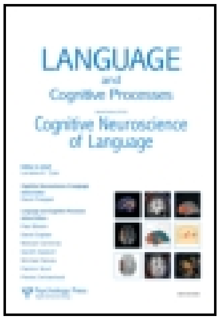

Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Temporal synchrony between speech, action and gesture during language production

R. Breckinridge Church[a], Spencer Kelly[b] & David Holcombe[a]

[a] Psychology Department, Northeastern Illinois University, 5500 N. St. Louis Avenue, Chicago, IL 60625, USA

[b] Psychology Department, Neuroscience Program and Center for Language and Brain, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA

Published online: 19 Nov 2013.

To cite this article: R. Breckinridge Church, Spencer Kelly & David Holcombe , Language and Cognitive Processes (2013): Temporal synchrony between speech, action and gesture during language production, Language and Cognitive Processes, DOI: 10.1080/01690965.2013.857783

To link to this article: http://dx.doi.org/10.1080/01690965.2013.857783

Language, Cognition and Neuroscience, 2013
http://dx.doi.org/10.1080/01690965.2013.857783

Routledge Taylor & Francis Group

Temporal synchrony between speech, action and gesture during language production

R. Breckinridge Church[a]*, Spencer Kelly[b] and David Holcombe[a]

[a]*Psychology Department, Northeastern Illinois University, 5500 N. St. Louis Avenue, Chicago, IL 60625, USA;* [b]*Psychology Department, Neuroscience Program and Center for Language and Brain, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA*

(*Received 18 September 2012; accepted 8 October 2013*)

Researchers have theorised that speech and gesture are integrated in communication. We ask whether this integrated relationship is indexed by a unique temporal link between speech and gesture. University students performed videotaped tasks that elicited: (1) speech and action descriptions about how to act on objects and (2) speech and gesture descriptions about how to act on objects. Integration was indexed by measuring the onset of speech with actions and speech with gestures (in milliseconds), with smaller differences reflecting a greater degree of synchrony. One hundred per cent of the subjects gestured in the gesture condition and performed actions in the action condition. Speech and gesture were more tightly synchronised than speech with action. The greater synchrony between gesture and speech suggests that the two could be uniquely designed to work together for the purpose of communication.

Keywords: gesture; action; speech; language; synchrony

Hand gestures are a natural part of spoken language. Although gestures are similar to actual actions on objects, it is unknown whether these two manual movements have the same or different relationship with accompanying speech. The present study is a first step at addressing this question.

A common type of co-speech gesture is an iconic gesture, which visually represents information about object attributes, relations and actions. These gestures temporally co-occur with speech and often provide information that supplements or complements the accompanying speech to provide a more complete message than either modality alone (McNeill, 1992). For example, imagine making a basketball shooting gesture while asking your friend if he wants to go to the gym after work. Note that speech or gesture alone is not sufficient to convey the complete *what, when and where* of the message – only the combination of two modalities reveals the whole meaning.

Gestures not only help the speaker convey a complete idea but it can aid in the listener's comprehension of communication. Research has shown that both children and adults detect and process information conveyed in gesture and that gesture enhances the processing and memory for the speech it accompanies (Church, Garber & Rogalsky, 2007; Goldin-Meadow, Wein, & Chang, 1992; McNeill, Cassell, & McCullough, 1994; Thompson, Driscoll, & Markson, 1998; Thompson & Massaro, 1994). Moreover, listeners (observers) are likely to detect, process and remember information from gesture that accompanies speech (Church, Kelly, & Lynch, 2000; Kelly, Barr, Church, & Lynch, 1999; Kelly & Church, 1997, 1998). Thus, there is evidence that co-speech gestures and accompanying speech reflect an integrated or combinatory system of communication intertwined in the process of language production as well as comprehension (Goldin-Meadow, McNeill, & Singleton, 1996; Kelly, Ozyurek, & Maris, 2010; Loehr, 2007; McNeill, 1992).

Evidence that speech and gesture are conceptually integrated comes from a variety of perspectives. From the perspective of neuroscience, hand gestures influence semantic (and pre-semantic) stages of the brain processing of speech (Kelly, Kravitz, & Hopkins, 2004) and the two modalities both activate Broca's area, an area of the brain used in language processing (Willems, Ozyurek, & Hagoort, 2007). Moreover, there is some evolutionary evidence suggesting that speech and gesture have been connected forms of communication since the dawn of human communication (Armstrong & Wilcox, 2007; McNeill, 2012). Armstrong and Wilcox (2007) suggest that the origin of language is gestural and depictive, coming out of manual actions and shared perceptions of our surroundings. Originally, however, gesture was the predominant modality of communication while speech in the form of accentuating vocalizations was the secondary communication modality. Eventually, according to this account, speech evolved as the primary modality for communication. McNeill (2012) takes the evolutionary argument a step further and theorises that speech and gesture are such a unique and natural communicative

*Corresponding author. Email: rbchurch@neiu.edu

pairing that they must have evolved *together* as a semiotic image-language system. The point to be made here is that the integral connection between the speech and gesture modalities has had a long evolutionary history (Armstrong & Wilcox, 2007; McNeill, 2012).

Temporally, speech and gesture have a tight connection as well (Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Loehr, 2007; McNeill, 1992; Wachsmuth, 2002). For example, Loehr (2007) videotaped four groups of friends in a conversational setting. Speech and gesture was recorded and coded for timing. Loehr's study showed that speech and gesture were temporally synchronised with the onset of speech to a high degree in all utterances, and this was taken as strong evidence that the two modalities comprise a tightly integrated system (see McNeill, 1992, for more on this).

The present study investigates whether gestures have a special temporal relationship with speech or whether any manual action that accompanies speech is similarly coupled. One manual movement that may rival a hand gesture for its tight coupling with speech is a co-speech action on an object. For example, imagine someone folding an actual piece of paper while verbally explaining how to make an origami object. Note how the physical presence of the actual object being described directly and concretely *scaffolds* the accompanying speech. In this way, producing actions on objects while speaking might be very similar to gesturing while speaking. Indeed, some researchers have argued that hand gestures are simulations of actual actions on real objects (Cook & Tanenhaus, 2009; Goldin-Meadow & Beilock, 2010; Hostetter & Alibali, 2008). For example, Hostetter and Alibali (2008) note that simulating an action (e.g., pantomiming) activates neural areas that are active during the real production of that action (see Jeannerod, 2001; Rizzolatti & Craighero, 2004). So it is conceivable that when someone produces a gesture (e.g., folding), it could be a simulated version of the actual act of folding. If this were true, producing a gesture could be virtually the same as producing an actual action on an object. Thus, one might hypothesise that actions are no different than gestures in terms of their temporal connection to speech.

In the present experiment, participants described everyday activities (e.g., sweeping the floor, combing hair, throwing a dart; see Appendices 1 and 2) with either speech and gesture or speech and actions on objects. These video-recorded descriptions were coded for the extent to which the speech and gesture/action temporally coincided. Using this synchrony as an index for speech and gesture being a uniquely integrated communication system (McNeill, 1992), we advanced two hypotheses: (1) if gesture and speech have a privileged relationship in language production (Armstrong & Wilcox, 2007; McNeill, 1992), gestures should be more temporally aligned with speech than actions on objects; and (2) if gesture and speech are functionally equivalent during language production, gesture and actions should have a similar temporal relationship with speech.

Method

Participants

Fifteen university students (1 male, 14 females) aged 19–32 participated. The participants were volunteers who were solicited through the Sona system for voluntary research participation. The participants signed the consent forms and were given extra course credit points to participate.

Design and procedure

The within subject variable was the type of movement paired with speech (speech-gesture and speech-action), and the dependent variable was the timing of the *onsets* of speech, gesture and action. That is, we measured the difference in timing between a gesture onset (or *growth point*; see McNeill, 1992) and the onset of its lexical affiliate, in addition to the onset of an action on an object and its lexical affiliate. The *growth point* of gesture signals the inception of an idea, and therefore, the point at which we would want to determine synchrony. Others have pointed to the importance of measuring synchrony at this conceptual planning phase of communication rather than the stroke phase (Hagoort & Chu, 2012; Kita, Ozurek, Allen, Brown, Furman, & Ishizuka, 2007). The advantage of measuring onsets at the preparation phase is that the physical demands of preparing for a gesture compared to an actual action are more comparable than when producing the stroke (e.g., more dexterity and effort is required to actually put on a coat than merely gesture it). Thus, the closer that gesture or action preparations were aligned in time with the accompanying lexical affiliate was our operational definition of a system with *tight synchrony.*

The participants were told that they would be performing physical scenarios based on the use of simple everyday objects and that these scenarios would be video-recorded. Once the participants appeared comfortable and appeared to understand the task, they performed seven speech-gesture and seven speech-action paired descriptions, counterbalanced to control for order effects. Two scripts containing random sequences of the item task scenarios were used. Participants were randomly assigned to one of the two orders. That is, the seven speech-action and seven speech-gesture tasks were randomly ordered such that in some cases, a subject would do a speech-action vignette before the comparable speech-gesture vignette and vice versa.

All of the objects were present in the testing laboratory. For each object-action scenario, the object was set on a table in front of the subject. For the *speech-action* condition, participants were prompted with the following question, 'Please describe to me how to throw a dart. Please *use* the object in front of you'. Participants were prompted with the following question for the *speech-gesture* condition, 'Please describe to me how to throw a dart. Please do *not use* the object in front of you' (see Appendix 1 for all seven task requests). The video camera recorded the responses, and the relative onsets of speech and gesture and speech and action were coded at the level of milliseconds (see below).

Materials

A Sony video camera capable of capturing video with a time code was used to capture the movements of the subjects. The video was captured in standard NTSC video at a rate of 29.97 frames per second (fps). A Dell computer capable of video review was used to code the videotapes. The video was encoded from videotape to MPEG video file using a Hauppage video encoding device. Props included a dart, comb, broom, spoon, rubber bowl, a jacket and a pool stick.

Coding and analysis

For purposes of this study, temporal synchrony is defined as the difference in time between the onset at the preparation phase (not the stroke; see McNeill, 1992) of the *primary* gestured or manual action and the onset of the related verb produced in response to the request by the experimenter about a particular object[1] (e.g., 'describe to me how to throw a dart…'). Each video was reviewed and temporal synchrony was recorded for each speech-gesture and speech-action scenario in the following way. First, a coder watched the video until the particular target verb, gesture or action was produced. Once the target verb, action or gesture was identified, the coder watched the video in slow motion. When looking at a speech-gesture description, the coder first focused on the onset of speech, then the onset of gesture. This order of coding was used for the speech-action descriptions as well. The ELAN system for coding video speech and gesture data (Sloetjes & Wittenburg, 2008) was used to code the timing of the speech-action and speech-gesture descriptions. This system allows for a fine-grained time code using milliseconds so we could precisely identify onset times. To accurately mark the onset times, the coder would have to rewind and play the identified verb, gesture and action many times to precisely identify onset times. Once the onset was identified, the coder hit the pause button and marked the time provided by the time code in the ELAN system. Onset for speech was determined by marking the articulation of the initial consonant of the target verb. The onset of gestures was determined by marking the initial preparatory movement prior to the stroke of the gesture. This was identified by looking at the moment when the gesturing hand became tense in preparation for creating the fully formed gesture. In the example of the gesture for throw, the onset time would be marked when the arm moved from a rest position to the position where the arm moved back just prior to the release or throwing action. The onset time for the action of throwing a dart would be similarly marked when the arm moved back just prior to the actual throwing action (which some subjects actually did). Note that we did not code with a particular hypothesised direction for synchrony patterns in mind.

To determine reliability, we brought in a third coder who was not familiar with the purpose of the study to recode 21 speech-action and 21 speech-gesture descriptions (15% of the total descriptions; three subjects). This coder was trained in how to identify target verbs, gestures and actions (those related to the actions associated with the props) and how to determine the onsets of the target verbs, actions and gestures. We first wanted to establish reliability for identifying the particular actions, gestures and verbs that were to be marked for onset times. For the most part, the verbs that were related to the requests (throw, sweep, mix, comb hair, remove lid, use a pool stick, put on a jacket) would be coded as the target verbs. In rare cases, subjects would deviate slightly by using alternative verbs like 'twist' instead of 'remove' or 'stir' instead of 'mix', but even these were easy to identify. Reliability for identifying target verbs (those accompanied by gestures and those accompanied by actions, $N = 42$) within a subject's description was 95% for identifying target gestures ($N = 21$) and 95% for identifying target actions ($N = 21$).

Reliability coding for onset times in milliseconds was extremely difficult. Given the fine level of resolution, we decided that a discrepancy between coders that was less than 50 ms would be considered agreement in onset time. Any discrepancy larger than this was considered as lack of

[1]The descriptions often included statements about the sequence of steps leading to the primary action (e.g., when asked to describe how to throw a dart, a subject said, 'first I would pick up the object and then I would place my right foot forward and then throw the dart'. In addition, some subjects felt compelled to comment on the result saying something like, '…in order to hit the target'. We focused on the primary target action, gesture and verb related to the request given by the researcher. In the example just described, we coded the onset of the verb 'throw' and the onset of the accompanying action or gesture for throwing (see Appendices B and C for examples of these descriptions).

agreement. Given this criterion, inter-rater reliability was 86% for coding the onset time of speech when it occurred with an action (mean difference in onset times between coders = 31 ms; SD = 12 ms), 93% when coding the onset of speech when it occurred with a gesture, (mean difference in onset times between coders = 19 ms; SD = 28 ms), 86% for coding the onset time of gesture (mean difference in onset times between coders = 16 ms; SD = 26 ms) and 86% for coding the onset time of action (mean difference in onset times between coders = 43 ms; SD = 104 ms).

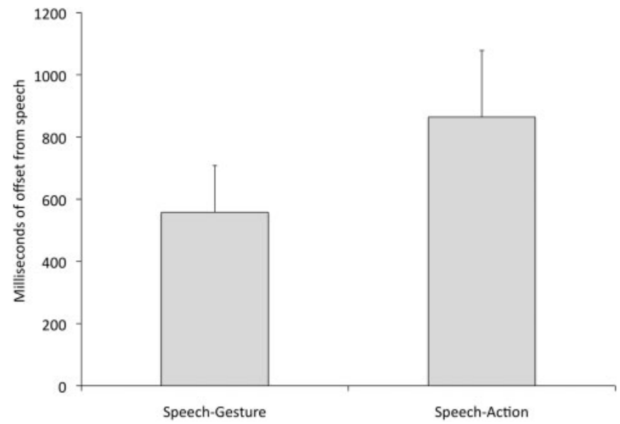

Figure 1. Average onset differences in milliseconds between speech with gesture and speech with action. The means represented in this graph are the average difference across all task actions between the onset time of speech target verbs and the onset time of those gestures paired with the verbs, compared with the onset time of speech target verbs and the onset times of those actions paired with the verbs. The results indicate that the difference in timing of the onsets for speech target verbs and paired target gestures is smaller (showing tighter synchrony) than the difference in timing of the onsets of speech target verbs and paired target actions.

Results

Comparing speech-gesture and speech-action synchrony

Comparing the difference in the onset of speech and gesture versus speech and action, we found that speech and gesture was significantly more tightly linked than speech and action. On average, the difference between speech-gesture onsets was 593 ms (SD = 320 ms), whereas the difference between speech-action onsets was 863 ms (SD = 440 ms). This was significant by analysing with subjects as the unit of analysis, $t(14) = 2.17$, $p < .05$, (see Figure 1) and by items, $t(6) = 2.45$; $p = .05$ (see Figure 2).

Does the onset of gestures and action precede or follow the onset of speech?

Note that across the speech-gesture or speech-action pairings, synchrony was somewhat variable. This prompted us to examine whether, as McNeill (1992) might predict, the visuo-spatial depiction of the underlying conceptual representation anticipated the speech. We found that for the vast majority of speech-gesture and speech-action pairings, the onset of the action depiction (gestural or manual) anticipated the onset of speech: 81% of the 105 speech-gesture pairings and 89% of the speech-action pairings. This pattern was significantly different from chance (binomial test $p < .001$ for both gesture-speech and action-speech).

Are gestures and action related?

Finally, to explore how consistent participants were in producing actions and gestures, we explored whether *within the same vignette*, there was a correlation between the two in terms of whether they both preceded or followed the onset of speech. We found that gesture and action were highly correlated in their temporal relationship to speech with vignette pairs. That is, when the onset of an individual's gesture occurred before speech for a given vignette, the onset of that individual's comparable action for that same vignette also occurred before speech, and vice versa. For the 85 instances where the onset of gesture *preceded* the onset of speech, the onset of action for that same item also *preceded* the onset of speech 93% of the time and for the 20 items where the onset of gesture *followed* the onset of speech, the onset of action also *followed* the onset of speech 70% of the time, $\chi 2$ (1, $N = 105$) = 8.42, $p < .001$).

Qualitative analysis of speech with action versus gesture

To get a better qualitative sense of the verbal, gestural and actional descriptions our participants produced, we provide some examples in Appendix 2 and 3. Appendix 2 provides just the verbal accompaniments to the gestures and actions for descriptions of throwing a dart, sweeping with a broom, mixing and opening a bottle. Note that the descriptions provided in the speech accompanying action were very similar to the speech accompanying gesture regardless of how many scenarios occurred between the speech-action and speech-gesture productions. Appendix 3 provides examples of two subject's descriptions of throwing a dart and putting on a jacket using a combination of speech + action and then speech + gesture. The gestures and actions are also very similar, with the only difference being the absence of the object in the gesture condition.

Discussion

These results support the first hypothesis that there is tighter synchrony between speech and gesture than between speech and action. In fact, the preparation for a

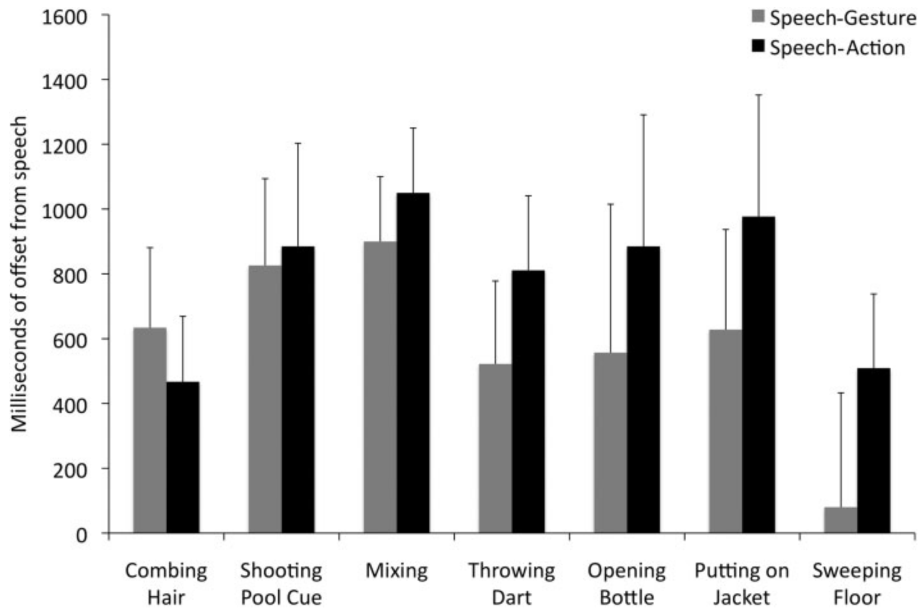

Figure 2. Average onset differences in milliseconds between speech with gesture and speech with action for each task action. The means represented in this graph are the average difference for each of the tasks actions, between the onset time of speech target verbs and the onset time of those gestures paired with the verbs, compared with the onset time of speech target verbs and the onset times of those actions paired with the verbs. The results indicate that the difference in timing of the onsets for speech target verbs and paired target gestures is smaller (showing tighter synchrony) than the difference in timing of the onsets of speech target verbs and paired target actions across all task actions with the exception of the task of combing hair.

gesture was over a quarter of a second closer to speech – a vast span in the timescale of online language processing – than the preparation for an action. The pattern of tighter synchrony between speech and gesture compared to speech and action was also predominant across the task items. Thus, the gestures produced in this study showed a unique and privileged temporal coupling with speech, as suggested in the literature (Loehr, 2007; McNeill, 1992).

Despite the differences in timing, gesture and action did have some significant commonalities. First, the fact that both action and gesture preparations predominantly occurred *before* speech is consistent with past studies measuring the timing of gesture and speech (Loehr, 2007; McNeill, 1992). This relative timing has been taken as evidence that non-verbal/perceptual-motor representations – in this case, gestural and action-based – fundamentally underlie production of linguistic expressions (Hostetter & Alibali, 2008; McNeill, 1992). Second, subjects were remarkably consistent with how they produced gestures and actions relative to speech within items, giving credence to claims that iconic gestures indeed may be simulations of actual actions on objects (Armstrong & Wilcox, 2007; Hostetter & Alibali, 2008).

Possible explanations for tighter speech-gesture than speech-action synchrony

There are a number of explanations for the synchrony results. One possibility is that speech-gesture onset was more synchronised than speech-action onset simply because the quality of the speech-accompanying gesture was different from the quality of speech-accompanying action. However, our qualitative descriptions (Appendix 2) suggest that this is not the case: when gestures and actions were used to accompanying speech descriptions of task performance, the speech was strikingly similar.

A second explanation is that because gesture is more likely to occur with speech, it is more practiced than action occurring with speech. The more natural occurrence of gesture with speech might therefore result in a tighter temporal synchrony than speech with action. This explanation, of course, is consistent with the literature suggesting that gesture and speech form an integrated system.

A third explanation is that the physical and cognitive processes by which subjects generated actions and gestures were fundamentally different. That is, perhaps it was more physically and cognitively demanding to produce an action on an object – after all, an actual object was being manipulated which required more attention and dexterity – and this made its relationship with speech much more variable and unpredictable. Although we cannot rule out this possibility, it is worth noting that rather than actions on objects looking inconsistent and stochastic, the correlation analyses showed that the relative timing between actions and speech was systematically related to the timing between gestures and speech. As mentioned earlier, this suggests that there are at least a few important shared processes involved in producing both actions and gestures with speech.

A fourth explanation is that gesture and speech are indeed a tightly integrated communication system. That is, gesture and speech may share the same representational production mechanisms in a way that actions on objects and speech do not (McNeill, 1992; McNeill & Duncan, 2000). Note that this explanation does not really contradict the explanations provided here. Indeed, this explanation accounts for *why* speech might operate differently when accompanied by action versus gesture, and *why* gestures are less physically and cognitively demanding than actions on objects. Below, we discuss theoretical implications for this fourth possibility.

Implications for theory

The present results are interesting in light of the recent *Gesture as Simulated Action* (GSA) framework proposed by Hostetter and Alibali (2008). The GSA claims that gestures are simulations of real actions on the world – indeed, the strong correlations between action and gesture onsets suggests that they are at least close family members. Given this close relationship, it is interesting to ask why actions are less temporally coupled with speech than gesture. After all, in some circumstances actions can afford equal if not clearer information than gesture. For example, actually showing how to tie a shoelace may be more informative than gesturing how to tie a shoelace. Armstrong and Wilcox (2007) have suggested that original human communication was probably oriented around showing how to perform actions like killing a buffalo, preparing a fire or making an arrow – all of which likely used physical props in their execution. It is also interesting to note that there are many modern circumstances where *actions do speak louder than words* as in the case of teaching an activity like weaving (Greenfield, 1984).

Using a gestural simulation of an action, however, could confer advantages. In addition to the fact that gestures are extremely convenient (no props required), gestural simulations are not bound by the physics of actual things in the world. This may liberate speakers from the responsibilities of cumbersome actions and allow flexibility in communication across time and contexts (Armstrong & Wilcox, 2007; Cartmill, Beilock, & Goldin-Meadow, 2012; Ping & Goldin-Meadow, 2008). This flexibility may allow gestures to travel more freely with accompanying language. Moreover, gestures may be more communicatively pure – that is, more dedicated to communication – than actions on objects. Consider that actions on objects have two possible functions: they allow someone to accomplish instrumental goals by physically manipulating the world *or* they communicate to *someone else* about how to accomplish those goals. In contrast, gestures serve only the communicative function. Therefore, although an underlying action representation may be the original source that generates language (including gesture; Hostetter & Alibali, 2008), hand gestures – more than actual actions on objects – may be best suited to go with the speech that springs from that representation.

There is recent evidence to support the claim that gestures may play a more advantageous role with speech than action in the realm of problem solving (Beilock & Goldin-Meadow, 2010; Goldin-Meadow & Beilock, 2010). Goldin-Meadow and Beilock (2010) pitted gesture production against action production while people solved the Tower of Hanoi (TOH) and found that producing gestures actually reflected a more accurate representation and, consequently, a more sophisticated understanding of the problem than actual actions on the TOH objects. This suggests that gestures play a special role compared to actual actions not only in speaking but in how people think about and solve conceptual problems in the absence of speech.

Conclusion

This study examined the relationship between two aspects of multimodal communication, one manual and the other verbal. Building on previous research showing that hand gestures reliably accompany speech and significantly impact comprehension, we have shown – for the first time – that gestures are more tightly linked to speech than actions are. We suggest that this new finding provides additional evidence that gestures may be *uniquely designed* to accompany speech.

It should be noted that the context of these results are limited; asking individuals to show how they act and gesture on a narrow set of objects is a very constrained task. Indeed, it is possible that other more complicated language contexts would result in different temporal patterns. Moreover, although our measure of an *integrated system* – which focused on the temporal synchrony of preparations of gesture and action relative to speech – is well grounded in the literature, future research should investigate other indices of integration before we can conclude for certain that gesture and speech have a unique and privileged relationship during language production. These findings are, however, a first step in this direction and should serve an initial building block towards better understanding how language is grounded in action, and how all actions may not be created equal.

Acknowledgements

We would like to acknowledge all those college students who participated in this study, the Psychology departments at Northeastern Illinois and Colgate Universities and the Center for Language and Brain at Colgate University for their generous support. We would also like to thank David McNeill who was the inspiration for this study and provided continuous support with the interpretation of the data and many other individuals who have helped shape our thinking about gesture, in particular, Susan Goldin-Meadow, Martha Alibali, Sotaro Kita and Melissa Singer. We would like to particularly thank Theodora Koumoutsakis for her invaluable help with the preparation of this manuscript.

References

Armstrong, D. F., & Wilcox, S. (2007). *The gestural origin of language*. Oxford: Oxford University Press.

Beilock, S., & Goldin-Meadow S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*, 1605–1610. doi:10.1177/0956797610385353

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 129–143. doi:10.1098/rstb.2011.0162

Church, R. B., Garber, P., & Rogalsky, K. (2007). The role of gesture in social communication and memory. *Gesture*, *7*, 137–157. doi:10.1075/gest.7.2.02bre

Church, R. B., Kelly, S. D., & Lynch, K. (2000). Multimodal processing over development: The case of speech and gesture detection. *Journal of Nonverbal Behavior*, *24*(2), 151–174. doi:10.1023/A:1006610013873

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied understanding: Speaker's gestures affect listener's actions. *Cognition*, *113*, 98–104. doi:10.1016/j.cognition.2009.06.006

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*, 664–674. doi:10.1177/1745691610388764

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*(1), 34–55. doi:10.1037/0033-295X.103.1.34

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction*, *9*(3), 201–219. doi:10.1037/0033-295X.103.1.34

Greenfield, P. M. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 117–138). Cambridge, MA: Harvard University Press.

Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, *23*, 1845–1854. doi:10.1162/jocn.2010.21462

Hagoort, P., & Chu, M. (2012). *The synchronization of speech and gesture*. Paper presented at the International Society of Gestural Studies, Lund, Sweden.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*, 495–514. doi:10.3758/PBR.15.3.495

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, *14*, S103–S109. doi:10.1006/nimg.2001.0832

Kelly, S. D., & Church, R. B. (1997). Can children detect conceptual information conveyed through other children's nonverbal behaviors? *Cognition and Instruction*, *15*(1), 107–134. doi:10.1207/s1532690xci1501_4

Kelly, S. D., & Church, R. B. (1998). A comparison between children's and adult's ability to detect conceptual information conveyed through representational gesture. *Child Development*, *69*, 85–93.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577–592. doi:10.1006/jmla.1999.2634

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain & Language*, *89*, 253–260. doi:10.1016/S0093-934X(03)00335-3

Kelly, S. D., Ozyurek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*, 260–267. doi:10.1177/0956797609357327

Kita, S., Ozurek, A., Allen, S.Brown, A.Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes*, *22*, 1212–1236. doi:10.1080/01690960701461426

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, *7*(2), 179–214. doi:10.1075/gest.7.2.04loe

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. New York, NY: Cambridge University press.

McNeill, D., Cassell, J., & McCullough, K. (1994). Communicative effects of speech- mismatched gestures. *Research on Language and Social Action*, *27*, 223–237.

McNeill, D., & Duncan, S. (2000). Growth points in thinking for speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, *44*, 1277–1287. doi:10.1037/0012-1649.44.5.1277

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. doi:10.1146/annurev.neuro.27.070203.144230

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). http://www.lrec-conf.org/lrec2008/

The LREC is a conference that publishes it's proceedings on-line. There is no reference to a publisher, the website is the following: http://www.lrec-conf.org/lrec2008/

Thompson, L. A., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, *22*(3), 167–187. doi:10.1023/A:1022914521401

Thompson, L. A., & Massaro, D. W. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology*, *57*, 327–354. doi:10.1006/jecp.1994.1016

Wachsmuth, I. (2002). *Communicative rhythm in gesture and speech*. Amsterdam: John Benjamins Publishing Company.

Willems, R. M., Ozyurek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*, 2322–2333. doi:10.1093/cercor/bhl141

Appendix 1. Task Requests

1. Please describe to me how to throw a dart, please [do not]* use the object in front of you.
2. Please describe to me how to comb hair, please [do not]* use the object in front of you.
3. Please describe to me how to sweep the floor, please [do not]* use the object in front of you.
4. Please describe to me how to mix ingredients in a bowl, please [do not]* use the object in front of you.
5. Please describe to me how to remove a lid from a bottle, please [do not]* use the object in front of you.
6. Please describe to me how to put on a jacket, please [do not]* use the object in front of you.
7. Please describe to me how to use a pool stick, please [do not]* use the object in front of you.

*Bracketed words were added for eliciting gesture

Appendix 2. Examples of Speech-Action and Speech-Gesture Descriptions of Actions on Objects *

Subject 1

Dart (speech-gesture description immediately followed speech-action)

1. **Speech with action**: I would pick up the dart, place it in my right hand because I am right handed, aim at the object, and place my right arm back and [throw] it.
2. **Speech with gesture**: Grabbing it in my right hand because I'm right handed, pick my arm up, aim at the object and step forward and [throw] it.

Subject 2

Dart (speech-gesture followed by speech-action description 10 scenarios later)

1. **Speech with action**: This my dart feather side facing me, pointy side facing away from me, holding between thumb and fore fingers and [throw].
2. **Speech with gesture**: I put it between my fingers to stabilise it and then [throw] it.

Subject 3

Broom (speech-gesture followed by speech-action description 7 scenarios later)

1. **Speech with action**: Pick up the broom. [Sweep] the dirt from one side of the room to the other side.
2. **Speech with gesture**: First you would pick up the broom and put it on the floor and then what you would do is [sweep] the dirt from one side of the room until you reach the other side.

Subject 4

Mixing bowl (speech-action followed by speech-gesture 6 scenarios later)

1. **Speech with action**:This is what you do, you pick this up and [stir] it all around and mix it all together.
2. **Speech with gesture**: Pick up the spatula [mix] all the ingredients all together until its all blended.

Subject 5

Bottle (speech-gesture followed by speech action description 6 scenarios later)

1. **Speech with action**: I grab the bottle, grab the lid and [turn] it in the counter-clockwise motion.
2. **Speech with gesture**: I take the bottle and then I [twist] it in a counter-clockwise motion and take the top off.

* Bracketed words are the target verbs

Appendix 3. Speech-action versus speech-gesture examples

<u>Subject 1: Dart</u>

Speech-action

[Action]	Picks up dart and holds it, moves dart in front of right eye, arm goesback and throws dart
[Speech]	"I would pick up the dart place it in my right hand because I am right handed, aim at the object, and place my right arm back and throw it like that…"

Speech-gesture

[Gesture]	Picks up dart holds dart at waist, hold dart rotates up, holds dart the front of the right eye, body leans forward, throw dart
[Speech]	"Grabbing it in my right hand because I'm right handed, pick my arm up, aim at the object and step forward and [throw] it."

<u>Subject 2: Jacket</u>

Speech-action

[Action]	Picks up the jacket, puts one arm through, puts other arm through, zips up jacket
[Speech]	"You pick it up, put one arm through, put the other arm through, then zip it up."

Speech-gesture

[Gesture]	Put one arm through, right hand swings over chest, zip up jacket
[Speech]	"Put one arm through, swing it over, then you zip it up."