

John Benjamins Publishing Company



This is a contribution from *Why Gesture? How the hands function in speaking, thinking and communicating*.

Edited by R. Breckinridge Church, Martha W. Alibali and Spencer D. Kelly.

© 2017. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Exploring the boundaries of gesture-speech integration during language comprehension

Spencer D. Kelly
Colgate University

The present review explores the integrated relationship between gesture and speech during language comprehension. Taking a broad view, it presents a conceptual framework that approaches the comprehension of gesture and speech along three different dimensions: (1) components of language (semantics, pragmatics, phonetics and syntax), (2) levels of analysis (social, cognitive, and biological) and (3) timeframes of integration (online, moment-to-moment, developmental). The evidence suggests that some linguistic components (e.g., concrete semantic and pragmatic) are deeply connected to gestures, but others (e.g., abstract semantic, syntactic and phonetic) are less so. In this way, the hands help to delineate what aspects of language function as part of the body and what aspects operate independently of it.

General introduction

There has been a shift over the past three decades from viewing language as a relatively encapsulated and modular ability (Chomsky, 1980; Fodor, 1983) to an embodied and grounded one (Gibbs, 2006; Glenberg & Kaschak, 2002; Lakoff & Johnson, 1999; Zwaan, 2004). Indeed, it is now quite in vogue to approach language not as some abstract and disembodied ability, but as a concrete process that is deeply rooted in the body. In the present chapter, I explore the possibility that both views are correct for different aspects of language. Some components may be fundamentally grounded in the body, whereas others may operate somewhat independently from it.¹ From this perspective, the role of gesture in communication

1. Of course, technically, all of language is tied to the body through its production and comprehension. The distinction here concerns the extent to which the physical constraints of the body – through the systems of perception and action – are a relatively small or large part of what fundamentally *constitutes* language (for more on this distinction in cognition generally, see Barsalou, 1999).

may not be as straightforward as previously assumed: It may function as part of – but also separate from – language.

To appreciate why language may have deep roots in the body, it is useful to consider the evolutionary context for how language came into existence. It is now widely accepted that modern language emerged as part of a more distributed bodily communication system, with hand gestures in particular receiving attention as the potential starting point for human language (Armstrong & Wilcox, 2007; Corballis, 2003; Tomasello, 2008). One major reason for this focused attention on gesture is the growing research in neuroscience demonstrating a strong neural relationship between language and action in the human brain (for reviews, see Bates & Dick, 2002; Kelly et al., 2002; Pulvermüller, 2005; Rizzolatti & Arbib, 1998).

If language and gesture have a deep evolutionary relationship, it makes sense to explore the remnants of that connection in present day communication (Povinelli, 1993). In fact, many researchers have argued that speech and gesture remain tightly bound in current language use (Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992, 2008, 2012). Perhaps the most thoroughly developed and widely investigated claim is David McNeill's theory of gesture-speech integration. For McNeill, gesture and speech are part of the same communication system, with the two modalities dividing up a message in different but complementary representational formats. One way to think about this relationship is to consider that gesture adds what is *newsworthy* to speech (McNeill, 2008). That is, gestures imagistically index what is novel or relevant in a spoken utterance within a given context. For example, making a "tall" gesture would add newsworthy information to the utterance, "No, it was other guy," by providing an important context for interpreting the accompanying speech, thereby highlighting the relevant content of the message. Thinking along these lines, it is productive to ask what aspects of spoken language are open to this "news" delivered through gestures.

McNeill's theory began in the realm of language production, but my colleagues and I have advanced the "Integrated Systems Hypothesis" (ISH) as an explicit and formal test of the theory in the domain of language comprehension (Kelly, Özyürek & Maris, 2010; Özyürek, 2014). The present review will adopt the general spirit of the "ISH" and ask the following question: To what extent is gesture a fundamental part of language comprehension, and to what extent does language comprehension operate independently from gesture?

Converging methods

Understanding the extent to which gesture and speech are integrated during language comprehension requires taking a "robust" approach (Wimsatt, 1981). That

is, in order to thoroughly understand the mechanisms for this integration, it is necessary to look for converging evidence along multiple dimensions.² In this review, I will consider three dimensions in particular: (1) different components of language (semantics, pragmatics, phonetics and syntax), (2) different levels of analysis (social, cognitive, and biological) and (3) different timeframes of integration (online, moment-to-moment, developmental).

The first dimension approaches language by breaking it down to its traditional linguistic components. The majority of research on gesture in language comprehension has focused on semantic and pragmatic processes, but other key components of language, such as phonetics and syntax, have received much less attention. Isolating these sub-components of language allows for an exploration of the potential upper and lower boundaries of the integrated relationship between gesture and speech as a communicative system.

Focusing on level of analysis, much of the early research on gesture-speech integration used behavioral techniques, such as transcript analyses of spoken and gestured narratives, descriptions or explanations (Goldin-Meadow, Alibali & Church, 1993; Kendon, 1997; McNeill, 1992). Although we have learned a great deal about gesture from measuring it in these ways, there are many interesting questions that cannot be addressed using overt behavioral measures. Indeed, by looking beneath the behavioral surface of gesture and more directly at its neural processing, we might gain new insights that would be missed by focusing only on higher levels of analysis (e.g., social and cognitive). For example, does the brain process gesture the same way it processes speech during comprehension? When and where are gestures integrated with speech in the neural comprehension of language? Does the brain treat gestures as a special type of visual information? Indeed, cognitive neuroscience techniques, such as fMRI and ERPs, often reveal aspects of processing that are missed by traditional behavioral measures.

Regarding the third dimension, timeframe, traditional claims that gesture and speech form an integrated system in language production focused on a relatively narrow timespan, within an utterance or from one utterance to the next. For example, how might a particular gesture be related to a particular phrase when describing a scene or explaining a concept? However, it is interesting to consider the relationship between gesture and speech over shorter and longer timeframes as well. For example, how might gestures impact the immediate online processing of a word (Kelly, Kravitz & Hopkins, 2004) or how might gesture's relationship with speech change over a developmental time period reflecting differences in the way

2. This approach is inspired by the "Four Dimensions Framework" first articulated by Norman Adler and Randy Gallistel at the University of Pennsylvania (Goldin-Meadow, McClintock & Wimsatt, unpublished manuscript).

children mentally represent conceptual information (Alibali & Goldin-Meadow, 1993; Church, 1999; Perry, Church, & Goldin-Meadow, 1992)? Thus, the question of how much gesture and speech are integrated during language comprehension can be approached from multiple timeframes. I will focus on three such timeframes in this chapter. The first is *online integration* that investigates the immediate perceptual and cognitive processing – on the order of milliseconds and seconds – of when gestures and speech interact during the earliest stages of language comprehension. The second is *moment-to-moment integration* that investigates the immediate consequences – on the order of seconds and minutes – of gesture and speech integration in comprehending a message. And third, *developmental integration* refers to how gesture and speech form unified representations over the period of days, weeks, months and years in long-term memory as the result of some sort of sustained learning. A fourth timeframe, the evolutionary timespan, is also very important but is not the central focus of the remainder of this chapter (for more on this, see McNeill, Chapter 5).

The four linguistic components

Semantics

The vast majority of research on gesture-speech integration focuses on semantic processing. The semantic component of language is so rich for gesture researchers because gesture and speech reveal meaning in two very different, but complementary, ways. Whereas speech is inherently arbitrary and conventional, co-speech gestures are naturally imagistic and idiosyncratic (McNeill, 1992). Given these two different ways of representing meaning, it is not surprising that theorists and researchers have been interested in how they come together in language comprehension.

Online and moment-to-moment integration

There are already thorough reviews on online and moment-to-moment semantic integration published elsewhere (Hostetter, 2011; Kelly, Manning, & Rodak, 2008; Özyürek, 2014; Willems & Hagoort, 2007), but I will present some highlights. By now, it is now well established that gestures do communicate information above and beyond speech. In particular, listeners/viewers are sensitive to iconic gestures depicting concrete objects, actions and events – especially when conveying complementary information to speech – and the communicative power of gesture is particularly pronounced for children (Hostetter, 2011).

Although there is much behavioral research showing that people glean information from iconic gestures, much less is known about how people *integrate* gesture and speech during comprehension. Summarizing research from my lab, we have learned three things. One, relative to speech alone, incongruent iconic gestures make people slower and less accurate – and congruent gestures make people faster – at processing speech (Kelly et al., 2010). Two, the content of iconic gestures influences comprehension of speech to the same extent that the content of speech influences the comprehension of gesture – their semantic influence is bi-directional. And three, iconic gestures that complement speech (e.g., producing a drinking gesture while saying, “The researchers were up late at the conference”), make speech content more memorable compared to when no gestures are present (Church, Garber & Rogalsky, 2007; Kelly, Barr, Church & Lynch, 2001).

In the past decade, researchers have made great progress exploring the possible neural mechanisms for how and when gestures and speech are semantically integrated. In Özyürek’s (2014) recent review, she identifies a network of brain regions that integrates gesture and speech, starting with the medial temporal gyrus (MTG) and superior temporal sulcus (STS) that initially process multimodal input, and then feed into the inferior frontal gyrus (IFG) for a more holistic conceptual integration over larger discourse units (Hagoort, 2005). This fMRI research is nicely complemented by the temporal sensitivity of event-related potentials (ERPs), showing that this integration occurs as early as 200 to 300 ms (Kelly, Kravitz, & Hopkins, 2004; Wu & Coulson, 2010), a time window well documented to reflect the earliest stages of semantic integration (Kutas & Hillyard, 1980). Moreover, the neural processing of gesture and speech are temporally overlapping during language comprehension, which suggests that the brain treats gesture and speech similarly during semantic integration (Özyürek, Willems, Kita, & Hagoort, 2007).

Developmental integration

Researchers have also explored the neural mechanisms for long-term integration of gesture and speech as the result of learning. For example, my lab has shown that observing congruent iconic gestures when learning novel verbs in Japanese not only helps people learn words better than no gesture, but ERP data also reveals that words learned with gesture produce a large LPC – an ERP component that reflects strength of imagistic memory traces (Klaver et al., 2005) – than words without gesture (Kelly, McDevitt, & Esch, 2009). With regard to localization of brain regions, Macedonia and colleagues have done fMRI research showing that iconic gestures help people learn novel vocabulary items in an invented language, and they argue that a neural mechanism is strengthened networks in the premotor cortex (Macedonia, Müller, & Friederici, 2011). One conclusion from these studies is that gestures deepen sensorimotor traces and make long-term memories

for newly learned linguistic items less likely to decay. Indeed, research focusing on long-term memory consolidation shows that recalling sentences learned with metaphoric gestures is correlated with increased hippocampal activation (Straube, Green, Weis, & Chatterjee, 2009).

Further exploration

Although researchers have learned much about gesture and speech on the semantic level, there are a number of intriguing questions that remain. Here are four. First, as Hostetter (2011) points out in her review, the strongest evidence for the semantic function of gesture in language comprehension involves *concrete* things (e.g., objects attributes, dynamic actions, spatial relations, etc.). However, there are many semantic aspects of language that are much more abstract (ideas, principles, inner states, laws, etc.), and barring the use of metaphor, this information is difficult to capture in gesture. If a significant chunk of language is indeed “off-limits” to gesture, it suggests a boundary where gesture and speech do not have an integrated relationship (I will return to this in the Conclusion).

Second, according to McNeill (2013), gestures are a special type of bodily action, and their relationship with speech is privileged. If this is the case, gestures and speech should be integrated during language comprehension differently than other manual actions and speech. Some preliminary behavioral research suggests that this is indeed the case (Kelly, Healey, Özyürek, & Holler, 2014), but further work should explore this on the neural level as well, especially given the suggestion that gestures are simulations of actual actions (Hostetter & Alibali, 2008).

Third, regarding the issue of the extent to which language is *constituted* by the body, it is still unclear just how much the motor system is a necessary *part* of the semantic processing of gesture. Indeed, there is a larger debate in the literature on manual actions between researchers who believe that motor simulation is a central part of how action and language are understood (Gallese & Lakoff, 2005; Pulvermüller, 2005; Rizzolatti & Arbib, 1998) and researchers who believe that the motor system is just a down stream by-product of an earlier spread of activation of abstract meaning (Mahon & Caramazza, 2008; Hickok, 2014). Much of the work on this issue has used spatial imagining techniques such as fMRI – which show brain activation on the order of seconds – and the debate would benefit from tools that have good spatial *and temporal* resolution, such as magnetoencephalography (MEG), in order to pull apart the location and millisecond time-course of the motor simulations during the “integration” of the two modalities (see Masumoto et al., 2006).

I have put integration in scare quotes because it is not clear that the word adequately captures the process of how gesture and speech actually interact in

the brain.³ In the traditional use of the term (e.g., McNeill, 1992), gesture-speech integration is very much a *conceptual* blending of two different representational formats. However, as Skipper (2014) suggests with his “Natural Organization of Language in the Brain” (NOLB) model, the process by which gesture and speech (and any other multimodal input) come together may be less conceptual and more statistical in nature (this is similar to Bayesian learning neural models; see Knill & Pouget, 2004). That is, rather than processing the meaning of a gesture and a word and then blending the two “meanings,” the model claims that gesture is just one of many probabilistic cues that simply constrain what the auditory cortex is perceiving (or expecting to perceive). These probabilistic prediction models are promising in other domains of language processing/learning (Tenenbaum et al., 2011), and they deserve more attention in the gesture literature.

Pragmatics

McNeill’s concept of *newsworthiness* is useful in thinking about the pragmatics of gesture-speech integration during comprehension. Not only can gesture imagistically communicate what is semantically new and relevant within the linguistic context of speech, but it can also communicate what aspects of the extra-linguistic context are relevant to that speech. For example, a deictic gesture can indicate a point in time (e.g., in the past, present or future) or space (e.g., an object near or far) that, together with the immediate spoken context, allows the intended meaning of an utterance to materialize.

Moment-to-moment and developmental integration

Young language learners are very good at interpreting the intentions of gestures, especially pointing gestures (Murphy & Messer, 1977). In addition, children can differentiate the pragmatic meaning of declarative pointing gestures (“That is a doggie.”) from those that are imperative (“Get your ball.”) (Baron-Cohen, 1989). Indeed, Tomasello (2008) argues that pointing gestures are one of the earliest tools that humans use to explore shared intentionality and build mutual understanding, and this lays the social groundwork for all of language development.

The developmental literature on the role of gesture (and multimodal input more generally) adds an important dimension to traditional linguistic theories in pragmatics (Austin, 1975; Grice, 1975). Traditional pragmatic theories are built on the assumption that we can infer what people mean based on knowledge

3. I thank Jeremy Skipper for suggesting this at the 2010 Neurocognition of Gesture Processing workshop hosted by the Max Planck Institute for Cognitive and Brain Sciences.

concerning kinds of speech acts (Austin, 1975), conventions governing the flow of discourse (Grice, 1975), and common ground or “mutual knowledge” (Clark & Marshall, 1981). Note that this approach assumes that information about a speaker’s intention lies somewhere *outside* of what is communicated – which, in turn, is traditionally considered as equivalent to the *spoken* portion of the message.

Viewing gesture and speech as a fundamentally integrated system turns this traditional view on its head. If gestures function to add what is newsworthy in the context, the classic problem in pragmatics of speech “under-determining” meaning goes away. For example, consider the pragmatic intent of the spoken utterance, “It’s almost time for dinner,” accompanied by a pointing gesture and glance to a messy pile of toys in front of a child playing. In this case, the deictic information indexes an aspect of the physical environment – a mess to be cleaned up – that makes the intentions of the utterance crystal clear: The child should put her toys away before dinner. Children as young as 3-years-old understand the intentions of these sorts of speech-gesture utterances much better than speech or gesture alone (Kelly, 2001). In fact, it is hypothesized that it is precisely the combination of speech and gesture that allows young children to “break into” the complex social world of pragmatic communication. Moreover, even adults benefit from the combination of speech and pointing gestures to help them understand the intended meaning of these sorts of unconventional pragmatic utterances (Kelly et al., 1999). Additionally, as with semantic processing, deictic gestures have a bi-directional relationship with speech during pragmatic processing. Without speech, the intended physical referent of a deictic gesture can be quite ambiguous. For example, understanding the intention behind pointing to an open window is easier in the presence vs. absence of saying “It’s getting cold in here.” In this way, speech adds “news” to gesture as much as gesture informs speech.

Online integration

The neural mechanisms underlying gesture-speech integration at the pragmatic level are similar to mechanisms for semantic integration. For example, the left IFG is not only a neural unification site for semantic information across speech and gesture, it also binds pragmatic information (e.g., intentions, background knowledge, relevant aspects of context) to speech content during language comprehension (Hagoort, 2005; Hagoort, Hald, Bastiaansen, & Petersson, 2004). But as with any sort of complex social communication, there is a network of neural mechanisms at play. One line of research has used fMRI to explore how other pragmatic indicators of intentions, such as eye gaze, may modulate the neural integration of gesture and speech (Holler et al., 2014; Straube, Green, Jansen, Chatterjee, & Kircher, 2010). For example, Holler et al. (2014) showed that the right MTG was more active for speech with gesture than without it during language comprehension – but only

when a speaker was making eye contact directly with the listener. This finding is interesting because the right lateralized effect in the MTG is consistent with other research on the neural processing of pragmatic information in the production of pointing gestures (de Langavant et al., 2011), and it suggests that the right MTG might be influenced by higher-order pragmatic processes involved in multimodal processing in a similar way as how the left IFG integrates semantic context with spoken information during comprehension.

In addition to exploring what particular brain regions are involved in the pragmatic integration of gesture and speech, it is interesting to consider the timing of that integration. In work from my lab, I have argued that pragmatic attributions of communicative intent may modulate the integration of gesture and speech at very early semantic stages of processing (Kelly, Ward, Creigh, & Bartolotti, 2007). When listeners were presented with an utterance conveying incongruent gesture-speech combinations, there was a bi-lateral N400 effect (indexing difficulties with semantic processing) in frontal scalp regions when subjects believed that the gesture and speech were intentionally coupled. However, when subjects were told that two different people produced the gesture and speech (i.e., one person's speech was dubbed onto another person's gesture), the N400 effect was present only in the right hemisphere. This suggests that neural mechanisms in the right hemisphere may be relatively obligatory in how they integrate gesture and speech, but mechanisms in the left hemisphere may be more strategic in nature. The possibility that there are different hemispheric mechanisms for gesture-speech integration helps to explain how it can be both simultaneously automatic (Kelly, Creigh, & Bartolotti, 2010) and controlled (Holle & Gunter, 2007; Holler et al., 2014).

Further exploration

There are still many interesting questions that deserve attention in understanding the pragmatics of gesture comprehension. For example, one good question is what are the neural mechanisms that keep track of semantic meanings and pragmatic intentions of gesture? Hagoort and et al. (2004) showed that although semantic (“Dutch trains are sour”) and pragmatic (Dutch trains are white”) violations produced similar N400 effects and overlapping left IFG activation in contrast to baseline true statements (“Dutch trains are yellow”), EEG oscillations in the gamma range – which reflect post-integration binding across modalities – differentiated the semantic and pragmatic violations. Thus, it is likely that executive processes in the frontal lobe may monitor when gestures add semantic vs. pragmatic “news-worthy” information to speech.

A second productive question is what sorts of extra-linguistic pragmatic variables modulate gesture-speech integration? As mentioned above, researchers are already considering the role of eye gaze as a modulator of gesture comprehension,

but there are a whole host of other extra-linguistic variables not currently receiving much attention. For example, emotional information conveyed through facial expressions and body posture index how speakers are feeling while they are speaking. Indeed, these two variables interact, with emotional bodily language (EBL) influencing perception of facial expressions very quickly (100ms) in emotional processing brain regions (e.g., the cingulate) (de Gelder, 2006). Although there are a handful of studies considering how affective signals conveyed through the hands and body influence social and pragmatic variables – such as perception of power between men and women (Bailey & Kelly, 2015) and racial in/out-group processing (Hinzman & Kelly, 2013) – no study, to my knowledge, has investigated how these emotional displays of gesture are integrated with pragmatic intentions of speech.

Phonetics

For all the research on the semantic and pragmatic integration of gestures and speech, there is far less attention to the phonetic function of gesture. This dearth of research is notable because the phonetic stage of language comprehension is the first possible place where gestures could connect with speech. Moreover, it is well established that other multimodal inputs, such as face and mouth movements, play a significant role in how people comprehend phonetic information during speech perception (as far back as Sumbly & Pollack, 1954, and of course, McGurk & MacDonald, 1976). Coming back to McNeill's theory, just how newsworthy are hand movements in perceiving speech sounds? Does integration occur all the way down to individual phonemes in individual spoken words?

One likely candidate to play a role at the earliest stages of phoneme processing might be *beat* gestures, which are rhythmic movements that highlight specific parts of discourse. Indeed, naturalistic research has documented that beats are tightly timed – along with other behaviors like eyebrow movements and head nods – with certain points of phonetic emphasis (Loehr, 2007). In fact, if you ask someone to produce a beat with a particular word in a sentence, the acoustics of that word are exaggerated (Krahmer & Swerts, 2007). These studies suggest a tightly coupled integration of speech and gesture at the level of phonemes in language production. What about comprehension?

Online and moment-to-moment integration

It is well established that even very early in development, infants integrate phonemic information across modalities (Dodd, 1979; Kuhl & Meltzoff, 1984). For example, Kuhl and Meltzoff (1984) showed that when infants see a face making a sound, they prefer sounds that match versus mismatch the face. To my knowledge, there

is no comparable research exploring how hand gestures aid phonemic perception at the earliest stages of development, but there is evidence that as children get older, they increasingly use deictic gestures to help constrain ambiguous speech sounds. For example, Thompson and Massaro (1986, 1994) showed that pointing to objects (e.g., a ball vs. a doll) helps to clarify speech that was synthesized along a continuum of “ba” to “ga” in children as young as 3-years-old. Further, they showed that older children (4-, 5- and 9-year-olds) and adults increasingly used input from pointing gestures to disambiguate speech, suggesting that gesture-speech integration on the phonemic level improves over development.

More recently, there has been a rise of research investigating how beat gestures are integrated with speech during comprehension. Returning to Krahmer and Swerts (2007), not only do beat gestures affect the acoustics of speech production, but they also cause listeners to hear spoken words differently. They showed this in two ways: (1) when people were auditorily presented with words originally produced with beats – but the visual information was removed – listeners heard words as louder than when the speech was not produced with beats, and (2) when people heard the same word with the same spectral acoustics, they nevertheless perceived the word to be louder when a visual beat was added to it compared to when there was no visual beat present. Moreover, recent research measuring ERPs to naturalistic spoken discourse shows that the timing of the visual influence of beats on phoneme perception occurs as early as 100 ms (Biau & Soto-Faraco, 2013). This suggests that beat gestures, like other visual inputs such as lip movements (van Wassenhove, Grant, & Poeppel, 2005), are tightly tied to the earliest stages of phonemic processing.

As with semantic integration, the brain regions involved in this integration of gesture and speech are widely distributed. In the first fMRI study on the subject, Hubbard and colleagues scanned participants while they watched a naturalistic video of someone speaking with and without meaningful beat gestures (Hubbard, Wilson, Callan, & Dapretto, 2009). The presence of beats with speech increased activation in bi-lateral nonprimary auditory cortex compared to a speech only baseline; moreover, in the right planum temporale, speech plus beat gestures produced multisensory activation that went beyond the sum of the individual contributions of speech and gesture alone. Based on basic neural mechanisms for rhythm processing, the authors speculate that the right planum temporale may play a significant role in synthesizing the rhythmic properties of speech and gesture. Finally, it was not just any sort of visual information that influenced speech processing: The STS showed more activation when speech was accompanied by beat gestures than meaningless hand movements.

Recent research, however, has presented a different picture of the neural network involved in integrating gestures and speech sounds. As mentioned earlier, Skipper's NOLB model (2014) approaches the processing of gesture and speech from the perspective of a prediction-making network. Specifically, he explains that the brain does not process speech sounds in only a bottom up or feed-forward fashion. That is, in addition to the auditory cortex (A1) sending outputs to "higher" brain regions involved in meaning comprehension (e.g., MTG and STS) or motor simulation (e.g., frontal regions), there are top down influences of these areas back to A1 as well. These higher brain regions use contextual information to feed *backwards* to modulate processing in A1 depending on how much work it needs to do. To provide support for this model, Skipper did a high-density (256 electrode) EEG experiment, using sLORETA dipole source modeling estimates, to measure when and where iconic gestures affect the processing of spoken lexical affiliates. The main finding was that the presence of gesture actually *decreased* activation in posterior frontal regions (which function to make motor predictions) and posterior temporal regions (which function to multimodally constrain meaning) compared to when no gestures preceded speech. This decrease of activity was interpreted to mean that the higher brain regions were using gesture (and other aspects of context) to confirm or predict what was, or would be, happening in A1. As a consequence, in accordance with the model, the top-down information ultimately reduced the amount of effort A1 needed to accurately identify the "auditory object" it was phonetically processing. In this way, the presence of gesture serves as a semantic context that makes it easier – metabolically and cognitively – to predict and identify speech sounds during comprehension.

Developmental integration

In addition to work on the online and moment-to-moment phonemic integration of gesture in speech in one's native language (L1), research from my laboratory has used second language (L2) training studies to investigate longer-term phonemic integration of the two modalities (Hirata & Kelly, 2010; Hirata, Kelly, Huang, & Manansala, 2014; Kelly & Lee, 2012; Kelly, Hirata, Manansala, & Huang, 2014). We have chosen a language, Japanese, which poses particular challenges to native English speakers. In Japanese, vowel length is phonemic, which means that the length of a vowel changes the meaning of a word, as in the final phoneme in "rika" ("science") versus "/rikaa/" ("liquor"). In a series of experiments, we used short and long sweeping gestures to visually represent the length of spoken vowel sounds while teaching novice L2 learners words containing the novel vowel length distinctions. We found that both observing (Hirata & Kelly, 2010) and producing (Hirata et al., 2014) gestures do not help to perceptually distinguish long and short Japanese vowels compared to a speech only training baseline. Moreover, the

presence of gestures during phoneme training does not ultimately help learning novel Japanese vocabulary items either (Kelly et al., 2014), and this was still the case even when iconic gestures accompanied instruction (Kelly & Lee, 2012). This is notable because when L2 phonemes are not novel (as in Kelly et al., 2009), iconic gestures *do help* with word learning.

These findings are interesting in light of Skipper's NOLB model. One possible explanation is that the gestures used in our L2 training were too vague and weak of a context to constrain the processing of the novel L2 phonemes. However, Kelly and Lee (2012) used very clear iconic gestures and found that even they did not help with learning – in fact, the iconic gestures in that study actually *hurt* word learning, at least when the phonemes were novel. A second possibility is that because the Japanese speech sounds were so novel, A1 needed to work extra hard to process it, and this could have made any top down influences from motor or conceptual areas unhelpful, or worse, distracting.

Finally, an intriguing third possibility concerns whether gestures are designed to connect to the *form* or to the *content* of speech. For example, consider someone saying, “I was so hungry that I ate the whole thing,” while making an eating gesture just before saying the lexical affiliate, ate. The NOLB model would predict that posterior temporal regions (such as STS) would analyze the *meaning* of the gesture to guess the meaning of the lexical affiliate (the action of eating something) and then that would get mapped onto a likely phonological form, such as the first phoneme of ate. However, the results from the L2 training studies suggest that learners do not use information from the *form* of a gesture (such as a long or short sweeping movement) to directly guess the *form* of the lexical affiliate (i.e., a long or short vowel). In other words, listeners/viewers may find it unnatural to expect that a gesture visually represents upcoming sounds, but quite natural to expect that a gesture visually represents upcoming *meanings* that will ultimately map onto those sounds. In this way, gesture and speech may not have a directly integrated relationship on the phonemic level.

Further exploration

One area for further investigation is to explain why some studies show that gestures produce more activation in the neural processing of speech sounds (Hubbard et al., 2009) and other studies show less activation (Skipper, 2014). One possibility is that the type of gesture matters. Whereas Hubbard and colleagues focused on mostly beat gestures, Skipper focused on iconics. Perhaps beat gestures produce higher activation in auditory areas to make the semantics of an utterance more clear (i.e., making a beat on a word makes it clear that the word is semantically newsworthy). In contrast, iconic gestures might have a different function: They may use imagistic meaning to help constrain and simplify expectations about

upcoming speech sounds *themselves*, thus reducing the amount of work in A1. Considered from this perspective, it makes sense that beat gestures might increase processing of auditory cortex, whereas iconic gestures might decrease it.

A second interesting point is that whereas beat gestures function well to highlight the phonemic properties of words within a *sentence* (i.e., suprasegmental processing, as in Krahmer & Swerts, 2007), they do not seem well suited to highlight novel speech sounds within a *word* (i.e., segmental processing, as in Hirata & Kelly, 2010). This suggests that for novel speech sounds – e.g., sounds not within one’s normal native repertoire – gesture and speech may not be integrated at the lowest levels of phoneme processing. Perhaps because gestures are so well suited for highlighting semantically relevant information at the suprasegmental utterance level, it is unnatural for them to draw attention to lower level phonemic information at the segmental timing level, especially when auditory processing demands are high. Addressing this issue will help to delineate just where gesture-speech integration begins and ends at the earliest stages of phonemic processing.

Syntax

Syntax is the least explored level of language in the research on gesture-speech comprehension. There are some accounts of the evolution of language suggesting that a gestured system paved the way for a spoken system (e.g., Armstrong & Wilcox, 2007; Corballis, 2003; Rizzolatti & Arbib, 1998; Tomasello, 2008), and from this perspective, gestures would likely have played at least an early syntactic role with speech in language production. Indeed, the fact that hand gestures can take on the entire burden of communication in conventional sign languages suggests that they can – when necessary – take on such a function. However, other accounts (e.g., McNeill, 2012, Chapter 5) suggest that speech and gesture evolved together as a tightly coupled system *from the start*, and from this perspective, it is possible that gestures may have left much of the syntactic work to the spoken modality. After all, from the perspective of gesture and speech adding what is “newsworthy” to one another, gestures are much better suited to contribute iconic semantic information to the syntactic organization of speech than the other way around – speech providing iconic semantic information to the syntactic organization of gesture. Although this relationship is mostly considered in language production, studying it from the perspective of language comprehension may provide new insights.

During comprehension, emblematic gestures (e.g., peace sign, thumbs up, OK sign) can play a syntactic role and fill in for word classes such as nouns, adjectives and verbs (e.g., saying, “You” and then making a “come here” emblem with a

beckoning upturned index finger). In fact, ERP research shows that people comprehend these “mixed syntax” gestures (Gullberg, 2006) in a similar fashion as actual words (Gunter & Bach, 2004). However, as with sign language, these are cases in which gestures substitute for words, and they do not speak to the question of whether gesture and speech are syntactically *integrated* online during language comprehension.

Moment-to-moment and developmental integration

There is some evidence in very young children that speech and gesture play a complementary syntactic role during early language development (Bates & Dick, 2002; Goldin-Meadow, 2005). Although much of this work has focused on language production, there are a few developmental studies that have explored this during comprehension. For example, Morford and Goldin-Meadow (1992) showed that children who were transitioning from one- to two-word speech best understood utterances that split words and gestures across syntactic categories (e.g., a “give me” gesture while saying “cup” was understood better than saying, “Give me cup.”). This finding suggests that it may be cognitively optimal to distribute syntactic work across modalities – rather than to pack it all within speech – in the earliest stages of language development. Of course, one could argue that gestures were playing more of a semantic, and less of a syntactic, role because they were not necessarily helping children to understand syntactic rules such as word order or predicate structure.

Online integration

To my knowledge, there is only study that has explored the online processing of gesture and speech at the syntactic level during comprehension (Holle et al., 2012). This study used ERPs to explore the role that gestures play in clarifying syntactic ambiguity in German. The German language allows for an SOV word order (“the woman the men hired *has*”) and also a less frequent OSV word order (“the woman the men hired *have*”). Because the second sentence is less common, it produces some confusion (indexed by the P600, a measure of syntactic ambiguity) when people reach the target word, *have*. However, if a beat gesture accompanies the second noun predicate – “men,” which in this case is the subject of the sentence – the P600 to the target word, *have*, is reduced. This suggests that people more easily understood that “men” was the subject of the sentence when a gesture occurred with it. Moreover, the fact that there was a decreased P600 effect rather than a decreased N400 effect suggests that the processing was most likely syntactic and not semantic in nature. Interestingly, this syntactic disambiguation effect was not present when a dot on the computer screen replaced the beat gesture,

suggesting that gesture may be a special type of visual input that differentiates itself from other visual information (see Kelly et al., 2014, for a similar claim on the semantic level).

Further exploration

Given the paucity of research on the syntactic functions of gesture, there are several areas that are ripe for exploration. For example, Kaschak and Glenberg's Indexical Hypothesis (2002) suggests that gestures can illustrate novel affordances of objects (e.g., using a cane to press an elevator button or a cell phone to open a beer bottle), and this information can interact with the syntax of a sentence to highlight novel meanings. It would be interesting to explore whether adding iconic gestures to novel syntactic constructions (e.g., making a "cane push" gesture while saying, "He caned the elevator button") helps listeners more quickly and accurately comprehend the meaning of syntactically novel utterances.

In addition to the role of gesture in clarifying syntax in an L1, another interesting question is exploring the role of gestures in L2 syntax learning. After all, we know that in production, gestures reflect novel syntactic structures of an L2 (Kita & Özyürek, 2003; So, Kita, & Goldin-Meadow, 2009; Stam, 2006). For example, Kita and Özyürek (2003) showed that Turkish, Japanese and English speakers packaged information about motion events in gesture differently according to the syntax of each language. It would be interesting to do a training study to determine if novice learners of those L2 languages syntactically benefited from the presence of those linguistically appropriate – versus "inappropriate" – L2 gestures.

Because there is so little research on the topic, it is currently difficult to make firm conclusions about syntactic integration of gesture and speech during comprehension. In fact, one could even argue that the handful of studies on the topic reflect more of a semantic or pragmatic, and less of a syntactic, function of gesture. For example, one might interpret Morford and Goldin-Meadow's (1992) results as iconic gestures pragmatically disambiguating the accompanying speech: Producing a "give me" gesture while saying, "cup," could pragmatically clarify that a child was requesting a cup rather than commenting on it. Even the innovative ERP study by Holle and colleagues (2012) is open to interpretation. For example, it is possible that the beat gestures in that study initially shifted the pragmatic expectations of who was doing the action and who was being acted upon, and this discourse shift had down-stream syntactic implications (for more on gesture identifying linguistic referents in discourse, see So, Kita, & Goldin-Meadow, 2009). This is not to minimize these previous studies, but it does open up the possibility that gestures may be designed more for semantic and pragmatic purposes and are perhaps only coopted to serve more conventionalized syntactic functions.

Moreover, as with phonetic integration, there is evidence that gesture may actually disrupt learning of syntax when syntactic structures are too novel or demanding. For example, Post and colleagues used gesture to train 6th-grade children on how to transform active sentences to passive sentences (Post, van Gog, Paas, & Zwaan, 2013). In one condition, the children watched animations of words changing their location in a sentence – transforming it from an active to passive construction – and in the other condition, children watched and imitated a teacher “moving” the animations with her hand. Contrary to their hypotheses, children who were trained by imitating the deictic gestures were worse at learning the passive constructions than children who passively watched the animations with no gesture. Moreover, the detrimental effect of gesture was more severe for children with generally lower language skills. The authors concluded that in cases when cognitive demands are high and/or pre-existing skills are low, gestures may actually interfere with syntactic learning rather than help it, a finding that is consistent with research showing that high perceptual demands can disrupt phonetic learning in a novel L2 (Hirata & Kelly, 2010; Hirata et al., 2014).

Conclusion

The present review has aimed for a “robust” understanding of gesture-speech integration during language comprehension by analyzing it along three different dimensions (Wimsatt, 1981). This approach has produced a nuanced picture. On multiple levels of analysis – social, cognitive and biological – and across more than one timeframe – online, moment-to-moment, and developmental – the evidence has converged to demonstrate that gesture and speech have a highly integrated relationship in pragmatic, concrete semantic and suprasegmental phonetic processes during the comprehension of language. However, this integrated relationship between gesture and speech is much less evident – and in some case, non-existent – in the abstract semantic, segmental phonetic and syntactic processing of language.

This nuanced picture is interesting in light of the larger question of what gesture can tell us about how much of language is constituted by the body during comprehension. There has been much excitement over embodied perspectives on language comprehension over the last two decades. Perhaps part of the enthusiasm is that embodied accounts offer such a contrast to views of language as an abstract process that is mostly encapsulated and isolated from the body (Chomsky, 1980; Fodor, 1983). However, the present results suggest that it may be misguided to dismiss these more traditional “disembodied” views so quickly (see also, Binder & Desai, 2011; Hickok, 2014; Mahon & Caramazza, 2008). After all, there are some real benefits to “disembodied” linguistic processes. For example, having

an arbitrary semantic dimension to words not only helps capture meaning that is very abstract (e.g., concepts such as, infinity, privilege, truth, democracy, etc.), but it also flexibly allows for spread of activation across conceptual space that is unencumbered by the body. In addition, the automatized quality of phonemes and the conventionalized nature of syntax make processing (familiar) phonemic units and syntactic structures fast and easy. Grounding these processes in the body may allow for the best of both worlds: an abstract, fast and flexible spoken system that is housed in and constrained by a slower but much more concrete bodily framework (see Dingemanse et al., 2015, for a similar argument concerning the benefits of arbitrariness and iconicity *within* the speech signal).

It is interesting to think of this “grounding” in the context of McNeill’s (2012) claim that gesture and speech *began* as an integrated system at the evolutionary dawn of language (contra “gesture-first” claims). If McNeill is right, it is possible that over the course of evolution, some aspects of language – such as syntactic, segmental phonetic, and abstract semantic elements – may have slowly taken on more of a conventionalized and abstract nature, leaving the body to do more and more of the concrete and imagistic work. Closer to home, ontogeny may recapitulate phylogeny in how children develop language. Indeed, it is likely that some of those same elements of language progressively get more modularized (to use Karmiloff-Smith’s term, 1995) over the course of development in order to exploit the benefits of a disembodied linguistic system. Ultimately, this nuanced view makes hand gestures a unique and useful tool for exploring the embodied nature of language: The hands can help us delineate what aspects of language function as part of the body – and what aspects operate independently of it.

References

- Alibali, Martha W., & Goldin-Meadow, Susan 1993. “Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child’s state of mind.” *Cognitive Psychology* 25: 468–523. doi:10.1006/cogp.1993.1012
- Armstrong, David F., & Wilcox, Sherman E. 2007. *The Gestural Origin of Language*. Oxford; New York: Oxford University Press. doi:10.1093/acprof:oso/9780195163483.001.0001
- Austin, John L. 1975. *How To Do Things With Words* (Vol. 1955). Oxford University Press. doi:10.1093/acprof:oso/9780198245537.001.0001
- Bailey, April, & Kelly, Spencer D. 2015. “Picture power: Gender versus body language in perceived status.” *Journal of Nonverbal Behavior* 39: 317–337.
- Baron-Cohen, Simon 1989. “Perceptual role taking and protodeclarative pointing in autism.” *British Journal of Developmental Psychology* 7 (2): 113–127. doi:10.1111/j.2044-835X.1989.tb00793.x
- Barsalou, Lawrence W. 1999. “Perceptions of perceptual symbols.” *Behavioral and Brain Sciences*, 22(04), 637–660. doi:10.1017/S0140525X99532147

- Bates, Elizabeth, & Dick, Fredrick 2002. "Language, gesture, and the developing brain." *Developmental Psychobiology* 40 (3): 293–310. doi:10.1002/dev.10034
- Biau, Emmanuel, & Soto-Faraco, Salvador 2013. "Beat gestures modulate auditory integration in speech perception." *Brain and Language* 124 (2): 143–152. doi:10.1016/j.bandl.2012.10.008
- Binder, Jeffrey R., & Desai, Rutvik H. 2011. "The neurobiology of semantic memory." *Trends in Cognitive Sciences* 15 (11): 527–536. doi:10.1016/j.tics.2011.10.001
- Chomsky, Noam 1980. "Rules and representations." *Behavioral and Brain Sciences* 3 (01): 1–15. doi:10.1017/S0140525X00001515
- Church, Ruth B. 1999. "Using gesture and speech to capture transitions in learning." *Cognitive Development* 14: 313–342. doi:10.1016/S0885-2014(99)00007-6
- Church, Ruth B., Garber, Philip, & Rogalsky, Kathryn 2007. "The role of gesture in social communication and memory." *Gesture* 7: 137–157. doi:10.1075/gest.7.2.02bre
- Clark, Herbert H., & Marshall, Catherine R. 1981. "Definite reference and mutual knowledge." In *Elements of Discourse Understanding*, A.K. Joshi, B. Webber, I.A. Sag (eds), 10–63. Cambridge: Cambridge University Press.
- Corballis, Michael C. 2003. "From mouth to hand: gesture, speech, and the evolution of right-handedness." *Behavioral and Brain Sciences* 26 (02): 199–208. doi:10.1017/S0140525X03000062
- de Langavant, Laurent C., et al. 2011. "Behavioral and neural correlates of communication via pointing." *PloS one* 6 (3): e17719.
- de Gelder, Beatrice 2006. "Towards the neurobiology of emotional body language." *Nature Reviews Neuroscience* 7 (3): 242–249. doi:10.1038/nrn1872
- Dingemanse, Mark, Blasi, Damián E., Lupyan, Gary, Christiansen, Morten H., & Monaghan, Padraic 2015. "Arbitrariness, iconicity, and systematicity in language." *Trends in Cognitive Sciences* 19: 603–615. doi:10.1016/j.tics.2015.07.013
- Dodd, Barbara 1979. "Lip reading in infants: Attention to speech presented in-and out-of-synchrony." *Cognitive Psychology* 11 (4): 478–484. doi:10.1016/0010-0285(79)90021-5
- Fodor, Jerry A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press.
- Gallese, Vittorio, & Lakoff, George 2005. "The brain's concepts: The role of the sensory-motor system in conceptual knowledge." *Cognitive Neuropsychology* 22: 455–479. doi:10.1080/02643290442000310
- Gibbs, Raymond W. Jr. 2006. *Embodiment and Cognitive Science*. Cambridge University Press.
- Goldin-Meadow, Susan, McClintock, Martha, & Wimsatt, William Unpublished manuscript. "Solving psychological problems in four dimensions: Heuristics for the social and biological sciences."
- Goldin-Meadow, Susan 2005. *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press.
- Goldin-Meadow, Susan, Alibali, Martha W., & Church, Ruth B. 1993. "Transitions in concept acquisition: Using the hand to read the mind." *Psychological Review* 100 (2): 279–297.
- Glenberg, Arthur M., & Kaschak, Michael P. 2002. "Grounding language in action." *Psychonomic Bulletin & Review* 9 (3): 558–565. doi:10.3758/BF03196313
- Grice, H. Paul 1975. "Logic and conversation." In *Syntax and semantics* (Vol. 3), P. Cole and J.L. Morgan (eds.). New York: Academic Press.
- Gullberg, Marianne 2006. "Some reasons for studying gesture and second language acquisition (Homage à Adam Kendon)." *IRAL-International Review of Applied Linguistics in Language Teaching* 44 (2): 103–124. doi:10.1515/IRAL.2006.004
- Gunter, Thomas C., & Bach, Patric 2004. "Communicating hands: ERPs elicited by meaningful symbolic hand postures." *Neuroscience Letters* 372 (1): 52–56. doi:10.1016/j.neulet.2004.09.011

- Hagoort, Peter, Hald, Lea, Bastiaansen, Marcel, & Petersson, Karl Magnus 2004. "Integration of word meaning and world knowledge in language comprehension." *Science* 304 (5669): 438–441. doi:10.1126/science.1095455
- Hagoort, Peter 2005. "On Broca, brain, and binding: a new framework." *Trends in Cognitive Sciences* 9 (9): 416–423. doi:10.1016/j.tics.2005.07.004
- Hickok, Gregory 2014. *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition*. WW Norton & Company.
- Hinzman, Lindsay & Kelly, Spencer D. 2013. "Effects of emotional body language on rapid out-group judgments." *Journal of Experimental Social Psychology* 49: 152–155. doi:10.1016/j.jesp.2012.07.010
- Hirata, Yukari & Kelly, Spencer D. 2010. "Effects of lips and hands on auditory learning of second-language speech sounds." *Journal of Speech, Language, and Hearing Research* 53 (2): 298–310. doi:10.1044/1092-4388(2009/08-0243)
- Hirata, Yukari, Kelly, Spencer D., Huang, Jessica, & Manansala, Michael 2014. "Effects of hand gestures on auditory learning of second-language vowel length contrasts." *Journal of Speech, Language, and Hearing Research* 57 (6): 2090–2101. doi:10.1044/2014_JSLHR-5-14-0049
- Holle, Henning, & Gunter, Thomas C. 2007. "The role of iconic gestures in speech disambiguation: ERP evidence." *Journal of Cognitive Neuroscience* 19 (7): 1175–1192. doi:10.1162/jocn.2007.19.7.1175
- Holler, Judith, Schubotz, Louise, Kelly, Spence D., Hagoort, Peter, Schuetze, Manuela & Özyürek, Asli 2014. "Social eye gaze modulates processing of speech and co-speech gesture." *Cognition* 132: 696–697.
- Holler, Judith, Kokal, Idil, Toni, Ivan, Hagoort, Peter, Kelly, Spencer D., & Özyürek, Asli 2014. "Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG." *Social Cognitive and Affective Neuroscience* 10 (2): 255–261.
- Hostetter, Autumn B., & Alibali, Martha W. 2008. "Visible embodiment: Gestures as simulated action." *Psychonomic Bulletin & Review* 15 (3): 495–514. doi:10.3758/PBR.15.3.495
- Hostetter, Autumn B. 2011. "When do gestures communicate? A meta-analysis." *Psychological Bulletin* 137 (2): 297. doi:10.1037/a0022128
- Hubbard, Amy L., Wilson, Stephen M., Callan, Daniel E., & Dapretto, Mirella 2009. "Giving speech a hand: gesture modulates activity in auditory cortex during speech perception." *Human Brain Mapping* 30 (3): 1028–1037. doi:10.1002/hbm.20565
- Karmiloff-Smith, Anette 1995. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press.
- Kelly, Spencer D. 2001. "Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension." *Journal of Child Language* 28 (02): 325–349. doi:10.1017/S0305000901004664
- Kelly, Spencer D., Barr, Dale J., Church, Ruth B., & Lynch, Katheryn 1999. "Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory." *Journal of Memory and Language* 40 (4): 577–592.
- Kelly, Spencer D., Iverson, Jana M., Terranova, Joseph, Niego, Julia, Hopkins, Michael, & Goldsmith, Leslie 2002. "Putting language back in the body: Speech and gesture on three time frames." *Developmental Neuropsychology* 22 (1): 323–349. doi:10.1207/S15326942dn2201_1
- Kelly, Spencer D., Kravitz, Corinne, & Hopkins, Hopkins 2004. "Neural correlates of bimodal speech and gesture comprehension." *Brain and language* 89: 253–260. doi:10.1016/S0093-934X(03)00335-31

- Kelly, Spencer D., Ward, Sarah, Creigh, Peter, & Bartolotti, James 2007. "An intentional stance modulates the integration of gesture and speech during comprehension." *Brain and Language* 101: 222–233. doi:10.1016/j.bandl.2006.07.008
- Kelly, Spencer D., Manning, Sarah M., & Rodak, Sabrina 2008. "Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education." *Language and Linguistics Compass* 2 (4): 569–588. doi:10.1111/j.1749-818X.2008.00067.x
- Kelly, Spencer D., McDevitt, Tara, & Esch, Megan 2009. "Brief training with co-speech gesture lends a hand to word learning in a foreign language." *Language and Cognitive Processes* 24 (2): 313–334. doi:10.1080/01690960802365567
- Kelly, Spencer D., Creigh, Peter, & Bartolotti, James 2010. "Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing." *Journal of Cognitive Neuroscience* 22: 683–694. doi:10.1162/jocn.2009.21254
- Kelly, Spencer D., Özyürek, Asli, & Maris, Eric. 2010. "Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension." *Psychological Science* 21: 260–267.
- Kelly, Spencer D., & Lee, Angela L. 2012. "When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high." *Language and Cognitive Processes* 27 (6): 793–807. doi:10.1080/01690965.2011.581125
- Kelly, Spencer D., Healey, Meghan, Özyürek, Asli, & Holler, Judith 2014. "The processing of speech, gesture and action during language comprehension." *Psychonomic Bulletin & Review* 22 (2): 517–523.
- Kelly, Spencer D., Hirata, Yukari, Manansala, Michael, & Huang, Jessica 2014. "Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language." *Frontiers in Psychology* 5. doi:10.3389/fpsyg.2014.00673
- Kendon, Adam 1997. "Gesture." *Annual Review of Anthropology* 26 (1): 109–128.
- Kendon, Adam 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press. doi:10.1017/CBO9780511807572
- Kita, Sotaro, & Özyürek, Asli 2003. "What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking." *Journal of Memory and Language* 48 (1): 16–32. doi:10.1016/S0749-596X(02)00505-3
- Klaver, Peter, et al. 2005. "Word imageability affects the hippocampus in recognition memory." *Hippocampus* 15 (6): 704–712. doi:10.1002/hipo.20081
- Knill, David C., & Pouget, Alexandre 2004. "The Bayesian brain: the role of uncertainty in neural coding and computation." *Trends in Neurosciences* 27 (12): 712–719. doi:10.1016/j.tins.2004.10.007
- Krahermer, Emile, & Swerts, Marc 2007. "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception." *Journal of Memory and Language* 57 (3): 396–414. doi:10.1016/j.jml.2007.06.005
- Kuhl, Patricia K., & Meltzoff, Andrew N. 1984. "The intermodal representation of speech in infants." *Infant Behavior and Development* 7 (3): 361–381. doi:10.1016/S0163-6383(84)80050-8
- Kutas, Marta, & Hillyard, Steven A. 1980. "Reading senseless sentences: Brain potentials reflect semantic incongruity." *Science* 207 (4427): 203–205. doi:10.1126/science.7350657
- Lakoff, George, & Johnson, Mark 1999. *Philosophy In The Flesh: The Embodied Mind And Its Challenge To Western Thought*. Basic Books.
- Loehr, Daniel 2007. "Aspects of rhythm in gesture and speech." *Gesture* 7 (2): 179–214. doi:10.1075/gest.7.2.04loe

- Macedonia, Manuela, Müller, Karsten, & Friederici, Angela D. 2011. "The impact of iconic gestures on foreign language word learning and its neural substrate." *Human Brain Mapping* 32 (6): 982–998. doi:10.1002/hbm.21084
- Masumoto, Kouhei, *et al.* 2006. "Reactivation of physical motor information in the memory of action events." *Brain Research* 1101 (1): 102–109. doi:10.1016/j.brainres.2006.05.033
- Mahon, Bradford Z., & Caramazza, Alfonso 2008. "A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content." *Journal of Physiology-Paris* 102 (1): 59–70. doi:10.1016/j.jphysparis.2008.03.004
- McGurk, Harry, & MacDonald, John 1976. "Hearing lips and seeing voices." *Nature* 264: 746–748 doi:10.1038/264746a0
- McNeill, David 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press.
- McNeill, David 2008. *Gesture and Thought*. University of Chicago Press.
- McNeill, David 2012. *How Language Began: Gesture and Speech in Human Evolution*. Cambridge University Press. doi:10.1017/CBO9781139108669
- McNeill, David 2013. "Why we gesture." *On Research Gate*.
- Morford, Marolyn, & Goldin-Meadow, Susan 1992. "Comprehension and production of gesture in combination with speech in one-word speakers." *Journal of Child Language* 19 (03): 559–580. doi:10.1017/S0305000900011569
- Murphy, Catherine M., & Messer, David J. 1977. "Mothers, infants and pointing: A study of a gesture." *Studies in Mother-Infant Interaction*: 325–354.
- Özyürek, Asli, Willems, Roel, Kita, Sotaro, & Hagoort, Peter 2007. "On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials." *Journal of Cognitive Neuroscience* 19 (4): 605–616. doi:10.1162/jocn.2007.19.4.605
- Özyürek, Asli 2014. "Hearing and seeing meaning in speech and gesture: insights from brain and behaviour." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130296. doi:10.1098/rstb.2013.0296
- Perry, Michelle, Church, R. Breckinridge, & Goldin-Meadow, Susan 1992. "Is gesture-speech mismatch a general index of transitional knowledge?" *Cognitive Development* 7 (1): 109–122. doi:10.1016/0885-2014(92)90007-E
- Post, Lysanne S., Van Gog, Tamara, Paas, Fred, & Zwaan, Rolf A. 2013. "Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning." *Computers in Human Behavior* 29 (4): 1450–1455. doi:10.1016/j.chb.2013.01.005
- Povinelli, Daniel J. 1993. "Reconstructing the evolution of mind." *American Psychologist* 48: 493–509. doi:10.1037/0003-066X.48.5.493
- Pulvermüller, F. 2005. "Brain mechanisms linking language and action." *Nature Reviews Neuroscience* 6 (7): 576–582. doi:10.1038/nrn1706
- Rizzolatti, Giacomo, & Arbib, Michael A. 1998. "Language within our grasp." *Trends in Neurosciences* 21 (5): 188–194. doi:10.1016/S0166-2236(98)01260-0
- Skipper, Jeremy I. 2014. "Echoes of the spoken past: how auditory cortex hears context during speech perception." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130297. doi:10.1098/rstb.2013.0297
- So, Wing C., Kita, Sotaro, & Goldin-Meadow, Susan 2009. "Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand." *Cognitive Science* 33 (1): 115–125. doi:10.1111/j.1551-6709.2008.01006.x
- Stam, Gale 2006. "Thinking for speaking about motion: L1 and L2 speech and gesture." *IRAL-International Review of Applied Linguistics in Language Teaching* 44 (2): 145–171.

- Straube, Benjamin, Green, Antonia, Jansen, Andreas, Chatterjee, Anjan, & Kircher, Tilo 2010. "Social cues, mentalizing and the neural processing of speech accompanied by gestures." *Neuropsychologia* 48 (2): 382–393. doi:10.1016/j.neuropsychologia.2009.09.025
- Straube, Benjamin, Green, Antonia, Weis, Susanne, & Chatterjee, Anjan 2009. "Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance." *Journal of Cognitive Neuroscience* 21 (4): 821–836. doi:10.1162/jocn.2009.21053
- Sumby, Willaim H., & Pollack, Irwin 1954. "Visual contribution to speech intelligibility in noise." *The Journal of the Acoustical Society of America* 26 (2): 212–215. doi:10.1121/1.1907309
- Tenenbaum, Joshua B., Kemp, Charles, Griffiths, Thomas L., & Goodman Noah D. 2011. "How to grow a mind: Statistics, structure, and abstraction." *Science* 331: 1279–1285. doi:10.1126/science.1192788
- Thompson, Laura A., & Massaro, Dominic W. 1986. "Evaluation and integration of speech and pointing gestures during referential understanding." *Journal of Experimental Child Psychology* 42 (1): 144–168. doi:10.1016/0022-0965(86)90020-2
- Thompson, Laura A., & Massaro, Dominic W. 1994. "Children's integration of speech and pointing gestures in comprehension." *Journal of Experimental Child Psychology* 57 (3): 327–354. doi:10.1006/jecp.1994.1016
- Tomasello, Michael 2008. *Origins of Human Communication*. Cambridge: MIT Press.
- van Wassenhove, Virgini, Grant, Ken W., & Poeppel, David 2005. "Visual speech speeds up the neural processing of auditory speech." *Proceedings of the National Academy of Sciences of the United States of America* 102 (4): 1181–1186. doi:10.1073/pnas.0408949102
- Willems, Roel M., & Hagoort, Peter 2007. "Neural evidence for the interplay between language, gesture, and action: A review." *Brain and Language* 101 (3): 278–289. doi:10.1016/j.bandl.2007.03.004
- Wimsatt, William 1981. "Robustness, reliability and overdetermination." In *Scientific Inquiry and the Social Sciences*, M. Brewer and B. Collins (eds), 124–163. San Francisco: Jossey-Bass.
- Wu, Ying Choon, & Coulson, Seana 2010. "Gestures modulate speech processing early in utterances." *Neuroreport* 21 (7): 522.
- Zwaan, Rolf A. 2004. "The immersed experienter: Toward an embodied theory of language comprehension." *Psychology of Learning and Motivation* 44: 35–62. doi:10.1016/S0079-7421(03)44002-4