

John Benjamins Publishing Company



This is a contribution from *From Gesture in Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*.

Edited by Mandana Seyfeddinipur and Marianne Gullberg.

© 2014. John Benjamins Publishing Company

This electronic file may not be altered in any way.

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

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

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

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

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

Including facial gestures in gesture–speech ensembles

Janet Bavelas, Jennifer Gerwing and Sara Healing
Department of Psychology, University of Victoria

Conversational facial gestures fit Kendon's (2004) specifications of the functions of hand gestures. We illustrate how facial gestures in dialogue, like hand gestures, convey referential content as well as serving pragmatic, interpersonal and interactive functions. Hand and facial gestures often occur together, creating an integrated visual image in gesture–speech ensembles. A semantic features analysis demonstrates how speakers adjust their use of these visible versus audible expressive resources according to context. Speakers who were interacting face to face (compared to speakers who could not see their addressee) were significantly more likely to rely on their hand and facial gestures than on their words when describing key semantic features, and their gestures were more likely to convey information that was not in their words.

Part 1: “Ensembles of gestures and speech”

In this chapter, all Kendon quotations (including those used as section headings) are from his 2004 book.

We shall see that speakers create *ensembles* of gesture and speech, by means of which a semantic coherence between the two modalities is attained. This is not to say that speech and gesture express the same meanings. They are often different. Nevertheless, the meanings expressed by these two components *interact* in the utterance and, through a reciprocal process, a more complex unit of meaning is the result. (pp. 108–109)

We begin by extending Kendon's theory of gesture–speech ensembles to include more than hand gestures in these ensembles. An example from a video-recorded medical consultation between physician and patient (Healing, unpublished data) will illustrate the nature of these expanded ensembles. The physician had asked about the patient's symptoms since they last met.

Patient: “Usually first thing in the morning, I’m kind of, ‘Oh!’ Trying to get things going, ’n yesterday I just – ”

Physician: “So that’s good.”

Nothing the patient said seemed to answer the physician’s question. He did not say how he “usually [is] first thing in the morning” or what it meant for him to be “trying to get things going,” much less what he was like “yesterday.” Yet the physician’s reply (“So that’s good”) indicated that the patient’s answer was both informative and positive. As shown in Table 1, it was the patient’s animated and precisely timed gestures with hands, arms, shoulders, upper torso, face, head and eyes that conveyed all of this specific information. Thus, although gesture and speech expressed different meanings, they interacted in his utterances to create more complex units of meaning that were still semantically coherent.

Table 1. Patient telling his physician how his symptoms have been since his last visit

The patient began by depicting himself as he usually was when he woke up:

“Usually first thing in the morning, I’m kind of;” [Physician nods]
[hunching his shoulders forward, sitting stiffly, not moving– as if bracing himself; his face is frozen and tense]

Then he enacted more about his usual morning:

Moaning “Ohhhh.”
[leaning over to one side, grimacing]

He followed by showing how he usually had to try to begin moving:

“Trying to get things going.” [Physician says, “m-hm”]
[kneading his affected leg, as if trying to get it going]

Then he quickly contrasted this description with how he felt yesterday:

“’n yesterday, I just – ”
[sits up straight and begins to mime a vigorous marching motion with his arms and legs; head held high with a determined face]

The physician started to nod and to say “So that’s good,” and the patient simultaneously ceased these actions and smiled at the physician.

“Facial gestures”

While writing about the use of the kinesic medium in sign language, Kendon mentioned “facial gestures, such as eyebrow movements or positionings, movements of the mouth, head postures and sustainments and changes in gaze direction” (p. 310). His list includes most of what we consider facial gestures. More formally, *conversational facial gestures* are any configuration or movement of the face or of

the head (including the eyes) that is synchronized with speech in both timing and meaning. For example, the patient's grimace, leaning over, and the long "Ohhhh" all occurred together and all contributed to the picture of how he felt. Similarly, precisely with "n yesterday I just –," he held his head high with a determined look and marched with his arms and legs.

"Closing off further investigation"

Anyone who watches the faces of interlocutors in a dialogue will see a virtually constant succession of rapid and diverse facial gestures by both speaker and addressee. Yet conversational facial displays in dialogues are the subject of only three descriptive studies (Brunner 1979; Chovil 1989, 1991/1992; Ekman 1979) and two experiments (Chovil 1991; Bavelas, Gerwing & Healing in press). Given the abundance of these co-speech gestures in face-to-face dialogues, it is initially difficult to understand why there is so little investigation of them in the literature.

We propose that the answer is an unquestioned assumption that facial expression equals emotional expression, which has dominated interpretations of the face for centuries. Kendon (p. 31) illustrated this historical bias with sketches of fixed and stereotypic facial expressions of "contempt," published by Lebrun in 1734 as lessons on how to portray "the passions". These pictures are typical of the literature of the time, in which certain facial expressions were equated with specific emotions. Almost 300 years later, Ekman and Friesen's (1969) typology of non-verbal behavior included many functions of hand gestures but only one for facial expression: "the face is the primary site of affect displays" (p. 71). They narrowed affect displays even further, proposing that there is a fixed set of muscle configurations that correspond to innate, universally recognized emotions.

Although Birdwhistell (1970) almost simultaneously proposed that facial displays serve linguistic functions (and Ekman 1979, 1997) occasionally wrote about communicative facial expressions), conversational facial gestures have remained largely unrecognized, while the non-linguistic emotion approach has continued to motivate contemporary research. Even researchers who propose social ("audience") effects on facial expressions (e.g. Kraut & Johnston 1979) have limited themselves to emotional expressions such as smiling, fear or anger (see review in Chovil 1997). We affirm Kendon's observation that "the typology Ekman and Friesen presented... might almost be said to have had the effect of closing off further investigation" (p. 72) – even more so for facial than for hand gestures.

Although Ekman and Friesen's focus on emotional expressions may have hampered investigation of other possible functions of facial gestures, it is interesting to consider the results of their examination of nearly 6000 facial actions

of patients with affective disorders. These were depressed or bipolar individuals who were talking about their feelings, which could range from depression to mania. Yet *fewer than a third* of their facial actions were classifiable as emotional expressions (Fridlund, Ekman, & Oster 1987, pp. 160–161). Similarly, in the only systematic analysis of facial actions in dialogues in a non-clinical setting, Chovil (1989, 1991/1992; see also Bavelas & Chovil 1997) found that personal reactions of any kind were only about a quarter of the 720 meaningful facial gestures identified in the analysis; the remainder were semantically or syntactically related to the wide range of topics in the dialogues. Here we focus on this neglected majority, specifically on how conversational facial gestures share many of the characteristics and functions that Kendon outlined for hand gestures. The examples are from data gathered for Bavelas, Gerwing and Healing (in press).

A note on terminology

The literature is populated with a variety of terms for what the face does, each of which has underlying theoretical assumptions. The term *facial action* focuses on the formal musculature of the face instead of on the functions these actions might serve in interaction. *Facial* (or *emotional*) *expression* focuses on what the face reveals about an internal emotional state. Kraut and Johnston (1979) borrowed the ethological term *display* in order to distinguish between a social facial *display* and an emotional facial *expression*. Chovil (1989, 1991, 1991/1992) used “display” for the same reason. Although the above terms could convey important theoretical distinctions, they appear to be used interchangeably in the literature. Bavelas, Gerwing, and Healing (2014) have proposed *conversational facial gestures* as a term that emphasizes the close functional similarities to conversational hand gestures. The next sections document these similarities by mapping Kendon’s (2004: Chapter 9) outline of the features of hand gestures onto the features of facial gestures.

“The gestured component of an utterance”

“Referential content”

“The gestured component of an utterance can be a part of its *referential* content” (p. 158), either by pointing or representing. Facial gestures can also serve a pointing function: the simplest deictic is a quick sideways head motion toward something (“It’s over there”). More subtly, one person can simply shift his or her eyes to the side to indicate that the interlocutor should look at someone in that direction. The deictic function of gaze direction is apparent even in infants: by 12 months, human infants followed the direction of the experimenter’s eyes, whereas even adult great apes were less likely to do so (Tomasello, Hare, Lehmann & Call 2007).

Sherzer (1973) and Enfield (2001) described the more complex “lip-pointing” deictic, which has meanings that depend on the immediate conversational context.

The more common and varied way in which facial gestures convey referential content is by various techniques of *representation*, and it is here that the capacities of hand and facial gestures differ the most. Hand gestures often use the techniques of modeling (e.g. forming a shape) or depicting (e.g. sketching in the air), but the face is not well suited to either of these. Facial gestures excel at *enacting* any imaginable face, that is, demonstrating anything that any face can look like. In contrast to a small set of stereotypic affect displays, conversational facial gestures are virtually unlimited in number and kind. The rapidity and flexibility of many muscle groups enables the face to enact the way the speaker looked or might have looked in a past situation – or might look in a future or even a hypothetical situation.

Nor is the face limited to enacting oneself; it can just as easily represent someone else’s reaction in the past, present or future. Such facial gestures can represent a real person, a character in a story (human or not), or someone entirely hypothetical or generic. For example, in Figure 1, while retelling excerpts from the movie *Shrek 2*, the speaker described a scene in which Shrek had captured the cat who had attacked him. Frame 1 shows her own (non-representational) animated story-telling face; frame 2 shows her version of Shrek’s slightly fiendish triumph; and frame 3 depicts the suddenly concerned and apologizing cat. (The experiment described below focuses on this capacity of facial gestures to enact the face of a movie character.)



Figure 1. While describing a scene from the movie *Shrek 2*, the speaker made the hand gesture in frame 2 and the facial gestures in frames 2 and 3. The three frame shots cover a 6.75 second period. (The face of the addressee is inset at the upper right)

“Pragmatic” and “interactive or interpersonal functions”

Even further from the world of emotion are the pragmatic and interactive functions of facial gestures. These non-referential functions in dialogue (also called collateral communication or meta-communication) are about the dialogue itself rather than about its topic.

“*Modal functions*” are the first of Kendon’s three pragmatic functions of hand gestures. They alter the way in which “the utterance is to be interpreted” (pp. 158–159). An example of a facial gesture that serves a modal function is the *facial shrug*, which is analogous to a shoulder shrug. It typically involves a quick eyebrow flash and the retraction of a corner of the mouth; see Figure 2. Just as the shoulders can “shrug something off,” a facial shrug can convey that something does not matter (e.g. that enough has been said or that it has been said well enough). Both Ekman (1985) and Chovil (1989, 1991/1992) observed facial shrugs. Smiles can also serve a modal function; Coates (1991) found that smiles played a role in marking ironic humor.



Figure 2. The speaker was completing her description of a scene from *Shrek 2* when, instead of continuing on from frame 1, she made a *facial shrug* (in frame 2). Then she said “I guess” (in frame 3) and went on. The three frame shots cover a 1.3 second period

“*Performative functions*... indicate the kind of speech act or interactional move a person is engaging in” (p. 159). For example, speakers often raise their eyebrows to indicate a question, even if the syntax was not interrogative (Ekman 1979: 185). Brunner’s analysis (1979) showed how addressees’ smiles can function as back-channels, with the same timing, placement and function as verbal back-channels. Their performative function is to indicate that the addressee is following what the speaker is saying. Nods serve the same function so often that they are often treated as verbal rather than gestural.

“*Parsing functions*” are useful for “punctuating the spoken discourse or... marking out its different logical components” (p. 159). The most common group in Chovil’s (1989; 1991/1992) data were what she called *syntactic* displays, especially eyebrow movements that either emphasized a single word by a quick flash or underlined a whole phrase by staying up for the duration. Chovil also found that speakers who were relating a story or anecdote could use their smiles, not to show happiness, but to punctuate their narratives. For example, the patient in Table 1 smiled to mark a shift from telling about his symptoms to attending to the physician’s comment. Similarly, the speaker in Figure 1 smiled immediately after frame 3 to close off that part of her narrative.

“*Interactive or interpersonal functions*” include “the use of gestures as a way of indicating to whom a current utterance is addressed” (p. 159), for example by gaze and head direction. Another easily recognized facial gesture with an interactive function is the *thinking face*. As shown in Figure 3, this facial gesture usually involves shifting the gaze away from the addressee and looking thoughtful or searching, which indicates “that a current speaker, though not actually speaking, is nevertheless still claiming a role as speaker (still ‘holding the floor’)” (p. 159).



Figure 3. While recalling a scene from *Shrek 2*, the speaker made an extended (1.7 second) *thinking face* in frames 2 and 3

“*Two different kinds of expressive resource*”

Kendon emphasized that:

The gesture–speech relationship... is best understood in terms of a point of view that sees gesture and speech as two different kinds of *expressive resource* available to speakers, and that the gestures employed within an utterance, like the words that are employed, are components of the speaker’s *final product*. (p. 111)

By illustrating how facial gestures share Kendon’s characteristics of hand gestures, we have proposed that facial gestures are also a visible expressive resource that is a part of the speaker’s final product. The rest of this chapter focuses on flexibility between speech and gestures (of either kind), as a function of the “expressive resources ... existing within the context of the given moment of interaction” (p. 111).

Part 2: “There is flexibility in the gesture–speech relationship”

It appears that there is flexibility in the gesture–speech relationship... Both the gestures and the verbal expressions used are to be accounted for in terms of such factors as... the speaker’s knowledge of various expressive resources... as well as the constraints and possibilities existing within the context of the given moment of interaction. (p. 111)

To assess and test this flexibility, we applied a *semantic features analysis* (Beattie & Shovelton 1999, 2002; Gerwing & Allison 2009, 2011; Holler & Beattie 2002, 2003, 2004; Holler & Stevens 2007; Holler & Wilkin 2009) to speakers’ retelling of a video excerpt that could be described with speech, hand gestures and facial gestures. This method starts by identifying a set of semantic features that are specific to the material the speaker is describing. That is, the analysts stipulate in advance certain specific information in the stimulus material, then they assess whether words, gestures or both contribute information about each of these features.

Gerwing and Allison (2009) compared the semantic features method to two other ways of studying the relationships between gestures and speech (i.e. deictic references and redundancy) and found that the semantic features analysis had the advantage of identifying precisely how and when speakers distribute information between the two modes. The studies cited above have shown the utility of semantic features analysis for understanding the relationship between speech and hand gestures, but to the best of our knowledge, this method has not previously been applied to facial gestures.

The data are a subset of Experiment 1 in Bavelas et al. (in press), in which speakers retold several scenes from the movie *Shrek 2* in either a face-to-face dialogue, a telephone dialogue or a monologue. This new analysis focused on the face-to-face and telephone dialogues¹ in order to assess how visibility would affect the use of audible versus visible expressive resources. One scene was particularly suitable because it included features that could be described with words, hand gestures or facial gestures. In this scene, Puss in Boots (a cat) wants to join Shrek on his journey. The cat, who suddenly appears very small, clutches his hat under

1. Two experiments (Bavelas et al. 2008, in press) have shown that holding a phone does not significantly decrease the overall rate of hand gestures. These two studies found no difference between face-to-face and telephone conditions, which replicated the results of five similar experiments that compared face-to-face versus partition conditions: Rimé (1982), Bavelas, Chovil, Lawrie and Wade (1992: Exp. 2), Pine, Burney and Fletcher (2010), Holler, Tutton and Wilkin (2011), de Ruiter, Bangerter and Dings (2012). (Note that the de Ruiter et al. data on overall gesture rate were obtained from J. P. de Ruiter, personal communication, July 13, 2012.) See Bavelas and Healing’s (2013) review of visibility effects on hand gestures.

his chin and, with enormous eyes, looks up at the much bigger Shrek, silently using his cute face and huge eyes to manipulate Shrek into taking him along (see Figure 4). The comic elements of the scene are the cat's pose and especially his eyes, which he makes impossibly large and endearing – a strategy that any cat owner will recognize.



Figure 4. Puss in Boots silently pleading to be taken along in *Shrek 2*

The semantic feature analysis focused on how speakers conveyed three key pieces of information in the scene: the cat's clutching the hat, making very big eyes, and looking up at Shrek. We compared the proportions of visible versus audible means of conveying these three features when speakers were in two different conversational contexts: a face-to-face dialogue versus a telephone dialogue.

The hand or facial gestures that could describe the three features were all *enactments*, that is, first-person portrayals in which the speaker momentarily presented herself in the way the cat looked, clutching the hat or looking up with big eyes. First-person enactments of oneself or someone else are a common kind of hand or facial gesture. For example, the patient in the initial example used hand and facial gestures to portray himself, first as he was on a bad day, then on a recent good day. Recall that very little of this information was in his words; the visible information was not redundant with the audible information.

We predicted that the experimental condition would change the relative distribution of the information that speakers presented using visible versus audible means: in face-to-face dialogues, speakers would convey more information by visible means (hand and/or facial gestures), and in telephone dialogues, speakers would convey more information by audible means. This difference would demonstrate flexibility in the relationship between speech and gestures. We further predicted that the information conveyed by visible means would be redundant with speech less often in the face-to-face condition and more often in the telephone condition. In de Ruyter, Bangerter and Ding's (2012) terms, the visible information in the face-to-face condition would be *obligatory*, while in the telephone condition, it would be *non-obligatory*.

Method

Participants

Initially, 40 female undergraduate psychology students participated for course credit: 20 in the face-to-face condition (10 dyads) and 20 in the telephone condition (10 dyads). We randomly assigned both the experimental conditions and the participants' roles as speaker or addressee. We analyzed the 16 speakers who included the scene described above, 9 who were face to face and 7 who were on the phone.

Materials

The experimental stimulus was a 2 min., 45 s. video containing two excerpts from *Shrek 2*. The scene chosen for this analysis, in which the cat is silently appealing to Shrek, was 10 s. long.

Equipment

The experiment was held in the University of Victoria Psychology Department's Human Interaction Laboratory suite, using three Panasonic WV-CP474 color cameras to capture a large front view and narrow side view of the speaker, plus a smaller inset of the addressee. The speakers viewed the movie excerpts on a small color TV/VCR in an adjacent room within the suite. Speakers in the telephone condition used a handheld phone with the dialogue tapped into the audio track of the video recording. We digitized the videos with Broadway ProDVD (www.b-way.com) and analyzed them on an 18-inch ViewSonic G90fb color monitor using ELAN (www.lat-mpi.eu/Tools/Elan (Brugman & Russel 2004; Wittenburg, Brugman, Russel, Klassmann & Sloetjes 2006)).

Procedure

Before recording began, the participants met in the main recording room and provided written consent. In the telephone condition, the addressee then moved to a nearby office while the speaker remained in the main recording room, and they did the experimental tasks over the telephone. In both conditions, the participants spent a few minutes getting acquainted, then did two unrelated pilot tasks followed by the main task of viewing and describing the movie excerpts. The speaker watched these excerpts twice in an adjacent room, then returned to the main recording room to describe the scenes to the addressee either face to face or by telephone. Afterward, the experimenters debriefed the participants, answered questions and showed them the video of their participation. Each participant then signed a form indicating the permissible uses of their video (e.g. permission to view for analysis only, permission for viewing by professional audiences, permission to include a still photo in an academic journal).

Semantic features analysis

The first step was to identify the speakers who included this particular scene in their descriptions. In the face-to-face condition, nine speakers described it. In the telephone condition, eight did, but one participant was not analyzable because she had moved her head off-screen while describing the scene, leaving seven participants for analysis in that condition.

The analysis focused on three key semantic features of the cat's silent persuasion of Shrek during this scene: (1) the cat clutched his hat under his chin; (2) he made his eyes disproportionately large; and (3) he was looking up at Shrek, who was much taller. (Detailed operational definitions for these features are available from the authors.)

First, using only the audio, the analysts located when each speaker used words to refer to one of these features:

For the cat clutching his hat, the words could be any reference to "clutching," including synonyms (e.g. "holds his hat"). Verbal references to the hat alone, which did not include how the cat was holding it, were not sufficient to count for this feature.

For the cat's big eyes, verbal references counted only if they conveyed that the cat's eyes were unusually or disproportionately big (e.g. "huge" or "really, really big"). References to "big eyes" alone were not considered sufficient to count as a reference to this feature.

For the cat looking up, speakers had to convey the upwards direction of the cat's gaze, so the speaker's words counted only if they combined "looking" and "up". "Looking" alone was not sufficient because a key component of this feature was the height difference between the cat and Shrek.

The analysts then used the video to decide whether the participants used hand or facial gestures to depict any of the three features:

For clutching the hat, a reference counted as visible if the speaker's hands were in a clutching position (i.e. in fists) and held somewhere between under the chin and in front of the chest.

For the big eyes, visible references were widening the eyes, making them look bigger than the speaker's own baseline position, based on examining the speaker's eyes before and after the scene. Speakers could also use hand gestures to demonstrate big eyes by suggesting two large circles with curled thumbs and index fingers, then holding these up in front of their eyes.

For looking up at Shrek, the speaker had to look away from the addressee in an upwards direction (e.g. gazing upward and sometimes also tilting the head upward). The precise timing of the gaze was important. The analysts had to watch more than just this scene in order to ensure that this was in fact a portrayal of the

cat looking up and not the speaker raising her eyes to show that she was searching for a word (i.e. not a “thinking face” as in Figure 3).

Two analysts (JG and SH) conducted all of the analysis together. For reliability, a third analyst (JB) trained on a randomly selected 30% of the excerpts, then worked independently on a new randomly selected 30%. Agreement required identifying exactly the same word, hand gesture or facial gesture, and they agreed on 93% of their decisions.

Qualitative results

The words that speakers used for clutching the hat included “holding his hat,” “clutches, like, his hat,” or “holds his little hat in his hands.” Their most common hand gesture was to hold one or both fists close together, right under the chin (see Figure 5).

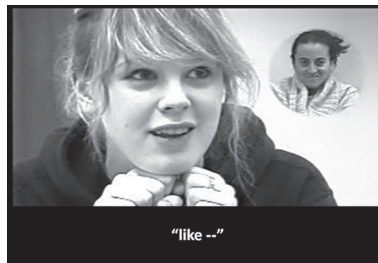


Figure 5. The speaker gestures the cat *clutching the hat* with her hands (.23 s) as well as the cat *looking up* by gazing upward (.43 s)

Speakers described the cat’s big eyes in a wide variety of ways: “big big black eyes,” “makes his eyes all big,” “he does this whole, like, big big eye thing,” “like big pussy cat eyes, they are huge,” “huge like puppy dog’s eyes.” They could also use a facial gesture to make their own eyes appear unusually big, for example, by opening their eyes more widely and looking far to the side, which combined to show much more white, as in Figure 6. Several speakers used a hand gesture that projected huge eyes in front of their face, as in Figure 7a.

For the cat looking up at Shrek, speakers said, for example, “peers up at Shrek” or “looks up to Shrek”. They gestured this feature by casting their own eyes up, as in Figure 5, or tilting their head up, as in Figure 7b.

These figures also illustrate two other patterns of the visible enactments. First, the speakers often demonstrated one or more features without any accompanying speech, just introducing the enactment with the discourse marker “like” (Figures 5

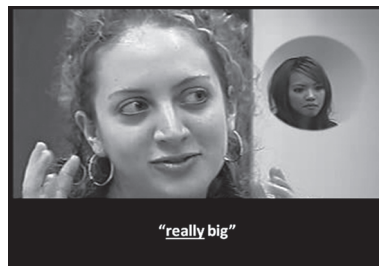


Figure 6. The speaker gestures the cat's *big eyes* by showing more whites of her own eyes (.40 s)



Figures 7a and 7b. The speaker gestures the cat's *big eyes* using a hand gesture in the first frame and then gestures the cat *looking up* with a facial gesture in the second frame. (Total duration =.81 s)

and 7b). Second, and perhaps most important, speakers frequently used their hand and facial gestures simultaneously (as in Figure 5) or in quick succession (as in Figures 7a and 7b). The result was to provide a more complete image of how the cat looked at that moment.

Quantitative results

Creating proportions

Table 2 shows the results for each speaker, aggregated across the three features. For each speaker, we first summed the number of references that this speaker made to any of the features using words, using hand gestures and using facial gestures. The sum of hand and facial gestures was the total number of visible references. The sum of visible and audible (verbal) references was the total number that the speaker made in any modality. The two key measures were the proportions of total references that were audible versus visible.

Table 2. Individual speakers' use of visible versus audible references to semantic features as a function of experimental condition

Dyad #	Face-to-face condition				Telephone condition				Total references	Total audible ^a	Total references	Total audible ^a
	Visible references		Total references		Visible references		Total references					
	Hand gestures	Facial gestures	Total visible ^a	Total	Hand gestures	Facial gestures	Total visible ^a	Total				
2	2	1	3 (.75)	4	3	0	4	4 (.80)	1 (.20)	5	1 (.20)	5
5	3	2	5 (1.0)	5	7	1	0	1 (.5)	1 (.5)	2	1 (.5)	2
9	3	4	7 (.875)	8	16	0	0	0	0	0	0	0
12	3	2	5 (.71)	7	17	0	0	0 (0)	1 (1.0)	1	1 (1.0)	1
18	2	0	2 (.67)	3	19	2	2	4 (.80)	1 (.20)	5	1 (.20)	5
21	2	0	2 (1.0)	2	28	1	0	1 (.5)	1 (.5)	2	1 (.5)	2
26	0	0	0	0	34	0	0	0	0	0	0	0
27	1	3	4 (.57)	7								
29	2	2	4 (.80)	5								
Mean proportions			.71	.18	Mean proportions			.37	.34			.34
Standard deviations			.30	.16	Standard deviations			.37	.36			.36

^a The numbers in italics are the proportions of total references.

Note that we aggregated hand and facial gestures for two reasons. First, the theoretical focus was on visible versus audible enactments as a function of visible and not-visible experimental conditions. Second, as illustrated in Figures 5 and 6, the hand, head, gaze and facial gestures often occurred together, portraying an integrated picture of the cat at a particular moment. Separating them would obscure this integration.

Use of speech versus gesture within experimental conditions

We first examined (a) whether speakers were more likely to use words or gestures to convey information about the three semantic features and (b) whether these distributions varied as a function of the experimental condition. As shown in Table 2, every speaker in the face-to-face condition used a higher proportion of gestures than words. The difference in the mean audible and visible proportions was statistically significant: within-subjects $t(8) = 4.530$, one-tailed $p < 0.002$. In contrast, speakers in the telephone condition were more varied, and the mean audible and visible proportions were identical to each other.

Use of speech versus gesture between experimental conditions

We also compared how the conversational context (i.e. speaking face to face or on the telephone) affected how much information speakers conveyed in each modality. Comparing across conditions in Table 2 shows that the mean proportion of visible references to the three semantic features in the face-to-face condition was almost double the proportion in the telephone condition, a difference that was statistically significant: between-subjects $t(14) = -.015$, one-tailed $p < 0.032$. That is, speakers who were talking face-to-face made visible references to the features significantly more often than speakers on the telephone made visible references to the same features. Conversely, the mean proportion of audible references was higher in the telephone condition than in the face-to-face condition. However, this difference was not significant, which led to our third analysis.

Redundancy between speech and gesture as a function of experimental condition

We hypothesized that, although the proportional use of speech to convey information about the semantic features did not differ between experimental conditions, the relationship between the speakers' words and gestures would differ in the two conditions. Specifically, the hand and facial gestures in the telephone condition would tend to convey the *same* information as the words and would therefore be redundant (i.e. not obligatory). As a result, information would be available to the addressee via the speakers' words, even though the gestures were not visible. In contrast, the hand and facial gestures in the face-to-face condition would convey

different information than was conveyed in words and would therefore be non-redundant (i.e. obligatory). The addressee would have to see the speaker's visual enactment to get all of the information the speaker was providing about the three features.

To test this hypothesis, we returned to each speaker's raw scores for each semantic feature and recorded whether the speaker's visible contribution was obligatory (i.e. the speaker did not convey the same information in words) or not obligatory (i.e. the speaker also conveyed the same information in words). We then collapsed the three semantic categories and recorded whether or not that speaker conveyed at least one semantic feature using an obligatory hand or facial gesture. Three speakers were excluded because, as shown in Table 2, they did not refer to any of the three features.

As shown in Table 3, seven of the nine speakers in the face-to-face dialogues conveyed information in hand or facial gestures that was obligatory, that is, not conveyed at all in the words. One speaker conveyed information about the semantic features in both words and hand/facial gestures. In telephone dialogues, four of the seven speakers conveyed information in hand or facial gestures that was redundant with the words or was non-obligatory. One participant conveyed information in at least one category using obligatory hand or facial gestures.

Table 3. Effect of experimental condition on obligatory vs. non-obligatory gestures

Relation of gestures to words	Experimental condition	
	Face-to-face	Telephone
At least one hand or facial gesture that was obligatory (not redundant with words)	7	1
No hand or facial gestures that were obligatory (all were redundant with words)	1	4

Note: One speaker in the face-to-face condition and two in the telephone condition did not describe any of the semantic features analyzed.

* $\chi^2 (1, N = 13) = 5.923; p < 0.05$. Note that two of the expected frequencies are lower than conservative practice recommends. However, the pattern is clear in the observed frequencies themselves.

These results are consistent with the only other semantic features analysis that has compared how speakers distributed information in speech and hand gestures in different conversational contexts. Gerwing and Allison (2011) found that speakers who were describing the shape of the skirt on an unusual dress conveyed significantly more of this information in their gestures than in their words when speaking to an addressee in a face-to-face dialogue. When the addressee was on the telephone, speakers conveyed significantly more of the information in their words.

Discussion

This chapter has provided two related illustrations of the lasting influence and relevance of Kendon's work, especially his magnum opus (2004). Part 1 used his extensive and detailed specifications of the characteristics, contributions and functions of conversational hand gestures as a framework and set of standards for including *conversational facial gestures* as another instance of "visible action as utterance". This framework made it possible to articulate and document the extensive similarities of facial to hand gestures, which offer an alternative to approaches that see the face as stereotypic configurations related to a few emotional expressions. Facial gestures include anything the face, head and eyes can do to convey any meaning related to the talk in progress: they can convey referential content either deictically or by direct representation. They can serve pragmatic functions, such as indicating the mode or frame of an utterance (e.g. the facial shrug), indicating the kind of speech act (e.g. eyebrows marking a question), parsing the utterance (e.g. a smile closing a narrative), or indicating the status of turns (e.g. a speaker's thinking face). In all of these functions, facial gestures are part of the speech-gesture ensembles that constitute language in dialogues.

Hand and facial gestures, as well as other bodily movements of the torso or legs, often act in concert with speech and each other. The gestural components of these ensembles present detailed and nuanced *images* of how someone looked (e.g. the cat) or acted (e.g. the patient). Therefore, in drawing attention to facial gestures, we are not suggesting that they should be studied in isolation from gestures made by other parts of the body. The challenge is to appreciate the parts while still keeping them in the context of the whole ensemble, in which the gestures "serve to create an image of the object that is the topic of the spoken component" (p. 161).

Part 2 added facial gestures to an experimental demonstration of Kendon's insights on the flexibility of the relationship between parts of the speech-gesture ensembles in different interactional contexts. A semantic features analysis showed how speakers represent the same material differently depending on changes in transmission conditions. When conversing face-to-face, speakers conveyed information about semantic features more in facial and hand gestures than in words. In some instances, speakers conveyed information about a feature entirely in gestures, making these references obligatory or necessary for accurate comprehension of the description. When conversing on the telephone, speakers conveyed information about the three features equally in words and gesture. However, the gestures that they used were also more likely to be redundant with their words and were therefore not necessary for accurate comprehension. In other words, "Speakers... can control these two components and can orchestrate them differently, according to the occasion" (p. 127).

References

- Bavelas, J., and Chovil, N. 1997. "Faces in dialogue." In *The Psychology of Facial Expression*, J. A. Russell and J. M. Fernandez-Dols (eds), 334–346. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659911.017
- Bavelas, J., Chovil, N., Lawrie, D. A., and Wade, A. 1992. "Interactive gestures." *Discourse Processes* 15: 469–489. DOI: 10.1080/01638539209544823
- Bavelas, J., Gerwing, J., and Healing, S. 2014. "Hand and facial gestures in conversational interaction." In *Handbook of Language and Social Psychology*, T. Holtgraves (ed.), 111–130. Oxford: Oxford University Press.
- Bavelas, J., Gerwing, J., and Healing, S. in press. "The effect of dialogue on demonstrations: direct quotations, facial portrayals, hand gestures, and figurative references." (accepted for publication in *Discourse Processes*)
- Bavelas, J., Gerwing, J., Sutton, C., and Prevost, D. 2008. "Gesturing on the telephone: Independent effects of dialogue and visibility." *Journal of Memory and Language* 58: 495–520. DOI: 10.1016/j.jml.2007.02.004
- Bavelas, J., and Healing, S. 2013. "Reconciling the effects of mutual visibility on gesturing. A review." *Gesture* 13: 63–92.
- Beattie, G., and Shovelton, H. 1999. "Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation." *Semiotica* 123 (1–2): 1–30. DOI: 10.1515/semi.1999.123.1-2.1
- Beattie, G., and Shovelton, H. 2002. "An experimental investigation of some properties of individual iconic gestures that mediate their communicative power." *British Journal of Psychology* 93: 179–192.
- Birdwhistell, R. L. 1970. "Masculinity and femininity as display." *Kinesics and Context: Essays on Body Motion*. Philadelphia, PA: University of Pennsylvania.
- Brugman, H., and Russel, A. 2004. "Annotating multimedia/multimodal resources with ELAN." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva (eds), 2065–2068. Paris: European Language Resources Association.
- Brunner, L. J. 1979. "Smiles can be back channels." *Journal of Personality and Social Psychology* 37: 728–734. DOI: 10.1037/0022-3514.37.5.728
- Chovil, N. 1989. *Communicative Functions of Facial Displays in Conversation*. Ph.D. dissertation, University of Victoria, BC.
- Chovil, N. 1991/1992. "Discourse-oriented facial displays in conversation." *Research on Language and Social Interaction* 25: 163–192. DOI: 10.1080/08351819109389361
- Chovil, N. 1991. Social determinants of facial displays. *Journal of Nonverbal Behavior* 15: 141–153. DOI: 10.1007/BF01672216
- Chovil, N. 1997. "Facing others. A social communicative perspective on facial displays." In *The Psychology of Facial Expression*, J. A. Russell and J. M. Fernandez-Dols (eds), 321–333. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659911.016
- Coates, L. J. 1991. *A Collaborative Theory of Inversion: Irony in Dialogue*. M.A. thesis, University of Victoria, BC.
- de Ruiter, J. P., Bangerter, A., and Dings, P. 2012. "The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis." *Topics in Cognitive Science* 4 (2): 232–248. DOI: 10.1111/j.1756-8765.2012.01183.x

- Ekman, P. 1979. "About brows: Emotional and conversational signals." In *Human Ethology*, M. von Cranach, K. Foppa, W. Lepenies and D. Ploog (eds), 169–249. Cambridge: Cambridge University Press.
- Ekman, P. 1985. *Telling Lies*. New York: Norton.
- Ekman, P. 1997. "Should we call it expression or communication?" *European Journal of Social Sciences* 10: 333–359.
- Ekman, P., and Friesen, W. V. 1969. "The repertoire of nonverbal behavior: Categories, origins, usage, and coding." *Semiotica* 1: 49–98.
- Enfield, N. J. 2001. "Lip-pointing: A discussion of form and function with reference to data from Laos." *Gesture* 1: 185–221. DOI: 10.1075/gest.1.2.06enf
- Fridlund, A. J., Ekman, P., and Oster, H. 1987. "Facial expressions of emotion: Review of literature, 1970–1983." In *Nonverbal Communication and Behavior* (2nd ed.), A. W. Siegman and S. Feldstein, eds. (143–224). Hillsdale, NJ: Erlbaum.
- Gerwing, J., and Allison, M. 2009. "The relationship between verbal and gestural contributions in conversation: A comparison of three methods." *Gesture* 9 (3): 313–336. DOI: 10.1075/gest.9.3.03ger
- Gerwing, J., and Allison, M. 2011. "The flexible semantic integration of gestures and words: Comparing face-to-face and telephone dialogues." *Gesture* 11 (3): 308–329. DOI: 10.1075/gest.11.3.03ger
- Holler, J., and Beattie, G. 2002. "A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk." *Semiotica* 142: 31–69.
- Holler, J., and Beattie, G. 2003. "How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?" *Semiotica* 146: 81–116.
- Holler, J., and Beattie, G. 2004. "The interaction of iconic gesture and speech." In *Lecture Notes in Computer Science, 5th International Gesture Workshop, Genova, Italy, 2003: Selected Revised Papers*, A. Cammurri and G. Volpe (eds), 63–69. Heidelberg: Springer Verlag.
- Holler, J., and Stevens, R. 2007. "An experimental investigation into the effect of common ground on how speakers use gesture and speech to represent size information in referential communication." *Journal of Language and Social Psychology* 26: 4–27. DOI: 10.1177/0261927X06296428
- Holler, J., and Wilkin, K. 2009. "Communicating common ground: How mutually shared knowledge influences the representation of semantic information in speech and gesture in a narrative task." *Language and Cognitive Processes* 24: 267–289. DOI: 10.1080/01690960802095545
- Holler, J., Tutton, M., and Wilkin, K. 2011. "Co-speech gestures in the process of meaning coordination." In *Proceedings of the 2nd GESPIN – Gesture in Speech and Interaction Conference*, Bielefeld, 5–7, September 2011.
- Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kraut, R. E., and Johnston, R. E. 1979. "Social and emotional messages of smiling: An ethological approach." *Journal of Personality and Social Psychology* 61: 743–745.
- Pine, K. J., Burney, D. J., and Fletcher, B. 2010. "The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener." *Journal of Nonverbal Behavior* 34: 169–178. DOI: 10.1007/s10919-010-0089-7
- Rimé, B. 1982. "The elimination of visible behavior from social interactions: Effects on verbal, nonverbal and interpersonal variables." *European Journal of Social Psychology* 12: 113–129. DOI: 10.1002/ejsp.2420120201
- Sherzer, J. 1973. *Verbal and Nonverbal Deixis: The Pointed Lip Gesture among the San Blas Cuna*. University of Texas at Austin: Institute of Latin American Studies.

- Tomasello, M., Hare, B., Lehmann, H., and Call, J. 2007. "Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis." *Journal of Human Evolution* 52 (3): 314–320. DOI: 10.1016/j.jhevol.2006.10.001
- Wittenburg, P., Brugman, H., Russel, A., Klassman, A., and Sloetjes, H. 2006. "ELAN: A professional framework for multimodality research." In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, Genoa.