Published in final edited form as:

Ecol Psychol. 2010 October 1; 22(4): 286–303. doi:10.1080/10407413.2010.517115.

Embodied, Embedded Language Use

Carol A. Fowler

Department of Psychology, University of Connecticut, Haskins Laboratories

Abstract

Language use has a public face that is as important to study as the private faces under intensive psycholinguistic study. In the domain of phonology, public use of speech must meet an interpersonal “parity” constraint if it is to serve to communicate. That is, spoken language forms must reliably be identified by listeners. To that end, language forms are embodied, at the lowest level of description, as phonetic gestures of the vocal tract that lawfully structure informational media such as air and light. Over time, under the parity constraint, sound inventories emerge over communicative exchanges that have the property of sufficient identifiability.

Communicative activities involve more than vocal tract actions. Talkers gesture and use facial expressions and eye gaze to communicate. Listeners embody their language understandings, exhibiting dispositions to behave in ways related to language understanding. Moreover, linguistic interchanges are embedded in the larger context of language use. Talkers recruit the environment in their communicative activities, for example, in using deictic points. Moreover, in using language as a “coordination device,” interlocutors mutually entrain.

In Noam Chomsky’s view, maintained consistently over several decades, there can be no science of public language use. In 1986, he wrote, “The concept of E (external)-language, *however construed*, appears to have no significance” (Chomsky, 1986, p. 31, italics added). More recently, he said,

It has been, and still is, conventional to regard language as a system whose function is communication. To the extent that that characterization has any meaning, which is not very great, it appears to be incorrect actually in fundamental ways. The overwhelming use of language is internal, for thought. (Presentation at Yale University, November 15, 2006)

This opinion would not be important if Chomsky had had little impact on how research among linguists and psychologists is pursued. However, he has had substantial impact. In particular, the vast majority of research studies conducted in the field of psycholinguistics focuses on the knowledge systems and mental processing of language users, not on the fundamentally social nature of public language use. Ironically, the influence that Chomsky has had on ideas and on research by others has been a result of his own prolific public language use in books, papers, and talks.

It is important and valuable to investigate knowledge systems and processing that supports language use (although one may question whether “mental” processing is necessarily internal, e.g., Ryle, 1949). However, there is also much to be learned about between-person language use and the behavioral, social, and physical settings in which it does work (e.g., Thibault, 2004). In what follows, I outline some of the domains in which the study of public language use has much to offer.

WHY STUDYING PUBLIC LANGUAGE IS IMPORTANT: PUBLIC PHONOLOGY

Vowel Inventories: (Starting Small)

Some properties of languages are very common, and others are rare. I focus on phonological language forms in this discussion. The “point vowels” /i/, /a/, and /u/ are very common across languages. In his characterization of properties of 317 languages, Maddieson (1984) mentions that variants of these three vowels are present in, respectively, 91.5%, 88.0%, and 83.9% of languages. He mentions as well that 93.5% of back vowels are lip rounded whereas about the same percentage of front vowels are unrounded. Why are these vowels and kinds of vowels in particular so common?

As for /i/, /a/, and /u/, they are far apart in acoustic space. Accordingly, they are popular most likely because they are easy to distinguish. A perceptual basis for the correlation between backing and rounding can be identified as well. Back vowels are produced with a tongue constriction toward the back of the oral cavity. The constriction partitions the oral cavity into two cavities, one in front of the constriction (toward the lips) and one behind it. When the constriction is toward the back of the oral cavity, the cavity in front of the constriction is relatively long, producing a low second formant (F2). Protruding the lips further lengthens that cavity, lowering F2 even more. Front vowels, produced with a tongue constriction toward the front of the oral cavity, are associated with high F2s. These remain high when the lips are spread or are unrounded, keeping the front cavity short.

Although it is very likely that perceptual distinctiveness affects the popularity of vowels in language communities (e.g., Diehl & Kluender, 1989), the characteristics of the auditory system do not provide a sufficient explanation for variations in the popularity of members of sound inventories. This is because auditory systems, that is, ears, sense organs, and neural networks, are incapable of selecting among candidate language forms.

Rather, auditory systems have to be embodied in agents who are embedded in a social environment in which the agents talk back and forth with the aim in mind of something like communication. When communication succeeds, which it only has a chance to do if language forms (most especially, word forms) are accurately identified, those language forms are maintained. When it fails, perhaps because some language forms are insufficiently distinguishable, the language forms may not be maintained. In short, language forms emerge and persist or do not persist in the course of public, social language use. Simulations by de Boer (2000; see also Oudeyer, 2005) show how this may work.

In de Boer’s (2000) simulations, agents interact socially by playing an imitation game. In the game, one agent, the initiator, produces a token vowel from its repertoire of vowel categories at random, and another agent, the imitator, attempts an imitation by using a token vowel from its inventory of categories. The initiator accepts the imitation if it is close enough to the vowel it emitted and signals success to the imitator. The imitator adds the perceived token to the vowel category from which it had chosen a token as an imitation. If the initiator perceives a member of a vowel category different from the category it had initially produced, then (under some conditions), the imitator adds a new vowel to its inventory of vowels under the assumption that the initiator is making a distinction that the imitator is not.

Depending on the amount of noise in the environment, the tolerance for differences between original productions and what count as successful imitations varies. Under very noisy conditions, tolerance is greater than under less noisy conditions. In noisy conditions, then, prototypes come to occupy large regions of acoustic space, and small vowel inventories

develop with prototypes far apart in the space. In one set of simulations with considerable noise, a three-vowel inventory emerged with formants like those of /i/, /a/, and /u/.

The Parity Constraint on Language Forms

In de Boer's (2000) simulations, vowel systems emerge with agents communicating under a "parity constraint," which is essential to any system involved in interpersonal communicative exchanges (Liberman & Whalen, 2000). One component of this constraint is a requirement that, with sufficient reliability, language forms (most especially word forms) produced by a talker and perceived by a listener must be the "same" in the sense of being identifiable as tokens of the same type. Otherwise communication must fail, and consequently speakers' communicative intents will not have a chance to be realized. If, for example, a speaker across the room from an open window says, "I'm cold," hoping that someone nearer to the window will close it, that hope only has a chance of being fulfilled if the listener hears that the speaker is cold. The likelihood that that will occur is increased if the consonants and vowels that compose word forms are sufficiently intelligible and distinctive. (Of course, more is required for a communicative intent to be realized than successful perception of language forms. I am referring here just to a "bottom-line" requirement.)

Humans' ability to use spoken language is an evolutionary adaptation of their species. A component of that adaptation is the ability to use speech to make linguistic communications available to others. We know that because, except in cases of profound hearing loss, languages are universally spoken. Manual gestural languages such as American Sign Language work very well, but they are not used among hearing members of any cultural communities as the primary language.

Given that speaking is part of the human adaptation to language use, one might expect languages to be adapted to the achievement of parity. That is, language forms should have a character that optimizes their successful recognition by listeners. Findings that distinctive phonological segments are popular constitute one indication that they do. Yet contemporary characterizations of speech production and perception suggest a different sense in which they do not have a character that optimizes their successful recognition by listeners.

What would it mean for languages to be adapted to the achievement of parity? Because language forms are the means by which language users can make their communications public, one adaptation to achievement of parity should be that language forms are themselves public things; they are available to be perceived rather than to be inferred. Another should be that the language forms are preserved throughout a communicative exchange; they should not, in the course of an event in which a speaker speaks and a listener perceives the speech, be lost and have to be recovered. Yet, by most accounts, neither of these adaptations to achievement of parity is part of our evolutionary adaptation for the spoken language.

Consonants and vowels are believed not to be public things (e.g., Pierrehumbert, 1990; Repp, 1981). They are held to be abstract categories in the minds of language users. Moreover, because talkers coarticulate (temporally overlap) the actions of the vocal tract that are meant to implement consonants and vowels when they speak, vocal tract actions are not supposed to be transparent to properties of those segments (e.g., Diehl, Lotto, & Holt, 2004; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Ohala, 1981). Further, because vocal tract actions for speech cause the acoustic signals that serve as the primary information source regarding the language forms being conveyed, on this account, the signals cannot be transparent to the segments either. Accordingly, speech perception must be a reconstructive process of analyzing the acoustic signal to identify its informative acoustic

“cues” and inferring language forms from the cues (e.g., Massaro, 1987; Sawusch & Gagnon, 1995). In this account, which I call a “mind-to-mind transmission” account, language forms are in the mind of the talker, but get distorted or lost in speech production, giving listeners a poor signal for perception. If, nonetheless, listeners manage to recover the language forms, the forms end up in the listener’s head as well. There is mind-to-mind parity but limited support for it out in the world. If that is accurate, then languages do not appear to be adapted to the achievement of parity in this respect.

But is it accurate? What prevents the idea that language forms are public actions? That is what several of my colleagues and I at Haskins Laboratories began to ask many years ago. My colleagues Browman and Goldstein (e.g., 1986, 1992) proposed that phonological entities are actions of the vocal tract that they called “phonetic gestures.” These actions create and release constrictions in the vocal tract in different locations and to different degrees. For example, producing /p/ requires two gestures, a complete constriction gesture at the lips and a devoicing gesture (a wide gestural constriction) of the vocal folds in the larynx. Properly phased, the gestures generate acoustic signals specifying to listeners the first consonant in the word *pad*.

Browman and Goldstein’s “articulatory phonology” has been described many times; I refer the reader to that literature (especially Browman & Goldstein, 1992, and commentaries in the same issue of *Phonetica*). However, for present readers, I provide just two clarifications. First, by proposing that phonological entities are linguistically significant actions of the vocal tract and not abstract categories in the mind, Browman and Goldstein are not proposing that knowledge systems supporting language use are unimportant. Rather, phonological entities and what language users know about the entities should not be confused.

Second, coarticulation is widely seen as a process that leads to distortion of properties of phonological segments (e.g., Ohala, 1981). If that is accurate, then there is not much to be gained by the idea that (intended) actions of the vocal tract are themselves linguistic things. If they are distorted in implementation, then a mind-to-mind transmission theory is still required. In my view, however, coarticulation is generally misunderstood. It is temporal overlap of gestures, not distortion of them. Browman and Goldstein’s (1992) phonetic gestures are coarse-grained actions. What matters for /p/’s lip gesture is that the lips close. It does not matter whether the jaw does most of the work or the lips do most of the work. It only matters that the lips close. Accordingly, when a tongue body gesture for the low vowel /a/ overlaps with a lip constriction gesture for /p/, it does not matter that the vowel pulls the jaw down. Due to a transiently established synergistic relation between jaw and lips (e.g., Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984), the lips take over some of the work done by jaw raising in other contexts, and the lips close.

Coarticulation resistance (e.g., Fowler, 2005; Recasens, 1985) further challenges the idea that coarticulation distorts gestural goals. Consonants such as /b/ and /p/ that do not use the primary articulator for vowels (the tongue) “permit” extensive coarticulatory encroachment into their domain. In contrast, a consonant such as /z/ that does use the tongue and has delicate requirements to achieve a constriction degree appropriate for generation of frication noise permits considerably less. Consonants only permit coarticulatory encroachments for which their associated synergies can compensate.

In short, coarticulation does not prevent the essential properties of phonetic gestures from being achieved in the public domain. It is important to note that this means that (undistorted) linguistic actions are the very actions that cause the acoustic speech signal. There is the chance, then, that acoustic information for them is better than acoustic information has been

thought to be for phonological entities and that the second proposed adaptation to parity (that language forms are preserved throughout a communicative exchange) is a true adaptation as well, obviating a mind-to-mind transmission account of communication of language forms.

Preservation of Language Forms in Communicative Exchanges: Perception of Gestures

There is another issue. By conventional accounts, listeners extract acoustic information about the mental phonological categories that compose the language forms in the talker's utterance. In contrast, by the present account, they extract information about phonetic gestures, the very actions that cause informative structure in the signal. Why is it not simpler to suppose that acoustic cues map onto mental phonological categories? One answer is that, if objects are gestural, the relation between acoustic structure and corresponding perceptual object is considerably more direct than it can be if language forms are mental categories.

However, the claim that listeners perceive phonetic gestures is highly controversial. After all, acoustic signals, not gestures, stimulate the listener's perceptual system. In my view, the claim that gestures are perceived should not be controversial for two reasons. First, it makes speech perception very like perceptual achievements in the same and other perceptual modalities,¹ and second, there is plenty of evidence for it.

The universal character of perception—In Gibson's perceptual theory (e.g., Gibson, 1966, 1979/1986), functionally speaking, perception works in the same way across the perceptual modalities. Our perceptual systems are the only means we have to know our environment, and, for our very survival, we need to know those components of our environment that constitute our ecological niche.

For vision, light reflects from objects, events, other organisms, and more and is lawfully and causally structured by them. Distinctive properties of these components of the niche structure the light in distinctive ways so that the light provides information for them. We intercept light structure over time, but, rather than perceiving it, we use the structure for what it is, namely, information for components of the niche. We see them, not the reflected light. And that is what we need to do.

Likewise, objects that we explore haptically causally deform the skin (among other things). Different objects (a pen, a blanket, a tree trunk) deform the skin differently. The pattern of deformations stimulates receptors in the skin. We use information in those patterns to perceive properties of the explored object, not to perceive the skin deformations themselves.

Audition should work in the same way and for the same reason: we need to know about the environment, not temporal patterns of structure in the air. When a door slams, we hear a door slamming, an event in the ecological niche, not an acoustic event (e.g., Carello, Wagman, & Turvey, 2003). As for speech, linguistic actions of the vocal tract, phonetic gestures, causally structure the air. In an ecological theory of speech perception (e.g., Best, 1995; Fowler, 1986, 1996), different actions structure the air differently so that patterning in the air over time provides information for the different gestures. We should expect perceivers to behave as they do as visual, haptic, and general auditory perceivers and use patterning in the relevant informational medium (here, primarily the acoustic signal) as information for its source in the world, phonetic gestures.

¹For Liberman and colleagues (e.g., 1967), perception of speech gestures reflected the special nature of speech perception. In their view, in auditory perception generally, the proximal acoustic signal, not its distal mechanical cause, is perceived. The view I present here is quite different: perceivers perceive the world of distal objects and events generally. Speech is not special in that respect.

If listeners do perceive phonetic gestures, then language forms are preserved throughout a communicative exchange and, in this sense, as in the idea that language forms are public things, languages are adapted to the achievement of parity.

Evidence for perception of (public) phonetic gestures—There is no crucial experiment that can be done to confirm the claim that gestures are perceived. There is only the weight of converging evidence. This is not the place to provide a complete review of the evidence. I provide just three examples. Every example I provide, indeed, every other piece of converging evidence, can be interpreted in some other way that does not invoke perception of gestures. However, alternative interpretations of the different pieces of converging evidence are different from each other, whereas there is just one, unifying, interpretation from the perspective presented here: Gestures are perceived.

1. “When articulation and sound wave go their separate ways, which way does the perception go?” (Liberman, 1957, p. 121). Liberman and colleagues obtained two complementary findings that they interpreted as evidence for perception of articulation. In the first (Liberman, Delattre, & Cooper, 1952), an invariant acoustic cue (a stop burst centered at 1440 Hz) presented before formants for the vowels /i/, /a/, and /u/ were heard as /p/ before /i/ and /u/ but as /k/ before /a/. In order for the same stop burst to be produced by a consonantal gesture coarticulating with a following vowel, a constriction at the lips is required when the coarticulating vowels are /i/ or /u/, but a constriction at the velum is required when the coarticulating vowel is /a/. Perception tracked the different articulations, not the invariant acoustic cue.

In the second finding (Liberman, Delattre, Cooper, & Gerstman, 1954) very different second formant transitions were found to specify the same consonant /d/ to listeners in the context of following /i/ and /u/. In this case, the same consonantal gesture, a constriction gesture of the tongue tip against the palate behind the teeth, gives rise, because of coarticulation with the following vowel, to very different transitions when it is released into the vocal tract configurations for /i/ and /u/. Here, perception tracked the invariant phonetic gesture, not the varying formant transitions.

2. Crossmodal speech perception. A second set of examples suggesting perception of gestures involves crossmodal speech perception. First, speech in noise is more intelligible if the listener can see the speaker (e.g., Sumby & Pollack, 1954). Why? Certainly, participants viewing a face see visible speech gestures. If listeners are striving to perceive phonetic gestures from a noisy acoustic speech signal, the additional information about gestures provided by the articulating face should be helpful, and it is.

A converging finding is provided by the well-known McGurk effect (e.g., MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). In this paradigm, one syllable, say /ma/, is provided acoustically, dubbed onto a face mouthing an appropriately selected other syllable, say /da/. In the example, listener-viewers usually report hearing /na/, a report that integrates the visible place of articulation with other acoustically specified phonetic properties (nasality, voicing). The basis for audiovisual integration must be some “common currency” across the perceptual modalities. A likely candidate is the information for phonetic gestures provided to both modalities.²

Finally, infants recognize this common currency. Presented with two videos of a speaker mouthing different vowels, 5-month-old infants will look at the monitor displaying speech

²The visual modality is not the only modality besides audition that provides listeners with information for gestures (e.g., Fowler & Dekle, 1991; Gick & Derrick, 2009).

gestures that are consistent with the vowel presented acoustically over a loudspeaker between the monitors (Kuhl & Meltzoff, 1984). Five-month-old infants also show a McGurk effect (Rosenblum, Schmuckler, & Johnson, 1997).

3. Compensation for coarticulation. A third source of evidence for perception of phonetic gestures is the phenomenon of compensation for coarticulation. Listeners intercept acoustic speech signals that are produced by coarticulating talkers. Therefore, the acoustic structure at nearly all points in time has been caused by the production of two or more phonetic gestures. Research suggests that listeners closely track talkers' gestures to "compensate" for coarticulation. A classic example is provided by Mann (1980). She synthesized an acoustic continuum of syllables ranging from /da/ to /ga/ by varying the third formant (F3) transition at the beginning of each syllable. Syllables were presented in the context of one of two precursor syllables /a/ or /ar/, and listeners' task was to identify the continuum members as "da" or "ga." She found that more of the ambiguous syllables in the middle of the continuum were identified as /ga/ if the precursor syllable was /a/ than if it was /ar/.

One account of this finding is that listeners literally compensated for the coarticulatory overlap of /l/ and /r/ with the consonant of the following syllable. The tongue tip constriction of /l/ overlapping with the velar constriction of /g/ would move the point of constriction forward during /g/. The (back) pharyngeal constriction of /r/ overlapping with the tongue tip constriction of /d/ might pull the tongue tip back during /d/. When listeners heard syllables from the ambiguous region of the continuum that specified a point of constriction too far forward for /g/ and too far back for /d/, in the context of /l/, they identified the syllable as "g"; in the context of /r/, they identified it as "d." By this account, listeners closely track talkers' speech gestures.

There is another account, however. /r/ has a very low ending F3, lower than the lowest F3 onset of /da-/ga/ continuum members. /g/ has a high ending F3, higher than the highest F3 onset of /da-/ga/ continuum members. Instead of true compensation for coarticulation, the outcome could be a spectral contrast effect (e.g., Lotto & Kluender, 1998) whereby the low F3 offset of /ar/ made following F3 onsets sound higher than they were and so more like /d/. The high F3 offset of /a/ made F3 onsets sound lower and more like /g/.

As different as these accounts are, they have proven difficult to dissociate. In most contexts, they make the same prediction. However, Viswanathan, Magnuson and Fowler (2010) found a dissociation in two Tamil liquid consonants. Whereas one had a front and one a back place of articulation like English /l/ and /r/, both had low F3s like English /r/. In a test involving both the English and the Tamil consonants, "g" responses in the context of the Tamil consonants patterned by place of articulation (i.e., more "g" responses in the context of the front than the back liquid) not by F3. Compensation for coarticulation, then, provides further evidence that listeners perceive gestures.

WHY STUDYING PUBLIC LANGUAGE IS IMPORTANT: MOVING UP AND OUT

Understanding how the segment inventories of languages come to have the properties that they do requires that we see the physical systems that produce and perceive language as embodied in agents who are embedded in a physical and social context in which the agents talk back and forth. Optimizing successful communicative exchanges requires that language primitives, consonants and vowels, the building blocks of words, themselves be embodied as vocal tract actions that are perceived by listeners because acoustic signals specify them.

However, embodiment and embedding have much wider scope than that in public language use. As for embodiment, evidence shows that language understanding itself does not reside only in the brain; it is at least a whole body achievement; it is embodied. As for embedding, language use is embedded in a human social niche in which other activities accompany it as integral parts of the activity of communicating. In that social niche, interlocutors become participants in systems larger than themselves (e.g., Fowler, Richardson, Marsh, & Shockley, 2008). They exhibit their participation in a variety of ways that reveal their mutual coordination.

Embodied Communications Involving Language

Comprehension—There is substantial evidence that understanding language manifests itself outside of the brain. One radical interpretation of this evidence is that the brain is not an autonomous seat of language understanding or cognizing. The findings to date do not wholly warrant that radical interpretation. Possibly, the bodily reflections of language understanding are merely epiphenomenal consequences of brain activity reflecting language understanding. That is, language understanding is known to be accompanied by activity in sensorimotor areas of the brain (e.g., Pulvermüller, 2005). Possibly, that activity is integral to language understanding; the bodily dispositions to act in particular understanding-related ways may be epiphenomenal. But I find it intriguing to speculate that there is more to it. Possibly brains are not “nearly decomposable systems” in the sense of Simon (1969). Perhaps language understanding is a whole-body achievement. Following I review some sample findings.

In research by Bargh, Chen, and Burrows (1996), participants in two groups were given sets of five words and were asked to make a sentence from four of the five words in each set. Critical words presented to the experimental group were meant to suggest older people, words such as *gray*, *Florida*, and *bingo*. Words presented to the control group did not suggest older people. The dependent measure in the experiment was obtained after sentences had been generated by participants. It was the time it took participants to leave the laboratory and cross a strip of silver carpet tape 9.75 m away on their way to the elevator to go home. Participants in the experimental group walked more slowly than those in the control group. Participants embodied their stimulus-word-provoked thoughts about older people.

Glenberg and Kaschak (2002) presented sentences to listeners who judged whether each sentence made sense or not. Participants placed a finger on a home button and, after hearing a sentence, moved to a response button. The response button to indicate that a sentence made sense was either beyond the home button, requiring a movement of the participant's hand away from the body, or it was closer to the participant than the home button, requiring a movement of the hand toward the participant's body. Critical sentences were sensible and implied movement either toward (“Andy delivered the pizza to you”) or away (“You delivered the pizza to Andy”) from the participant. Participants' responses were faster when they were consistent in direction with the movement implied by the sentence. (So, for the sentence “Andy delivered the pizza to you,” responses were faster if the required movement was toward the participant than if it was away.)

An interesting and somewhat surprising characteristic of these outcomes is that they suggest that embodiment of language understanding is simulation-like. People who are led to think of older people walk slowly like older people stereotypically do. People who understand sentences implying movement in some direction move more readily in that direction than in an opposite direction.

The finding is surprising (to me) because a more adaptive kind of embodiment might be one in which an utterance prepares listeners to behave in ways that are appropriately responsive to the information conveyed by the utterance. If Andy is going to give me a pizza, and I am going to take it, should I not reach out *toward* the pizza rather than, as in Glenberg and Kaschak's (2002) study, being disposed to move as the pizza does? Similarly, if a passenger in a car traveling down a country road were to say, "Watch out! A moose is coming at us up ahead," a disposition to move in the direction the moose is moving (i.e., a disposition to throw the vehicle into reverse) does not prepare the driver for a realistic response. A disposition to veer left or right might be more useful. Are embodiments of language understanding always simulations?

They are not; however, to my knowledge as yet, no investigations have been conducted to ask whether embodiments in more meaningful (mooselike) settings are adaptive rather than simulative. There is one experiment, however, suggesting that embodiments are not always simulative. Olmstead, Viswanathan, Aicher, and Fowler (2009) asked readers to make verbal responses ("yes") when sentences they read were sensible. They performed the sentence task at the same time that they swung a pair of handheld pendulums in an in-phase (0°) or anti-phase (180°) relation. Sentences either alluded to actions of the hand ("Todd is chopping the garlic") or referred to inanimate states of affairs ("The barn is housing the goats"). Findings were that understanding hand sentences led to shifts in the attractors of the hand-pendulum system (away from the pendulum's phase relations (near 0° or 180°) in the absence of a secondary task) with no increase in noise in the system over the level of noise present when no verbal task accompanied pendulum swinging. (An increase of noise might reflect the effect of the cognitive load of the verbal task on the motor task.) Understanding inanimate sentences did not shift the attractors or increase noise. The first finding suggests that sentences referring to the body (or perhaps specifically to the hands) had an impact on the organization of the dynamical system performing the hand-pendulum coordination task. There is nothing simulative about the effects, however. It is more that the performance of two tasks that are linked somehow in respect to the fact that the human body is relevant to both tasks leads to an organization of the pendulum swinger/language understander that accommodates both tasks (cf. Pelliccia & Turvey, 2001).

Embodiment in production: More than talking communicates—When we talk to other people, we gesture manually, our faces are expressive, and our heads and eyes move in ways that can be informative to interlocutors. It is controversial whether manual gestures communicate to anyone other than the speaker himself or herself (e.g., Krauss, Morrel-Samuels, & Colasante, 1991; McNeill, 1992). However, McNeill (e.g., 1985) suggests that manual gestures that accompany speech provide communicative information along a separate channel than speech that can augment the talker's communicative content. Sometimes the gestures provide information that the words do not convey.

As for facial expressions, they can indicate the talker's stance with respect to what she or he is saying. For example, a talker might say, "We are supposed to have a snowstorm tomorrow," expressing pleasure on the face (because school may be canceled) or concern (because a flight by air to a vacation resort may be canceled).

Remarkably, even the head movements that accompany speech improve its intelligibility (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). Finally, shifts in eye gaze can, for example, regulate conversational turn taking (e.g., Lerner, 2003).

Language Use as Embedded in the Ecological Niche of Language Users

I have suggested three senses (surely among others) in which language use is embodied. It is embodied when talkers generate utterances composed of phonetic gestures. It is embodied in

listener/readers when their actions show simulative or reorganizational reflections of utterance content. It is embodied when manual gestures, facial expressions, and head movements accompany language use.

Beyond embodiment, language use is also embedded in a social and physical context in which participants in a communicative exchange become participants in a larger cooperative, coordinated communicative system.

Using the environment to provide communicative meanings—When talk is about the here and now, talkers may exploit their embeddedness in a shared environment with their interlocutors to get their meaning across. For example, if two individuals are cooking dinner together and one asks the other, “Where is the spatula?” the other might answer, “It is over there,” pointing to the spatula’s location. In order for the first individual to know where the spatula is, she or he has to follow the point to the spatula’s location.

More interestingly, Hanks (1992) provides an example in which pointing is done in a different way. He describes a case in which two Maya men are working together outside and out of sight of one another. One man, searching for a tool, calls out to the other (in Mayan),

Is it over there?

The other replies,

Yes, it’s here (where I am).

In this exchange, the man in search of his tool can only know where the tool is if he follows the speaker’s voice, that is, if he makes use of his sound localization ability to locate the speaker and therefore the tool.

More embedding: Signs of mutual participation in events involving talk—The relevant embedding for talkers to provide deictic points and for listeners to understand them is the shared physical (and sometimes social—A: “Who has the spatula?” B: pointing “She does.”) environment. Embeddedness in the social environment in which talking occurs has been demonstrated in ways other than by pointing. In particular, talkers mutually entrain in their talk and in their actions.

When individuals talk cooperatively, they entrain in their dialect, their speaking rate, their vocal intensity, and their rate and duration of pausing (see, e.g., Giles, Coupland, & Coupland, 1991, for a review). The disposition to entrain is strong enough that it can even have negative social consequences as, for example, when mimicking another speaker’s regional dialect may be perceived as mockery. We (Sancier & Fowler, 1997) followed the speech of a bilingual speaker of Portuguese (her native language) and English (learned at age 15). In an environment in which only English was spoken, her accented English migrated (on the dimension of voiceless stop voice-onset times) toward that of her interlocutors. Unfortunately, although she was not speaking Portuguese in the English-language environment, her Portuguese also migrated English-ward to an extent that, on returning to Brazil, her father accused her of speaking “explosively” (with excessively long, English-like voice-onset times). Most of the time, however, vocal entrainments are not so noticeable as to offend; rather, they may foster feelings of rapport.

These vocal entrainments may reflect an important consequence of linguistic exchanges taking place in social contexts: when groups of individuals participate in a social setting involving talk, a larger social system encompassing them emerges (e.g., Fowler, Richardson, Marsh, & Shockley, 2008). Vocal entrainments provide one index of the emergence of these larger organizations.

There are other indices as well. In these kinds of settings, participants may exhibit posture mirroring (LaFrance, 1985) so that if, for example, a speaker crosses his or her arms, so do some listeners. In addition, talkers and listeners may show interactional movement synchrony (e.g., Bernieri, 1988; Bernieri, Reznick, & Rosenthal, 1988; Condon, 1976) and, remarkably, they may exhibit entrainment of their postural sway (Shockley, Santana, & Fowler, 2003).

The idea that, in social events involving language use, participants self-organize into larger systems to serve joint aims, suggests that in this sense, these events are not special. Marsh, Richardson, Baron, and Schmidt (2006) propose a similar view of participants in social events whether or not talk is involved.

CONCLUDING REMARKS

Language use is embodied in part in that fundamental language forms are public actions of the vocal tract that enable public communication by means of language forms to be as reliable as it can be. More than this, language users embody their communicative intentions in their manual gesturing, their facial expressions, their eye gaze, and more. As for listeners, they embody their language understanding in at least two senses: in laboratory contexts, they exhibit dispositions to simulate utterance meanings and they show evidence of self-organization that accommodates activities that are ongoing as utterances are being comprehended. My guess is that laboratory investigations may also show, though they have not yet, that, in more meaningful contexts than those used to date in the study of embodied language use, embodiment of language understanding can also reflect a disposition toward actions that are adaptively related to utterance content (as when an individual urged to “Duck!” is disposed to do so).

Language use is also embedded in the ecosocial contexts in which it occurs. Speakers recruit the environment to their communicative purposes when they use deictic points, manual or vocal. They also recruit interlocutors who exhibit their incorporation into a larger system organized to serve conjoint aims in their variety of entrainments one to the other.

Because of its emphasis in the past several decades on the human biological capacity for language and on “language use” as the covert mental processing that recruits linguistic competence for purposes of language production and comprehension, the study of language embodiment and its embeddedness in the physical and social environment has lagged. It is timely now to incorporate the study of public language use into the developing ecological approach to social perception and action (e.g., Marsh et al., 2006).

Acknowledgments

Preparation of the article was supported by Grants HD-01994 and DC-02717 to Haskins Laboratories. I thank Paul Thibault and an anonymous reviewer for constructive and stimulating comments on an earlier draft.

References

- Bargh JA, Chen M, Burrows L. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 1996;71:230–244. [PubMed: 8765481]
- Bernieri FJ. Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior* 1988;12:120–138.
- Bernieri FJ, Reznick JS, Rosenthal R. Synchrony, pseudosynchrony and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology* 1988;54:243–253.

- Best, CT. A direct realist view of cross-language speech perception. In: Strange, W., editor. *Speech perception and linguistic experience*. Baltimore, MD: York Press; 1995. p. 171-204.
- Browman C, Goldstein L. Towards an articulatory phonology. *Phonology Yearbook* 1986;3:219–252.
- Browman C, Goldstein L. Articulatory phonology: An overview. *Phonetica* 1992;49:155–180. [PubMed: 1488456]
- Carello, C.; Wagman, JB.; Turvey, MT. Acoustic specification of object properties. In: Anderson, J.; Anderson, B., editors. *Moving image theory: Ecological considerations*. Carbondale: Southern Illinois Press; 2003. p. 79-104.
- Chomsky, N. *Knowledge of language: Its nature, origin, and use*. New York, NY: Praeger; 1986.
- Condon W. An analysis of behavioral organization. *Sign Language Studies* 1976;13:285–318.
- de Boer, B. Emergence of sound systems through self-organization. In: Knight, C.; Studdert-Kennedy, M.; Hurford, J., editors. *The evolutionary emergence of language*. Chicago, IL: Chicago University Press; 2000. p. 177-198.
- Diehl R, Kluender K. On the objects of speech perception. *Ecological Psychology* 1989;1:121–144.
- Diehl RL, Lotto AJ, Holt LL. Speech perception. *Annual Review of Psychology* 2004;55:149–179.
- Fowler CA. An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* 1986;14:3–28.
- Fowler CA. We do hear sounds not tongues. *Journal of the Acoustical Society of America* 1996;99:1730–1741. [PubMed: 8819862]
- Fowler CA. Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics* 2005;33:199–213.
- Fowler CA, Dekle DJ. Listening with eye and hand: Crossmodal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 1991;17:816–828. [PubMed: 1834793]
- Fowler, CA.; Richardson, MJ.; Marsh, KL.; Shockley, KD. Language use, coordination and the emergence of cooperative action. In: Fuchs, A.; Jirsa, VK., editors. *Coordination: Neural, behavioral and social dynamics*. Berlin, Germany: Springer-Verlag; 2008. p. 261-279.
- Gibson, JJ. *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin; 1966.
- Gibson, JJ. *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum; 1986. (Original work published 1979)
- Gick B, Derrick D. Aero-tactile integration in speech perception. *Nature* 2009;462:502–504. [PubMed: 19940925]
- Giles, H.; Coupland, N.; Coupland, J. Accommodation theory: Communication, context, and consequence. In: Giles, H.; Coupland, J.; Coupland, N., editors. *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge, UK: Cambridge University Press; 1991. p. 1-68.
- Glenberg AM, Kaschak MP. Grounding language in action. *Psychonomic Bulletin & Review* 2002;9:558–565. [PubMed: 12412897]
- Hanks, W. The indexical ground of deictic reference. In: Duranti, A.; Goodwin, C., editors. *Rethinking context*. Cambridge, UK: Cambridge University Press; 1992. p. 43-76.
- Kelso JAS, Tuller BH, Vatikiotis-Bateson E, Fowler CA. Functionally-specific articulatory cooperation following jaw perturbation: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance* 1984;10:810–832.
- Krauss R, Morrel-Samuels P, Colasante C. Do conversational hand gestures communicate? *Journal of Personality and Social Psychology* 1991;61:743–754. [PubMed: 1753329]
- Kuhl P, Meltzoff A. The intermodal representation of speech in infants. *Infant Behavior and Development* 1984;7:361–381.
- LaFrance M. Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin* 1985;11:207–217.
- Lerner G. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 2003;32:177–201.
- Lieberman AM. Some results of research on speech perception. *Journal of the Acoustical Society of America* 1957;29:117–123.

- Liberman AM, Cooper F, Shankweiler D, Studdert-Kennedy M. Perception of the speech code. *Psychological Review* 1967;74:431–461. [PubMed: 4170865]
- Liberman AM, Delattre P, Cooper F. The role of selected stimulus variables in the perception of the unvoiced-stop consonants. *American Journal of Psychology* 1952;65:497–516. [PubMed: 12996688]
- Liberman AM, Delattre P, Cooper FS, Gerstman L. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied* 1954;68:1–13.
- Liberman AM, Whalen DH. On the relation of speech to language. *Trends in Cognitive Sciences* 2000;4:187–196. [PubMed: 10782105]
- Lotto AJ, Kluender KR. General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics* 1998;60:602–619. [PubMed: 9628993]
- MacDonald J, McGurk H. Visual influences on speech perception. *Perception & Psychophysics* 1978;24:253–257. [PubMed: 704285]
- Maddieson, I. *Patterns of sounds*. Cambridge, UK: Cambridge University Press; 1984.
- Mann VA. Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics* 1980;28:407–412. [PubMed: 7208250]
- Marsh KL, Richardson MJ, Baron RM, Schmidt RC. Contrasting approaches to perceiving and acting with others. *Ecological Psychology* 2006;18:1–38.
- Massaro, D. *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum; 1987.
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature* 1976;264:746–748. [PubMed: 1012311]
- McNeill D. So you think gestures are nonverbal? *Psychological Review* 1985;92:350–371.
- McNeill, D. *Hand and gesture: What gestures reveal about thought*. Chicago, IL: University of Chicago Press; 1992.
- Munhall KG, Jones JA, Callan DE, Kuratate T, Vatikiotis-Bateson E. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science* 2004;15:133–136. [PubMed: 14738521]
- Ohalá, J. The listener as a source of sound change. In: Masek, C.; Hendrick, R.; Miller, R.; Miller, M., editors. *Papers from the parasession on language and behavior*. Chicago, IL: Chicago Linguistics Society; 1981. p. 178–203.
- Olmstead AJ, Viswanathan N, Aicher KA, Fowler CA. Sentence comprehension affects the dynamics of bimanual coordination: Implications for embodied cognition. *Quarterly Journal of Experimental Psychology* 2009;62:2409–2417.
- Oudeyer PY. The self-organization of speech sounds. *Journal of Theoretical Biology* 2005;233:435–449. [PubMed: 15652151]
- Pellecchia GL, Turvey MT. Cognitive activity shifts attractors of bimanual rhythmic coordination. *Journal of Motor Behavior* 2001;33:9–15. [PubMed: 11265053]
- Pierrehumbert J. Phonological and phonetic representations. *Journal of Phonetics* 1990;18:375–394.
- Pulvermüller F. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 2005;6:576–582.
- Recasens D. Coarticulatory patterns and degrees of coarticulation resistance in Catalan CV sequences. *Language and Speech* 1985;28:97–114. [PubMed: 4087987]
- Repp B. On levels of description in speech research. *Journal of the Acoustical Society of America* 1981;69:1462–1464. [PubMed: 7240580]
- Rosenblum LD, Schmuckler M, Johnson JA. The McGurk effect in infants. *Perception & Psychophysics* 1997;59:347–357. [PubMed: 9136265]
- Ryle, G. *The concept of mind*. Chicago, IL: University of Chicago Press; 1949.
- Sancier M, Fowler CA. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 1997;25:421–436.

- Sawusch J, Gagnon D. Auditory coding, cues and coherence in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 1995;21:635–652. [PubMed: 7790838]
- Shockley K, Santana MV, Fowler CA. Mutual interpersonal postural constraints are involved in the supra-postural task of cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 2003;29:326–332. [PubMed: 12760618]
- Simon, H. *The sciences of the artificial*. Cambridge, MA: MIT Press; 1969.
- Sumby WH, Pollack I. Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America* 1954;26:212–215.
- Thibault, P. *Brain, mind, and the signifying body: An ecosocial semiotic theory*. London, UK: Continuum; 2004.
- Viswanathan N, Magnuson JS, Fowler CA. Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance* 2010;36:1005–1015. [PubMed: 20695714]