Introduction: Multimodal interaction

TANYA STIVERS and JACK SIDNELL

That human social interaction involves the intertwined cooperation of different modalities is uncontroversial. Researchers in several allied fields have, however, only recently begun to document the precise ways in which talk, gesture, gaze, and aspects of the material surround are brought together to form coherent courses of action.[1] The papers in this volume are attempts to develop this line of inquiry. Although the authors draw on a range of analytic, theoretical, and methodological traditions (conversation analysis, ethnography, distributed cognition, and workplace studies), all are concerned to explore and illuminate the inherently multimodal character of social interaction. Recent studies, including those collected in this volume, suggest that different modalities work together not only to elaborate the semantic content of talk but also to constitute coherent courses of action. In this introduction we present evidence for this position. We begin by reviewing some select literature focusing primarily on communicative functions and interactive organizations of specific modalities before turning to consider the integration of distinct modalities in interaction.

1. Semiotic modalities

As conversation analysts, we begin by observing that social interaction is most 'at home' in face-to-face interaction. This is as Schegloff (2000a: 1) puts it a 'species-distinctive embodiment of the primordial site of sociality.' Levinson notes along similar lines that

> ... conversation is clearly the prototypical kind of language usage, the form in which we are all first exposed to language — the matrix for language acquisition. Various aspects of pragmatic organization can be shown to be centrally organized around usage in conversation. [For instance, the] unmarked usages of grammatical encodings of temporal, spatial, social discourse parameters are organized

Semiotica 156–1/4 (2005), 1–20 0037–1998/05/0156–0001

around an assumption of co-present conversational participants. (Levinson 1983: 284)

Face-to-face interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours. In this chapter we follow Enfield (2005) in distinguishing between the vocal/aural and visuospatial modalities. The vocal-aural modality encompasses spoken language including prosody. The visuospatial modality includes gesture, gaze, and body postures. As Enfield notes, these differ not only in terms of modality but also with respect to which semiotic ground plays a dominant role in their organization. Vocal-aural signs are prototypically symbolic whereas indexicality and iconicity are more important in the visuospatial modality.[2] We want to point out that by looking at interaction from a multimodal perspective we do not mean to privilege one modality over another (e.g., visuospatial over vocal/aural) but rather to suggest that much can be gained from examining a turn-at-talk for where it is situated vocally (e.g., sequentially, prosodically, syntactically) as well as visuospatially (e.g., body orientation, facial expression, accompanying gestures), and that different modalities should not, *a priori*, be treated as more or less important.

2. The vocal modality

Since the early 1960s conversation analysts have studied language as a part of talk-in-interaction. The focus of conversation analytic work is on those organizations of practice through which social interaction is accomplished (i.e., turn-construction, sequence organization, repair, etc.; see Schegloff 2004). A basic finding of CA is that the meaning or communicative function of a particular linguistic item (e.g., individual words, grammatical constructions etc.) is, more often than not, a product of the context within which it occurs. Thus, Schegloff et al. (1996) write,

The meaning of any single grammatical construction is interactionally contingent, built over interactional time in accordance with interactional actualities. Meaning lies not with the speaker nor the addressee nor the utterance alone ... but rather with the interactional past, current and projected next moment. (Schegloff et al. 1996: 40)

Research by conversation analysts focusing on the verbal modality has shown that lexical selection is shaped by features of the interactional

context (e.g., Sacks and Schegloff 1979; Schegloff 1972; Schegloff 2000b) and that particular lexical items are often used as resources for accomplishing particular interactional tasks. For instance, in particular sequential environments, 'okay' is an interactional practice for initiating sequence closure (Beach 1993; Schegloff, in press); turn-initial 'well' can foreshadow impending disagreement (Pomerantz 1984); and freestanding 'oh' typically registers a change-of-state in the speaker from not-knowing to knowing (Heritage 1984).[3]

Many CA studies have focused not on particular lexical items but on the organization of turns-at-talk and sequences of turns. Cut-offs, sound stretches, and 'uh's,' for instance can (among other things) alert the recipient to the possibility of imminent self-repair (Schegloff et al. 1977). Moreover, both the turn's design and the turn's placement in relation to a preceding turn, carry communicative import. An excellent example is provided by recent studies of question design. Heritage (2002) argues that a question formatted as a negative interrogative such as, 'Isn't it beautiful down there?' conveys an assertion to be agreed or disagreed with rather than posing an information seeking query as the format 'Is it pretty down there?' does.

Taken as a whole, conversation analytic studies have described not only a very wide range of actions accomplished through the vocal modality, but also, important for many of the studies in the present volume, the systematic organization of such practices (for example those associated with the orderly distribution of opportunities to speak; see Sacks et al. 1974). Some of the studies in the present volume are particularly concerned to extend this kind of analysis to the multimodal character of social interaction and so ask, for instance, about the relations between gesture and turn design or sequence organization (see papers by Hayashi and Sidnell, for instance).

In addition to the lexico-syntactic channel, the vocal modality also includes the prosodic channel. That these are discreet channels is evidenced by the fact that a speaker can produce a particular word with upward or downward intonation or, in certain languages, with different tones. A speaker cannot, of course, produce two words (or grammatical constructions) at the same time, suggesting that lexico-grammatical structure comprises one channel. Recent studies have also suggested that the prosodic channel allows speakers to perform discrete communicative functions which laminate onto the lexico-semantic channel and the visuospatial modality. For example, with respect to turn taking, intonation can work for or against the syntactic construction of a turn in order to facilitate or block turn transfer (Ford and Thompson 1996; Local et al. 1986; Local et al. 1985; Local and Walker 2004; Sacks et al. 1974; Schegloff, in press;

Wells and Peppé 1996). It has also been suggested that in same speaker re-sayings such as 'no no no,' prosody is a primary resource for hearers to understand the re-saying as comprising a single unit of talk rather than multiple discrete units (Stivers 2004).

Intonation has also been shown to communicate an invitation to display recognition such as when a reference to a person is delivered with upward or 'try marked' intonation (Sacks and Schegloff 1979). The prosodic contour of a particular token has been shown to alter what sort of stance it communicates (e.g., Local 1996; Müller 1996). Although this research certainly offers evidence that prosody has communicative functions, most of the work thus far has focused on issues of turn taking to the exclusion of other possible functions (cf. Selting, 1996). Although none of the papers in this volume deal exclusively with prosody, prosody is nonetheless an important interactional resource.

The prosodic and lexico-syntactic channels can be seen to work together. For instance as Ford and Thompson (1996) show with respect to turn taking, while syntax 'nominates' an utterance as complete, intonation can 'second the nomination'. However, relatively little is understood about how these two channels work together. Moreover, although work on prosody certainly takes into account what else is going on in the vocal modality, the visuospatial modality is often neglected in such studies (cf. C. Goodwin and M. H. Goodwin 1987; M. H. Goodwin and C. Goodwin 1986; M. H. Goodwin 1996). In addition, research focusing on lexico-syntactic practices of social interaction has most commonly relied on telephone calls which, as a context, is mono-modal (though multi-channel).

3. The visuospatial modality

Manual gestures, facial expressions and body posture may together be understood as constituting a visuospatial modality. Gestures have been studied extensively by researchers from a number of different fields but especially psychology and psycholinguistics. Although certain types of gesture may be connected with language production (e.g., Goldin-Meadow 2003; Kita and Özyürek 2003; McNeill 1992), strong evidence exists to suggest that gestures have basic communicative import for interaction (e.g., Bavelas et al. 1992; C. Goodwin 1986; M. H. Goodwin 1980, 1983; M. H. Goodwin and C. Goodwin 1986; Heath 1982, 1986; Kendon 1994, 2004; Streeck 1993; Mondada 2004). For instance, M. H. Goodwin and C. Goodwin (1986) showed that a gesture could be treated as something to be recognized and confirmed. Heath showed that in particular

contexts, gestures could engender a display of co-participation from the hearer (Heath 1992; see also C. Goodwin 1986).

For our purposes here, it is critical that we understand the interactional work accomplished via the visuospatial modality. For instance, the way people organize their bodies when interacting with one another has been shown to be important for such issues as facilitating a common focus of attention (Kendon 1990). This is communicative in the sense that when a person enters what Kendon called an 'F-formation' (i.e., when speakers are co-oriented to each other around a central space such as three speakers in a triangle), they rely, at least initially, on the visuospatial modality to convey their willingness to participate in a given interaction (or, conversely, if they are outside the F-formation, they communicate, among other things, that they are not to be oriented to as core participants). In Kendon's terms this is a means for establishing interactional 'withness' (Kendon 1990: 250). Additionally, participants may display, through their body posture, whether they are in a stable or unstable position relative to a particular activity such as a conversation (Schegloff 1998). When conversing in a 'torqued' body posture (i.e., one part of the body oriented in one direction and another in an alternative direction as happens when you turn your head or upper torso away from an interlocutor to greet a passing colleague), the speaker communicates that the activity is unstable and that it is a shorter term or 'subordinate' involvement relative to another activity (Goffman 1963).

The visuospatial modality may also be used to communicate a preparedness to transition to a next activity. For instance, in a medical visit, a physician may propose to close the activity of history taking and initiate a transition to the relevant next activity through the action of setting down a pen (Robinson and Stivers 2001). Whereas gestures such as nods and points are designedly communicative, putting a pen down becomes communicative only in a particular sequential context.

Gaze, another channel within the visuospatial modality, has also been shown to contribute in a range of ways to the unfolding interactive situation (and can thus be said to 'communicate'). One important job routinely accomplished by gaze is that of selecting a recipient in multiparty interaction (C. Goodwin 1979). At the same time, gaze is also used by recipients to show that they are attending to the talk of the moment. Goodwin's pioneering study of interactive sentence construction showed that speakers orient to this use of gaze by recipients. In the example he discusses, a current speaker seeks out an attending recipient — a recipient who is gazing at him — and modifies his talk in the course of its delivery so to as to make it hearable as appropriate, indeed specifically designed for, just that participant who is displaying co-participation through gaze.

Moreover, participants can recognize talk that is directed specifically to them by virtue of the speaker's gaze direction. This becomes particularly relevant in cases where a question is asked in the presence of multiple possible recipients but does not contain any verbal address — a particular participant in such a situation can recognize that they have been selected to speak next by virtue of the questioner's gaze direction (Lerner 2003).

4. Multiple modalities

Thus far, we have drawn attention to the fact that different channels within different modalities can perform discrete communicative functions. The studies in this volume are concerned with the integration and co-operation of such different modalities in social action and interaction. As noted at the outset, face-to-face social interaction is necessarily multimodal and typically involves the cooperation of vocal and visuospatial modalities. There are many ways in which these modalities work together. The communicative work that is performed by one modality may be supported or extended by the work of another modality. Gesture, for instance, can behave in this way in conjunction with deictics (e.g., Enfield 2001; Kendon and Versante 2003) or descriptions (e.g., Enfield 2003). However, it is also possible for work by one modality to be modified by work of a different modality. Kendon argues that in Italian, for instance, particular gestures can transform the action accomplished by a particular utterance (Kendon 1995).[4]

We now want to turn to two very short examples in order to make two separate points relevant to the goals of this volume. The first instance offers an example of how the lamination of the visuospatial modality onto the vocal modality, alters the communicative impact of the turn. The second instance offers an example of what can be gained by looking at the integration of different modalities in their sequential and interactional context.

4.1. *Example (1)*

A relatively simple yet illustrative case of multimodality is provided by certain emblematic gestures (see Haviland 2004 on the distinction used here between points, illustrators and emblems). For instance, in a skit from the American TV show, *Saturday Night Live*, a character played by the late Chris Farley ('Bennet Brauer') admits that,

(1) maybe I'm not "the norm." I'm not "camera friendly." I don't "wear clothes that fit me." I'm not a "heartbreaker."

Figure 1. *Chris Farley as Bennet Brauer (reprinted by permission of Lions Gate Films)*

In the course of producing "the norm," "camera friendly," "wear clothes that fit me," and "heartbreaker," the actor raises his hands up on either side of his head, and, at the beginning and end of each word, rhythmically draws index and pointing finger into an otherwise closed fist before releasing them again. The gesture thus produced is vernacularly known as "airquotes" and has a significance available to any member of the society in which it is routinely used: the words so encapsulated by the gesture are to be understood as belonging to someone other than their immediate speaker.[5]

Airquotes are then an emblematic gesture which have a clearly metalinguistic function. We may note that the shape of the hand in this case bears no obvious semantic relation to the content of the talk (i.e., there is no connection between quotation marks and not being 'the norm') rather the gesture is a facsimile of an orthographic symbol. The two modalities in this case are thus related not in semantic but, rather, pragmatic terms — one providing the context within which the other is to be understood.

It is worth asking how different modalities work together. How is it that a recipient can know which gestures and other movements go with a

particular sequence of sounds? Temporal relations within an unfolding course of talk appear to be important in this respect. In the case we have been considering, it is by virtue of a temporal connection of virtual simultaneity that the recipient can see the relation between this gesture (the air-quotes) and the particular part of the talk (not "the norm") to which it is directed. The recognition of this relation is aided also by the observable connection between the prosodic character of the talk and the prosodic character of the gesture — so, movement of the hands is coordinated with the syllabic and lexical structure of the talk.

Airquotes are used to achieve what have often been characterized as 'footing' shifts (see Goffman 1980; Levinson 1988; Schegloff 1996; C. Goodwin and M. H. Goodwin 2004) and this gesture appears, like other emblems, to be specifically adapted to this single purpose. But footing shifts may in fact be accomplished by a range of non-emblematic gestures. A very basic way in which gesture differs from the vocal modality lies in the fact that gestures occupy space and consequently have 'fronts' and 'backs' ('tops' and 'bottoms,' etc.). As such gestures represent objects (actions, events, etc.) from a particular perspective and by virtue of this propose the way in which they are to be understood by a recipient. As a result, gestures can convey an organization of participation as well as perspectival shifts in the course of a telling. For instance, a subtle repositioning of the head and widening of the eyes can work to position the teller as someone who *witnessed* the events being described. A subsequent return of the head to its former position accompanied by another gesture may indicate that what is now being said (or 'reenacted') belongs to a participant in the scene rather than a witness. Such shifts of footing are routinely accomplished within a single turn-at-talk.

The temporal coordination of gesture, gaze and talk are crucial since it is this co-occurrence which indicates just what gesture is being linked (semantically, pragmatically, referentially ...) to what strip speech. It is well established that iconic gestures typically precede their lexical affiliates (Butterworth and Beattie 1978; Morrel-Samuels and Krauss 1992; Schegloff 1984). With respect to face-to-face interaction, Schegloff (1984) suggests that this ordering might provide evidence of the extent of the 'projection space' — the point at which something not yet articulated can be understood as interactionally 'in play.' Specifically, once a gesture is produced, it is available to interlocutors for any number of actions just as a word is available once articulated.

Certainly, the co-occurrence of gesture and lexical affiliate suggests a semantic relation of mutual elaboration (or emphasis). A number of analysts have examined the ways in which gestures are produced so as to elaborate what is being said or even substitute for what is not (or cannot

be) said (see example [2]). However, there is more to temporal coordination between talk and gesture than simply the setting up of semantic relations such that they can be recovered by a recipient. Specifically, the coordination of different modalities serves an important interactional function.

Crucially, as Schegloff (1984) noted, a gesture's production is coordinated not only with semantic and/or lexical units of the kind typically defined by linguistic analysis, but also with interactional ones (of the kind defined by conversation analytic methods). Schegloff noted then that the initiation of a gesture is routinely coordinated with the beginning of a turn-at-talk. Turn-beginnings are interactionally sensitive in a number of ways and for a number of reasons. For one thing, a turn beginning — in its design and composition — reflects the contingencies of the turn transition from which it emerged and which it proposes, by its occurrence, to close. In multiparty, face-to-face interaction, where several speakers may simultaneously self-select, a gesture can serve to solicit a public display of co-participation in the gesturer's talk. Specifically, where a speaker produces a gesture at the outset of their turn, they invite co-participants (or some subset of them) to redirect their gaze so as to be able to see it (See C. Goodwin 1986). As such, gestures can play an important role in turn-transfer via self-selection (Mondada 2005).

A gesture can be inspected by its recipient in much the same way as a turn-at-talk to find that it is now beginning, now continuing, now reaching completion (C. Goodwin 2000, 2002). To the extent that the beginning, continuation and completion of a gesture are coordinated with the beginning, continuation and completion of a turn (or sequence), gestures play a role in providing for the recognizability of turn completion (and continuation, etc.). It is instructive in this respect to note that gestures are frequently held or suspended in the course of their production (Enfield 2004). Such suspensions — which delay the progressivity of the turn/gesture — may play a role similar to pauses occurring at points of maximal grammatical control: they delay the progress of the talk while maintaining the current speaker's exclusive right to the turn (see Sidnell, this volume).

In studying the relations between different modalities it is thus crucial that we keep in mind the distinct semiotic properties of each. Language is distinguished from other modalities by the extreme development of arbitrary, semantic, referential qualities in morphology and other aspects of linguistic form. (Spoken) language (or at least the lexico-syntactic channel) is fundamentally linear — each semantic unit necessarily occurring either before or after every other.[6] Gesture clearly differs in this respect. Although recent work by Enfield (2004) has shown that manual signs

can enter into linear and essentially grammatical combinations, gestures typically present to a recipient multiple aspects of a represented thing simultaneously — that is 'imagistically' (i.e., location. direction, shape, size, texture, movement, etc.).

We now turn to a second example. In this case we focus on a sequence of interaction. Although the focus of this discussion will be on a gesture that comes into 'interactional play' prior to its projected lexical affiliate with critical consequences, we will also examine other relevant aspects of the interaction from a multimodal perspective.

4.2. *Example (2)*

In the example below, Andy, Tim, and Joe are sitting together outside of a barbershop where Tim and Joe work and which Tim owns. They are chatting together. At line 1, Andy announces that he saw a Mexican at a local beach who had a pit bull. Through the use of 'I seen ...' to introduce this report, Andy projects that what is remarkable about the dog is something observable. What we want to focus on in this example is when and how the gesture in line 6 is used as a resource in this description. However, it is first necessary to work through the sequential context in terms of both the vocal and the visuospatial modalities.

(2) So Cut (TS HS5)

```
1 AND:  I seen this Mexican (out there at (Phoenix) beach ma:n_
2 AND:  (fella he) had=uh pit bull so: (0.2)/((head shake))
3 TIM:  Just lookin' [vicious.
4 AND:               [I'm talkin' about- (dat no nuh nuh   so:)
5 AND:  I'm talkin' a[bout                [so cu:t.=
6 AND:               [((fists clenched))  [
7 TIM:                                    [muscled up.
8 AND:  =ma[:n=he look- (0.5)
9 TIM:     [Whoa:.
```

First, notice that at line 2 Andy projects that some sort of adjective descriptor of the dog, which would possibly complete this turn constructional unit, is due next. His recipient(s) are provided with several indicators of what type of descriptor this will be. First, note that the person reference is an ethnic category label 'this Mexican.' By introducing the person with this reference, Andy indicates that this is not a person his recipients would be able to recognize among the people they know (Sacks

and Schegloff 1979; Schegloff 1996). The use of 'this' further highlights that the person is to be heard as a category of person and not as an individual. Through this formulation then, Andy highlights, the ethnicity of this person as relevant to the rest of his description. Next, he offers a place reference: '(Phoenix) beach.' Although the exact reference is not something we as analysts can be certain of, it is designed for the interlocutors as a known place and one which would suggest what kind of event this would be (Schegloff 1972). At least one possibility is that, given that the event took place on a beach in southern California, it was done 'for show'; that it was designedly public. Another possibility is that the event is newsworthy (at least in part) because it occurred in this place.

Third, the formulation of the dog is by its breed, 'pit bull.' Like the other references in this turn ('this Mexican' and '(Phoenix) beach'), 'pit bull' also invokes particular characteristics. Pit bulls are widely known to be aggressive, are frequently bred for fighting, and they are exceptionally strong for their size. All of this is culturally available knowledge. Indeed, Tim owns a pit bull himself, and Andy knows this. Coupled with the person who had the dog being Mexican and the location being at a particular beach, the interlocutors are looking for an adjective which could focus more on the aggressive nature, the strength, or the appearance of the dog. Further, this could be designed as either a positively or a negatively valenced assessment.

In addition to the lexical resources, note that syntactic resources may be doing work to position the speaker in terms of dialect (e.g., 'I seen' and not 'I saw'). And the prosodic channel is a resource for inhibiting his recipient from initiating a turn at points of possible syntactic completion (e.g., before 'ma:n' at the end of line 1). Moreover, we can see in Figure 2 the positioning of the interlocutors bodily. Although Andy, Tim, and Joe are all seated on a wall that to some extent inhibits their ability to fully orient their torsos towards one another, note that Tim torques his head towards Andy and appears to gaze towards Andy. Additionally, his torso is upright enough to allow easy access to Andy's body movements. In this way, Tim positions himself, through the visuospatial modality, as a recipient prior to any vocal contribution. Thus, he can make use of this modality to align with Andy's action initiation by taking up a stance of recipient.

By contrast observe that Joe, by gazing downwards and bodily positioning himself as oriented to the ground, places himself outside of the participation framework (C. Goodwin 1981, 1988, 2000; Kendon 1990) and thus as a non-focal participant.

At the end of line 2, Andy hesitates completing his turn with the stretch on 'so:' and the subsequent 0.2 second silence. Both interrupt the

2 (fella he) had=uh pit bull so: (0.1) [(0.1)

Figure 2. *From left to right: Andy, Tim, and Joe*

progressivity of the turn (Schegloff 1979). However, just following the stretch, Andy does a head shake (in the silence), and this head shake appears to function as an intensifier of the term which is to follow (Kendon 2002). Thus, this is yet another projection of the next term suggesting that it will be an extreme formulation. Thus, the projection is, once again making use of both the lexico-syntactic and prosodic channels of the vocal modality and multiple channels of the visuospatial modality.

It is in this location that Tim attempts an anticipatory completion of Andy's turn (Lerner 1996, 1991). Although not completely achieving 'contiguous placement of the affiliating utterance' (Lerner 1996: 246), he does show an orientation to this insofar as 'Just lookin' vicious' is fitted to 'had a pit bull.' Moreover, Tim's 'looking' complements Andy's introductory framing of the telling with 'I seen . . .' Thus, it is an alternative construction but one which is apparently trying to affiliate with the basic idea that Andy has projected.

'Vicious' is certainly in line with the references Andy made as well. 'Dog fighting' is an activity that, at least in Los Angeles, tends to be associated with the ethnic category 'Mexican' and a vicious fighting dog at the beach would surely constitute a notable scene. However, it turns out that this is not what Andy was going for. In line 4, he competes with Tim to actually complete the telling with 'I'm talking about,' but when 'vicious' is audible, Andy cuts his turn off and initiates self repair (Schegloff et al. 1977) in order to overtly reject Tim's anticipatory completion as incorrect ('dat no nuh nuh so:'). Although the exact form of the rejection is difficult to hear, it is clearly a rejection. And, it is subsequent to this that,

at line 5, Andy returns to his description as evidenced by his reuse of the language he had used at the outset of line 4: 'I'm talkin' about' (Schegloff 1987).

It is at this point that gesture comes to play a particularly important role in this sequence. The still below shows the participants just following the head shake immediately prior to Tim's anticipatory completion in line 3. It is from this posture that, at line 5, Andy moves up to a gesture of clenched fists — apparently iconic of the description to follow. This is shown in Figure 3 below.

What is particularly interesting about this example has to do with the location of the gesture and what it is used to do. It is sequentially positioned at a place just prior to a third effort to reach a proper descriptor of the pit bull (the first being at line 2; the second following the failed anticipatory completion in line 4). Secondly, it is positioned following a first anticipatory completion (Lerner 1991, 1996), by Andy's primary recipient Tim, which has been rejected. Both dimensions of this gesture's positioning appear important. This position may communicate that it has been added as an additional resource for Tim, and he makes use of it. However, this second of Tim's attempts also fails as the two participants produce divergent descriptions in overlap. While Andy produces 'so cut' (a

5		I'm talkin' a	bout	[so cu:t.=
6	AND:		((fists clenched))	[
7	TIM:			[muscled up.

Figure 3.

term which is about the definition of the muscles), Tim produces 'muscled up' (a term about the bulk of the muscles).

AND: I'm talkin' a[bout [so cu:t.[7]=
AND: [((fists clenched)) [
TIM: [muscled up.

In this example a gesture is used after a progressivity failure. In this sequential position, a gesture may be designed to help co-participants anticipate what is coming next and as such provide additional resources for recipients to collaboratively complete utterances. In this case, although the gesture does appear to be designed in just this way, it fails. We may speculate that this failure results from a mismatch between what the gesture appears to represent and the description ultimately provided by Andy. That is, the clenching of the fists appears to represent musculature or strength. Coupled with the vocal resources Andy has provided (discussed earlier) and the rejection of vicious as a descriptor, Tim's anticipatory completion 'muscled up' seems on target. It moves from behavioral to corporeal but is nonetheless an extreme characterization, one that would be consistent with the person who had the dog being Mexican, the scene being on a beach and thus a publicly visible and notable event, the breed of the dog being a pit bull, and the head shake. Laminated on top of all of that is the gesture. That Tim's anticipatory completion is wrong only further highlights the use he made of the gesture since the gesture served to restrict the range of descriptors which would be consistent with all of the other projections. And further, Tim treats that gesture as inviting another effort at anticipatory completion in that he begins this immediately subsequent to the gesture.

Just following the overlapping turn completions, Tim receipts Andy's completion with 'whoa.' When anticipatory completions are successful it is usually the original speaker/teller who confirms the anticipatory completion (Lerner 1987). But here the interlocutor, having been incorrect (again), displays an orientation to this by immediately receipting and treating as news Andy's completion ('so cut') and thus his description. Note that if the completion had been successful (i.e., the same as Andy's descriptor) this would have likely had different implications for how the sequence would have been completed.

The gesture clearly elaborates the semantic content of the co-occurring talk. But this gesture, like many others, seems to have a more obviously interactive purpose. As seen in the example we have examined here (also see papers by Hayashi and Sidnell in this volume) a gesture projects talk that is yet to be articulated and thus can is available to participants as a resource for anticipating just what that projected talk will be. More

instances would need to be investigated to explore whether the use of gesture in this position might have to do with either or both the progressivity failure and/or the unsuccessful anticipatory completion. But both of these dimensions appear important for understanding this case. For this reason and others, both the visuospatial and vocal modalities provide important resources in the collaborative production of emergent turns-at-talk.

5. Discussion

In this introduction we have provided a select review of the work that has been done that is most relevant to multimodality research on interaction of the type included in this volume. As will be overviewed in the final section of this introduction, the papers cover a broad range of current approaches to multimodal interaction research. One area that we found to be somewhat under-represented in the literature is work which relies equally on sequence structural vocal resources and visuospatial resources in order to provide us with further insight into how the modalities work together *structurally* (cf. C. Goodwin 1979 as an illustration of research that does integrate structural vocal resources and visuospatial resources). If we are to move towards a theory of social interaction, we will need to understand not only how the vocal modality works but how the different channels and modalities work together as well as the mechanics that underlie such co-operation. The brief look at a single case here provides evidence that when sequence structure and gesture are looked at together, we may gain insight into how interaction is organized.

6. The papers

The papers by Hayashi and Sidnell consider the use of gesture and gaze in conversational interaction. Hayashi looks specifically at the ways in which illustrators project talk that has yet to be articulated and thus become resources in co-participants efforts to produce anticipatory completions. Sidnell considers the use of illustrators and points in references to persons. He suggests that these gestures are organized in relation to the unfolding sequences of talk of which they are a part. Murphy's paper examines the use of gesture by a group of architects who collectively imagine the specific characteristics and properties of a building that has yet to be constructed. In his analysis, gestures are used to bring a third dimension to the building drawings. Moreover the gestures are used to convey a sense of how users of the building might experience it. Becvar examines talk and gesture in a biochemistry teaching lab. Here gestures are used to talk about and represent the shape, movement and dynamic character

of theoretical constructs crucial to the work being done. Gestures come to circulate in the lab and embody a theory about the behavior of proteins. Phillabaum considers the integration of gesture, talk and gaze in a photography class. His analysis shows that learning to see the technical details constitutive of the photographer's work is the result of multimodal interaction between teacher, student and the photographs themselves. Alač's paper examines the coordination of gesture and talk in the work of examining brain images. Finally, Thompson, Graham, and Russo's distinctive contribution on music performance illustrates ways in which studies of gesture and facial expression can be extended to other forms of human interaction.

Notes

1. There are, of course, exceptions including for instance early work by the Goodwins (C. Goodwin 1979, 1984; M. H. Goodwin 1980) and Heath (1986) as well as a paper by Sacks and Schelgoff (2002 [1975]). Other early work from a rather different perspective includes Kendon (1972, 1975, and 1980), Pittenger (1960), and Birdwhistell (1972). Clearly, the availability of video technology has had a massive impact in encouraging researchers to consider the multimodal character of interaction.
2. For discussion see Enfield (2004, 2005) and Meier, Cormier, and Quinto-Pozos (2002).
3. The reader will note that each of thee claims is qualified ('often,' 'may,' 'typically,'). Such qualification is necessary since there is no one-to-one mapping between surface form and interactional function. For instance, turn-initial 'well' may be used also to resume a sequence which has been stalled and diverted, 'okay' may be used to initiate a sequence (as when a teacher enters the room and announces 'okay!').
4. Gesture has been shown to play a very wide range of communicative roles and functions. Earlier research often assumed a functional categorization of gesture. Ekman and Friesen for instance write that 'facial and body behavior involve a number of quite different kinds of behavior which will be described in terms of five categories distinguished by particulars of usage, origin and coding' (Ekman and Friesen 1969: 63). Ekman and Friesen then go on to describe the differences between emblems, illustrators, affect displays, regulators, and adaptors. Such categorizations are problematic since functions are not discrete. The analyses of Hayashi and Sidnell (the volume) indicate that gestures typically thought of as illustrators can have a clear regulative function. For recent discussions of gesture see Haviland (2004), C. Goodwin (1986), and Enfield (2004).
5. Unlike other forms of reported speech, the original speaker is not specified and thus air-quotes are subject to various kinds of inference which make possible certain extensions — e.g., satire, a diffuse source etc. It is this that allows the gesture to be used in a way that could be translated as 'so-called.' Ekman and Friesen write:

> ... emblems are those nonverbal acts which have a direct verbal translation, or dictionary definition, usually consisting of a word or two, perhaps a phrase. This verbal definition or translation of the emblem is well known by all members of a group, class or a culture ... An emblem may repeat, substitute, or contradict some part of the concomitant verbal behavior; a crucial question in detecting an emblem is whether it could be replaced with a word or two without changing the information conveyed. (Ekman and Friesen 1969: 63)

6. The phenomenon which linguists term 'co-articulation' represents an exception to this overwhelmingly accurate generalization.

References

Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse Processes* 15, 469–489.

Beach, W. A. (1993). Transitional regularities for casual 'okay' usages. *Journal of Pragmatics* 19, 325–352.

Birdwhistell, R. (1972). A kinesic-linguistic exercise: The cigarette scene. In *Directions in Sociolinguistics*, J. H. Gumperz and D. Hymes (eds.), 381–404. New York: Holt Rienhart.

Butterworth, B. and Beattie, G. W. (1978). Gesture and silence as indicators of planning in speech. In *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, R. Campbell and P. Smith (eds.), 347–360. New York: Plenum.

Ekman, P. and Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1, 49–58.

Enfield, N. J. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture* 1 (2), 185–212.

—(2003). Producing and editing diagrams using co-speech gesture: Spatializing nonspatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology* 13 (1), 7–50.

—(2004). On linear segmentation and combinatorics in co-speech gesture: A symmetry-dominance construction in Lao fish trap descriptions. *Semiotica* 149 (1/4), 57–123.

—(2005). The body as a cognitive artifact in kinship representations: Hand gesture diagrams by speakers of Lao. *Current Anthropology* 46 (1), 1–26.

Ford, C. E. and Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In *Interaction and Grammar*, E. Ochs, E. A. Schegloff, and S. A. Thompson (eds.), 134–184. Cambridge: Cambridge University Press.

Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gathering*. New York: Free Press.

—(1980). Footing. In *Forms of Talk*, 124–159. Pennsylvania: University of Pennsylvania Press.

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help us Think*. Cambridge: Belknap Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In *Everyday Language: Studies in Ethnomethodology*, G. Psathas (ed.), 97–121. New York: Irvington Publishers.

—(1981). *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

—(1984). Notes on story structure and the organization of participation. In *Structures of Social Action: Studies in Conversation Analysis*, John M. Atkinson and John C. Heritage (eds.), 225–246. Cambridge: Cambridge University Press.

—(1986). Gesture as a resource for the organization of mutual orientation. *Semiotica* 62 (1/2), 29–49.

—(1988). Participation frameworks in children's argument. In *Growing Into A Modern World: Proceedings from An International Interdisciplinary Conference on the Life and Development of Children in Modern Society*, K. Ekberg and P. E. Mjaavatn (eds.), 1188–1195. Trondheim: Norwegian Centre for Child Research.

—(2000). Action and embodiment within situated human interaction. *Journal of Pragmatics* 32, 1489–1522.

—(2002). Time in action. *Current Anthropology* 43 (supplement).
Goodwin, C. and Goodwin, M. H. (1987). Concurrent operations on talk: Notes on the interactive organization of assessments. *IPRA Papers in Pragmatics* 1, 1–55.
—(2004). Participation. In *A Companion to Linguistic Anthropology*, Alessandro Duranti (ed.), 222–244. Oxford: Blackwell.
Goodwin, M. H. (1980). Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry* 50 (3/4), 303–317.
—(1983). Searching for a word as an interactive activity. In *Semiotics 1981*, J. N. Deely and M. D. Lenhart (eds.), 129–138. New York: Plenum Press.
—(1996). Informings and announcements in their environment: Prosody within a multiactivity work setting. In *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 436–461. Cambridge: Cambridge University Press.
Goodwin, M. H. and Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–75.
Haviland, J. (2004). Gesture. In *A Companion to Linguistic Anthropology*, A. Duranti (ed.), 197–221. Oxford/Malden: Blackwell.
Heath, C. (1982). The display of recipiency: An instance of sequential relationship between speech and body movement. *Semiotica* 42 (2/4), 147–161.
—(1986). *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press.
—(1992). Gesture's discreet tasks: Multiple relevances in visual conduct and in the contextualization of language. In *The Contextualization of Language*, P. Auer and A. di Luzio (eds.), 101–127. Amsterdam: Benjamins.
Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In *Structures of Social Action*, J. M. Atkinson and J. Heritage (eds.), 299–345. Cambridge: Cambridge University Press.
—(2002). The limits of questioning: Negative interrogatives and hostile question content. *Journal of Pragmatics* 34, 1427–1446.
Kendon, A. (1972). Some relationships between body movement and speech. In *Studies in Dyadic Communication*, Aaron W. Siegman and Benjamin Pope (eds.), 177–210. Elmsford, NY: Pergamon Press.
—(1975). Gesticulation, speech and the gesture theory of language origins. *Sign Language Studies* 9, 349–373.
—(1980). Gesticulation and speech: Two aspects of the process of utterance. In *Relationship Between Verbal and Non-verbal Communication*, M. R. Key (ed.), 207–227. The Hague: Mouton.
—(1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.
—(1994). Do gestures communicate?: A review. *Research on Language and Social Interaction* 27 (3), 175–200.
—(1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics* 23, 1–31.
—(2002). Some uses of the head shake. *Gesture* 2 (2), 147–182.
—(2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
Kendon, A. and Versante, L. (2003). Pointing by hand in 'Neapolitan'. In *Pointing: Where Language, Culture and Cognition Meet*, S. Kita (ed.), 109–137. Mahwah, NJ: Erlbaum.
Kita, S. and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48, 16–32.

Lerner, G. (1987). Collaborative turn sequences: Sentence construction and social action. Ph.D. dissertation, University of California (Irvine).
—(1991). On the syntax of sentences in progress. *Language in Society* 20, 441–458.
—(1996). On the 'semi-permeable' character of grammatical units in conversation: Conditional entry into the turn-space of another speaker. In *Interaction and Grammar*, E. Ochs, E. A. Schegloff, and S. Thompson (eds.), 238–276. Cambridge: Cambridge University Press.
—(2003). Selecting next speaker: The context sensitive operation of a context-free organization. *Language in Society* 32, 177–201.
Levinson, S. C. (1983). *Pragmatics.* Cambridge: Cambridge University Press.
—(1988). Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation. In *Erving Goffman: Exploring the Interaction Order*, P. Drew and A. Wooton (eds.), 161–227. Boston: Northeastern University Press.
Local, J. K. (1996). Conversational phonetics: Some aspects of news receipts in everyday talk. In *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 177–230. Cambridge: Cambridge University Press.
Local, J. K., Kelly, J., and Wells, W. H. G. (1986). Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics* 22, 411–437.
Local, J. K., Wells, W. G., and Sebba, M. (1985). Phonology for conversation: Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9, 309–330.
McNeill, D. (1992). *Hand and Mind.* Chicago: University of Chicago Press.
Meier, R. P., Cormier, K., and Quinto-Pozos, D. (eds.) (2002). *Modality and Structure in Signed and Spoken Languages.* Cambridge: Cambridge University Press.
Mondada, L. (2004). Temporalité, sequentialité et multimodalité au fondement de l'orginisation de l'interaction : Le pointage comme pratique de prise du tour. *Cahiers de Linguistique Française* 26, 169–192.
—(2005). Multimodal resources for turn-taking: Pointing and the emergence of the next speaker. Paper presented at a Colloquium, Nijimegen, the Netherlands.
Morrel-Samuels, P. and Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18, 615–623.
Müller, F. E. (1996). Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 131–176. Cambridge: Cambridge University Press.
Pittenger, Robert E. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis.* Ithaca, NY: P. Martineau.
Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In *Structures of Social Action: Studies in Conversation Analysis*, J. M. Atkinson and J. Heritage (eds.), 57–101. Cambridge: Cambridge University Press.
Robinson, J. D. and Stivers, T. (2001). Achieving activity transitions in primary-care encounters: From history taking to physical examination. *Human Communication Research* 27 (2), 253–298.
Sacks, H. and Schegloff, E. A. (1979). Two preferences in the organization of reference to persons and their interaction. In *Everyday Language: Studies in Ethnomethodology*, G. Psathas (ed.), 15–21. New York: Irvington Publishers.
—(2002 [1975]). Home position. *Gesture* 2 (2), 133–146.
Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.
Schegloff, E. A. (1972). Notes on a conversational practice: Formulating place. In *Studies in Social Interaction*, D. Sudnow (ed.), 75–119. New York: Free Press.

—(1979). The relevance of repair for syntax-for-conversation. In *Syntax and Semantics 12: Discourse and Syntax*, T. Givon (ed.), 261–288. New York: Academic Press.
—(1984). On some gestures' relation to talk. In *Structures of Social Action*, J. M. Atkinson and J. Heritage (eds.), 266–296. Cambridge: Cambridge University Press.
—(1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In *Talk and Social Organisation*, G. Button and J. R. E. Lee (eds.), 70–85. Clevedon: Multilingual Matters.
—(1996). Some practices for referring to persons in talk-in interaction: A partial sketch of a systematics. In *Studies in Anaphora*, B. Fox (ed.), 437–485. Amsterdam: John Benjamins.
—(1998). Body torque. *Social Research* 65 (3), 535–596.
—(2000a). Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29 (1), 1–63.
—(2000b). On granularity. *Annual Review of Sociology* 26, 715–720.
—(2004). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. Paper presented at the Wenner-Gren conference on the Interactional Foundations of Society, Duck, North Carolina.
—(in press). *A Primer for Conversation Analysis: Sequence Organization*. Cambridge: Cambridge University Press.
Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382.
Schegloff, E. A., Ochs, E., and Thompson, S. (1996). Introduction. In *Interaction and Grammar*, E. Ochs, E. A. Schegloff and S. A. Thompson (eds.), 1–51. Cambridge: Cambridge University Press.
Selting, M. (1996). Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished' questions in repair initiation. In *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 231–270. Cambridge: Cambridge University Press.
Sidnell, J. (forthcoming). Language games and the diversity of signs. In *The Semiotics of Language and Writing*, A. A. Iannucci and M. Danesi (eds.). Madison: Atwood Press.
Streek, J. (1993). Gesture as communication: Its coordination with gaze and speech. *Communication Monographs* 60 (4), 275–299.
Stivers, T. (2004). 'No no no' and other types of multiple sayings in social interaction. *Human Communication Research* 30 (2), 260–293.
Wells, B. and Peppé, S. (1996). Ending up in Ulster: Prosody and turn taking in English dialects. In *Prosody in Conversation*, E. Couper-Kuhlen and M. Selting (eds.), 101–130. Cambridge: Cambridge University Press.

Tanya Stivers (b. 1970) is a Staff Scientist at the Max Planck Institute for Psycholinguistics ⟨Tanya.Stivers@mpi.nl⟩. Her research interests are in the structure of social interaction. Her recent publications include 'Negotiating who presents the problem: Next speaker selection in pediatric encounters' (2001); '"No no no" and other types of multiple sayings in social interaction' (2004); 'Parent resistance to physicians' treatment recommendations: One resource for initiating a negotiation of the treatment decision' (2005); and 'Modified repeats: One method for asserting primary rights from second position' (2005).

Jack Sidnell (b. 1969) is an Associate Professor of Anthropology at the University of Toronto ⟨jack.sidnell@utoronto.ca⟩. His research interests include conversation analysis, social interaction, and Wittgenstein. His recent publications include 'Conversational turn-taking in a Caribbean English Creole' (2001); 'Constructing and managing male exclusivity in talk-in-interaction' (2003); 'An ethnographic consideration of rule-following' (2003); and '"There's risks in everything": Extreme case formulations and accountability in inquiry testimony' (2004).