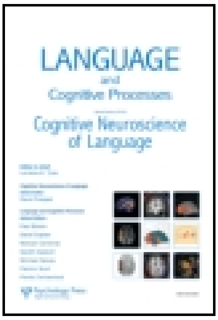

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high

Spencer D. Kelly [a] & Angela L. Lee [a]
[a] Department of Psychology, Colgate University, Hamilton, NY, USA

Available online: 25 Oct 2011

To cite this article: Spencer D. Kelly & Angela L. Lee (2011): When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high, Language and Cognitive Processes, DOI:10.1080/01690965.2011.581125

To link to this article: http://dx.doi.org/10.1080/01690965.2011.581125

iFirst

LANGUAGE AND COGNITIVE PROCESSES
0000, 00 (00), 1–15

Psychology Press
Taylor & Francis Group

When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high

Spencer D. Kelly and Angela L. Lee

Department of Psychology, Colgate University, Hamilton, NY, USA

It is now widely accepted that hand gestures help people understand and learn language. Here, we provide an exception to this general rule—when phonetic demands are high, gesture actually hurts. Native English-speaking adults were instructed on the meaning of novel Japanese word pairs that were for non-native speakers phonetically hard (/ite/ vs. /itte/, which differ by only a geminate) or easy (/tate/ vs. /butta/, which differ by a geminate and also their segmental composition). The words were presented either with or without congruent iconic gestures, for example, "Ite means stay" (with a STAY gesture). After instruction, participants were given phonetic and vocabulary tests for the words they had learned. Although performance for the phonetic task was above chance for all conditions, gesture played different roles in the semantic task for easy and hard word pairs—it helped word learning for easy pairs, but it hurt for hard pairs. These results suggest that gesture and speech are semantically integrated during word learning, but only when phonetic demands are not too high.

Keywords: Hand; Gesture; Iconic; Beat; Speech; Language; Semantic; Word; Phoneme; Phonetic; Prosodic; Geminate; Foreign; L2; Learning; Memory; Vocabulary.

Correspondence should be addressed to Spencer D. Kelly, Department of Psychology, Colgate University, 13 Oak Dr., Hamilton, NY 13346, USA. E-mail: skelly@mail.colgate.edu

The authors would like to thank Yukari Hirata for consulting with us on early phases of this project, reading drafts of the manuscript and helping us with the revision. We also thank Satomi Nunokawa for serving as the speaker (gesturer) in our instructional stimuli. We also thank the members of the Nijmegen Gesture Center at the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands) for feedback on the data.

http://www.psypress.com/lcp DOI: 10.1080/01690965.2011.581125

There has been a proliferation of research in recent years illustrating the beneficial role that co-speech hand gestures play in language comprehension and learning. This work has bolstered theoretical claims that gesture and speech comprise a tightly integrated semantic system (Bernardis & Gentilucci, 2006; Clark, 1996; Goldin-Meadow, 2003; Hostetter & Alibali, 2008; Kendon, 2004, Kita & Özyürek, 2003; McNeill, 1992, 2005). However, much less is known about how gesture influences other levels of language, such as basic speech processing. In the present study, we demonstrate that when phonetic demands are high, gesture actually hinders—rather than helps—language comprehension and learning.

Although theories on the integration of gesture and speech originated in the realm of language production, researchers have provided empirical support for an integrated relationship in language comprehension as well (for a review, see Kelly, Manning, & Rodak, 2008). In the first study to explicitly relate theories on gesture production to gesture comprehension, Kelly, Özyürek and Maris (2010) have advanced the "Integrated Systems Hypothesis" (ISH) to explain *how* gesture and speech interact during comprehension. Specifically, they demonstrated that speech and iconic gestures (which visually convey information about the meaning of objects, actions, and spatial relationships) mutually and obligatorily interact during comprehension—so much so, the researchers claim, that gesture and speech should be viewed together as mutually constituting language (see also, Clark, 1996; McNeill, 1992).

Strong claims such as this require a more thorough examination of what exactly *is integrated* between speech and gesture. Indeed, the common thread running through much of this previous work is that the "integrated relationship" is fundamentally a *semantic* one involving the connection of speech and iconic gestures. However, it is possible that other types of gestures, such as beats, may interact with different aspects of language, such as processing speech sounds. Beat gestures are different from iconic gestures in that the former are thought to convey temporal and rhythmic properties of language (e.g., stressing a word in a sentence), whereas the latter convey semantic information related to the meaning of language (e.g., depicting how someone acted on an object). Empirical support for this distinction comes from recent research demonstrating that beat gestures play a role in the global prosodic processing of words in sentences (Hubbard, Wilson, Callan, & Dapretto, 2008). Hubbard and colleagues found that low-level auditory areas, such as the planum temporale in the superior temporal gyrus, are more active during language comprehension when words are versus are not accompanied by beat gestures. The investigators concluded that beats function to focus listeners' (viewers') attention to global prosodic elements in a sentence, and this ultimately makes certain words clearer and more intelligible (see also Krahmer & Swerts, 2007).

In addition to the role that gesture plays in globally processing words in a sentence, what about the role in processing particular speech sounds within an individual word? To be sure, we know that lip and mouth movements fundamentally change how people process phonemic information, as the classic McGurk effect illustrates (Green, Kuhl, Meltzoff, & Stevens, 1991; Massaro & Cohen, 1983; McGurk & MacDonald, 1976; Skipper, van Wassenhove, Nusbaum, & Small, 2007). Moreover, recent research has shown that this visual information helps in learning non-native phonemes (Hardison, 2003; Hazen, Sennema, Iba, & Faulkner, 2005). However, although lip and mouth movements clearly play a local phonetic role—that is, they visually influence how individual speech sounds in a word are heard or learned—hand gestures might be a different story. Hirata and Kelly (2010) investigated the role that lip movements and hand gestures played in how native English-speakers perceived and learned long and short vowels in Japanese—vowel length is phonemic in Japanese, meaning that the length of a vowel *alone* changes the meaning of a word, as in /rika/ ("science") versus /rikaa/ ("liquor"). Compared to auditory training alone, the addition of lip movements helped people learn these new contrasts, but adding beat gestures did not. Apparently, this is a place in which one type of gesture, beats, do not benefit language processing.

The fact that beat gestures do not help in discriminating L2 speech sounds is interesting in light of research demonstrating that learning new words in an L2 (with familiar phonetic contrasts) is enhanced by the presence of semantically congruent *iconic* gestures (Kelly, McDevitt, & Esch, 2009; Sueyoshi & Hardison, 2005). For example, Sueyoshi and Hardison (2005) showed that Korean and Japanese learners of English (especially those with lower proficiency) understood English lectures best when iconic hand gestures accompanied those lectures compared to audio alone. So it seems that whereas beat gestures do not help vowel learning in an L2, iconic gestures do help vocabulary learning.

The present study combines these two lines of work—the role of gesture in the comprehension of words versus the processing of speech sounds—by exploring how iconic gestures help people learn non-native L2 phonetic contrasts, and subsequently, new word meanings comprised of those novel phonemes. There are at least two possible roles that iconic gestures could play in this learning context. On the one hand, perhaps the rich imagistic meaning of iconic gestures would allow learners to "hook" attention on difficult-to-perceive non-native speech contrasts. Indeed, we know that beat gestures can enhance prosodic stress and make words more intelligible and clear (Hubbard et al., 2008; Krahmer & Swerts, 2007). Perhaps iconic gestures could serve a similar function and go a step further by providing useful semantic information as well. On the other hand, iconic gestures may detract from perceiving non-native contrasts, and also subsequent learning of new words containing those contrasts. For example, in the domain of L2 phoneme learning, there is recent evidence that beat gestures actually hinder learners' ability to pay attention to more helpful information conveyed through lip movements (Hirata & Kelly, 2010). Thus, iconic gestures may actually disrupt learners' ability to hear novel phoneme contrasts in an L2, and make it difficult to attach meanings to words containing those contrasts.

The present study tests between these two possibilities by exploring the role that iconic gestures play in helping (or not helping) English speakers to learn the distinction between geminates and singletons in the Japanese language. Similar to long and short vowels, the length of time in which a speaker holds a consonant changes the meaning of a word in Japanese. For example, holding the stop "t" in the word, /itte/, produces a geminate, but releasing it immediately in the word, /ite/, produces a singleton, and this distinction results in two different meanings: "go" versus "stay." This phonemic contrast is extremely difficult for English speakers to perceive, even with hours of acoustic training (Harada, 2006; Hayes, 2001; Hayes-Harb, 2005; Hirata, 2004).

To explore the role that iconic gestures play in learning these novel contrasts, we taught native English-speakers Japanese word pairs that were phonetically difficult or easy to distinguish. The hard pairs differed only by the absence or presence of a geminate (/ite/ vs. /itte/), and the easy pairs differed by a geminate and also their segmental composition (/tate/ vs. /butta/).[1] Half the pairs were taught with iconic gestures ("/Ite/ means stay," while producing a STAY gesture), and half were taught without gesture. Following instruction, participants were tested on how well they could acoustically identify words with geminates versus singletons (phonetic task) and remember the meanings of the new words (semantic task).

Based on previous research on the role of iconic gestures in L2 vocabulary learning (Kelly et al., 2009; Sueyoshi & Hardison, 2005), we hypothesised that participants should learn and remember the meaning (semantic task) of easy words better when instructed with versus without gesture. However, there were two competing hypotheses regarding the hard words. If iconic gestures provide a facilitative context for learning novel phoneme contrast in an L2 (Hubbard et al., 2008; Krahmer & Swerts, 2007), participants should perform better on the phonetic and semantic tasks for hard words when

[1] For simplicity, the paper will refer to "hard" and "easy" word pairs. However, it should be noted that these word pairs are easy or hard only for non-native speakers of Japanese who do not use geminates to make a phonological distinction; whereas for native Japanese speakers—who makes these distinctions on a regular basis—these two types of word pairs are equally easy.

instructed with versus without gesture. In contrast, if iconic gestures do not help learning novel phoneme contrasts in an L2 (Hirata & Kelly, 2010), participants should perform no better, or even worse, on the phonetic and semantic tasks for hard words when instructed with versus without gesture.

METHOD

Participants

Forty-two college-aged participants (30 female) participated in the study. None had proficiency in a second language or any formal experience with Japanese. All completed an informed consent and received course credit for their participation.

Materials

The instructional video was comprised of a female native Japanese speaker teaching the meanings of 16 Japanese words. The words were taught in pairs that had two different levels of phonetic difficulty for English speakers. For the "hard" pairs, the two words were phonetically identical except for the absence or presence of a geminate—for example, /ite/ (stay) contains a singleton in which the "t" is not held, but /itte/ (go) contains a geminate in which the "t" is held. To the untrained ear of a native English speaker, words that differ only by a geminate are perceived as virtually identical and are very difficult to tell apart without lengthy training (Hayes, 2001; Hayes-Harb, 2005; Hirata, 2004). For the "easy" pairs, the words differed by a geminate and also their segmental composition (e.g., /tate/ [stand] vs. /butta/ [hit]). That is, in addition to containing a geminate, /butta/ phonetically differs from /tate/ at the start and end of the word. In this way, the word pairs are easier to distinguish because there are two additional cues (that are very salient to English speakers) not present in the hard pairs. The rationale for these two levels of difficulty was to determine whether gesture would play similar or different roles when phonetic demands were low (easy pairs) versus high (hard pairs). See Appendix 1 for all 16 words.

The items were chosen from a pilot study using 24 items (34 participants). Of these 24 items, we removed items for which gesture clearly did not facilitate learning compared to speech alone, leaving 16 items for the present study. In addition to this pilot, we asked a separate group of 15 participants to explicitly rate (on a scale from 1 to 7, ranging from not congruent to most congruent) the semantic relationship between the gestures and the English translations for all 16 items. The purpose of this control study was to compare whether the semantic relationship between gesture and speech was comparable for easy and hard pairs. The outcome of this "semantic relationship control" will be presented in the Results section.

In order to highlight the difference between words with geminates and singletons, the pairs of words were always presented together (i.e., /ite/ was always presented with /itte/), but the order within the pairs was randomised. In addition, the order in which the pairs were presented within the easy and hard conditions was randomised across the experiment.

The spoken content of the training was of the same format for all items. For example, to teach the "/Ite/" item, the speaker said, "/Ite/ means stay." Within each video, there were two versions of instruction. In the Gesture condition, the speaker produced iconic gestures with the Japanese word and English translation (e.g., "/Ite/ [STAY GESTURE] means 'stay' [STAY GESTURE]"), and in the No Gesture condition, the speaker produced no gesture with either word. Refer to Figure 1 for multiple frames of the gesture training videos for the /ite/ and /itte/ pair. Other than the presence or absence of gesture, the videos were designed to be as similar as possible. To help orient the participant before a new trial, the English translation ("stay", "go", etc.) was presented on the screen to participants for 3 seconds prior to the onset of each video. The purpose of this label was to avoid any confusion about the English translation of the word. The complete video for each word, edited with iMovie software, had the following format: English translation (3 seconds), instruction clip (3 seconds), 500 ms of black screen, and a repeat of the instruction clip. Each of these segments was repeated five times over the course of the training, for a total of 80 instructional videos.

There were two different versions of the entire training video (participants received one or the other)—Set 1 presented one half of the word pairs (hard and easy) with gesture and the other half without gesture, and Set 2 inverted this pattern. In this way, all words were presented with and without gesture across the two sets, but because participants received only one set, no one ever saw the same word with and without gesture. Refer to Appendix 1.

Procedure

We conducted the instruction and testing in a single day with groups of two and three participants. The experimenter introduced the concept of the geminate and explained it to participants with examples of easy and hard pairs (not drawn from the experimental pairs). Following this, half of the participants watched the hard video first, and the other half watched the easy video first (the "easy" and "hard" videos were presented in separate blocks). We broke the videos into blocks of easy and hard pairs because we were interested in possible order effects. After the first training block, participants were given a two-minute break and then were tested on what they learned (see below). Following the first test, participants watched the second

Figure 1. Sample stimuli for the hard word pairs, "Ite" and "Itte," in the Gesture condition. The three frames on the left show "stay," and the three on the right show "go." The gestures occurred twice in the instructional videos, once during the Japanese word and once during the English translation. [To view this figure in colour, please visit the online version of this Journal.]

instructional video ("easy" if "hard" first, or "hard" if "easy" first), and were subsequently tested on those items in the same fashion as the first block. Both videos (40 segments each) were approximately 17 minutes long, making for roughly 35 minutes of instruction.

Testing

Participants performed two tasks during the testing phase. In the forced choice test (semantic task), they were auditorily presented (over speakers) with the Japanese words they had heard during instruction (the words were taken directly from the instructional videos), and had to choose between two possible answers in English: the correct translation and the translation of the other word in the pair. In the geminate/singleton identification test (phonetic task), participants were auditorily presented with the same words, but this time, they had to indicate whether the word contained a geminate or a singleton. Note that in the first test, participants actually had to do a semantic and phonetic analysis of the words—that is, the only way to attach the correct meaning was to hear the phonetic distinction between the words. In the second test, participants could perform the task without any semantic processing of the words. Also note that for the hard word pairs, participants had to distinguish between the geminate and singleton (because the pair differed *only* on that dimension) in order to do the semantic task, but this was not necessarily the case for the easy words, which differed also in their segmental properties.

The two blocks of testing took approximately 15 minutes, for a total of 50 minutes for the whole experiment.

RESULTS

We ran repeated-measures ANOVAs on the arcsine transformed proportions of correct responses for the forced choice questions (semantic task) and geminate/singleton identification responses (phonetic task), with Instruction and Difficulty as within subjects factors and Order of Difficulty as a between groups factor. Order of Difficulty refers to whether participants got the "hard" or "easy" block first. In addition to running the ANOVAs across participants, we also conducted ANOVAs across items. As a further test of learning, we ran one sample T-tests to determine if performance on both tasks within each of the four conditions exceeded chance (50%). Finally, as a control, we ran a paired T-test comparing the strength of semantic relationship between gesture and speech for easy and hard pairs.

Forced choice

There was a significant main effect of Difficulty, $F(1, 40) = 63.14$, $p < .001$, by participants; $F(1, 14) = 56.32$, $p < .001$, by items, with participants learning more words in the easy pairs ($M = 0.93$, $SEM = 0.02$) than hard pairs ($M = 0.61$, $SEM = 0.04$).[2] Although there was no main effect of Instruction, $F(1, 40) = 0.12$, *ns*, by participants; $F(1, 14) = 0.28$, *ns,* by items, there was a significant interaction of Difficulty by Instruction, $F(1, 40) = 5.02$, $p = .031$, by participants; $F(1, 14) = 7.81$, $p = .014$, by items. Figure 2 shows that

[2]The means are presented in actual proportions even though arcsine transformed proportions were used for the ANOVAs.

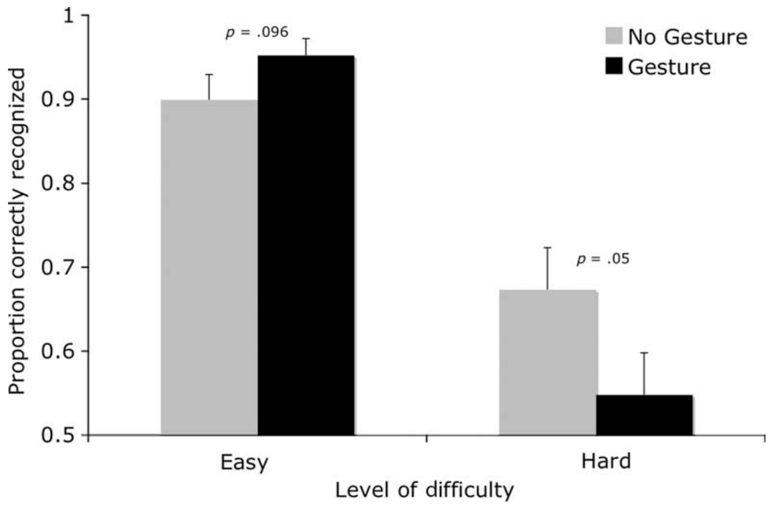

Figure 2. Proportion of correct answers for the forced choice question. The interaction shows that the Gesture versus No Gesture condition produced better learning for the easy pairs, but worse learning for the hard pairs.

participants learned more items in the easy pairs with gesture, $t(41) = 1.33$, $p = .096$ (one-tailed), but more items in the hard pairs *without* gesture $t(41) = 2.03$, $p = .05$ (two-tailed). In fact, participants were no better than chance at correctly remembering the meaning of the hard items that were in the Gesture condition, $t(41) = 0.89$, *ns*, but they were significantly above chance in the No Gesture condition, $t(41) = 3.50$, $p = .001$. Participants were also significantly above chance for easy pairs in both the Gesture, $t(41) = 23.20$, $p < .001$, and No Gesture conditions, $t(41) = 11.68$, $p < .001$. There were no significant effects of Order of Difficulty.

Geminate/singleton identification

Performance in all four conditions was significantly above chance: Gesture Easy ($M = 0.58$, $SEM = 0.03$), $t(41) = 2.47$, $p = .018$; Gesture Hard, ($M = 0.71$, $SEM = 0.05$), $t(41) = 4.47$, $p < .001$; No Gesture Easy, ($M = 0.63$, $SEM = 0.04$), $t(41) = 3.43$, $p = .001$; No Gesture Hard, ($M = 0.76$, $SEM = 0.04$), $t(41) = 6.99$, $p < .001$.

The ANOVA revealed a significant main effect of Difficulty, $F(1, 40) = 11.36$, $p = .002$, by participants, but not by items $F(1, 14) = 1.92$, *ns.*[3] However, there was no main effect of Instruction, $F(1, 40) = 1.28$, *ns*, by

[3] Because the analyses by participants and items produced inconsistent results, we do not treat this main effect as reliable.

participants; $F(1, 14) = 1.76$, *ns,* by items, and no significant interaction of Difficulty by Instruction, $F(1, 41) = 0.09$, *ns*, by participants; $F(1, 14) = 0.02$, *ns,* by items. There were no significant effects of Order of Difficulty.

Semantic relationship control

For the separate group of 15 control participants, the paired *T*-test revealed that the semantic relationship between gesture and speech was stronger for the hard items ($M = 6.02$, $SEM = 0.16$) than the easy items ($M = 4.63$, $SEM = 0.18$), $t(14) = 3.40$, $p = .004$.

DISCUSSION

The results of the vocabulary test revealed that participants learned and remembered the meaning of words in the easy pairs better when instructed with versus without gesture. In contrast, gesture hindered learning for the hard pairs. Interestingly, all conditions produced better-than-chance performance on the geminate/singleton identification task, but there were no differences between the Gesture and No Gesture conditions for this task.

The findings with the easy word pairs in the vocabulary task are in keeping with previous research demonstrating that iconic gestures help L2 vocabulary learning, at least when phonetic demands are low (Kelly et al., 2009; Sueyoshi & Hardison, 2005). In both of those studies, gestures that conveyed congruent information to the accompanying spoken instruction enhanced learners' ability to remember new word meanings with familiar phoneme constructions. However, it is worth noting that the effects of the present study were not as strong as past work. Because there were fewer easy words to learn in the present study, it is likely that there was a ceiling effect for these items—indeed, the Gesture condition had very little room to improve from the 90% correct performance in the No Gesture condition. With more words or less instruction, it is likely that this difference would have been comparable to previous work demonstrating gesture's effectiveness for learning phonetically "easy" foreign words.

The results from the phonetic task are also consistent with previous research. Training participants to discriminate between geminates and singletons resulted in performance above chance in all four conditions (see Hirata, 2004). Moreover, the fact that training with iconic gestures was no better than training with speech alone fits with previous work on the role of beat gestures in L2 phoneme learning (Hirata & Kelly, 2010). Similar to the present study, Hirata and Kelly found that Audio + Gesture training was no better than Audio Alone instruction at helping English-speakers distinguish between Japanese words with long and short vowels. So it appears that in the context of L2 learning, both beat (Hirata & Kelly, 2010) and iconic gestures (present study) are not particularly effective in training people to hear novel phoneme contrasts at the local phonetic level. This is interesting in light of work showing that beat gestures do influence speech processing at the global prosodic level in sentence contexts (Hubbard et al., 2008; Krahmer & Swerts, 2007). We will return to this issue and discuss it further in the context of a working memory model of L2 learning.

The surprising result in the present study is that gesture did not help, but actually *hurt*, participants' ability to learn the meaning of words in phonetically hard pairs. In fact, not only did it detract from learning these words, the addition of gesture produced no better than chance performance in the phonetically hard condition. There are a couple of possible explanations for this finding. One possibility is that the gestures were semantically more difficult to understand in the hard condition. However, that possibility can be ruled out because the control analysis demonstrated that if anything, the gestures for the hard items were semantically *more transparent* than the easy items. Another possibility is that participants were simply distracted by gesture and did not hear the geminate/singleton distinctions for the hard words in the Gesture condition. However, this seems unlikely in light of the results from the geminate/singleton identification task. Recall that both the No Gesture and Gesture conditions produced performance above chance in the identification task for the hard pairs, but did not statistically differ from one another.

A more likely explanation is that participants perceived the geminate/singleton distinctions for items in the hard pairs equally well in both conditions, but just had more difficulty in the Gesture condition attaching meaning to these highly similar phonetic forms. That is, perhaps iconic gestures added too much additional semantic content to the spoken input, and this visual information interfered with the ability to attach meaning to the newly-learned phonetic forms. This possibility receives some support from the literature. For example, Hayes-Harb (2007) found that learning to distinguish novel phonetic contrasts does not necessarily translate to adults being able to use those contrasts to differentiate words on a lexical task. This disassociation between phonetic and lexical processes is also evident in children's acquisition of L1. For example, Stager and Werker (1997) found that even though 14-month-olds could easily discriminate between native phonemes in a strictly auditory task, they failed to make these distinctions when they also had to attach meanings to the newly mastered phoneme constructions. In other words, when speech is phonetically novel—as in L2 learning or early L1 learning—too much meaning may tax the system, and learning breaks down.

This interpretation fits well a model of second language learning proposed by Baddeley, Gathercole, and Papagno (1998). According to the model, the phonological loop in working memory is specifically dedicated to the job of learning a new language, either as a child acquiring a first language or an adult learning a foreign language. Difficulties in learning a language arise when the phonological loop is taxed with novel speech sounds, and this disrupts the encoding of those novel sounds into permanent memories for new words. For example, when adults are asked to remember words in a language that contains novel speech sounds, the phonological loop is heavily taxed (i.e., people need to create new memories for these sounds), and performance is weak (Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992). Given the taxing load of encoding new speech sounds into long-term memory, it is possible that the multimodal addition of iconic gestures added a visually distracting dimension to the task of learning, and although it did not compromise the ability to learn these sounds, it left inadequate resources in working memory to attach meaning to them. It would be interesting to explore whether hand gestures that convey no semantic information (e.g., beat gestures) also distract learners, or whether only iconic gestures that carry semantic information are a burden.

In general, the present results help to elucidate theories on the relationship between gesture and speech. As outlined earlier, most theories take as their starting point that gesture and speech comprise a fundamentally integrated system. However, almost all of the empirical work on this topic has focused on the semantic relationship between the two modalities. Only recently have researchers extended the theory to other levels of language comprehension, such as the processing of basic speech sounds. Indeed, certain gestures, such as beats, appear to be integrated with speech in the global prosodic processing of words in sentences (Hubbard et al., 2008; Krahmer & Swerts, 2007). In this context, the present results provide an interesting twist on the integrated systems theory. Iconic gestures, at least in the domain of learning an L2, do not enhance more local processing of new phonetic forms within single words (nor do beat gestures, as shown by Hirata & Kelly, 2010). Thus, it is possible that gesture facilitates local processing of speech sounds only for familiar phonemes in one's native language (consistent with Baddeley et al., 1998). One possible explanation for why gesture is integrated only with familiar phonetic forms has to do with the very nature of gesture and speech: because gesture is holistic and global and speech is segmental and hierarchical (McNeill, 1992), gesture may have to wait for the speech processing system to do the nitty-gritty phonetic analysis of phonemes (which is over-burdened for novel phonemes) before it can do what it does best, which is to combine with other levels of language, such as semantic or global prosodic.

In sum, the results from the present study suggest that theories of gesture will need to be more sensitive to a number of additional issues not currently within their scope—that is, they will need to go beyond the semantic system and more thoroughly consider how different types of gestures are related to speech across every level of language. Incorporating this information into theories of gesture will help build new bridges to other relevant literatures on language comprehension and learning. For example, by focusing on the local processing of speech sounds, there will be exciting opportunities to link to the long-established literature on how speech is multimodally integrated with other prominent visual cues, such as the lips, mouth, and face (Green et al., 1991; Massaro & Cohen, 1983; Massaro & Jesse, 2007; Skipper et al., 2007). Indeed, Skipper has argued that different types of multimodal input (e.g., lip movements versus hand gesture) engage different neural networks that optimally distribute the work of processing language in a highly orchestrated fashion (Skipper, Goldin-Meadow, Nusbaum, & Small, 2009). Future research should explore the time-course of this neural processing and determine when different levels of language (e.g., phonetic vs. semantic) interact and influence one another. Exploring these sorts of new and stimulating questions will ultimately help to develop a theory of language comprehension and learning that goes well beyond the linguistic stream and thoroughly incorporates the natural and ubiquitous sorts of multimodal conditions that make up everyday face-to-face interactions.

Manuscript received 25 August 2010
Revised manuscript received 11 April 2011
First published online month/year

REFERENCES

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*, 178–190.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Belknap Press of Harvard University Press.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics*, *50*(6), 524–536.

Harada, T. (2006). The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. *Studies in Second Language Acquisition*, *28*(4), 601–632.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522.

Hayes, R. L. (2001). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers: Language in Cognitive Science*, *12*, 17–27.

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics*, *4*(1), 1–29.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*, 65–94.

Hazen, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*, 360–378.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, *116*(4), 2384–2394.

Hirata, Y., & Kelly, S. D. (2010). The effects of lips and hands on auditory learning of second language speech sounds. *Journal of Speech, Language and Hearing Research*, *53*, 298–310.

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

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2008). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, *30*, 1028–1037.

Kelly, S. D., Manning, S., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 1–20.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Neural correlates of learning Japanese words with and without iconic gestures. *Language and Cognitive Processes*, *24*, 313–324.

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

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (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*, 16–32.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*, 396–414.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(5), 753–771.

Massaro, D. W., & Jesse, A. (2007). Audiovisual speech perception and word Recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 19–35). Oxford: Oxford University Press.

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

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

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign language vocabulary learning. *Journal of Memory and Language*, *30*, 331–347.

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effect of phonological similarity and item length. *Quarterly Journal of Experimental Psychology*, *44*, 47–67.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, *19*, 661–667.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*, 2387–2399.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*, 661–699.

APPENDIX 1. EASY AND HARD WORD PAIRS

EASY PAIRS

- tate—stand (Set 1: gesture; Set 2: no gesture)
- butta—hit (Set 1: gesture; Set 2: no gesture)
- ute—shoot (Set 1: gesture; Set 2: no gesture)
- notta—rode (Set 1: gesture; Set 2: no gesture)
- oto—sound (Set 1: no gesture; Set 2: gesture)
- nikki—diary (Set 1: no gesture; Set 2: gesture)
- michi—street (Set 1: no gesture; Set 2: gesture)
- kokki—flag (Set 1: no gesture; Set 2: gesture)

HARD PAIRS

- haite—sweep (Set 1: no gesture; Set 2: gesture)
- haitte—enter (Set 1: no gesture; Set 2: gesture)
- ite—stay (Set 1: no gesture; Set 2: gesture)
- itte—go (Set 1: no gesture; Set 2: gesture)
- bachi—drum stick (Set 1: gesture; Set 2: no gesture)
- bacchi—badge (Set 1: gesture; Set 2: no gesture)
- saka—slope (Set 1: gesture; Set 2: no gesture)
- sakka—writer (Set 1: gesture; Set 2: no gesture)