–363.
- Moore, C., Angelopoulos, M., Bennett, P., 1999. Word learning in the context of referential and salience cues. *Developmental Psychology* 35, 60–68.
- Morales, M., Mundy, P., Rojas, J., 1998. Following the direction of gaze and language development in 6-month olds. *Infant Behavior and Development* 12, 373–377.
- Morales, M., Mundy, P., Delgado, C.E.F., Yale, M., Messinger, D., Neal, R., et al., 2000. Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology* 21, 283–298.
- Morris, J.P., Pelphrey, K.A., McCarthy, G., 2005. Regional brain activation evoked when approaching a virtual human on a virtual walk. *Journal of Cognitive Neuroscience* 17, 1744–1752.
- Mottron, R., Calvert, G.A., Jaaskelainen, I.P., Matthews, P.M., Thesen, T., Tuomainen, J., et al., 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage* 30, 563–569.
- Muller, R.-A., Kleinhans, N., Courchesne, E., 2001. Broca's area and the discrimination of frequency transitions: a functional MRI study. *Brain and Language* 76, 70–76.
- Mundy, P., 2003. Annotation: the neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry* 44, 793–809.
- Mundy, P., Gomes, A., 1998. Individual differences in joint attention skill development in the second year. *Infant Behavior and Development* 21, 469–482.
- Mundy, P., Card, J., Fox, N., 2000. EEG correlates of the development of infant joint attention skills. *Developmental Psychobiology* 36, 325–338.
- Mundy, P., Fox, N., Card, J., 2003. EEG coherence, joint attention and language development in the second year. *Developmental Science* 6, 48–54.
- Nakamura, A., Maess, B., Knosche, T.R., Gunter, T.C., Bach, P., Friederici, A.D., 2004. Cooperation of different neuronal systems during hand sign recognition. *NeuroImage* 23, 25–34.
- Neville, H.J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., et al., 1998. Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proceedings of the National Academy of Sciences* 95, 922–929.
- Noens, I.L., vanBerckelaer-Onnes, I.A., 2005. Captured by details: sense-making, language and communication in autism. *Journal of Communication Disorders* 38, 123–141.
- Noguchi, Y., Kaneoke, Y., Kakigi, R., Tanabe, H.C., Sadato, N., 2005. Role of the superior temporal region in human visual motion perception. *Cerebral Cortex* 15, 1592–1601.

- Ohnishi, T., Matsuda, H., Hashimoto, T., Kunihiro, T., Nishikawa, M., Uema, T., et al., 2000. Abnormal regional cerebral blood flow in childhood autism. *Brain* 123, 1838–1844.
- Paul, R., Cohen, D.J., 1985. Comprehension of indirect requests in adults with autistic disorders and mental retardation. *Journal of Speech and Hearing Research* 28, 475–479.
- Pellicano, E., Gibson, L., Maybery, M., Durkin, K., Badcock, D.R., 2005. Abnormal global processing along the dorsal visual pathway in autism: a possible mechanism for weak visuospatial coherence? *Neuropsychologia* 43, 1044–1053.
- Pelphrey, K., Adolphs, R., Morris, J.P., 2004. Neuroanatomical substrates of social cognition dysfunction in autism. *Mental Retardation and Developmental Disabilities Research Reviews* 10, 259–271.
- Pelphrey, K.A., Morris, J.P., 2006. Brain mechanisms for interpreting the actions of others from biological-motion cues. *Current Directions in Psychological Science* 15, 136–140.
- Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., McCarthy, G., 2003. Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *Journal of Neuroscience* 23, 6819–6825.
- Pelphrey, K.A., Viola, R.J., McCarthy, G., 2004. When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science* 15, 598–603.
- Pelphrey, K.A., Morris, J.P., McCarthy, G., 2005a. Neural basis of eye gaze processing deficits in autism. *Brain* 128, 1038–1048.
- Pelphrey, K.A., Morris, J.P., Michelich, C.R., Allison, T., McCarthy, G., 2005b. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cerebral Cortex* 15, 1866–1876.
- Petitto, L.A., Zatorre, R.J., Gauna, K., Nikelski, E.J., Dostie, D., Evans, A.C., 2000. Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Academy of Sciences* 97, 13961–13966.
- Peuskens, H., Vanrie, J., Verfaillie, K., Orban, G.A., 2005. Specificity of regions processing biological motion. *European Journal of Neuroscience* 21, 2864–2875.
- Pierce, K., Müller, R.-A., Ambrose, J., Allen, G., Courchesne, E., 2001. Face processing occurs outside the fusiform ‘face area’ in autism: evidence from fMRI. *Brain* 124, 2059–2073.
- Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience* 18, 2188–2199.
- Redcay, E., Courchesne, E., 2005. When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biological Psychiatry* 58, 1–9.
- Rimol, L.M., Specht, K., Weis, S., Savoy, R., Hugdahl, K., 2005. Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. *Neuroimage* 26, 1059–1067.
- Rivera, S.M., Koldewyn, K., 2005. Unraveling the mystery of motion perception impairments in autism: some further consideration. *Current Psychology of Cognition* 23, 189–197.
- Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. *Trends in Neuroscience* 21, 188–194.
- Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience* 2, 661–670.
- Roder, B., Stock, O., Neville, H., Bien, S., Rosler, F., 2002. Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: a functional magnetic resonance imaging study. *Neuroimage* 15, 1003–1014.
- Sabbagh, M.A., 1999. Communicative intentions and language: evidence from right-hemisphere damage and autism. *Brain and Language* 70, 29–69.
- Sadato, N., Yamada, H., Okada, T., Yoshida, M., Hasegawa, T., Matsuki, K.-I., et al., 2004. Age-dependent plasticity in the superior temporal sulcus in deaf humans: a functional MRI study. *BMC Neuroscience* 5, 56.
- Sakai, K.L., 2005. Language acquisition and brain development. *Science* 310, 815–819.
- Sakai, K.L., Tatsuno, Y., Suzuki, K., Kimura, H., Ichida, Y., 2005. Sign and speech: amodal commonality in left hemisphere dominance for comprehension of sentences. *Brain* 128, 1407–1417.
- Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else’s belief. *Nature Neuroscience* 7, 499–500.
- Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia* 42, 1435–1446.
- Schmahmann, J.D., Pandya, D.N., 1991. Projections to the basis pontis from the superior temporal sulcus and superior temporal region in the rhesus monkey. *Journal of Comparative Neurology* 308, 224–248.
- Schultz, J., Imamizu, H., Kawato, M., Frith, C.D., 2004. Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience* 16, 1695–1705.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123 (Part 12), 2400–2406.
- Scott, S.K., Rosen, S., Lang, H., Wise, R.J., 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. *Journal of the Acoustical Society of America* 120, 1075–1083.
- Seltzer, B., Pandya, D.N., 1989. Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *Journal of Comparative Neurology* 281, 97–113.
- Seltzer, B., Cola, M.G., Gutierrez, C., Masee, M., Weldon, C., Cusick, C.G., 1996. Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. *Journal of Comparative Neurology* 370, 173–190.
- Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., et al., 2006. “what” versus “where” in the audiovisual domain: an fMRI study. *Neuroimage* 33, 672–680.
- Specht, K., Reul, J., 2003. Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage* 20, 1944–1954.
- Spencer, J., O’Brien, J., Riggs, K., Braddick, O., Atkinson, J., Wattam-Bell, J., 2000. Motion processing in autism: evidence for a dorsal stream deficiency. *Neuroreport* 11, 2765–2767.
- Spiers, H.J., Maguire, E.A., 2006. Spontaneous mentalizing during an interactive real world task: an fMRI study. *Neuropsychologia* 44, 1674–1682.
- Tager-Flusberg, H., Joseph, R.M., 2003. Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences* 358, 303–314.
- Tomasello, M., Todd, J., 1983. Joint attention and early lexical acquisition style. *First Language* 4, 197–212.
- Tomasello, M., Barton, M.E., 1994. Learning words in nonostensive contexts. *Developmental Psychology* 30, 639–650.
- Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., Belliveau, J.W., 2001. Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences* 98, 11656–11661.
- Villalobos, M.E., Mizuno, A., Dahl, B.C., Kemmotsu, N., Muller, R.-A., 2005. Reduced functional connectivity between v1 and inferior frontal cortex associated with visuomotor performance in autism. *NeuroImage* 25, 916–925.
- Vouloumanos, A., Kiehl, K.A., Werker, J.F., Liddle, P.F., 2001. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience* 13, 994–1005.
- Wang, A.T., Lee, S.S., Sigman, M., Dapretto, M., 2006. Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain* 129, 932–943.

- Wang, L., Goodglass, H., 1992. Pantomime, praxis, and aphasia. *Brain and Language* 42, 402–418.
- Wetherby, A., Prizant, B.M., Schuler, A.L., 2000. Understanding the nature of communication and language impairments. *Autism spectrum disorders: a transactional developmental perspective*, 109–142.
- Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al., 2005. Identification of emotional intonation evaluated by fMRI. *NeuroImage* 24, 1233–1241.
- Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within 'wernicke's area'. *Brain* 124, 83–95.
- Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex* 13, 1034–1043.
- Xu, J., Kemeny, S., Park, G., Frattali, C., Braun, A., 2005. Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage* 25, 1002–1015.
- Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., et al., 2001. Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience* 4, 651–655.
- Zacks, J.M., Swallow, K.M., Vettel, J.M., McAvoy, M.P., 2006. Visual motion and the neural correlates of event perception. *Brain Research* 1076, 150–162.
- Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 11, 946–953.
- Zilbovicius, M., Boddaert, N., Belin, P., Poline, J.B., Remy, P., Mangin, J.F., Thivard, L., Barthelemy, C., Samson, Y., 2000. Temporal lobe dysfunction in childhood autism: a pet study. *Positron emission tomography. American Journal of Psychiatry* 157, 1988–1993.
- Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., Boddaert, N., 2006. Autism, the superior temporal sulcus and social perception. *Trends in Neuroscience* 29, 359–366.