

Revised paper submitted to  
*American Journal of Sociology*

**Solidarity and Conflict: Prosody as a Transactional Resource in Intra-  
and Intercultural Communication Involving Power Differences**

Wolff-Michael Roth, *University of Victoria*

Kenneth Tobin, *The Graduate Center, City University of New York*

RUNNING HEAD: Solidarity and Conflict

## Solidarity and Conflict: Prosody as a Transactional Resource in Intra- and Intercultural Communication Involving Power Differences

### ABSTRACT

This ethnographic study of teaching and learning in urban high school science classes investigates the ways in which teachers and students talk, gesture, and use space and time to produce the emotional climate necessary to be successful. In situations where teachers coteach as a means to learn to teach in inner-city schools, successful teacher-teacher collaborations are characterized by prosodic expressions that converge over time, adapting to match the prosodic parameters of students' talk; unsuccessful collaborations are associated with considerable differences in pitch between consecutive speakers participating in turns-at-talk. Situational conflicts are co-expressed by increases in pitch levels, speech intensities, and speech rates; and conflict resolution is accelerated by the coordination of pitch levels.

*Keywords:* Prosodic variation; Power; Conflict; Emotions

Conversations in and oriented toward the concrete realization of cultural-historically developed forms of activities constitute the crucial location where individual members contribute to producing and reproducing society as we know and experience it in our everyday lives. That is, society comes to live in and through talk in institutions and institutional talk (Boden 1994). In this, face-to-face encounters and the conversations they give rise to are important because they are the sites that mediate between individual and collective emotions that serve as the fuel for social life (Turner 2002). Face-to-face encounters incorporate interaction rituals, “in which participants develop a mutual focus of attention and become entrained in each other’s bodily micro-rhythms and emotions” (Collins 2004, p. 47).<sup>1</sup> During conversations emotions are co-articulated and co-communicated; and they therefore can be transmitted to new generations using *explicit* emotional talk, which includes facial (Schimmack 1996) and other bodily displays (Goodwin, Goodwin and Yaeger-Dror 2002) and, important here, a variety of prosodic means (Pittam and Scherer 1993). This suggests that the outcomes of (conversational) actions may be pragmatic resources for co-articulating and co-communicating the emotions that make face-to-face transactions possible and intelligible.

In this study, therefore, face-to-face conversations are taken to be mesolevel sites where macrolevel structures are produced and reproduced as participants draw on microlevel resources that mediate the degree of success of transactions. Both verbal and nonverbal means are deployed in face-to-face communication to distinguish group membership from non-membership, such as, for example, in the gay culture where young and older gay men discriminate against each other by means of comportment (e.g., flaunting sexuality, body features) and choice of words (e.g., “gym rat”) (Hajek and Giles 2002). Some research shows that the speakers’ culturally contingent accents are part of the discriminative means (Giles et al. 1995) and yet, though they may mediate the affect of communication participants, accents in and

---

<sup>1</sup> From a phenomenological perspective, human beings come to recognize emotions because of the alterity of the Self rather than because they cognize these within themselves and then transfer this recognition to the emotion of others (Franck 1981).

by themselves do not communicate affect. How discrimination occurs still is not well articulated theoretically; but the authors suggest hypotheses that communicated affect has to be included in the theoretical mechanisms that provide explanations for differentiation to the outside and solidarity and bonding on the inside of group differences.

In explaining such differentiation and cohesion, one needs to understand that conversation participants not only are oriented towards their current context but also to macrolevel structures, even though they may not talk *about* them. For example, a teacher and her students participating in a science class concerning chemical valence of certain elements ask questions, articulate models, argue over differing perspectives, and use resources such as gestures toward a periodic table to support their viewpoints. In so doing they enact familiar roles of teacher and student, and reproduce power differences that reflect race, gender, and social class, even though a transcript of the unfolding verbal transactions shows that the focus remains on chemical valence for the entire time. On the other hand, a teacher walking into the principal's office or into a school board office will act differently, and this difference is related to the different ways in which the current field is embedded in the societal macrostructure generally and the structure of educational systems in particular (e.g., hierarchy, power). The observable negotiations about "who is in the know" and what constitutes legitimate knowledge for the task at hand—the struggle for establishing differences in power and status—are mediated by emotions that give rise to a sense of solidarity or conflict (Kemper and Collins 1990). There is strong evidence that prosody generally and pitch ( $F_0$ ) levels and pitch contours specifically provide resources that enable and constrain the trajectory of a transaction, including the establishment of power and status (Gregory 1999). Also, the production of social alignment—Durkheim's "social similitude"—through alignment of prosody and speech amplitude has been reported to occur in dyadic interviews (Gregory, Webster and Huang 1993) and in multi-activity work settings, allowing participants to disentangle the voices pertaining to different conversations in the same setting (M. H. Goodwin 1996).

Whereas there are considerable studies of face-to-face transactions and prosody, there are few studies on the pragmatic use of prosody as part of a dynamic social structure that frames everyday encounters; in particular, there is a dearth of research concerning the use of prosody in situations characterized by institutional differences in power crossed with other differences of sociological import such as gender and culture. The purpose of the present study is to report on the role of prosody in intra- and intercultural communication in science classes taught in low performing, inner city high schools. In this study we provide evidence from the analysis of ethnographic data and prosody that (a) successful lessons—as indicated by outcome measures such as quizzes and tests—are associated with the production and reproduction of prosodic alignment, (b) conflict and resolution in classroom transactions are correlated with increases and decreases of pitch registers, and (c) prosody leads to entrainment phenomena not only at the level of sound frequencies but also in a variety of other physical movements (rocking legs, head movements, tapping with hands). Combined, these three dimensions lead to a better understanding of the resources used in the production and reproduction of successful and unsuccessful lessons; and it provides a better understanding of how new teachers become entrained into the ways of successful teachers when the former work “at the elbows” of the latter. Ultimately, therefore, our study shows not only how schooling is reproduced but also how—at the heart of an often-dysfunctional educational system—positive emotional climate is produced, providing an underpinning for successful teaching and learning.

#### THE MICROSOCIOLOGY OF FACE-TO-FACE ENCOUNTERS

Consistent with the definition provided by the Merriam-Webster Online Dictionary and its usage in the research literature (e.g., Barab and Roth 2006, Corno et al. 2002), we use the term *transaction* rather than *interaction* to emphasize the fact that the different actors and situations cannot be modeled independently. Rather, a transaction, like all dynamical situations, is better understood as “a communicative action or activity that involves two parties that reciprocally

affect or influence each other.”<sup>2</sup> When verbal exchanges occur in a cultural field, a great deal can be at stake and participants likely have multiple goals that are salient concurrently. For example, if an African American student chooses to describe to the whole class a model for chemical valence she is probably concerned with more than testing whether or not her model is robust. Earning and maintaining the respect of her peers is usually a central concern (Anderson 1999); and oppositional cultures and identities are expressed in linguistic codes (Smardon 2004). Accordingly, if and when challenges to a teacher’s articulated model arise, the prosody of her talk is likely to reflect her goals of earning and maintaining respect as much as, or even more than, illustrating that her knowledge of all chemistry is robust (Roth and Middleton 2006). At the same time communication involves much more than what is said (semantic content). How it is said is important too. Hence, transactions are sites for the production of social capital in that material and schematic structures dynamically unfold, simultaneously structuring agency and passivity. During transactions, participants therefore accrue capital consciously and unconsciously—they are empowered to act and they are positioned without being consciously aware of it—or they may lose face or they *simultaneously* may gain social capital among their friends and lose it with respect to the school and schooling (Roth et al. 2004).

#### *MICROSOCIOLOGY*

In this study, we take the same theoretical point of departure that characterizes other researchers of cultural-historical activities. Accordingly, the relevant unit for understanding human actions, including the analysis of prosody, discourse, and emotion is not the individual or language (as a semantic system) but the sequential organization of actions that produce and reproduce social life (Goodwin and Goodwin 2000). Early anthropological work suggests that when persons engage each other, they tend to regulate, synchronize, and adapt their conduct—that is, they act consciously in relation to one another (Hall 1976). Sociologists, too, have held

---

<sup>2</sup> If such situations are modeled mathematically, then it is impossible to write independent differential equations but rather there are sets of equations with shared parameters.

that transaction participants tend to harmonize their actions both at macroscopic (observable) and microscopic (unconscious) levels: “established by the reciprocal sharing of the Other’s flux of experience in inner time, by living through a vivid present together” (Schutz 1971, p. 173), which leads to an experiencing of the togetherness as a “We.” In his classic article on making music together, Schutz uses a metaphor of a band establishing a “mutual tuning-in relationship” as the members warm up establishing a beat and rhythmic entrainment. Human emotions regulate the “focus and flow of encounters with others, while developing commitments to the culture and structure of corporate and categoric units and the macrolevel institutional systems built from these mesolevel structures” (Turner 2002, p. 76).

Emotions are the key to increasing sociality and bonding and as participants orient to ongoing events within an activity, they afford the micro- and macrolevels of society becoming integrated. Emotions are constitutive of participants’ orientations, the conscious goals they pursue, and the unconscious operations that realize these goals. A subject takes up a position in the world of his or her meaning, and the various forms of expression are ways to articulate this positioning for others. “The phonetic ‘gesture’ brings about, both for the speaking subject and for his hearers, a certain structural co-ordination of experience, a certain modulation of existence, exactly as a pattern of my bodily behavior endows the objects around me with a certain significance both for me and others” (Merleau-Ponty 1945, p. 225). But speakers always take up position in a world that already is shared with others, so that force of speech is something real that comes to speakers from the audiences to which they are oriented and whose solidarity therefore is of a “mechanical sort” (Durkheim 1893, chap. 3); various non-conscious and unconscious<sup>3</sup> aspects of communicative actions are possible carriers of social alignment and solidarity (Gregory et al. 1993). Meaning therefore neither is in the word nor in the gesture: word

---

<sup>3</sup> Some aspects of communication are non-conscious, which means they could be conscious, such as nodding the head while listening, a form of “back-channeling” to which we normally do not attend to but which may be raised to conscious awareness. Some aspects of communication inherently are unconscious, and cannot be raised to conscious awareness, such as intentions and the particular words that we come to use without consciously selecting them.

and gesture are one-sided and even contradictory expressions of some higher-level meaning unit that sublates (both overcomes and preserves) all of its expressions. Emotions constitute an integral part of this higher-order meaning-making unit, which many researchers working in the spirit of Soviet social psychology refer to as the “activity system” (e.g., Roth and Lee 2007). Accordingly, this system oriented toward societal motives subsumes thinking, speaking, acting, gesturing, emoting, and so on.

*SOCIOLOGY, EMOTIONS, PROSODY*

Emotions are understood as complex response dispositions for engaging in certain classes of adaptive conduct. The different emotions are characterized by distinctive states of physiological arousal—negative and positive feelings associated with affective states, stimulation, and patterns of expressive reactions. For example, fear and anger are said to energize persons to engage in urgent actions when they are facing physical dangers and bodily harm or threats from other people; and satisfaction not only allows people to rest but also to strive for important goals (Kemper 1987). In both these examples, emotions (a) have a short-term, momentary regulative function of facilitating the unconscious elements that constitute goal-oriented actions and (b) are cultural resources—which we understand in terms of their role in a structure | agency dialectic (Sewell 1992)—that mediate the selection of long-term goals and participation in particular forms of societal activities. Even the most rational and calculative individual, transacting with others because of the potential payoffs, is bound by emotions in at least two ways: first, there is a two-way back channeling of affect that mediates any encounter; and second, the calculation of payoffs is mediated by the emotional valences attached to the outcomes (Damasio 1994/2000). Face-to-face encounters not only presuppose emotions but also produce them. For example, in power/status models of emotional mediation of micro-transactions, the activation of negative or positive emotions is related to whether anticipated relative status or power relations are realized (Kemper and Collins 1990). A person may become defensive to negative responses from others, and activate a defense mechanism because the self is threatened. These emotions not only drive



but also exhibit themselves in face-to-face encounters and thereby constitute structural resources and constraints available to the agency of any or all participants in an encounter. That is, we (the authors) see in emotion an aspect of human culture that squarely fits within existing approaches to cultural sociology (cf. Sewell 1999).

There is considerable psychological research on the recognition of emotion. One meta-analytic study—involving 87 articles, 97 studies, and 182 independent samples—shows that emotions are recognized above chance within and across cultures (Elfenbein and Ambady 2002) and with different levels of accuracy across channels (e.g., modes such as happiness and anger). For example, happiness is most accurately identified in facial expression, and anger is the emotion most accurately recognized in the voice. From our cultural-historical perspective, extant studies have two major shortcomings: First, they do not address irremediably contingent, ongoing praxis; and second, they are considered independent not only of the current activity but also of social action. These limitations do not acknowledge that since the early part of the last century emotions have been recognized as *constitutive* of societal, motive-driven activities, individual goal-directed actions, and unconscious, contextually conditioned operations (e.g., Bakhtin 1993; Weber 1994). This early work is supported by recent neurophysiologic research in which the incidence of inappropriate behavior was shown to be related to particular forms of frontal lobe damage to the pathways linking emotions to (social, material) actions (Damasio 1994/2000). Thus, emotionality also is linked up with the goal-formation of individuals who participate in societally mediated human activities that embody collective needs.

From a phenomenological perspective of the individual social actor, too, emotions are integral to the way in which human beings orient to the (social, natural) world, so that the ethical relation to others presents itself under the light of an affective multivalence where the emotional relief never is given a priori but always is the result of a continually renewed appreciation of the situation (Depraz 2006). Emotions are the very foundation of alterity, defining at once some of the most fundamental properties of social systems and some of the differences that define social structure (Scheler 1933). But this world is not independent from the acting individual, as is clear

from the widely used concept of the person-specific lifeworld to which individual persons orient. Bringing cultural-historical and phenomenological perspectives together has led some psycholinguists to adopt a dialectical perspective, according to which, in the case of one research group, speech and gesture are two one-sided, but mutually constitutive aspects of the ongoing communicative activity as a whole (e.g., McNeill 2002). More recent work has expanded the unit of analysis to include other semiotic resources continuously produced and reproduced in situation, including structures in the local setting, body orientations, and prosody (Roth and Pozzer-Ardenghi 2006).

#### *PROSODY, POWER AND CONTROL*

It is well established that speech carries—in addition to linguistic content—information about the speaker's intentions and emotional state and that listeners are capable of perceiving this information without having to stop and reflect about it (Scherer 1989). The fundamental frequency of phonation (in the literature referred to as *pitch* or  $F_0$ ), speech intensity, and speech rate individually and in combination constitute information about the current emotional state of the speaker. Pitch related parameters, including short-term perturbations, long-term variability, and mean value, are among the measures often reported to correlate with levels of speaker emotional stress, either task-induced or in real-life emergencies (e.g., Roth 2007). Intact pitch information, including gross changes and fine temporal structure, is crucial for the accurate identification of emotions (Protopapas and Lieberman 1997). There are indications that mean and maximum pitch levels and pitch contours are the most salient pitch-related measures that convey emotional information. Anger in particular appears to be encoded through the fundamental frequency (pitch), though speech intensity often correlates with anger (Pittam and Scherer 1993), such as in shouting. The research suggests that anger also is associated with increases in the incidence of high-frequency speech components, downward-directed  $F_0$  contours, and increasing articulatory rates.

Prosody is a transactional resource for arguing and negotiating differences. In a study of children playing hopscotch, prosodic features such as increased loudness and significant increases in pitch were resources for articulating opposition and difference (Goodwin, Goodwin, and Yaeger-Dror 2002). Opposition was marked by extended vowel length, raised pitch on negatives, and distinctive pitch contours. There were differences between African American and Latina girls, whereby the latter used dramatic intonation contours and the former used less extreme pitch maxima, less pronounced pitch contours, and less durational expansions.

Particular phenomena of societal transactions are associated with prosodic features of communication. Thus, abrupt-joins—i.e., within-turn changes in the trajectory of conversational topics—are associated with increased pitch, increased speech rate (temporal compression), continuity in articulation and close temporal proximity, and increased loudness across the pre-abrupt-join and post-join talk (Local and Walker 2004). Pragmatically, the abrupt-join constitutes work, the outcomes of which are resources that secure a speaker space for more talk beyond the completion of the unit with the pre-abrupt-join utterance. As a result, a speaker can achieve a multi-unit turn at talk. The prosodic features also can contribute to securing the shift in the transition to another topic within the same turn at talk.

One study shows how during the production and reproduction of interviews—involving a variety of celebrities and CNN's journalist Larry King—differences in the power and status of the interviewee are associated strongly with convergent features of their pitches ( $F_0$ ) (Gregory 1999). Interviews rated (by undergraduate students) as high in power and potency—on the basis of such factors as loudness, dominance, toughness, aggressiveness, hardness, activity, strength, and intensity—also show the highest convergence values with respect to the pitch levels of the participants' speech utterances. These results suggest that as the interviewer and interviewee adjust to relative power or status differences, their  $F_0$  (pitch) levels converge. It appears that pitch levels and contours contain important social information that transaction participants use as resources for conducting activities, and accommodation to status differences. In a similar vein, the analyses of tape-recorded interviews with enlisted airmen show not only alignment of

prosodic parameters, most notably pitch levels, but also an increase in the quality of the interviews, which generally was associated with the dominance of one interview partner (Gregory et al. 1993).

## METHODS

In this study, our aim is to understand how certain social events (i.e., science lessons of different degree of success) arise from face-to-face transactions (Turner 2002) and interaction ritual chains (Collins 2004). We follow others interested in the emergence of sociological features from mundane face-to-face transactions in combining careful ethnographic work with conversation analysis (Schegloff 1996) with a specific extension of this work through the computer-aided analysis of prosodic features (M. H. Goodwin 1996). The data for our study derive from a 7-year ethnographic effort to understand teaching, learning, and learning to teach in an innovative approach of preparing teachers for working in inner-city neighborhood schools. In the particular situation, we investigated coteaching, an organization of the usually solitary teaching job such that two or more teachers take full responsibility for all aspects of teaching a particular course. In our case, coteaching was used at an Ivy League school as the model for science teacher preparation. New teachers, as the teacher interns are referred to, were assigned to not-too-distant inner-city schools, and in especially large numbers to City High, an urban school located near our university in a large city of the northeastern US, where they taught in configurations from two to four, sometimes on their own, but often in direct arrangements with a resident teacher.

### *INSTITUTIONAL AND ETHNOGRAPHIC CONTEXT*

City High has more than 2,000 students, 98 percent of whom are of African American descent and more than 90 percent are from poverty-stricken or working class families. City High is organized into small learning communities, schools within the school, each including about 200 students and organized around a different core idea that organizes the curriculum (e.g.,

health, sports, or science and technology). The average daily attendance rate is 72 percent, that is, on any given day more than one-quarter of the students are not at school for a variety of reasons often directly related to poverty and other home conditions. Achievement on state-level examinations is consistently below average.

The research at City High, which is part of a larger study of science education in urban high schools, commenced in 1998 when a relatively small group consisting of the authors of this paper and several graduate students began to study the teaching and learning of science in urban high schools. The research group was dynamic, consisting of a core of university- and school-based researchers and others who joined the group for periods ranging from weeks to as long as a year. Steadily the size of the research group grew to include different categories of investigators—university professors, visiting scholars, post-doctoral scholars, doctoral students, and numerous school-based researchers including resident teachers, new teachers, high school students, and, in one instance, a school administrator. The methods employed included ethnography and microanalysis using digitized video.

#### *LEARNING TO TEACH THROUGH COTEACHING*

Coteaching is a method to induct new teachers into teaching by working at the elbow of a more senior teacher. This study was designed to understand the microprocesses that lead to successful and unsuccessful coteaching, where we generally observe fluidity after the participants have worked closely together for several months.

#### *DATA SOURCES*

In this interdisciplinary study, we draw on a variety of qualitative research methods appropriate in school contexts, including ethnography, conversation analysis, and microsociological approaches to studying social practices. In addition to writing the usual observational, methodological, and theoretical field notes, we videotape lessons and dialogue with the teachers and students, interview students and (new) teachers, audiotape interviews

conducted by high school student research assistants among their peers, and collect the teaching-related discussions new teachers held using an online internet forum. Teachers often are equipped with recorders to ensure that their talk is captured as they move about in the classroom and we also place recorders on various student desks to allow us to record as many contributions to whole-class conversations as possible.

In this article we draw our specific examples from the studies of two sets of new teachers assigned in consecutive years to coteach at City High with the same resident science teacher who, in this paper, is referred to by the pseudonym Alex. Of Cuban-African origin, Alex was in his second and third years of teaching at City High—years in which he was regarded by students and colleagues as an effective science teacher. However, he was not always effective in this way. During his first year at City High Alex experienced significant difficulties in teaching his classes despite five years of successful science teaching experience in urban schools in the Southeast of the US, and his cultural history that included living in large metropolitan areas in the Northeast (Roth et al. 2004).

Chris (white), the new teacher involved in the first ethnographic study featured below, was seeking certification in biology and chemistry. He had a B.Sc. in biology and after two years of graduate work in biology, decided to pursue a teaching career. The class involved in this ethnography was a general chemistry class for tenth-grade students. The class was situated in a small learning community in which most students were low performing and not seeking university admission. Passing the course satisfied one of four science credits required by the school district for high school graduation. The class had 24 students (12 males and 12 females), but the actual numbers on any particular day varied considerably given the high rates of absenteeism. These students had already had science with Alex during the previous year. Perhaps for this reason, a minimal number of behavior problems was observed in this class. Nevertheless—likely mediated by poor attendance related to home and other out-of-school situations—student grades generally were poor, and most students were quite weak in mathematics and reading, as indicated by their very low reading and mathematics scores on state

standardized tests. The excerpts presented in this paper derive from a lesson that occurred about three months into the coteaching of Chris and Alex. They had planned to have a brief preparation period, in which they would prepare students for laboratory activities in which they would construct and conduct measurements with electrochemical cells.

The second ethnography involves Victoria (Filipina-American) and Jessica (white), both with strong backgrounds in science and seeking certification to teach chemistry. This research occurred in the year following the ethnography with Alex and Chris. The small learning community in this study was for students with an interest in college admission after their completion of high school. Coteaching occurred in the chemistry class Alex taught during the first period of the day, which, scheduled in a block, met daily for 96 minutes during the fall semester. The class of 29 students had approximately equal numbers of males (15) and females (14), and consisted mostly of African American students (24) in their junior year. The classroom was spacious, the result of combining two formerly separate classrooms. The students' individual desks were arranged centrally to face what was originally the rear of both rooms during instruction time and laboratory tables were situated to the left and right of that central area. A demonstration desk and chalkboards were behind the students and inaccessible as instructional tools. At the new front of the class, there were two large rectangular whiteboards adjacent to one another and next to them Alex strategically positioned an equally large periodic table. An overhead projector and a pull-down screen were situated between the two whiteboards. This structural arrangement of resources afforded a particular style of whole class transactions that invited the coteachers to coordinate their verbal utterances and gestures to what they wrote on the whiteboards, projected on to the screen, and found salient on the periodic table.

#### *DATA ANALYSES: GENERAL*

Each researcher of the larger research team had access to a basic tool kit that included a computer loaded with video editing software, video and audio recorder, digitizer to convert analogue video to digital files, and external hard drives on which to store large files. All

researchers had access to data resources from a databank that was maintained by a project manager in a central research office. A password protected website was maintained for storing video and audiotapes, transcripts, analytic memoranda and drafts of papers. Transcripts of video vignettes and audio files were shared upon request and to facilitate cross-school comparisons a meeting of the full research group occurred for two hours once a week. As data resources were produced they were made available to all members of the research group. For example, Alex has a hard drive with copies of all videotapes produced in his classroom and he uses these resources in his own ongoing research (conducted to obtain a doctoral degree).

Any member of the research group could identify excerpts from the videotape that were worthy of further investigation. These segments were produced as individuals and groups watched tapes and saved salient vignettes that varied from approximately 30 seconds to three minutes. The resources used in this study were identified as salient in different ways. Those involving Chris and Alex arose from ongoing ethnography in which it was observed that when teachers cotaught they became like one another by being with one another (Roth and Tobin 2002). Some of the common features included verve and the way the teachers spoke. Ethnographically it was evident that Chris and Alex were becoming like the other (Roth et al. 2005) and we endeavored to produce compelling data through microanalysis to test the assertion that Chris and Alex had similar prosodic features. From our field notes we had observed that Alex used terms such as “really really” and “right right” in distinctive ways. We also observed that Chris began to use these terms in what sounded like the same way. Accordingly we searched the database for segments in which Alex and Chris used these terms, while teaching together and separately.

In the analyses below, an episode involving the student Mirabelle and the new teacher Victoria features prominently. We selected it because two student researchers identified the vignette as an example of what they considered to be effective student learning. The two students extracted the vignette and made it available for the research group who undertook intensive analyses. The reasons offered by the student researchers to support their selection were that



Mirabelle had a “trick” to work out chemical valence using the periodic table. By sticking to her guns she showed the class that her theory of valence did not always work. The teacher and the class collaborated to produce opportunities for all students to learn more about chemical valence and periodicity.

Until 2003, the researchers from City High, including Alex, a school administrator, several student researchers the university professors and graduate students, met at least three days a week at City High and in the research office. Since then, because the research at the City High site is ongoing, the researchers (including several of those who were student researchers) communicate as necessary using email, telephone, and videoconferencing. Face-to-face meetings also continue. Because Alex and the school administrator from City High are completing doctoral dissertations based on their research on the teaching and learning of science at City High, they meet several times a week to collaborate on their research and they also meet regularly with the second author of this paper, who is their supervisor.

#### *DATA ANALYSES SPECIFIC TO THIS STUDY*

Following the precepts of interaction analysis (Jordan and Henderson 1995), we began by analyzing data collectively, in sessions lasting about three hours, working second by second through the tapes. Our analyses proceed in the following way. The researcher currently at the controls runs the video until someone (we allow others to participate in such sessions for the purpose of learning to analyze data sources) requests halting the video to talk about a feature or episode—usually the episodes have a duration of somewhere between several hundred milliseconds to the order of ten seconds. The person requesting the halt points out what he finds salient, describes and interprets the episode, and generates hypotheses to be tested in the remainder of the same tape and in the remainder of the database as a whole. The other author also provides his description and interpretation. The episode is discussed until both analysts feel that there is nothing more to say about it at that moment—though subsequent periods of writing often turn up additional features, which are discussed during the next meeting on the following day. In this way, we work image-by-image through the video and, correspondingly, line-by-line

through the transcript. When appropriate—e.g., when there is transactional trouble—we hypothesize what might happen next before moving on to confirm or disconfirm our hypothesis. We then spend the following hours writing individual analyses, which we share with one another, comment upon, confirm or disconfirm in the remainder of the database. On the following day, we continue our collective analysis, both generating new hypotheses and categories and, simultaneously, (dis-) confirming existing ones. Once we have constructed hypotheses about possible features, we test them in randomly selected materials from the database as a whole. If there is contradictory evidence, then we revise the hypothesis and test it again in the data source materials.

Consistent with hermeneutic theories concerning the interpretation of texts and actions (e.g., Ricœur 1991), we know that our understandings gained by living in the situations (as ethnographers, teachers) enable our interpretive efforts; and our interpretive efforts lead to explanations that enhance and elaborate our (intuitive) understandings. The two processes enacted in interpretation—i.e., understanding and explaining—stand in a dialectical relationship, each presupposing the other. Ethnomethodological studies in particular showed that without an understanding of the situation from which data are culled, their classification and measurement would be impossible (e.g., Garfinkel 1967, Lynch 1991).

#### *CONVERSATION ANALYSIS*

All relevant video are digitized to make them available for analysis using iMovie HD and the professional version of QuickTime Player (Macintosh OS X). The software allows us to slow down and speed up the recording, which we interpret image by image to capture phenomena at the microlevel, where we often observe patterned actions that the speed of everyday activity do not allow us to observe in real time. The recorded events are transcribed and are augmented with selected salient video frames. The audiotapes of classroom events, interview sessions, and cogenerative dialoguing meetings are also transcribed and made available for analysis. In the school setting, the first transcriptions are often completed by the high school student research

assistants, because of the high fidelity with which they capture student contributions to the conversations in the science lessons.

Given the central role of face-to-face transactions as sites of society-producing interpersonal behavior, conversation analysis is a method of choice for sociologists to study the pragmatics of the production and reproduction of societal formations—such as school science classes. Conversation analysis as a method of inquiry has had a considerable history (Schegloff 1996); the inclusion of prosody in the analysis of transactions, however, is more recent and much less common (e.g., Couper-Kuhlen and Selting 1996). Face-to-face transactions are ritualistic in nature and integrally tied up with the emotional states of the participants; this ritualistic nature and the attendant emotions appear to be produced and reproduced by means of rhythmic synchronization that can occur at different temporal scales (Collins 2004). Within-group synchrony and between group asynchrony may in fact constitute those resources other researchers seek (e.g., Hajek and Giles 2002) in building theoretical models to understand how affect mediates cohesion within a group of people and their differentiation to the outside. Because prosody—together with facial and other bodily displays—constitutes an important avenue for articulating and making available emotions to others, the analytic method that integrates traditional conversation analysis with precise measurements of prosody has become an important tool to sociologists who study the pragmatics of the production and reproduction of culture (Roth 2007).

*DATA ANALYSES SPECIAL: PROSODY*

The analysis of speech parameters is a powerful tool in the hand of ethnographers who study naturally occurring situations including the emotions and changes thereof that transaction participants make available to one another. To conduct the quantitative/qualitative analysis of speech (and its relation to other perceptual features in the setting), the transcripts of selected lessons are enhanced to contain information regarding sequencing (overlap/latching), timed intervals, characteristics of speech production (stress, lengthening of phoneme, intonation,

loudness), and comments. At this stage, the software packages Peak DV 3.21 and PRAAT<sup>4</sup> are used to work with the soundtrack, to amplify the volume to improve the hearing of doubtful words, to measure the pauses using the waveform display of the sound, and to establish pitch levels ( $F_0$ ), means of the first formant ( $F_1$  mean), and pitch contours, which all are clearly visible in the PRAAT display of a particular sound. When errors in the determination of pitch by the PRAAT algorithm are apparent, we obtained the correct pitch by zooming into the sound spectrum until the pitch can be determined by means of visual inspection that identifies the repeating structure and determination of its length by hand measurement.

To understand pitch, consider that each sound is produced by a superposition of a range of frequencies. Pitch ( $F_0$ ) is the fundamental and dominating frequency produced by the vocal cords. It is the strongest correlate with how a listener perceives the intonation and stress given to words by the speaker; it also is the strongest correlate with psychological and sociological features of a situation.<sup>5</sup>  $F_1$  mean is the frequency of the first formant (constituting significant energy concentration in the spectrum) averaged over an utterance.  $F_1$  mean increases and decreases correlate well with a variety of emotions, including irritation and anger (Scherer 1989), which prominently feature in one of our analyses. Following ‘t Hart et al. (1990), a perceptual approach was chosen to the study of intonation both with respect to categorizations of natural speech and visual characterizations of pitch contours. Not unlike experimental studies, the speech in the situations of interest is categorized by judges (often university students) in terms of a set of emotions that have been explored in psychophysiological studies (Scherer 1989). Quantitative information about speech parameters is obtained independently through the analyses using PRAAT. Our observations and categorizations then are compared with the results of

---

<sup>4</sup> PRAAT is a widely used, multi-platform (PC, Macintosh, UNIX) software package for analyzing a variety of speech related feature including speech frequencies, Fourier spectrums, volume, intensity (in air), rate, and energy. ([www.praat.org](http://www.praat.org))

<sup>5</sup> More technically, pitch ( $F_0$ ) is obtained when the sound stream is submitted to an auto-correlated analysis of the main frequencies yield through a Fourier analysis of the sound spectrum that composes each sound. This yields estimates for the pitch ( $F_0$ ) and the other main frequencies (formants  $F_1, F_2 . . .$ ), that is, those where most of the speech energy is concentrated.

existing research made available in literature review studies. Interpretative convergence means that the quantitative and categorical assessments in this study are consistent with those provided by experimental (psychophysiological) studies.

To conduct a direct analysis between the pitch contours of two speakers, several occurrences of the two forms of the same words for both speakers were normalized for time of articulation and adjusted for the different beginning pitch level (which speakers always adjust to context) in the few instances that this was the case. The values for each person and each form of the word were averaged and plotted. Normalization is necessary for both time and pitch to be able to compare pitch contours across contexts and speakers (Gubrynowicz 2002). Time-normalization is necessary to adjust for the different speech (delivery) rates of different speakers or the same speaker in different contexts. Normalization in terms of beginning pitch is necessary because absolute pitch levels change according to the context—for example, when speakers adjust their absolute pitch to that of the previous speaker. For statistical comparison of two pitch contours, the resulting curves are sampled at or near the points on the temporal axis where PRAAT had extracted pitch from the sound spectrum.

To conduct comparisons of speech articulation rates of the same speaker within and across utterances and situations, the standard measure of syllables per seconds (Verhoeven and Mariën 2004) were used. Information about speech intensity was culled from PRAAT, with special attention to changes in speaker orientations and distances, which would change the intensities that reach the camera microphone, making impossible any inter-situation and inter-word comparisons. Speech energy and speech power are measures that the software package provides.

## PRODUCTION AND REPRODUCTION OF PROSODIC ALIGNMENT

With Émile Durkheim and the beginning of their field, sociologists have recognized the central role individual and collective emotions play in producing and reproducing society. Fundamental to understanding emotion (affect) as a central moment in the making of society is the double role prosody plays in face-to-face transactions: it is both resource for acting in

situation and outcome of actions produced by others and oneself.<sup>6</sup> It is through prosody that transaction participants communicate their orientations and the effect that another's prior action had on the Self. In this section, we show how, after working together for several months, participants in coteaching produce and reproduce similar prosodic patterns, which are new features at least for one of the participants (both new and resident teacher). Two main features with different temporalities were observed in the reproduction of prosodic patterns: inter-individual reproduction of prosodic patterns over time (of working together) and alignment of prosodic patterns within speaking turn pairs. Both features play significant roles in the communication of emotions but do so at different temporal scales. The first phenomenon was observed as part of the reproduction of (classroom) culture, as newcomers to the field of teaching come to speak like those who have been teaching in a particular school for longer periods of time. The second phenomenon is characteristic for the process of producing specific societal formations by means of transactional face-to-face work, and therefore, for the moment-to-moment production of conflict and alignment.

*INTER-INDIVIDUAL REPRODUCTION OF UTTERANCE/PROSODY FEATURES OVER TIME*

As part of our ethnographic observations, we noted that new teachers working for some time with experienced teachers tend to “pick up” expressions, grammatical features, and intonations. Our interviews reveal that participants are not normally conscious that these parameters of their communicative actions are changing. These observations are particularly striking when we note that different new teachers working with the same teacher (same or subsequent year) change in similar ways, becoming more like the experienced resident teacher. But simultaneously we observed that the resident teachers also changed in the direction of the new teachers they were

---

<sup>6</sup> There are other expressive resources as well, including those that in popular language are denoted “body language,” though in a strong sense, there is no “body language” because of the lack of a system of semantic and syntactic codes equivalent to languages (including the English language, computer languages [C++], or American Sign Language).

working with. Here we provide two typical cases selected from a chemistry lesson to show that, after working together for three months, teachers exhibit similarities in their ways of speaking.

First, Alex uses certain phrases and an oral grammar characteristic of the students' African American culture but which the new teachers in our database generally do not use when they arrive at the school. However, as they coteach with Alex, the new teachers, independent of their cultural background, begin to speak using these phrases and grammar. For example, in communicating with his students, Alex frequently uses "really, really" instead of "extremely" (e.g., "really, really hot" when referring to the temperature of a copper coin being heated with a blow torch). Using these speech features as tracers, we find that in any given lesson Alex uses the phrase frequently and it is customary for coteachers in the class to begin using the term with increasing frequency in their transactions with the students. That is, although they are not aware of the process, new teachers "pick up" this way of speaking from Alex during their coteaching experience: the coteachers' oral texts become similar to one another and are of a type that encapsulates efforts to communicate, educate, and maintain the energy and rhythm of their teaching.

A concrete example in the following two transcripts shows how Chris, who is white, somewhat overweight and laid back, acts like Alex, who is dark skinned, lithe, athletic, very bubbly, and energetic—exhibiting similar levels of verve to many of the students. Here, the speaker makes an assertion, pauses and seeks affirmation from the class with a rising and then horizontal inflection and rising volume (‘RIGHT-). The pitch rises sharply, and the utterance stops suddenly and sharply. Pauses, generally longer before than after, surround the utterance. This pattern can easily be discerned (a) on the graphically displayed pitch spectrum of the sound and (b) in the speech intensity display, where higher amplitudes characterize louder talk (Figure 1a).<sup>7</sup> This figure also shows that the surrounding context—longer pause preceding the word, a

---

<sup>7</sup> All episodes have been transcribed following the conventions of conversation analysis developed for the analysis of prosodic features (Selting et al. 1998):

[            beginning of overlapping talk or gesture;  
]            end of overlapping talk or gesture;





ethnographic work, we have heard this pattern of speaking in Alex's teaching prior to him coteaching with Chris. It is apparent that in Alex's successful coteaching experiences, new teachers adopted this pattern of speaking while coteaching with him. In this lesson, Chris exhibits a similar way of teaching and exhorting students to think about what was asserted.

There is a second type of context in which the word "right" appears. In the following two transcripts, characterized by a less sharp rise and less sharp stop, the utterance either immediately follows a statement or immediately is followed by a continuation of the sentence; again, there is a high degree of similarity between the two pitch contours,  $r(30) = 0.86$ ,  $p < 0.0001$  (Figure 1b). The pitch begins much higher than for the second type of articulation of "right," rises less sharply, but generally ends at the same pitch level (about 144 Hz). This pattern marks the end of a declarative statement ("you see this and that's an element all by itself (0.12) 'right'") and is an indication that the utterance is going to continue even though there may be an extended pause (e.g., "'right (1.52) and anything else"). This pattern, too, is one that has characterized Alex's speech for many years—though it has become accessible to us quantitatively only recently after finding suitable software—and Chris uses this pattern now without, as he indicated repeatedly during interviews, having been aware of acquiring this manner of speaking (Figure 1b). The prevalence of this feature is evident in the following two extended turns at talk that each teacher had in a chemistry lesson concerning elements, chemicals, and compounds.

#### Episode 3

C: 01 <<rall><<h>well no-no>, that=s not true, that=s not true>.  
 02 they combine to form other chemicals, other compounds. (0.46)  
 → 03 'rIGHT- (0.44) but you can=t have, (0.38) you know, (0.27)  
 04 liquid copper. ((Watches Alex write "Cu" on the chalkboard.))  
 → 05 (0.79) 'rIGHT- (0.20) it=s a metal. (1.33) <<dim>things like  
 06 that right?> (0.50) so you ha:ve (0.15) copper (0.18)  
 07 with=some=othe:r elements:, (0.85) <<all>for instance what we  
 08 are using today is copper sulfate> ((Alex writes CuSO4)) one of  
 09 the ones we are using. (0.32) so copper sulfate is an example,  
 → 10 (0.22) 'rIGHT of something in a liquid form that has copper,  
 → 11 (0.70) 'right=but=it's; this is combi:ned with other elements  
 12 (1.40) you can break it down further.

#### Episode 4

A: 01 that=s how we define elements=↑elements=something that you  
 02 cAN=t break down (0.16) 'in=anything=else. (0.47) just by  
 → 03 doing chEMistry. (1.33) 'rIGHT? (0.29) so:: COPper (0.49)  
 04 is=n=ELEment. how do you know? ↓i ↑think in this case it i:s,

05 (0.28) you gotta have somebody tELl you (0.21) or look at the  
 06 chart. (0.53) <<all>y=can=see=it=up> the:re ((points to  
 07 periodic table on the wall)), is an element. (0.62) <<f>↑if  
 08 you see it by itself>, this=s what Langston does, (0.30)  
 09 Langston sees it by itself; (0.22) he says, <<dim>↑that=s got  
 → 10 to be an element, i see it all by itself.> (0.44) `rIGHT  
 11 (0.49) now this u ((The "u" in "Cu")) might throw you `OFF  
 → 12 (0.55) ↓but that=s an element `rIGHT (0.27) if you see ↑thIS  
 13 (0.47) ((writes "N")), ↓you see ↑thIS: (0.38) ((writes an "S"  
 14 and "O2")) (0.41) ↓you see ↑thIS: (0.40) that=s an ELEment all  
 → 15 by ITself (0.12) `right- (1.52) and anything else, is mixed up  
 16 with something else, got it?

Our ethnographic study, which has followed many coteaching configurations in the course of the 7-year project, shows general convergence in these and other patterns of speech. This convergence can be observed even in the case of cultural differences between participant teachers. In the few instances where we observed conflict-laden relations between coteachers—often instances where we have had to reassign the new teacher to work in a different class—such convergence did not occur.

These findings are consistent with earlier research, according to which individuals with less institutional power and status adjust their speech parameters generally and their prosody in particular to the person with more power and status. The results also are consistent with the research that shows a lower perceived potency of person–person transactions when pitch convergence was not observed (Gregory 1999). However, in the case of coteaching, the situation is more complex as there are teacher–teacher, teacher–student, and teacher–teacher–student transactions. Accordingly, a research hypothesis we pursued in this study was that new teachers become like their coteachers because they are entrained into the same rhythmic patterns that are characteristic of the exchanges between the regular classroom teacher (here Alex) and his or her students. A likely candidate explanation to be explored by further research is that the students and regular teachers are and have been part of the school culture for longer periods of time, whereas the new teachers newly arrive. It is more likely that the newly arrived will lock into existing frequencies, because, to use an analogy from physics, his or her “inertial mass” is much

smaller than that of the existing group (class, school culture).<sup>9</sup> We explore this hypothesis in the next subsection.

*ALIGNMENT OF UTTERANCE/PROSODY FEATURES WITHIN SPEAKING TURN PAIRS*

When new teachers first come to the inner-city schools—even if they are African American like the students but from a different area in the US—they frequently find themselves in conflictual situations with students (Tobin 2006). Numerous studies undertaken in Philadelphia and New York City show that teachers and students experience difficulties in producing environments that are conducive to learning science. A challenge for all participants is to anticipate the culture of the other and produce forms of practice that expand collective agency (e.g., Emdin 2007). Understanding others’ culture appears to take time and effort and in some instances classes can become dysfunctional as teachers and students give up on trying to produce interstitial culture (Bhabha 1994) that might otherwise produce success. Even Alex struggled in his first semester at City High. He described himself as an “Anglo New Yorker” who had developed “amalgams of himself,” the person he considers truly to be, and the Black, Hispanic teacher that he became while teaching in an urban school in Miami. Although he had initially experienced difficulties (“With the Black kids from the ’hood, my ‘black man’ was a total fiasco”), Alex eventually learned ways of teaching and disciplining effectively. As Alex prepared to teach at City High he considered himself ready for his new assignment. He did not anticipate the struggle that eventual success would entail. However, the presence of new teachers in his classes proved beneficial to Alex, who commented that:

The presence of the two new teachers from the university saved me as I had colleagues with whom I could coteach. They gave me insight into the teaching, planning, and structure of the classes. They gave me breathing space. With them teaching, I could look at the class and attempt to solve the discipline and teaching problems. I was able to forge relationships with

---

<sup>9</sup> Two coupled harmonic oscillators, like two pendulums, experience frequency shifts that take into account their relative masses, where the lighter oscillator moves more than the heavier one.

the students during the class; I would take students aside and ask them why they were behaving in a given manner, or if they were learning. I was able to see the weaknesses in our approach and in our content. I would not have survived without the new teachers.

Our mesolevel ethnographic descriptions in problematic classrooms often contain the everyday terms “out of sync,” “disharmony,” and “head butting.” It turns out that the pitch levels of these teachers are different from those that their students display; one does not observe the continuity or similarity in other prosodic parameters that is observed once more harmonious relations are established. For example, Figure 2 displays the comparison of the pitch levels involving Victoria, immigrant new teacher of Hawaiian origin, and Mirabelle, an African American student. Typically (and depending on the base pitch of the two speakers), a mean difference of greater than 100 Hz ( $D_{mean} > 100$  Hz) is observed between the two speakers. Such differences also are observed in a harmonious classroom when individuals articulate differences of fact or opinion (see below). In the present instance, there is no convergence, even though, based on research linking power and status to prosodic features, there is a possibility that the lower power and status students ought to align with the teacher, who institutionally has more power (e.g., getting a student suspended). In a functioning, “harmonious” coteaching relationship, however, we do observe different patterns of development.

«««««« Insert Figure 2 about here »»»»»»»»

The combined evidence from our ethnographic work and the analysis of speech provides evidence that functioning and low conflict or conflict-free coteaching configurations, coteachers tend to become like one another—including in their prosody. This is particularly observable in the way a seasoned teacher accommodates students by enacting pitch continuation to match the students’ pitch rather than the other way around—as one would have expected from earlier research (Gregory 1994). When difference occurs in this type of situation, the pitch remains about the same from one speaker to the next, or moves into a lower register.

In teacher-student transactions, both within and between lessons, Alex (nearly) always seems to find the “right tone” in diverse situations. In addition to the semantic (word choices) and

grammatical features of his language, which align with those of the students' culture, an analysis of pitch in consecutive turns shows that Alex matches prosodic features, including speech intensity and pitch to the preceding turn produced by a student. An example of such matching is provided in the following episode, which pertains to the conversation about the relative hotness of candle and welding-torch flames.

**Episode 5**

01 A: <<cresc><h>when we did the ca=andle?>>  
 02 S: yea?=  
 03 A: =↓<<cresc>we did the candle>

Following a student utterance, Alex queried to understand the student contribution, “when we did the candle?” Alex’s pitch has been rising from 150 Hz to end at 230 Hz. However, the student has answered at a much lower level, moving from 110 Hz to 130 Hz in the course of uttering “yea.” When Alex continues after the student, Alex (likely unconsciously) first matches his own pitch to that of the student at 130 Hz before returning to his own preferred pitch level in the present context between 180 Hz and 190 Hz. Furthermore, the intensity levels show a similar pattern. The peaks of Alex’s intensities moved from 64 dB to 74 dB but, after the student has responded by speaking with a much softer voice (60 dB), Alex decreases speech volume to 64 dB and moves to 67.5 dB for each of the three syllables of the utterance “ca=andle.”

In our database, the same patterns are observed for Chris after having cotaught at least one lesson per day with Alex for over two months. In episode 6 Chris engages a student about heating a penny. The student suggests that holding the penny in the flame for a long time would lead to a different result. Chris’s utterance of “right” (turn 02) matches the student’s pitch at 216 Hz both during the overlapping word “right” (an acknowledgment of understanding and an indication by the recipient that he or she is listening, while indicating that the speaker may continue) and in the subsequent response, where Chris suggests that the flame has a specific temperature and however long one holds the penny, it will not exceed the temperature of the flame. Pertaining to the intensity, he picks up at about the mean of intensity for the last four words of the students (71 dB) and finishes at about 63 dB.

**Episode 6**

01 S: hold in the fI[re] [fo]r=a=long ti:me.  
 02 C: [ri]gh[t ]  
 03 <<dim>right but=the pEnny=s only at five hundred degrees for that whole time.>

Here, both teachers adjust their volume and pitch to match the levels of the preceding student; such adjustment occurs even in those cases where students speak at a base or context-related pitch that differs by as much as 100 Hz from that of either teacher. This contrasts the observation exemplified in the case of Victoria, whose pitch *never* aligned with either that of the students or her fellow teachers. In the case of Chris, this ultimately leads to pitch continuity across multiple turns involving teachers and students; this phenomenon of multi-turn pitch matching is exemplified in episode 7 and the associated representation of pitch in Figure 3a. In this situation, Chris affirms the statement that a student has made, matching his pitch to the pitch with which the student ends. Chris descends with the pitch nearer to his normal range (turn 02), followed by Alex, whose pitch rises toward the end of the utterance (turn 03). Chris picks up from Alex, initially matching his pitch and then increasing the pitch as he emphasizes the terms “really” and “hot” (turn 04).

**Episode 7**

01 S: like this one.  
 02 C: yea well yea.  
 03 A: that gets really hot,  
 04 C: that gets really ↑hot

«««««« Insert Figure 3 about here »»»»»»»»

In this situation, we observe prosodic complementation as a form producing and achieving alignment in collaborative action. A first speaker (Alex) has produced a contour that is complete in itself; but we might expect it to be followed by a particular contour from the next speaker. Both contributions constitute complete turns. However, although the first participant’s turn signals turn completion prosodically, syntactically, and pragmatically—Alex has agreed with the student, ending on a descending pitch—the second contour complemented the first both in content (semantically) and prosody, so that the two together form a prosodic pair. Furthermore, when Alex chimes in again, he confirms the content of Chris’s utterance by repeating it, and signals and produces alignment and agreement by continuing the pitch contour. Discord and

difference, on the other hand, would have been indicated and produced by significant pitch and intensity differences with the previous speaker (e.g., Goodwin et al. 2002).

This matching process contrasts, for example, the earlier mentioned rapid rise in pitch when Chris changes from an initial agreement with a statement to a disagreement. That is, the sudden rise in pitch signals an opposition between two ideas. Alex and, being entrained to him in this feature, Chris exhibits patterns of pitch matching: unless they express displeasure and opposition to the current situation. Thus, at the beginning of the lesson, Alex called the class to order, “Excuse me! (0.20) Hey!” In this utterance, the pitch rises from about 90 Hz on “ex” to 290 Hz – 315 Hz on “cuse.” This enormous rise in pitch clearly dramatizes the utterance, emotionally coloring the social situation and foreshadows the kinds of transactions that may follow. The pitch then rises from 137 Hz to 215 Hz on “me,” which can be heard as an exclamation. The subsequent “Hey!” rises from 275 Hz to 348 Hz. In both situations, the high—when compared with Alex’s normal pitch level between 150 Hz and 180 Hz—is a resource for understanding him to express disapproval.

A final example of Chris speaking like Alex is seen in turns involving both teachers and some students. The pitches are matched to provide continuity between speakers. When difference is articulated semantically, the corresponding pitch levels tend to be much lower, in or near the speaker’s normal range. In the following episode—almost entirely represented in Figure 3b in terms of pitch and speech intensities—Alex engages in an exchange with students about whether the different elements in a compound can be separated. Several students call out words and fragments, some apparently negating others affirming the possibility of separating the elements by chemical means. Alex expresses reservation and opposition twice (turns 05, 07). In both instances, Alex’s pitch is much lower than that displayed by the student (Figure 3b).

#### **Episode 8**

- 01 A: could i (0.42) sep=rate that stuff out?
- 02 (0.50)
- 03 S1: no-
- 04 S2: but [you can]
- 05 A: [well ]
- 06 S3: (don=t?)

- 07 A: i know i [can]  
 → 08 S2: [uh ] you know you can=  
 → 09 A: =i can break it down, right? (0.57) now copper (0.86) copper  
 likes the ground, like Ronald said. (0.32) do you think I can  
 break down copper into anything other than copper?

When Alex eventually provides the correct answer to his question (turn 09), his pitch moves way out of his normal pitch range to over 200 Hz, which is where the last student speaker (S2, Figure 3b) has ended. The figure shows that after producing a pitch continuation above 200 Hz, Alex's pitch returns to his normal range, with occasional spikes while producing emphasized syllables. It is this production of pitch continuity that new teachers, with few exceptions, are beginning to reproduce after coteaching with Alex for some time. More so, Alex not only matches the pitch levels of students but also reproduces pitch contours that previous speakers produce. This, too, is evident from Figure 3b, where the student contours on "but you can" and "uh you know" find their correspondence in Alex's production of "I know I can" and "I can break it." Thus, in these data generally, there is convergence not only between the two teachers but also convergence and difference between students and the new teacher, which reproduces the continuation and differences in pitch levels the regular (experienced) teacher displays. Most importantly, therefore, teachers match the cultural patterns in the prosody; and new teachers, independent of their culture of origin, learn to produce these patterns in the course of coteaching with someone else already exhibiting this cultural feature.

Reviews of the literature point out that cultural similarity has a statistically reliable effect on emotion recognition accuracy (e.g., Gudykunst and Ting-Toomey 1988); cross-cultural differences in predictive accuracy of displayed emotions, which are highly but negatively correlated with the physical distance between the cultural groups expressing and perceiving the emotions (Elfenbein and Ambady 2003); and there may be cultural differences in the prosodic realization of disagreement across cultures (Yaeger-Dror 2002). The evidence from the present work suggests that culturally different individuals not separated in space and participating in the same activity system for longer periods of time may not be disadvantaged in the recognition of emotion. They learn to produce and reproduce prosodic patterns and, with these, the attendant



affective orientations to other participants co-constitutive of the setting and its social and societal structure.

These findings differ from the studies concerning the relation between prosodic (pitch) features and power and status: the supposedly more powerful teachers aligned themselves with the supposedly less powerful students. But these results are consistent with an unpublished study following a clinical psychologist through her work with 24 schizophrenic patients. The results showed that with training and experience, her pitch values increasingly converged with those of her patients (Gregory 1999). That is, as the quality of the doctor-patient transactions increased, so did the convergence of the pitch ( $F_0$ ) values. On the other hand, for doctors in the course of their training in specialties, the distances increased, which has been interpreted as a decreased focus on interpersonal relations and an increased focus on the technical aspects of the case at hand. A candidate explanation may be that in the service professions (teaching, healthcare), providers may increase their effectiveness when they attune themselves to the recipients of the service, who, having the sense that the service provider really listens (exhibiting empathy or solidarity), better respond to the service. The present results are consistent with the hypothesis that periodic features—pitch, pitch contours, and rhythm—support the phenomenon of social entrainment. Thus, when one speaker uses a certain pitch, pitch contour, or speech rhythm, theories of social entrainment predict that the subsequent speaker will “tune in,” at least in situations of agreement, and perhaps to differ in situations of disagreement (e.g., Goodwin et al. 2002) or lack of (cultural) attunement. Prosodic orientation signals solidarity and a common mood (Damasio 1999) and may create a bridge between two turns that could not be achieved by verbal content (semantic means) alone.

## HEATING UP AND COOLING DOWN CLASSROOM ENVIRONMENTS

There have been suggestions that convergence of long-term average spectra in general and those of pitch in particular is characteristic of conversations involving participants institutionally located such that they are said to have different degrees of power (Gregory 1994). In the

previous section, we present evidence that working together over longer periods of time (about 3 months) leads to a convergence of speech patterns and prosodic displays; prosody, as other nonverbal communicative means—is a resource for transaction deployed for pragmatically dealing with issues at hand. If this were the case, then the divergence and convergence of prosody would be associated with practical action in more complex ways than considering only differences in institutional positions. The event analyzed in this section exhibits and exemplifies precisely this feature: increasing pitch levels when differences appear in the content of talk—rather than the decreases we observed—then a “heating up” of the situation is paralleled by rising pitch levels, increased speech volume, and increased speech rates. When speakers increase their pitch level, speech intensity, and speech rates over the previous speaker, then they “heat up” the situation and “up the ante,” literally trumping the commitment made before with higher prosody values. On the other hand, speakers calm the situation when their conversational contribution is produced with lower speech volume, pitch, and speech rates. In fact, the changing nature of face-to-face transactions involving periods of conflict and periods of agreement within the same conversation is a natural laboratory for studying (in a repeated  $N = 1$  [single-case study] mode) the role of emotions expressed by prosodic and other nonverbal means on the production and reproduction of the specific situation at hand and, of society more generally. The changes in prosody with respect to conflict and its resolution are articulated and exemplified in the following episode—one of numerous that we could also have selected to exhibit the same patterns—in which Victoria entered into a conflict with another person (sometimes the regular teacher, sometimes with students).

In a nutshell, a student (Mirabelle)—always intending to be involved, checking on homework, wanting to pass the course, and avoiding detentions that would keep her off the basketball team—has an idea and attempts to explain it. Victoria, the new teacher in training, challenges the explanation and the student enacts an argument ritual that is similar to how she might argue if challenged by someone outside of the classroom. Although Alex is coteaching and in the classroom at the time of this episode, he does not centrally participate in this series of

episodes from the same event. The transactions involving students and Victoria are most salient in this vignette, especially transactions between Victoria, Mirabelle, and three of her peers. Mirabelle sits near the back so that most students cannot see her without turning around (Figure 4).

««««« Insert Figure 4 about here »»»»»

Just prior to the conflict beginning to articulate itself, Victoria has presented a “trick” to use a periodic table of elements to figure out the number of valence electrons an atom of a specific element has. Mirabelle then announces that she has figured out a systematic way to remember the valence. Victoria reiterates that the placement in the table determines the valence, but Mirabelle counters that this is not what she is talking about. There is another turn pair, in which Victoria points out that what she has said “*is* the trick.” The transcript in episode 9 picks up at this point (the underlined speech elements are represented in terms of pitch and speech intensity over time in Figure 5, each alphabetical label refers to the panel of the same letter).

#### Episode 9

a 01 M: <<p>there=s another way you can figure it 'out>  
 02 (0.96)

b 03 V: this □IS the way to do 'this  
 04 A: <<p>I hope so.>  
 05 (0.34) ((Sasha and Tracy turn their heads toward Mirabelle))  
 06 Ta: [.hh h[h  
 [((turns, smiles at Mirabelle))

c 07 M: [arright ((smiles))

In a quieter voice than she has used before, Mirabelle restates that there is another way to figure “it” (valence) out (turn 01). After a considerable pause—nearing the one-second mark (turn 02)—Victoria utters in a determined fashion, “this IS the way to do this” (turn 03). As Figure 5b shows, the pitch rises by almost 100 Hz during the utterance of “is” and is also accompanied by considerably higher than normal speech intensity. There is another stressed word: rather than falling to the end of the turn, the pitch rises again dramatically to 270 Hz on “this,” thereby reinforcing that her (Victoria’s) way, the one she just explained, is the only way to remember valence, that is, the topic at hand. Far in the background, Alex comments calmly and with very low, almost inaudible, speech intensity, “I hope so” (turn 04), thereby both

sustaining Victoria's claim but also providing a potential resource for avoiding a heating up of the conflict that appears to announce itself. First Tracy and Sasha, then Tasha turn their bodies and heads while Mirabelle is in the process of orienting and situating herself for the account of her method to be produced. She announces the intention to articulate her method uttering "awright" (turn 07), and then explains how subtracting the number "2" from the atomic number of the first row of elements generates the valences of the associated atoms. With the utterance of "awright," Mirabelle moves her body right to left and her hand into a forward position, and then erects the body again as if taking a position from which to launch the articulation of her method of recalling and remembering chemical valences.

«««««« Insert Figure 5 about here »»»»»»»»

In turning their bodies and heads to look squarely at Mirabelle, the three peers exhibit their expectation that something is to be forthcoming—explicitly articulating encouragement for this something to begin. That is, what has happened so far and how it has happened has led to the expectation that the situation requires some conversational action on the part of Mirabelle. The resources for such an expectation clearly have been produced in the immediately preceding exchange; within the culture of these students in this school, the teacher perhaps has articulated a challenge, which Mirabelle, to preserve her symbolic capital with the peers, has to take up. Until this point, however, Mirabelle has stayed calm despite the determination and irritation that Victoria displays in her voice. Contrasting Figure 5a and Figure 5c reveals that following the low pitch and low volume intervention on the part of the regular teacher, Mirabelle's pitch has decreased to below 200 Hz.

Mirabelle has been subtracting two from the atomic numbers in the first full row of the period table of elements (lithium, beryllium, boron, carbon, nitrogen, and oxygen). Tracy is the first to ask Mirabelle in a low-pitched and low-intensity voice, where the two was coming from. Episode 10 begins as Victoria repeats the question but, as shown in Figure 5d, her pitch rises to near 300 Hz and speech intensity increases above normal levels, especially as she utters "two" (turn 01). There is a brief pause, interrupted by Tasha's "from" uttered at low speech intensity.



violent) against the offender. And this emotional reaction, as we show here, is co-expressed and communicated in prosody.

Following the outburst, Tracy, who has turned around, to face Mirabelle, suggests in a very low speech intensity that this way does not work for “all of them”; her pitch is more than 200 Hz lower than the final syllable Mirabelle has uttered (Figure 5f). Victoria restates that Mirabelle’s method does not work for all instances. The pitch returns to 230 Hz and then descends to 185 Hz at the end of turn 09. That is, Tracy’s turn can be heard as a resource that has had a calming and defusing effect. Mirabelle, too, returns to a lower pitch level (near 300 Hz). Her speech rate has decreased to 3.5 syllables per second. The question “well, what?” already appears to be more conciliatory (consistent with an  $F_1$  mean of 1561 Hz), and this possibility for a resolution can be heard in the prosody. Yet irritation can be observed again toward the end of the utterance (turn 10), where  $F_0$  moves beyond 500 Hz when Mirabelle comes to the point of the number “two,” associated with an increase in speech intensity to above 80 dB (Figure 5g) and an increase in  $F_1$  mean to 2626 Hz.

The voice intensity and sharply rising pitch at the end of the previous episode 10 are resources for understanding (interpreting) that the conflict has not yet been resolved for Mirabelle. Further contributions, uttered at lower pitch levels and speech intensities appear to add to the calming effect. Thus, Mirabelle still begins turn 04 of episode 11 above 300 Hz but progressively descends, following the pitch of Gavin (moving considerably above his normal pitch range), which overlaps hers and then drops progressively to a value below 200 Hz (Figure 5h).

#### Episode 11

|         |  |
|---------|--|
| 01      | (0.30)   |
| 02 S:   | <<p>(then? then? youre gone be)  |
| 03 V:   | <<f>but once> <<dim> we start getting in here it doesnt work anymore>  |
| h 04 M: | <u>i aint talking about ((points to the left part of the periodic table))</u><br><u>[this i=m talking about that] ((points to the right part of the periodic table))</u> |
| 05 G:   | <u>[&lt;&lt;p&gt;no she talking about ] row,</u><br>and according to rows it says, the (simple?) rows it checks (??)   |
| 06 M:   | <<f>why, why do three then?>   |
| 07      | (1.55)   |

08 M: if that's the ca[se ]  
 09 V: [no,] no  
 10 M: i don=t know (0.46) <<f>well [i> ]  
 i 11 V: [
 asked for the (0.16) the best way to learn how to put valence  
electrons. each box counts as one as you go across the per- as  
 you go across the period. (0.64) one valence electron, (0.55)  
 two (1.06) then you jump over, (1.36) three, four, five, six,  
 seven, eight. (0.61) that=s what i=m talking about.  
 12 (0.46)  
 j 13 M: arrright, so n=a would be ^nine.  
 14 (1.51)  
 15 right? so then you have nine for a long [time. ]  
 16 V: [alright.] this is one  
 valence, one=

Following two brief, partially overlapping turns, Victoria begins to explain again that she has provided an explanation to assist Mirabelle in the *best* way to learn about valence electrons. In this, Victoria begins with a higher than normal pitch (just below 300 Hz) and high speech intensity (83 dB), but descends to near 200 Hz for the second part of the utterance and lower speech intensity (around 74 dB) (Figure 5i). As if she had realized during the brief 0.40s pause the potential of her prosody to rekindle the conflict, Victoria's prosody parameters change to take on lower values that—on a mesolevel—are heard as less aggressive and therefore are more conducive to conflict resolution. Mirabelle has not yet given up on her alternative explanation, as seen from the fact that she proposes a way of using it for the second row in the periodic table of elements, the row that begins with sodium (Na) (turn 13). The pitch level remains higher than normal, between 250 and 300 Hz and the speech intensity repeatedly goes to 80 dB and beyond (Figure 5j).

Additional students “come to help,” as is apparent from the final episode depicted here. Thus, sitting in front of Mirabelle, first Stacy (turn 01) and later Ivory (turn 03) and Gavin (turn 05) together with Victoria (turn 04) suggest that Mirabelle has to “skip the middle,” meaning the section of the periodic table where the metals are listed. Victoria answers Mirabelle's question about why this is the case by explaining that the elements in the middle constitute a special class, about which the course is not concerned at the moment (turn 08), and that they would return to these elements in more detail later in the course (turn 10). Mirabelle, though still not giving up on her method in the content of her talk (turn 13), has returned to around 200 Hz in her pitch

range and between 70 and 75 dB for the speech intensity (Figure 5k). In this turn, her speech rate has returned to her normal level at around 5 syllables per second. At this point, the discussion continues for another minute, taking as its content the fact that Mirabelle's method works for a particular section of the table and why only for this and not other sections. The tension, which has been apparent previously, has been resolved as seen in the return to normal levels of pitch level, speech intensity, and speech rates on the part of all participants.

#### Episode 12

01 S: ((oriented toward Mirabelle, who cannot see her))  
you sk[ip the middle. ]  
02 V: [and thats as far] as (0.25) you  
[skip the middle]  
03 I: ((oriented toward Stacy))  
<<p>[I wasn=t ? ] for skip[ping  
04 V: [you skip the middle]  
05 G: [you skip the ?ch ]  
06 M: why?  
07 (0.48)  
08 V: because=this is a special class and we dont care about them when  
we=re talking about this right now.  
09 (0.48)  
10 S: =right.  
11 (0.43)  
12 V: we=ll go into it in more detail later o[n. ]  
k 13 M: [but] what i was sayin  
about l=i b=e, b=e: whatever=s=there. what i=ve said works.  
14 (0.38)  
15 K: .hhh it wor[ks] for the fi[rst] [yea]  
16 V: [it] [few] it works for the first [few]  
places

The previous section already provides evidence that institutional position, normally said to wield control and power, and status *do not* determine the deployment of prosodic features. In fact, in a non-deterministic approach, position and status *cannot* determine social and societal phenomena, thereby allowing for successes or failures to emerge contingently. This has been shown in an interesting study where an undergraduate physics student interviewed physics professors about graphs—the dynamics of the transactions could not be predicted, as knowledge and power came to be the *products* of the events as much as being resources for the agency of the participants (Roth and Middleton 2006). Quite to the contrary, the evidence provided here is an indication that prosody generally and pitch in particular are expressions of current orientation, such as resistance to comply with the standard question–response sequence or exhibiting a stance that indicates the particular effort made of accommodating the other not only by responding but



also by aligning the pitch or repeating pitch contour. Sociological theories concerned with transactions and interaction rituals suggest that the preferred state in transactions is the alignment of rhythmic and prosodic features, because such alignment produces and intensifies positive emotional energy for participants (e.g., Collins 2004). Considerable changes in prosodic parameters (e.g., pitch, beat) on the part of one participant with respect to others means breaking out of the repetitions at the microlevel; and it indicates that the person is not responding to the cues of others, who are frustrated in the process.

These analyses, featuring the emergence and resolution of conflict, provide further support to the contention that prosody generally and pitch in particular are transactional resources available in face-to-face communication and used for pragmatic purposes. Conflict is associated with increasing pitch levels over the previous speaker, whereas conflict avoidance and conflict resolution are associated with lowered pitch levels. Thus, when Alex disagrees with a student statement, his pitch remains in his normal pitch range or, if outside, descends into a lower register. On the other hand, in conflict situations, pitch ( $F_0$ ) levels (as well as  $F_1$  means) characteristically move into higher registers. This is apparent here, as Victoria initially and Mirabelle subsequently both fuel the articulation of difference as their pitch levels move higher and higher as if they were pushing each other on a swing that moves higher and higher. Conversely, our study (exemplified here in the conflict between Victoria and Mirabelle) shows that repeated contributions at lower pitch levels and speech intensity are correlated with a cooling effect that led to the ultimate defusing of conflict.

Transactions between Victoria and Mirabelle produce what have been referred to as emotions with negative valence or negative emotions (Turner 2002). It is apparent from the inflection of the voice—in addition to pitch and speech intensity—that Victoria increasingly is annoyed with what appears to be Mirabelle's resistance to accepting the articulated trick as the best way to learn chemical valences. The annoyance (irritation, anger) is available to others in and through her expression that there is one correct solution; and the one Victoria has articulated is it. The unfolding transaction shows that Mirabelle takes the teacher's stand as an outright refusal or

rejection of her method, even before she had the time to articulate it. When Mirabelle hears her method being questioned, she bursts out, her pitch tripling in value, the speech intensity quadrupling, her body movements vigorously oriented toward the teacher, the arm and hand aggressively moving forward and pointing toward the front. In all of these productions—resources for subsequent actions—Mirabelle articulates for others her emotional stance, which incorporates anger and readiness to defend herself against the experienced danger, whatever it might be. These productions require energy, which, as our measurements show, increase tenfold in Mirabelle's prosodic articulation of frustration and irritation (anger). All of the productions require higher than normal levels of energy, so that one legitimately articulates the situation as charged and characterized by high levels of energy. Anger, in fact, is the capacity to mobilize the energy to overcome barriers to an ongoing effort (Frijda 1986); and in its most intense form, Mirabelle's anger is an explosive reaction against the frustrations experienced in gaining acceptance for her method for remembering and recalling chemical valences. Mirabelle can be seen and heard as producing high levels of emotional energy with negative valence; and negative emotional energy is expressed largely by vocal means and gesture.

In this situation that drains significant amounts of emotional energy, the quiet, low-intensity and low-pitched contributions several classmates produce are followed by Mirabelle's lowering of the values of the same parameters. These productions, therefore, can be considered resources that oriented the class as a whole toward the current state of the conflict and towards the continued effort of defusing it. Allowing the participants to come out of the loggerheads by providing resources may be a form of commitment to the group and the group processes, that is, a form of solidarity allows both parties in the conflict to cool down and resolve issues about understanding chemistry. In this school, conflictual situations often end differently, leading students to storm out or drive the altercation to such a level that they come to be sent from the classroom and may be suspended from school for a period of time (Roth et al. 2004). In such situations, the prosody parameters stay at their elevated levels until a student's removal from class abruptly ends the situation.

## PROSODY, ENTRAINMENT, AND SOLIDARITY

A major question about how humans are able to make happen and bring off the specially human forms of transaction that produce and reproduce everyday ordinary society has been answered in terms of the emotional valence of jointly lived situations and the emotional valence tied to the setting of goals and the likelihood of achieving them (Damasio 1994/2000). Accordingly, theorists in the sociology of emotion (e.g., Collins 2004) and face-to-face transaction (Turner 2002) suggest that emotions—articulated and therefore communicated by prosodic and other observable verbal and nonverbal means—constitute the essential feature that binds individuals into a collective. Social binding, cohesion, and solidarity derive from resemblances and “social similitude” (Durkheim 1893, chapter 2 section 4). “Gluing” and cohesion and the solidarity that results from it are said to be possible because multiple pitches and other rhythmic phenomena have the tendency to align themselves when the different instances are not too far apart. Physicists—who have observed already in the seventeenth century that two pendulum clocks close to each other on the same wall will become aligned in their swings—use the concept of *entrainment* to explain such alignments. Sociologists, social psychologists, and anthropologists concerned with time and temporality of human transactions have borrowed this notion to describe the synchronization of human behavior (McGrath and Kelley 1986). In the two preceding sections, we show how speakers align their pitch levels and pitch contours to come into tune with other speakers. During our ongoing ethnographic work we observe other rhythmic phenomena—for example, in the movement of various body parts—in classrooms that align themselves with rhythmic patterns articulated and exhibited by the speaker(s). This is the case even when a recipient does not or cannot see the speaker and therefore gestures and facial expressions that often constitute the primary resources for accessing the emotions of others. To exemplify the presence of collectively enacted rhythmic patterns that align with those of the current speaker, we return to the same classroom cotaught by Victoria and Alex.



In this situation, the ingredients for a focused encounter and the production of solidarity are all present: bodily copresence, barrier to the outside, mutual focus of attention (periodic table of elements), and shared mood (Collins 2004). There is a basic rhythm that underlies the production, coordination, and reproduction of social alignment (“similitude”) within the classroom. Mutual focus and shared mood are linked through a feedback of intensification through rhythmic entrainment, a rhythm set as shown by vocal means. But Mirabelle produces rhythmic patterns by other nonverbal means, too. These features therefore become available for those who are in a position to see her—in addition to expressing the basic rhythm characterizing her emotional investment. Thus, the hand moves forward and reaches the foremost position exactly at the stressed utterance or vowel in the numbers, falling together with the intensity and pitch peaks (Figure 8). That is, the cyclical hand movement visible in Figure 7 is patterned such that the foremost position and the stressed syllables fall together.

Throughout the classroom, there are signs that show how others are in synchrony with Mirabelle, even when students are seated and oriented such that they cannot see her. This “beat” of her verbal production is available to others in audible form. With each number in the series three (lithium) to six (carbon), Mirabelle briefly glances to the periodic table, as if verifying what the next number is. The heads and gazes of other students (e.g., Gavin, Tasha, Shawn [for seating see Figure 4]) also move their regard to focus on the periodic table of elements. They move their gaze simultaneously with Mirabelle although they do not see her and although there is no indication in the speech content itself that suggests others ought to look at the table of elements. That is, the resource for producing this synchrony in orienting gaze is made available in other ways than through visual coordination.

There are other signs of synchrony. For example, Gavin rocks his head slightly back and forth. As Figure 8 shows, even though Gavin cannot see Mirabelle (Figure 4), his rhythm perfectly (within the measurement error imposed by the  $1/30^{\text{th}}$  of a second video frame rate) reproduces the forward position of Mirabelle’s hand, which itself is aligned with the rhythm with which the account of her method is produced. Gavin also produces the identical rhythmic pattern

with his right leg, which swings in a left-to-right motion matching in its extreme left position the foremost position of his head. Thus, when we mark the point in time when his chin reaches the foremost position, these temporal positions coincide with the forward position of Mirabelle's hand, and the peaks in her speech intensity and pitch. When Mirabelle arrives at "and so on and so on," Gavin stops rocking his head, the last coincident movement having been a slightest movement with a coincident closing of eyes following an upward movement of head to direct gaze to the periodic table. As there are no visual means that could have provided him with the resources for the coordination, prosody is the likely candidate that allows synchrony to emerge. We hypothesize that the students as a whole are aligned with Mirabelle in the production of her alternative and that they can empathize with Mirabelle who experiences Victoria's actions as an affront.

In the following episode, the production of patterns of synchrony is evident when the students are entrained into Victoria's re-articulation of her "short-cut trick." The episode follows a first presentation by Victoria about how to find out the valence electrons from the columns of the periodic table. Mirabelle has proposed to continue counting after the end of the first row, attributing to sodium (Na) the number 9. Here, Victoria then reiterates that the maximum number for calculating the valence of an element is 8. In the course of episode 14, two students and Victoria suggest to Mirabelle that for the rows where there are metal elements in the center of the periodic table (for example, scandium to zinc), she has to "skip the middle," meaning that in this part of the periodic table, counting is suspended. The students, who assist Victoria in articulating an explanation for dealing with Mirabelle's problem, are in fact aligned with her semantically; and this alignment is articulated also at the nonverbal level where rhythms are enacted in unison with others.

#### **Episode 14**

01 V: the maximum amount is eight, thats why this is a short-cut  
trick. (0.24) one (0.21) this column is one; (0.48) this is  
two;  
02 (1.28)  
03 T: three  
04 V: [three, four, five, six, seven, eight

05 X: [three, four, five, six, seven, eight ((several students  
throughout the class in unison))  
06 S: ((oriented toward Mirabelle, who cannot see her))  
you sk[ip the middle. ]  
07 V: [and thats as far] as (0.25) you  
[skip ] the middle  
08 X: ((oriented toward Stacy))  
[I wasnt] ? for skip[ping  
09 V: [you skip the middle]  
10 G: [you skip the ?ch ]  
11 M: why?

As the teacher utters, “This column is one,” Tasha, who already has nodded repeatedly affirming Victoria’s statements immediately prior to the beginning of episode 14, moves slightly backwards than forwards, her lips silently forming “one” as her head reaches the foremost position (Figure 9). Similarly, her head movement and silent lip formation parallels the utterance of “two.” Her hand then moves away from the body, fingers stretch out and point forward toward the periodic table. In synchrony with the teacher’s hitting the chalkboard with the chalk, Tasha verbalizes “three” aloud. In this one instance, the teacher’s utterance of “three” actually is uncoordinated with her beat of the chalk against the chalkboard, but the student is coordinated with it precisely. As the teacher counts out “four,” “five,” “six,” “seven,” and “eight,” Tasha moves her lips forming the words and simultaneously enacts a beat gesture, the hand reaching the down position precisely (within the 1/30<sup>th</sup> second of accuracy that the video allows) with the words. Behind Tasha, Ivory and another student begin to count at “three” and Ivory sharply nods her head in unison with each uttered word, the chin reaching the forward position precisely at the emphasized syllable. Other students in the class also fall into unison counting beginning with the number three. It is as if the two initial utterances had provided the resources for students to capture the rhythm so that they could produce their articulations of the number in unison with Victoria. Here, too, there is a mutual focus object, the periodic table of elements. The beat may serve to communicate the attention to the common object (Collins 2004) and by communicating this common attention, teacher and students become mutually aware of their common focus. Apart from group solidarity and positive emotional energy, the situation produces shared cognitive experience, that is, understanding of a chemical concept as a sacred object (i.e., symbol





levels, as participants are oriented toward producing and reproducing a lesson recognizably *as a lesson*; and they do so by orienting to this lesson in ways that they are familiar with from their experience. This orientation organizes expressions produced across a great variety of bodily productions, not only including speech and gesture (McNeill 2002) but also, as shown here, rhythmic patterns produced by different body parts and by various prosodic parameters.

In these analyses, synchrony is a resource for all to experience and recognize alignment and agreement. Thus, the synchrony Tasha displays with respect to counting and movement with the rhythmic patterns her teacher Victoria produced is consistent with her repeated head nods following Victoria's statements about how to remember and recall atomic valences. Solidarity, that is, the fact or quality of being perfectly united in some respect with others, here is articulated publicly, available for all to use as a resource in subsequent actions. Solidarity, then, not only is communicated by means of agreements on the cognitive content of statements but also, and perhaps more importantly, by a variety of rhythmic bodily means. From an analytic perspective, therefore, individual participants orient themselves and each other to existing public exhibits of (collective) solidarity but also concretely realize individual exhibits of this solidarity. This phenomenon has positive emotional valence, which in turn is expressed and re-expressed in public exhibits involving prosodic means and a variety of body movements. It is out of this alignment that other students are positioned to produce resources that allow Mirabelle, for example, to calm down prior to producing an altercation of the kind so prevalent in inner-city schools, altercations that frequently result in the suspension of students and their exclusion from schooling.

## EMOTIONS, PROSODY, AND POWER/STATUS

Emotions are central to the ways in which human beings orient to situations; and they are modified by the articulation of emotions of other participants in face-to-face transactions (Turner 2002). Besides facial expression, prosody is, as we show here, a major resource for displaying and experiencing emotions in a setting. There have been suggestions that participants with less

power and status converge in their prosodic parameters to align with those of more power and status; and differences and conflict arising between transaction participants having roughly equivalent power and status (e.g., children participating in hopscotch games) are characterized by very high values and differences in the pitch levels of participants. The present study augments the existing literature by contributing descriptions of prosody as a transactional resource that is deployed *pragmatically* (rather than deterministically and causally) and subject to the purposes at hand. Following others, we view prosody as but one aspect of the production of communication all subordinated to the same activity at hand and, therefore, being different expressions of the same societal-psychological unit (Vygotsky 1986). That is, sounds (words), prosody, body position, hand gestures, and other communicative resources articulated at some point in time are different expressions of the same underlying orientation, emotional valence, and meaning unit (Roth and Pozzer-Ardenghi 2006). The meaning of vocal gestures therefore does not lie behind them but is “intermingled with the structure of the world outlined by the gesture” so that “the smile, the relaxed face, gaiety of gesture really have in them the rhythm of action, the mode of being in the world which are joy itself” (Merleau-Ponty 1945, p. 216).

Summarizing our ethnographic evidence, we construct four major empirically grounded claims that can be used as hypotheses to be tested further in cross-site and cross-setting “confirmatory ethnography” (e.g., Roth 2005) or by quasi-experimental studies. First, independent of the power and status differentials, speakers in non-conflictual situations use lowered pitch registers to express difference in content and reluctance to submit to the normal turn-taking routines (e.g., the question–answer sequence). This effect is especially visible when the normal pitch ranges of the two interlocutors are sufficiently different and do not overlap. On the other hand, accommodation and acceding to requests are associated with pitch matching. Second, when two or more teachers coteach the same classes over time, the new teachers tend to “pick” up speech forms, grammatical patterns, and prosodic features from the regular classroom teacher. Initial differences between the pitches produced by new teachers (who take over a speaking turn from students) tend to become continuous in the same way that they are for the

(often-seasoned) regular classroom teacher. In non-conflict-laden classrooms, teachers tend to express differences in terms of lowered pitch levels. Inexperienced teachers and teachers in conflictual relations with their classes, on the other hand, display differences in terms of raised pitch levels. Third, raised pitch levels and speech intensity when differences in the cognitive (semantic) content of utterances are apparent tend to “heat up” the classroom atmosphere, whereas contributions at lower pitch levels and speech intensities are associated with a “cooling down” of the situation, that is, they are resources for defusing conflict. Fourth, even in the absence of visual clues, rhythmic patterns of a speaker are reproduced by other participants in the classroom and displayed prosodically and in a variety of body movements (e.g., rocking head, hand beat, rocking leg, pen hitting desk, chalk hitting chalkboard).

The results of this study suggest the need to theorize the production and reproduction of prosody differently than in previous research. Rather than viewing power and status as factors that *determine* pitch levels and convergence, we find the production of pitch levels, pitch continuation, pitch level repetitions, and so on to be associated with difference/resistance and accommodation. The concept of social solidarity denotes high levels of unity or agreement. Convergence in prosodic parameters among two or more participants is an expression of, and serves as a resource for the further production of solidarity, itself an expression of the emotional alignment of participants. In cases of conflict, some members of the collective may, through the production of lower pitch levels and speech intensities, assist an angered, excited, or animated member to “cool” or calm down.

Rather than students matching the pitch of their teacher, which would have been expected based on existing power/status theory, teachers in non-conflict classrooms of our study tend to produce pitch continuations to match the pitches of their students prior to returning to their own, normal levels during extended turns at talk. New teachers in such classrooms—those in non-conflict relations with their master teacher—fall in line with the pitch matching pattern, frequently leading to long exchanges in which the ending pitch of one speaker flows into the opening pitch of the subsequent speaker. This usually unconscious move on the part of the

teacher may find its explanation in the same concept of social solidarity, here exhibited by the regular teachers with extended experience in the school, who tend to adapt their ways of speaking to align with the ways in which their students spoke (i.e., taking account of home and street cultures).

We find solidarity—as a phenomenon that participants continuously produce and reproduce—also within the student body. Thus, the synchronous rhythmic features simultaneously found at various places in the classroom suggest that the students are “in tune” or “in synch” with one another. They also express anticipation of particular events, such as when numerous students turn around to face Mirabelle as if they had predicted a likely pattern of response, that is, as if they “saw something coming.” Anticipation inherently means *understanding* the situation in a phenomenological sense, that is, an intuitive lived sense of what is happening and what dangers might loom ahead rather than a reflective understanding.<sup>10</sup> At the same time, these various signs, observable from Mirabelle’s position, may have been resources encouraging her to take up the challenge and propose her alternative description of remembering and recalling valences.

Entrainment is not describable in terms of a causal link, as the production of synchrony, other than in a mechanical system, requires anticipation. Thus, in one instance, Mirabelle produces a particular rhythm, which Gavin also displays. It is not that Mirabelle’s rhythm *causes* Gavin to rock in the same rhythm, because Gavin, looking toward the front of the classroom, has no other resource than Mirabelle’s voice. If he had to consciously attend to making his rocking coincide with her activity peaks (intensity, pitch), he would be out of synchrony by something on the order of a second or two.<sup>11</sup> Even if there were non-conscious ways of causing synchrony, he would still be behind her, for he could only know when Mirabelle’s pitch peaks after having

---

<sup>10</sup> For a discussion of Martin Heidegger’s related terms *Vorhabe* (“fore-having”), *Vorsicht* ([attention] fore-sight), and *Vorgriff* (fore-conception [anticipation]) see Dreyfus (1991).

<sup>11</sup> For example, research in cognitive science on the difference between epistemic and pragmatic actions shows that in the game of Tetris, which demands players to rotate, shift, and place falling geometric objects, the conscious processing of the geometrical shapes prior to rotating, shifting and placing would take a minimum of 1.2 seconds (Kirsh and Maglio 1994).

heard it. This means, Gavin has to anticipate peaks, which is a production of his own, requiring that he already is in tune. The synchrony of his movement with Mirabelle's prosody is a consequence of being in tune. Such an anticipation clearly is observable at the instance when Tasha utters "three," just as Victoria hits the board with the chalk but prior to her own utterance of the same number word, and inconsistent with all other instances where number word and the hitting of the board fell together. That is, Tasha anticipated the correct placement of the count with respect to the noise from the chalk, but Victoria, who produces the chalk noise, is out of alignment with her speech, which only follows the rhythm of the chalk beats.

### BEYOND CONVENTIONAL WISDOM

In this study, we use different data sources (ethnographic observation, speech analysis on different prosody parameters, videotape) to provide convergent evidence for the use of prosody as a resource in the production and reproduction of teaching and learning and therefore in the production and reproduction of a particular aspect of society. Clearly, given the conflictual situation we observed, our approach to social and societal phenomena is non-deterministic, allowing us to understand that fields and structures survive the occasional disturbances: one student-teacher argument over valence does not destroy institutional relations or the activity system but it produces modifications. In schooling, these modifications may be sufficient to lead to success and failure on both student (failing a course; dropping out of school) and teacher part (attrition; stress with costs to the health-care system). The ethnographic, video-graphic, and sound-analytic evidence supports assertions that prosody and other rhythmic phenomena are a means for signaling alignment, difference, and emotional stakes across seeming boundaries of cultural origin, gender, and institutional position and associated control/power dimensions (teacher/student and experienced/novice teacher). The phenomena are not merely events of interest to microsociologists, but in fact produce and reproduce societal forms of activity (here schooling), which we show in their emergence at micro- and mesolevel grains of analysis. We suppose that the phenomena we describe here can in fact be used for studying other settings of

interest to sociologists, such as the making and remaking of managerial authority (e.g., Vallas 2006) or the continuous production and reproduction of various forms of institutional talk through talk in institutions (Boden 1994). The methods we use work particularly well in the context of other microsociological approaches that draw on conversation analysis, because the different prosodic means constitute resources and constraints for making societally motivated forms of activity in and through face-to-face encounters. We therefore do not think of ourselves as microsociologists, but rather, following others (e.g., Smith 1987), we see microsociology generally and conversation analysis particularly as but one aspect of an approach that integrates micro-, meso-, and macrolevels in social analysis.<sup>12</sup>

The conventional wisdom about teaching and much of the research in teacher education and policy initiatives are consistent with the cliché “don’t smile until Christmas.” In essence teachers are exhorted and often required to establish control over their students and maintain a quiet class—assumed to be effective. However, especially in classrooms in which diverse forms of culture are enacted by students and their teachers, the alignments needed to produce productive learning environments might be difficult to attain if teachers are expected to exercise power over students, providing structures to afford optimal forms of participation. In fact, students may feel that a teacher’s practices “shut them down” and frustrate their efforts to meet their goals. As our ethnographic work shows, teaching practices that attempt to control students often are perceived by the latter as disrespectful and thereby become resources for resistance, including disruption and refusal to engage. Perhaps the fact that teachers are unsuccessful in their efforts to attain control over students is a reason so many studies in urban classrooms present portraits of dysfunction—students respond by showing their disrespect for the teachers who seek to control them.

---

<sup>12</sup> Part of our work in this and other urban schools involves students, teachers, administrators, and researchers in the transformation of classrooms and institutions to improve teaching and learning without reproducing the societal structures (e.g., failing, drop-out, poverty) normally reproduced in urban schools (see, e.g., Tobin, Elmesky and Seiler 2005)

In this study we show how successful transactions and the alignment of successive speakers' prosody go hand in hand. Salient in this study were pitch ( $F_0$ ), amplitude, and power in the air; and, to a smaller extent, there also is convergent evidence from the less-researched measure,  $F_1$  mean. Although what we have learned is no doubt contingent on larger segments of time, at a microlevel successful transactions appear to be related to a given speaker either aligning with or being less than the previous speaker in the magnitude of pitch, amplitude, and power in the air. We observe that when speakers speak with higher magnitude on these prosody indicators the environment generally becomes heated and provided resources for asynchrony in terms of emotion and cultural enactment. Hence, counter to the conventional wisdom, teachers might be encouraged not to raise their voice to assert power over students. On the contrary, they might be encouraged to "speak under" their students and in like manner, students might become conscious of the role of prosody in creating and maintaining an emotional climate and associated environments in which learning can occur most readily.

In summarizing what we have learned from this study we do not assert that speakers always should align their prosody or come under the previous speaker. There have been numerous situations in which Alex, for example, used prosody and body movement to establish a rhythm to the classroom that is high energy. In so doing he created structures for students who seemed to easily resonate with high energy teaching, while making it difficult for those who seemed to prefer less energetic forms of participation. We offer this nuance as a hedge against overgeneralizing what we have learned and to argue for a need for additional research into what might be thought of as classroom rhythms. Perhaps a preference for tuning into classroom rhythms will be yet another form of diversity that teachers and students need to be aware of as educators seek ways to improve the quality of teaching and learning.

Finally, it is clear that teaching, as a caring profession, necessitates forms of teacher–student transactions that afford participation and learning. Ultimately, we might expect more learning to occur when students and teachers find themselves in and collaboratively construct and control environments in which they have a sense that they are "in all of this together," that is, when there

is a sense of solidarity (Roth in press). The classroom is a field quite different than the hospital ward, for example. Those of us who have been patients in a hospital ward may have experienced doctors with good bedside manner and others who are more interested in dispensing their expertise, not taking the time to engage with patients—dispensing and moving on. It is not surprising that studies of the prosody of doctor–patient transactions in a ward might be quite different than teacher-student transactions in a classroom or those of a TV personality interviewing the President’s press secretary. What seems likely, however, is that prosody alignments are associated with the extent to which synchrony emerges within groups and the subsequent entrainment that can produce solidarity, identity changes, positive emotions, and increased success on meeting collective goals.

## REFERENCES

- Anderson, Elijah. 1999. *The Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: W.W. Norton.
- Bakhtin, Mikhail M. 1993. *Toward a Philosophy of the Act*. Austin: University of Texas Press.
- Barab, Sasha, and Wolff-Michael Roth. 2006. “Curriculum-based Ecosystems: Supporting Knowing from an Ecological Perspective.” *Educational Researcher* 35(5):3–13.
- Bhabha, Homi. 1994. *The Location of Culture*. New York: Routledge.
- Boden, Deirdre. 1994. *The Business of Talk: Organization in Action*. Cambridge: Polity Press.
- Collins, Randall. 2004. *Interaction Ritual Chains*. Princeton, N.J.: Princeton University Press.
- Corno, Lyn, Lee J. Cronbach, Haggai Kupermintz, David F. Lohman, Ellen B. Mandinach, Ann W. Porteus, and Joan E. Talbert for the Stanford Aptitude Seminar. 2002. *Remaking the Concept of Aptitude: Extending the Legacy of Richard E. Snow*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Couper-Kuhlen, Elisabeth, and Margaret Selting. Eds. 1996. *Prosody in Conversation*. Cambridge: Cambridge University Press.
- Damasio, Antonio R. 1999. *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. San Diego, CA: Harcourt.
- Damasio, Antonio R. 2000. *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York: HarperCollins. (First published in 1994)
- Depraz, Natalie. 2006. *Comprendre la phénoménologie: Une pratique concrète*. Paris: Armand Colin.



- Dreyfus, Hubert L. 1991. *Being-in-the-World: A Commentary on Heidegger's 'Being and Time,' Division I*. Cambridge, Mass: MIT Press.
- Durkheim, Émile. 1893. *De la division du travail social*. Paris: Presses Universitaires de France.  
(Available in the original from a website maintained at the University of Quebec at Chicoutimi:  
[http://classiques.uqac.ca/classiques/Durkheim\\_emile/durkheim.html](http://classiques.uqac.ca/classiques/Durkheim_emile/durkheim.html))
- Elfenbein, Hillary A., and Nalini Ambady. 2002. "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-analysis." *Psychological Bulletin* 128:203–35.
- Elfenbein, Hillary A., and Nalini Ambady. 2003. "Cultural Similarity's Consequences: A Distance Perspective on Cross-Cultural Differences in Emotion Recognition." *Journal of Cross-Cultural Psychology* 34:92–100.
- Emdin, Christopher. 2007. "Exploring the Contexts of Urban Science Classrooms. Part 1: Investigating Corporate and Communal Practices." *Cultural Studies of Science Education*, 2, 319–41.
- Franck, Didier. 1981. *Chair et corps: Sur la phénoménologie de Husserl*. Paris: Les Éditions de Minuit.
- Frijda, Nico H. 1986. *The Emotions*. Cambridge: Cambridge University Press.
- Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.
- Giles, Howard, Angie Williams, Diane M. Mackie and Francine Rosselli. 2005. "Reactions to Anglo- and Hispanic-American-Accented Speakers: Affect, Identity, Persuasion, and the English-Only Controversy." *Language & Communication* 15:107–20.
- Goodwin, Charles, Marjorie H. Goodwin, and Malcah Yaeger-Dror. 2002. "Multi-Modality in Girls' Game Disputes." *Journal of Pragmatics* 34:1621–49.
- Goodwin, Marjorie H. 1996. "Informings and Announcements in their Environment: Prosody within a Multi-Activity Work Setting. Pp. 436–61 in *Prosody in Conversation: Interactional Studies* edited by Elizabeth Couper-Kuhlen and Margret Selting. Cambridge: Cambridge University Press.
- Goodwin, Marjorie H., and Charles Goodwin. 2000. "Emotion within Situated Activity. P. 239–57 in *Linguistic Anthropology: A Reader* edited by Alessandro Duranti. Malden, Mass: Blackwell.
- Gregory, Stanford W. Jr. 1994. "Sounds of Power and Deference: Acoustic Analysis of Macro Social Constraints on Micro Interactions." *Sociological Perspectives* 37:497–526.
- . 1999. "Navigating the Sound stream of Human Social Interaction." Pp. 247–85 in *Mind, Brain, and Society: Toward a Neurosociology of Emotion* edited by David D. Franks and Thomas S. Smith. Stamford, Conn: Jai Press.
- Gregory, Stanford, Stephen Webster and Gang Huang 1993. "Voice pitch and amplitude convergence as a metric of quality in dyadic interviews." *Language & Communication* 13:195–217.
- Gubrinowicz, Ryszard. 2002, April. "A Study of Speech Prosody of Subjects with Profound Hearing Loss Recorded at Child Age and 20 Years Later." In *Proceedings of the Speech Prosody 2002 International*

- Conference, edited by Bernard Bel and Isabelle Marlien. Available at: <http://aune.lpl.univ-aix.fr/sp2002/pdf/gubrynowicz.pdf> (Accessed July 24, 2007)
- Gudykunst, William B., and Stella Ting-Toomey. 1988. "Culture and Affective Communication." *American Behavioral Scientist* 31:384–400.
- Hajek, Christopher, and Howard Giles. 2002. "The Old Man Out: An Intergroup Analysis of Intergenerational Communication Among Gay Men." *Journal of Communication* 52:698–714.
- Hall, Edward T. 1976. *Beyond Culture*. New York: Doubleday.
- Jordan, Brigitte, and Austin Henderson. 1995. "Interaction Analysis: Foundations and Practice." *Journal of the Learning Sciences* 4:39–103.
- Kemper, Theodore D. 1987. "How Many Emotions are there? Wedding the Social and the Autonomic Components." *American Journal of Sociology* 93:263–89.
- Kemper, Theodore D., and Randall Collins. 1990. "Dimensions of Microinteraction." *American Journal of Sociology* 96:32–68.
- Kirsh, David, and Paul Maglio. 1994. "On Distinguishing Epistemic from Pragmatic Action." *Cognitive Science* 18:513–49.
- Local, John, and Gareth Walker. 2004. "Abrupt-Joins as a Resource for the Production of Multi-Unit, Multi-Action Turns." *Journal of Pragmatics* 36:1375–403.
- Lynch, Michael. 1991. "Method: Measurement—Ordinary and Scientific Measurement as Ethnomethodological Phenomena. Pp. 77–108 in *Ethnomethodology and the Human Sciences* edited by Graham Button. Cambridge: Cambridge University Press.
- McGrath, Joseph, and Janice Kelley. 1986. *Time and Human Interaction: Toward a Social Psychology of Time*. New York: Guilford Press.
- McNeill, David. 2002. "Gesture and Language Dialectic." *Acta Linguistica Hafniensia* 34:7–37.
- Merleau-Ponty, Maurice. 1945. *Phénoménologie de la perception*. Paris: Gallimard.
- Pittam, Jeffrey, and Klaus R. Scherer. 1993. "Vocal Expressions and Communication of Emotion." Pp. 185–97 in *Handbook of Emotions* Michael Lewis and Jeannette M. Haviland. New York: Guilford Press.
- Protopapas, Athanassios, and Philip Lieberman. 1997. "Fundamental Frequency of Phonation and Perceived Emotional Stress." *Journal of the Acoustical Society of America* 101:2267–77.
- Ricœur, Paul. 1991. *From Text to Action: Essays in Hermeneutics, II*. Evanston, Ill: Northwestern University Press.
- Roth, Wolff-Michael. 2005. *Doing Qualitative Research: Praxis of Method*. Rotterdam: SensePublishers.
- . 2007. "Motive, Emotion, and Identity at Work: A Contribution to Third-Generation Cultural Historical Activity Theory." *Mind, Culture and Activity* 14 (1/2):40–63.

- . In press. “Toward Solidarity as the Ground for Changing Science Education.” *Cultural Studies in Science Education* 2.
- Roth, Wolff-Michael, and Yew Jin Lee. 2007. “‘Vygotsky’s neglected legacy’: Cultural-historical activity theory.” *Review of Educational Research* 77:186–232.
- Roth, Wolff-Michael and David Middleton. 2006. The making of asymmetries of knowing, identity, and accountability in the sequential organization of graph interpretation. *Cultural Studies of Science Education* 1:11–81.
- Roth, Wolff-Michael, and Lilian Pozzer-Ardenghi. 2006. “Tracking Situated, Distributed, and Embodied Communication in Real Time.” Pp. 237–61 in *Focus on Cognitive Psychology Research* edited by Michael A. Vanchevsky. Hauppauge, N.J.: Nova Science.
- Roth, Wolff-Michael, and Kenneth Tobin. 2002. *At the Elbows of Another: Learning to Teach through Coteaching*. New York: Peter Lang.
- Roth, Wolff-Michael, Kenneth Tobin, Cristobal Carambo, and Chris Dalland. 2005. “Coordination in Coteaching: Producing Alignment in Real Time.” *Science Education* 89:675–702.
- Roth, Wolff-Michael, Kenneth Tobin, Rowhea Elmesky, Cristobal Carambo, Ya-Meer McKnight, and Jennifer Beers. 2004. “Re/making Identities in the Praxis of Urban Schooling: A Cultural Historical Perspective.” *Mind, Culture and Activity* 11:48–69.
- Schegloff, Emanuel A. 1996. “Confirming Allusions: Toward an Empirical Account of Action.” *American Journal of Sociology* 102:161–216.
- Scheler, Max. 1933. *Schriften aus dem Nachlass*. Berlin: Der Neue Geist Verlag.
- Scherer, Klaus R. 1989. “Vocal correlates of emotion.” Pp. 165–97 in *Handbook of Psychophysiology: Emotion and Social Behavior* edited by Hugh L. Wagner and Anthony S. R. Manstead. London: Wiley.
- Schimmack, Ulrich. 1996. “Cultural Influences on the Recognition of Emotion by Facial Expressions: Individualistic or Caucasian Cultures?” *Journal of Cross-Cultural Psychology* 27:37–50.
- Selting, Margaret, Peter Auer, Birgit Barden, Jörg Bergmann, Elizabeth Couper-Kuhlen, Susanne Günthner, Christoph Meier, Ute Quasthoff, Peter Schlobinski and Susanne Uhmman. 1998. “Gesprächsanalytisches Transkriptionssystem.” *Linguistische Berichte* 173:91–122.
- Sewell, William H. 1992. “A Theory of Structure: Duality, Agency and Transformation.” *American Journal of Sociology* 98:1–29.
- . 1999. “The Concept(s) of Culture.” Pp. 35–61 in *Beyond the Cultural Turn* edited by Victoria E. Bonnell and Lynn Hunt. Berkeley: University of California Press.
- Schutz, Alfred. 1971. “Making Music Together: A Study in Social Relationship.” Pp. 159–78 in Alfred Schutz, *Collected Papers, Vol. II*. The Hague: Martinus Nijhoff.

- Smardon, Regina. 2004. "Streetwise Science: Toward a Theory of the Code of the Classroom." *Mind, Culture, and Activity* 11:201–23.
- Smith, Dorothy E. 1987. *The Everyday World as Problematic: A Feminist Sociology*. Toronto: University of Toronto Press.
- Tobin, Kenneth. 2006. "Aligning the Cultures of Teaching and Learning Science in Urban High Schools." *Cultural Studies of Science Education* 1:219–52.
- Tobin, Kenneth, Rowhea Elmesky, and Gale Seiler. Eds. 2006. *Improving Urban Science Education: New Roles for Teachers, Students, and Researchers*. Lanham, MD: Rowman & Littlefield.
- 't Hart, John, Rene Collier, and Antonie Cohen. 1990. *A Perceptual Study of Intonation: An Experimental-phonetic approach to Speech Melody*. Cambridge: Cambridge University Press.
- Turner, Jonathan H. 2002. *Face to Face: Toward a Sociological Theory of Interpersonal Behavior*. Stanford, CA: Stanford University Press.
- Yaeger-Dror, Malcah 2002. "Register and Prosodic Variation, a Cross Language Comparison." *Journal of Pragmatics* 34:1495–536.
- Verhoeven, Jo and Peter Mariën. 2004. "Prosody and Foreign Accent Syndrome: A Comparison of Pre- and Post-Stroke Speech." In *Proceedings of Speech Prosody* edited by Bernard Bel and Isabelle Marlien. Available at [http://www.isca-speech.org/archive/sp2004/sp04\\_123.pdf](http://www.isca-speech.org/archive/sp2004/sp04_123.pdf) (Accessed July 24, 2007)
- Vygotsky, Lev S. 1986. *Thought and Language*. Cambridge, Mass: MIT Press. (First published in 1931)
- Vallas, Steven P. 2006. "Empowerment Redux: Structure, Agency, and the Remaking of Managerial Authority." *American Journal of Sociology* 111:1677–717.
- Weber, Max. 1994. *Sociological Writings*. New York: Continuum.

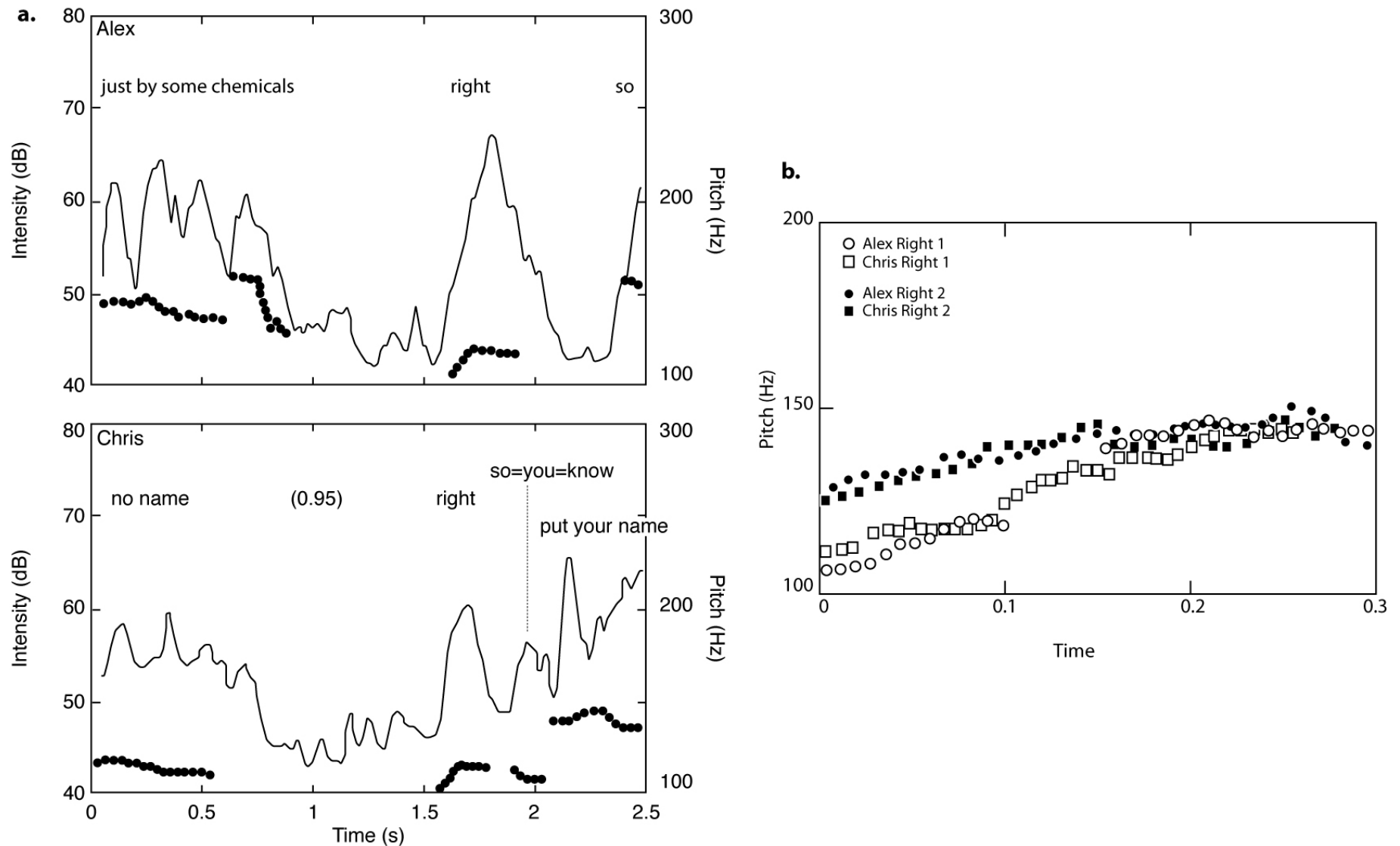


Figure 1. After working with Alex for a while, Chris has “picked up” some of the utterance features characteristic of Alex’s speech. a. Alex utters “right” preceded by a pause and followed by brief pause and a jump in pitch (●) with the next utterance; Chris exhibits the same speech-pause-speech and pitch contour pattern (“right”). b. A direct comparison of time- and pitch-normalized articulations of the two forms of “right” shows how similar the speech productions of the two have become.

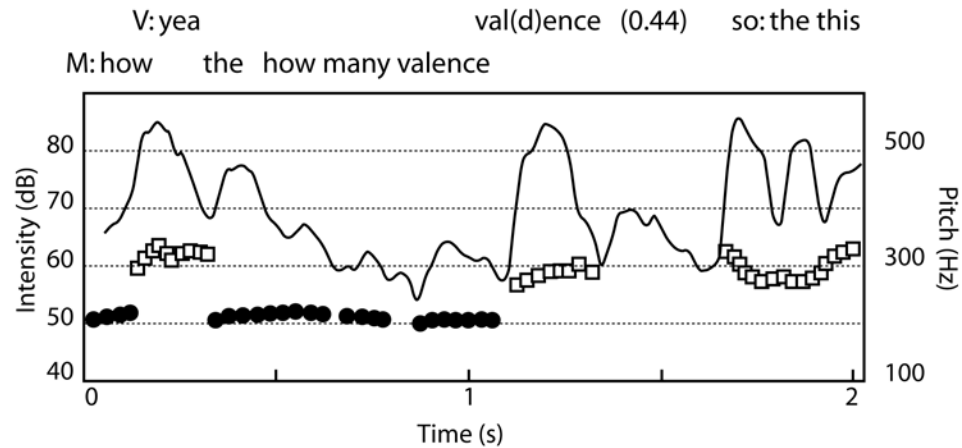


Figure 2. Victoria, an immigrant new teacher, is in conflict-laden relationship with her mentor teacher Alex (Cuban African American). Her pitch ( $\square$ ) does not align with his, nor does she “pick up” features of his speech—the mean difference between the two speakers exceeds 100 Hz. The differences are also observable in interactions with students, here Mirabelle ( $\bullet$ ), where pitch continuity observable in successful classes does not appear.

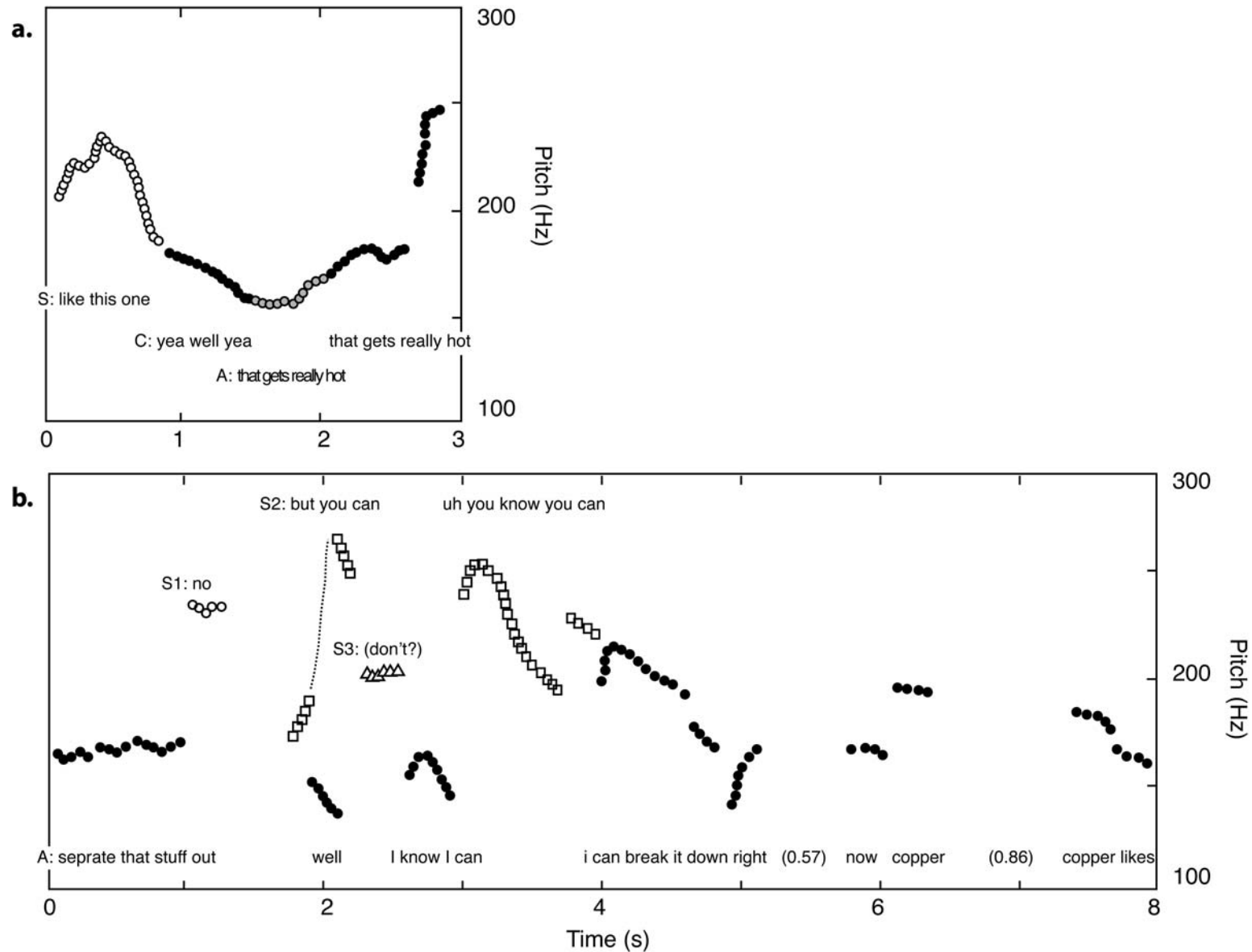


Figure 3. a. In the course of coteaching with Alex, Chris came to align pitch levels with students and Alex alike. b. Here, pitch discontinuity signals disagreement on matters of fact and opinion against a background of pitch levels always adjusted to the preceding speaker.

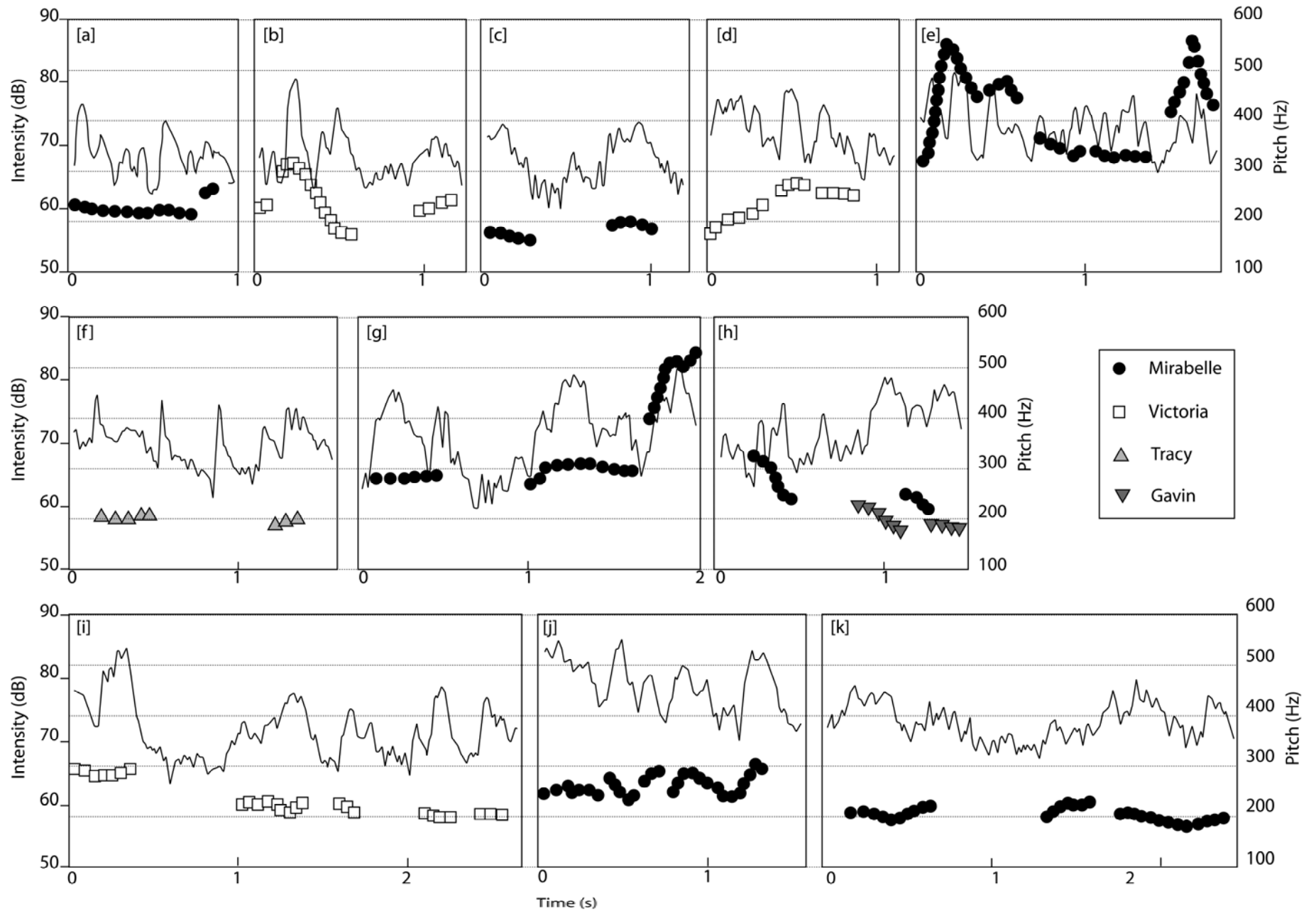


Figure 5. Classroom conflict and resolution are correlated with rising (“heating up”) and falling pitch levels (“cooling down”).



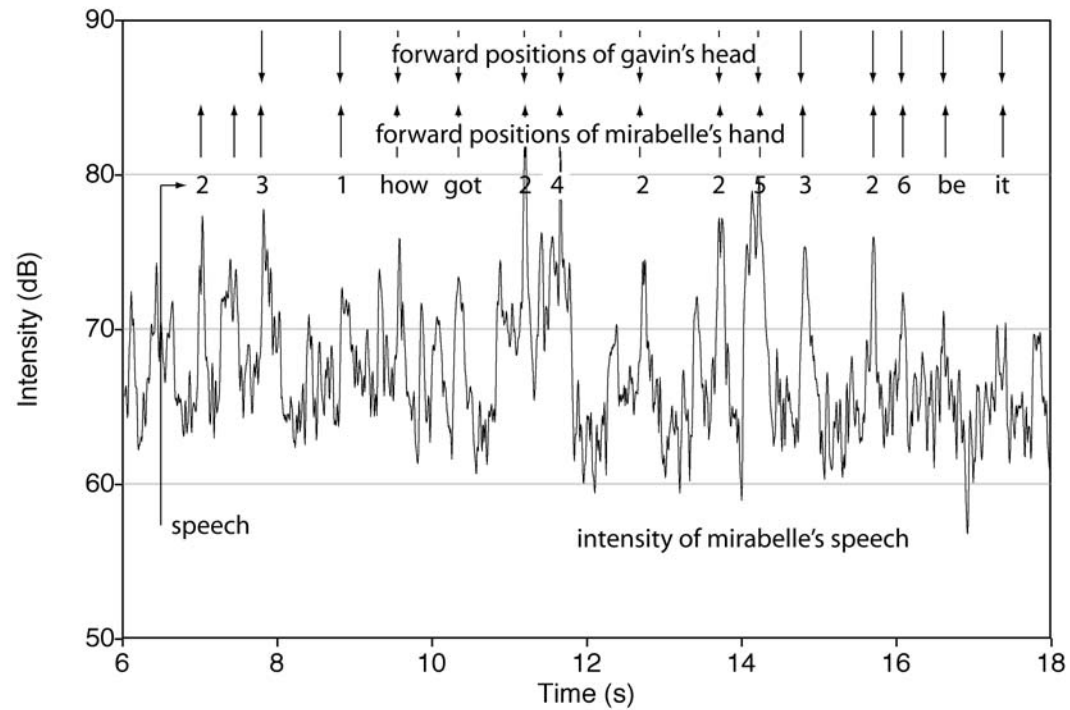


Figure 8. Mirabelle vocally produces a rhythm that she also produces gesturally; Gavin, who cannot see her, precisely reproduces the same rhythm.



Figure 4. Seating arrangement of some of the key players in the episode.



Figure 6. Mirabelle's emotional engagement can be read, among others, from the way she uses her body, arms and hands, to direct the attention of others.



Figure 7. Mirabelle produces a beat gesture ending in the forward position precisely with the utterance of the result of each calculation.



Figure 9. Tasha enacts a beat gesture in synchrony with the teacher's counting and action of hitting chalkboard with chalk.