Solidarity and conflict: Prosody as interactional resource in intra- and intercultural communication involving power differences

Wolff-Michael Roth, University of Victoria
Kenneth Tobin, City University of New York
Solidarity and conflict: Prosody as interactional resource in intra- and intercultural communication involving power differences

Abstract

Prosody has shown to be an important though perhaps underresearched resource interaction participants produce and use to make their face-to-face interaction work. Here we investigate the role of prosody in intra- and intercultural communication in situations of institutional differences that are resources for the production of power differences. The corpus was constituted from two large databases: (a) seven years of research in large inner-city school of a major U.S. metropolitan area, where teachers of different cultural backgrounds taught together in different cultural combinations science classes attended by African American students; and (b) an interview study involving two research assistants with 37 research scientists. The data were transcribed using the GAT transcription convention designed for prosodic analysis and analyzed in terms of pitch (F0), speech intensity, and speech rates using the PRAAT software. The use and display of prosodic resources along three major dimensions is reported. (a) Resistance to comply with the standard conversation formats—e.g., during hedging moves—is associated with diverging pitch levels; solidary efforts are associated with special efforts in matching pitch levels, even if these are out of the normal range of the speaker. (b) In situations where new teachers learn to teach in difficult inner-city schools, successful teacher-teacher collaborations are characterized by prosodic expressions that converge over time, which is simultaneously a matching of prosody characteristic of students. (c) Both within and across culture, conflict is both expressed in and accelerated by considerable increases in pitch levels, speech intensities, and speech rates; the resolution of conflict is both expressed in and accelerated by the coordination of pitch levels. These results are discussed in terms of pathos in interaction and the sociology of emotion, conflict, and solidarity.

Keywords: Prosodic variation; Power; Conflict; Emotions
1. Introduction

Conversations in and oriented toward 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. Face-to-face encounters and the conversations they give rise to are so important, because they are the sites that mediate between individual and collective emotions; these serve as the glue that make anything collective such as society possible (Turner, 2002). Face-to-face encounters are based on 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: 47). Emotions thereby are articulated and communicated—and transmitted to new generations—during conversations in a variety of ways in addition to explicit emotion talk, including facial (Schimmack, 1996) and other bodily displays (Goodwin, Goodwin and Yaeger-Dror, 2002) and, important here, a variety of prosodic means (Pittam & Scherer, 1993). That is, these bodily means are pragmatic resources for articulating and communicating the emotions that make face-to-face interaction possible and intelligible.

Face-to-face conversations, therefore, are the (mesolevel) sites where macrolevel institutions are produced and reproduced as participants draw on microlevel resources that mediate and even determine the interaction at hand. Conversation participants are oriented towards these macrolevel institutions, even though they may not talk about them: scientists and interviewers participating in an expert/expert study concerning graphs and graphing in their own discipline may never speak about their situation during the think-aloud session, participating in a research project or interview—because participation has been settled at a different time and place—so that the conversation recorded contains nothing but talk about graphs. Yet all participants are oriented to producing and reproducing a think-aloud or interview session. 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 & Collins, 1990). There is strong evidence that prosody generally and pitch (F0) levels and contours specifically provide resources that enable and constrain the trajectory of any interaction including the establishment of power and status (Gregory, 1999). The production of alignment via prosodic means has been shown to be an important conversational resource in multi-activity work settings, because these means support disentangling the voices pertaining to different conversations in the same setting (M. H. Goodwin, 1996).

Given the central role of face-to-face interactions as sites of interpersonal behavior, conversation analysis constitutes a method of choice for sociologists to study the pragmatics of the production and reproduction of societal formations—school classes and interviews, among others. Conversation analysis as a field and method of inquiry has had a considerable history; the inclusion of prosody in the analysis of interactions, however, is much less common and much more recent (e.g., Couper-Kuhlen and Selting, 1996). Face-to-face interactions are ritualistic in nature and integrally tied up with emotional state of the interaction participants; this ritualistic nature and the attendant emotions are produced and reproduced by means of rhythmic synchronization at different temporal scales (Collins, 2004). Because prosody—together with facial and 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 cultural-historical forms of culture and societal institutions.

There exists considerable psychological research on the recognition of emotion. One metaanalytic study—analyzing 87 articles that describe 97 studies and a total of 182 independent samples—showed that emotions are recognized above chance within and across cultures (Elfenbein and Ambady, 2002). Across studies, an interaction was observed such that emotions are recognized with different levels of accuracy across channels (modes). For example, happiness is the most accurately identified emotion in the facial expression, whereas anger is the
most accurately understood emotion in the voice. From a cultural-historical perspective, such studies have at least two major shortcomings: First, they are disconnected from real, once-occurrence, ongoing praxis, and second, they are considered independent not only of the current activity but also of the actions. Thus, it has been recognized already in the early part of the last century that emotions are *constitutive* of societal, motive-driven activities, individual goal-directed actions, and unconscious, contextually conditioned operations (Bakhtin, 1993; Vygotsky, 1931/1986). This early work received confirmation in recent neurophysiological work that showed how particular forms of frontal lobe damage is associated with inappropriate behavior; the damage was exactly to the pathways that linked emotions to (social, material) actions (Damasio, 1994/2000). Thus, emotionality is linked up with the goals of individuals who participate in societally mediated human activities that embody collective needs.

From a phenomenological perspective, too, emotions are integral to the way in which human beings orient to the world (Merleau-Ponty, 1945). But this world is not independent from the acting individual, as is clear from the concept of the person specific lifeworld to which individuals 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 identification of semiotic resources—including structures in the setting, body orientations, prosody, and the like—in the unit of analysis (Roth and Pozzer, 2006).

2. Prosody, emotion, and the microsociology of face-to-face encounters

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 the societal activity at hand (Goodwin and Goodwin, 2000). Early anthropological
work suggests that when persons interact, they tend to regulate, synchronize, and adapt their behaviors to one another (Hall, 1976). Sociologists, too, have held that interaction participants tend to harmonize their actions both at macroscopic (observable) and microscopic (unconscious) levels: “sharing of the other’s flux of experience in inner time, this living through a vivid present in common, constitutes . . . the mutual tuning in relationship, the experience of the ‘We,’ which is at the foundation of all possible communication” (Schutz, 1971: 171). Humans are emotional because this is (and always has been) a way of regulating 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: 76). Emotions are the key to increasing sociality and bonding. In this way, the micro- and macrolevel of society come to be integrated, as participants in societal forms of activity—e.g., schooling, interviewing—orient to the event as a whole, and emotions are constitutive of this orientation, those conscious goals they pursue, and the (un-, non-) conscious operations that realize these goals. The subject takes up a position in the world of his or her meaning, and the various forms of expression are means of articulating 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, 1962: 225). The meaning neither is in the word nor in the gesture; but word and gesture are one-sided and even contradictory expressions of the same meaning unit.

Emotion is understood as a complex response disposition to engage in certain classes of adaptive behavior; the different emotions are characterized by distinctive states of physiological arousal, feelings, 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; satisfaction not only allows people to rest but also to strive for important goals (Kemper, 1987). In both these examples, emotions (a) have short-term, momentary regulative function determining the (un-, non-)
conscious elements that constitute goal-oriented actions and (b) are resources that mediate the selection of long-term goals and participation in particular forms of societal activities. Even the most rational and calculative individual, interacting with others because of the potential payoffs, are 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 emotions. For example, in power status models of emotional mediation of micro-interactions, 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, which activates a defense mechanism because the self is threatened. These emotions not only drive but also exhibit themselves in behavior.

It is well established that the speech signal 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. The nature of speech production and the human vocal apparatus articulate emotional and other nonlinguistic information in several ways. The fundamental frequency of phonation (in the literature referred to as pitch or F0), 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 elevated 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 correct identification of simulated emotion (Protopapas and Lieberman, 1997). There are indications that mean and maximum pitch levels and pitch contours are the most important pitch related measures that convey emotional information. Anger in particular appears to be encoded through frequency (pitch), though often speech intensity also is correlated with this form of emotion (Pittam and Scherer, 1993). This review of
studies also provided evidence that associated with anger are increases in high-frequency energy, downward-directed F0 contours, and increasing articulatory rates.

Particular phenomena of social interaction are correlated with prosodic features of communication. Thus, abrupt-joins—i.e., within-turn changes in the trajectory of conversational topics—are associated with increased pitch, increase in 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 outcome of which secures a speaker space for more talk beyond the completion of the unit with the pre-abrupt-join utterance. As a result, the speaker achieves a multi-unit turn. The prosodic features contribute to securing the shift in the transition to another topic within the same turn.

One study showed how during the production and reproduction of interviews—in particular interviews with a variety of celebrity and CNN’s journalist Larry King—differences in the power and status of the interviewee are associated strongly with convergent features of their pitches (F0) (Gregory, 1999). Interviews rated (by undergraduate students) high in power and potency—including such factors as loudness, dominance, toughness, aggressiveness, hardness, activity, strength, and intensity—also show the highest convergence values with respect to the pitch levels. Whereas this may appear counterintuitive, the results have been interpreted in terms of how well interviewer and interviewee are able to accommodate the relative power or status relations between themselves and as a result their F0 (pitch) levels converge. One partner is thought to set the pace, whereas the other tends to comply. Pitch (levels, contours) therefore comprises important social information that interaction participants may use as resources for conducting the societal activity at hand, including the management of status.

Prosody also appears to be an interactional means 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 in 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 use dramatic intonation contours whereas the former use less extreme pitch maxima, less pronounced pitch contours, and less durational expansions.

3. Ethnographic backgrounds, materials, and analytic methods

In this study, we draw on two data corpuses. The first is part of a research program on graphs and graphing, involving practicing, mostly university-based scientists invited to do expert interpretations. The second corpus was established in the course of a seven-year ethnographic effort to understand enculturation into, knowing, learning, and learning to teach science in urban (inner-city) schools.

3.1. Data corpus 1: Induction of new teachers by means of coteaching

Coteaching is a method by means of which new teachers are inducted into teaching by working at the elbow of a more senior teacher. This study was designed to understand the microprocesses that lead to fluid coteaching, which we generally observe after a couple of months of working closely together.

3.1.1. Institutional and ethnographic context

As the director of teacher education of a university located in an urban area, the second author employed coteaching as the model for science teacher preparation at his institution. The new teachers were assigned in large numbers to a nearby urban school, City High School (CHS), where they taught in configurations from two to four, sometimes on their own, but often in direct arrangements with a resident teacher. The data sources we draw on in this study derive from a seven-year ethnographic effort of coteaching configurations conducted at CHS. Together with ethnographic observations, we have established an extensive database containing videotaped lessons, written productions of participating new teachers, and tape-recorded interviews with various stakeholders.
CHS is an urban school located in a large city of the northeastern US. The school has more than 2,000 students, 98 percent of whom are of African American descent and more than 90 percent from poverty-stricken or working class families. The school is organized into small learning communities (SLCs), 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 do not come to school for a variety of reasons often directly related to poverty and other home conditions. Achievement levels on state-level examinations are consistently below average. New teachers, even with previous teaching experience, initially find it very difficult to work at CHS.

3.1.2. Classroom contexts

One ethnographic study and the associated recordings pertain to Alex, the science teacher of Cuban-African origin in one of these small learning communities, to whom were assigned Victoria (Hawaiian American) and Jessica (white), new teachers involved in a graduate program in science teacher education at the University of Pennsylvania. They were assigned to coteach with Alex because, like him, they have strong backgrounds in science and are seeking certification to teach chemistry. Coteaching in this study was situated in the first-period-of-the-day chemistry class, which 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 (24). The physical classroom was spacious, created by 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 lab 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 he 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 interactions 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.

The second ethnographic study and the associated recordings also involve Alex and a new teacher (Chris, white) assigned to coteach with him. Alex has five years of experience teaching science in urban schools in Miami prior to joining the CHS staff. This study was conducted in his third year of teaching at CHS. During his first year, he had experienced significant difficulties despite his previous successful experience as a teacher in another urban school in the southeastern US and his cultural history that included living in large metropolitan areas in the northeastern United States. Chris, enrolled in the same teacher preparation program as Virginia and Jessica, previously had received a B.Sc. in biology, and after two years of graduate work in biology, decided to pursue a teaching career. The excerpts 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 ready students for the laboratory activities in which they would construct and conduct measurements with electrochemical cells.

3.1.3. Data sources

As part of our research, 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 cogenerative dialogue sessions, 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 at all times and recorders are placed on various student desks to assure that all contributions to whole-class conversations are recorded and available for analysis.
3.2. Data corpus 2: Think-aloud study of expertise in graphing

This study draws on one particular example from a database containing interviews and think-aloud protocols with thirty-seven scientists working either in a university or in the public service in British Columbia. There were fourteen ecologists and twenty-three from physics-related departments (e.g., geophysics, astronomy, physical chemistry). Four individuals had previously obtained a masters degree and were currently enrolled in a Ph.D. program. The remaining individuals had obtained a Ph.D. degree between 0 and 42 years prior to our interviews. All but two research associates were involved in teaching graduate, undergraduate, or laboratory courses. An undergraduate research assistant as part of a work term in his cooperative program in physics and a Ph.D. student (M.Sc. in biology, MA in sociology) conducted the interviews with the scientists in their respective fields. For the purposes of this study, we exemplify our results with data from the session conducted by the undergraduate physics student Daniel with Anne, an associate professor in a department of physics and astronomy who had received her Ph.D. more than 30 years prior to our study. At the time, she regularly published in academic journals and was acknowledged as an excellent teacher in the science faculty as evidenced by teaching awards. Daniel attended a coop program. Seeking employment and being interested in anthropology, he was hired and trained specifically for eliciting graph interpretations from physicists pertaining to graphs from within and outside their field (i.e., ecology). In his program, he was near the median in terms of achievement. He liked the experience doing social science research so much that he first took courses in qualitative research methods and then changed his major to anthropology.

To allow comparison of expert performance in different settings, we have adopted a standard set of graphing tasks that are used both in our laboratory and field studies. The graph featured in this study displays a plate displaying birthrate and death rate as functions of population size (density); the two curves intersect twice (Figure 1). As inscription, the graph constitutes a model of the type that ecologists began using during the 1950s and 1960s. Along with 25 percent of the
biologists in a previous study on graphing, the physicists had trouble correctly interpreting the population dynamic to the right of the second (upper) intersection. Thirteen (62 %), including Anne, incorrectly suggested that the population would crash and become extinct rather than merely decrease until the equilibrium stage.

3.3. Data analyses

All relevant video are digitized to make them available for analysis using iMovie HD (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 enhanced by 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 research assistants, because of the high fidelity with which they capture student contributions to the conversations in the science lessons.

The transcripts of selected lessons are then 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 are used to work with the soundtrack, for increasing the
gain to amplify the volume, which improves the hearing of doubtful words, for measuring the
pauses using the waveform display of the sound, and for establishing pitch (F0) levels and
contours, which are clearly visible in the display of a particular sound.

Following the precepts of interaction analysis (Jordan and Henderson, 1995), we begin by
analyzing data collectively, in sessions lasting about three hours, working second by second
through the tapes. Our analysis proceeds in the following way. The researcher currently at the
controls runs the video until someone requests to stop 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 stop points out what is salient to him, 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. The other person also provides his description and
interpretation. The episode is discussed until both analysts feel that there is nothing more to say
about it at the 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 was interactional 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, commented 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.

4. Prosody as marker of resistance and accommodation

There are suggestions that during conversations, participants become aligned in speech and
prosodic features. For example, during interviews, some research suggests that interviewer and
interviewees become aligned in the way energy is expended in the sound spectrum generally but
with respect to pitch (F0) in particular, which in fact is below the range of articulated sound (Gregory, 1994). More so, alignments and misalignments of prosody have shown to be related to power and status. Accordingly, lower power and status persons become entrained into the pitch frequency of more powerful persons so that there exists a convergence between these levels (Gregory, 1999). Yet the data from our different corpuses suggest a different picture. Thus, pitch (F0) alignment and misalignment rather than being signs of power and status in fact are situated resources used for practical purposes at hand. As the following, typical example from the corpus shows (substantiated by other episodes in the subsequent section), resistance and difference are characterized by pitch differences, noticeable in particular when the base frequencies of the participants are noticeably different, and pitch comes to be aligned when there is agreement or one person exhibits (e.g., at verbal level) a particular intent to accommodate the other.

In the same way as her peers, the physics professor Anne is completely oriented toward the graph. She rarely looks up, even when she asks a question. While she talks, both making descriptive and interpretive statements and questioning her two hands are positioned on the graphical feature or in the area of the sheet that constitutes the current topic of talk (Figure 1). Sometimes she follows a feature using pencil or finger, thereby iconically reproducing the shape of the indexed entity. At other moments, she merely points to a specific feature or toward a general area on the graphical display. Our first episode follows a lengthy first articulation of the sense Anne made of the graph. Subsequent to a brief pause, Anne asks whether she is right (turn 01). In the context of the think-aloud session that Anne has agreed to participate in, asking whether she is right certainly is unexpected and perhaps surprising. There is a 1.10-second pause, followed by Dan’s production of sounds that serve to both acknowledge his hearing of the question and stall on providing an immediate response.

**Episode 4.1 (250)**

01 A:  is that right?
02  (1.10)
03 D:  uh[m:: ] [khm ]
04 A:  [am=I loo[k]=i]’m looking at this region here ((pencil circles region below intersection 1)) ↓now. ((200 to 180 Hz))
05  (0.30)
Anne articulates what can be regarded as a clarification of the question, specifying which of the three regions—below, between, and above the two intersections (Figure 1)—she is referring to (turn 04). There is a brief pause (turn 05), before Daniel provides a qualifying “I think that’s correct” (turn 06) only to question whether what she has described is what is says on the graph. A long pause—during which Anne orients her hand to the graph—follows; Daniel tags on another querying “or,” as if prodding Anne to talk. The latter takes Daniel to mean that she has to reorient (turn 10) and after an exchange about what has been said and what has been meant (turns 11–12), Daniel produces two turns designed to create uncertainty about the correctness of the answer: First he emphatically responds no, when asked about the meaning of what he has said, as if he realized that has said more than he is supposed to in this kind of activity. Second, he created uncertainty by asking Anne whether she has expressed understanding or “personal opinion” (turn 17). Anne then begins another attempt at interpreting the curve in the light of the mathematical equation for the birthrate curve. The effect of the episode is one in which Daniel either hedges or creates uncertainty about the correctness of what Anne has said despite her request for confirmation of having the right answer. These hedging and uncertainty-producing moves consistently are associated with displayed pitch levels, which tend to diverge; they can be taken by the recipient as an indication that the interaction ritual and failed, and with it the achievement of intersubjectivity and solidarity (Collins, 2004). That is, in addition to the pauses
and interpolations, the pitch levels are resources that Anne senses or interprets as resistance or accommodation. However, in contrast to the rising pitches associated with disagreement and conflict among hopscotch playing children (Goodwin et al., 2002), Daniel’s pitch remained or moved into his normal range, much lower than that which is characteristic of Anne’s voice.

Anne, who has spoken with a mean frequency of 235 Hz (max = 274 Hz, min = 174 Hz, SD = 26 Hz), drops on the final word (“now”) to 180 Hz. In responding, Daniel stays within his range (min = 121.5 Hz, max = 154.7 Hz, mean = 139 Hz, SD = 7.5 Hz). This range clearly is exhibited throughout this episode, including turn 17, which continues to seed uncertainty because it raises questions about Anne’s articulation of the graph as alternative classifications of it under “understanding” or “personal observation” are proposed. Anne, reorienting to the graph, makes an attempt at grasping a formula given for the parabolic birthrate curve in which the parameters are b₀, b₁, and b₂. As Figure 2 shows, Daniels’ discursive move, which expresses uncertainty about the correctness status, is produced with a pitch considerably below Anne. For the entire two-hour recording session, these pitch levels correspond to the normal range for Dan and are near the normal 210–240 Hz range for Anne in the early part of her continuation, with high pitch spikes on “B” and “are” (Figure 2).

«««« Insert Figure 2 about here »»»»»

In episode 4.2, we see first resistance to respond, expressed in terms of pauses and interpolations that project continuing recipientship rather than an incipient move to assume the next turn. This has the effect on the person who poses the question of creating uncertainty, which might arise from a problem in the articulation of the question or its domain of application (Roth and Middleton, 2006). Thus, when Daniel does not respond directly to Anne’s question about being right in her graph interpretation, she re-specifies the domain of applicability of the explanation (“round this region?”), an assessment of correctness she now requests (turn 07). Dan hesitates, there is a pause, then produces another interpolation, before launching into a rapidly unfolding explanation about how to go about reading the graph by subtracting the death rate from the birthrate (turn 10).
Prosody as interactional resource

**Episode 4.2**

01 A: presumably that means that the population is in increasing.
02  (0.80)
03  <<insistent>>is that right then?>
04  (0.78)
→  05 D: u:m:
06  (0.40)
07 A: round [this] rE:ption?
→  08 D: [khum]
09  (0.75)
→  10  <<f>u:m (0.18) ↑yea if <<all, dim>you take well should I think I can stay off it if you take the birth minus the death>>

The bottom panel of Figure 2 shows the changes in Daniel’s pitch level, which initially (for the interpolations) remain in his normal range (“um,” “khum”: min = 143 Hz, max = 150 Hz, mean = 147 Hz, SD = 2 Hz). Then, when he actually responds, his pitch level jumps into a higher range, in fact, into the normal range of Anne. His pitch (F0) begins at 211 Hz and then progressively decreases into this normal range in the course of turn 10 for a mean of 121 Hz (SD = 4 Hz) over the last two seconds of the entire turn (min = 112 Hz, max = 131 Hz). The broken line in Figure 2 (bottom) shows how Daniel’s descending pitch can be extended backwards, where it meets up with the pitch level on which Anne ends her utterance. That is, in his pitch, Daniel expresses approximation and accommodation only to return to his normal, probably more comfortable pitch range.

Existing research suggested that convergence of prosodic parameters (long-term average spectrum [LTAS], but specifically F0) in face-to-face interaction is observed when persons with less power or status accommodate others with more power and status (Gregory, 1999). The analyses of this data corpus generally and the examples exhibited here more specifically make salient a somewhat different picture, one consistent with the idea of prosody as an interactional resource produced and used for the purposes at hand. Thus, whenever this interviewer—who at the time also is an undergraduate student in the same department as the interviewees—exhibits hesitancy to respond, any utterances are associated with his normal pitch range, which is also lower than that of most interviewees. Whenever the interviewer actually does respond—i.e., quickly and without tokens that delay the response—his pitch moves into the range of the other
person, often joining up to produce pitch continuation. That is, acceding to and accommodating
the question is associated with a pitch that meets that of the other. In the course of responding,
Daniel’s pitch falls back into its normal range. It is evident that the effect clearly is observable
when the normal pitch ranges of the two speakers are apart rather than overlapping or even the
same.

In this situation, the question of who is more powerful is not easy to answer, as shown by the
negotiations around who knows what and what mathematical knowledge is pertinent to the
situation. Thus, Anne, enlisted in the study as an expert, asks questions about something that she
does not know; and she asks a person whom she knows to be a moderately successful
undergraduate student in her own department. Because of the association of knowledge and
power, then, Daniel is the one in the know at that point in time and in control over providing or
not providing the sought-for answer. The issue of power, however, is more complex. On the one
hand, the university-based physicists interviewed by Daniel (as the biologists interviewed by the
PhD student) institutionally are in positions of power over Daniel. On the other hand, being part
of the research team, interviewees often hold him to be knowledgeable about the graphs they are
asked to think-aloud about; here, then, the institutional relation places Daniel in the position of
more power. But as an additional dimension, the physicists have been recruited as experts; as
such, they might be expected to know more about the attendant issues than the researchers asking
for expertise. Although in this situation it is Daniel who accommodates the other more than
others accommodate him in terms of pitch alignments, it is also Daniel who withholds
judgments, evaluations, and other responses, all instances during which his pitch remains in his
normal range and does not meet that of his interlocutors.

In summary, therefore, the analysis of this data corpus—here exemplified by excerpts from
the meeting between Anne and Daniel—suggests that it may be better to think of and theorize
pitch levels as (a) expressing and being correlated with the intent to resist or accommodate and
(b) as pragmatic resources that allow others to feel or make inferences about the positioning of
the speaker. Divergence in the periodic features during face-to-face encounters run against the
ritualistic aspect of interactions and therefore are associated with misalignment and conflict, whereas convergence—e.g., pitch continuity and other alignment of periodic features—enhances mutual focus and the creation of solidarity and positive emotional energy in the individual (Collins, 2004). We may ask the question whether these findings about the relation between pitch levels and institutional relations of power hold up in other settings where teachers and students interact.

5. Production and reproduction of prosodic patterns: cultural alignment

In our research in urban (inner-city) schools, important questions are (a) what and how do teachers learn from one another if they have joint responsibility for a course as a whole and (b) which resources enable rather than constrain communicative action when there are cultural differences between the participating teachers and between teachers and students. Coteaching is a model for training new teachers, who spend a year teaching at least one course per day alongside another teacher. In the course of a seven-year ethnographic effort of coteaching within several urban schools in a large U.S. metropolis, we found that the teachers came to resemble each other in the way they comported themselves, covered the available physical space, and even, as reported below, in the deployment of particular utterances, pitch contours, and ways of joining up with students’ pitch levels. Concerning positioning in classroom, the teachers become like each other because they have to take complementary positions, which entrains them into the same rhythm but with a phase delay. Convergences therefore have been observed in the way teachers ask questions (content and structure of query) or the way they use space (Roth, Tobin, Carambo, and Dalland, 2005). Concerning teacher–student interactions, our ethnographic research shows that some classrooms in this school are characterized by less conflict than others. Difficulties and tumultuous classrooms are observable especially when there are cultural differences between teacher and students, even if the difference is as small as that between African American and Cuban African. Conflicts—as shown, among others, by misaligned prosody—were observable especially in the first year of a teacher’s appointment in the school
and subsided especially quickly when a new teacher cotaught alongside another teacher with experience in the school. In this study, we are concerned with the developments observable at the prosodic level.

5.1. **Alignment of utterance/prosody features**

As part of our observations, we noted that new teachers working for some time with experienced teachers tended to “pick up” expressions, grammatical features, and intonations. Our interviews reveal that participants are not normally conscious that these parameters of their communicative actions have changed. 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 resident teacher. Here we provide two typical cases; the examples come from a chemistry lesson and it exhibits the similarities in ways of speaking that are apparent after working together for three months.

First, Alex uses certain phrases and an oral grammar characteristic of the students’ 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 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 found that in a given lesson Alex often uses the phrase and that it is customary for coteachers in the class to begin using the term with increasing frequency in their interactions 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: their 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 the 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 very bubbly and energetic in the same
way his students are. 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 3).

**Episode 5.1**

01 A: An element is not something you can just break down (0.24) in anything else (0.55) just by using CHEMicals. (0.72) `RIGHT- (0.28) so copper (0.61) is n element

**Episode 5.2**

01 C: A lot of times when we do labs we leave our papers no name (0.95) `RIGHT- (.). <<all>so=you=know> put=your=name keep it for- for the test.

The direct comparison provides evidence for the pausing immediately prior and subsequent to the utterance of “´RIGHT-”; the pitch and intensity contours are nearly identical (Figure 3). As part of our ethnographic work, we had noted this pattern of speaking several years ago in Alex’s teaching. In that study it was apparent that in his successful coteaching experiences, interns began using it while teaching with him. In this lesson, Chris exhibits a similar way of teaching and exhorting students to think about what was asserted. There is actually 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. This pattern both 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 them now without having been aware of acquiring this manner of speaking. 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 5.3 (375s)**

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 5.4 (884s)**

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=ca:n=see=it=up> the:re ((points to
07 periodic table on the wall)), is an element. (0.62) <<f>i
08 if 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
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 research, which followed many coteaching configurations in the course of
seven years, shows general convergence in these and other patterns. 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 had to reassign the new teacher to work in a different class—such convergence was not
observed.

This finding is 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. However, in the case of coteaching, the
situation is more complex as the teachers do not merely interact with one another but in fact interact with students in the presence of one another. The research hypothesis we therefore pursued was that the new teachers become like their partners 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.

5.2. Alignment of utterance/prosody features

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—frequently find themselves in conflictual situations especially with students. Our mesolevel ethnographic descriptions often contain the terms “out of sync,” “disharmony,” and “head butting.” It turns out that the pitch levels of these teachers are different from those students display; there is no continuity or similarity in other prosodic parameters. For example, Figure 4 displays the comparison of the pitch levels involving Virginia, a Hawaiian American new teacher and Mirabelle, an African American student. There is no convergence, even though, based on research linking power and status to prosodic features, suggest 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.

In functioning and low conflict or conflict-free coteaching configurations, the new teacher tends to become like the more like their more seasoned peer. This is particularly observable in the way the 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 existing research (Gregory, 1994). When difference is articulated in this type of situation, it is by means of the same device that Daniel displayed in the interview situations: the pitch remains or moves into a lower register.
In his interactions with the students, both within and between lessons, Alex always seems to find the “right tone” with students whatever the situation. In addition to the word choices and grammatical features of his language, which are those of the students’ culture, an analysis of pitch in consecutive turns shows that Alex matches prosodic features, including loudness and pitch to the preceding turn produced by a student. An example of such matching is provided in the following episode, pertaining to the conversation about the relative hotness of candle and welding-torch flames.

Episode 5.5
01 A: <<cresc><h>when we did the ca=andle?>>
02 S: yea?=
03 A: =&downarrow<<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 had answered at a much lower level, moving from 110 to 130 Hz in the course of uttering “yea.” When Alex continues after the student, Alex (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 and 190 Hz. Furthermore, the intensity levels show a similar pattern. The peaks of Alex’s intensities moved from 64 to 74 dB but, after the student has responded speaking with a much softer voice (60 Hz), Alex decreases speech volume to 64 dB and moves to 67.5 dB for each of the three syllables of the utterance “ca=andle.”

Episode 5.6 (626)
01 S: hold in the fI[re] [fo]r=a=long ti:me.
02 C:               [ri]ght
03  <<dim>right but=the pEnny=s only at five hundred degrees for that whole time.>

The same patterns are observed for Chris after having cotaught at least one lesson per day with Alex. In episode 5.6 he interacts with a student about heating a penny. The student suggested that holding the penny in the flame for a long time would lead to a different result. Chris’ 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 listener
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 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.

Here, both teachers adjust their volume and pitch to match the levels of the preceding student. This ultimately leads to pitch continuity across multiple turns involving teachers and students; this phenomenon is exemplified in episode 5.7 and the associated representation of pitch in Figure 5.b. In this situation, Chris affirms the statement that a student has made, matching in his pitch level the pitch with which the student has ended. 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, matching his pitch only to rise and then displays a jump in pitch, which emphasizes the terms “really” and “hot” (turn 04).

<table>
<thead>
<tr>
<th>Episode 5.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 S:</td>
</tr>
<tr>
<td>02 C:</td>
</tr>
<tr>
<td>03 A:</td>
</tr>
<tr>
<td>04 C:</td>
</tr>
</tbody>
</table>

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’ 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, Chris exhibit a preference for 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 290–315 Hz on “cuse,” to drop descending contour from 105–95 Hz. The pitch then rises from 137 to 215 Hz on “me,” which can be heard as an exclamation. The subsequent “Hey!” rises from 275 to 348 Hz. In both situations, the high—when compared with Alex’s normal pitch level between 150 and 180 Hz—pitch signals disapproval.

A final example for the fact that Chris has been becoming like Alex can be seen from the fact that in turns involving both and some students, the pitches are matched such as to constitute a continuity in the parameter. When difference is articulated verbally, the corresponding pitch levels tend to be much lower, in or near the speaker’s normal range, similar to the pattern we observe in Daniel’s hedging moves produced whenever Anne queries him early in the session. In the following episode—almost entirely represented in Figure 5.a 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 5.a), consistent with the way in which Daniel expresses difference (reluctance) in the previous section.

**Episode 5.8**

→ 01 A: could i (0.42) sep=rate that stuff out?
→ 02 (0.50)
→ 03 S1: no-
→ 04 S2: but [you can]
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 5.a) 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 him for some time. More so, Alex not only matches the pitch levels of students but also reproduces pitch contours that previous student speakers produce. This, too, is evident from Figure 5.a, 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 this corpus 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 (seasoned) teacher displays. Most importantly, therefore, teachers match the cultural patterns in the prosody; and new teachers, independent of their culture, 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 are highly but negatively correlated with the physical distance between the nations of the emotion expressing and perceiving individuals (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 data corpus suggests that culturally different individuals not separated in space and part of the
same activity system may not be disadvantaged in the recognition of emotion despite cultural differences.

These findings differ from the studies concerning the relation between prosodic (pitch) features and power and status. But they are consistent with an unpublished study following a clinical psychologist through her work with 24 schizophrenic patients showed that with training and experience, her pitch values increasingly converged with those of her patients (Gregory, 1999). That is, as the quality of her interactions with patients increased, so did the convergence of the pitch (F0) values. On the other hand, for doctors in the course of their training in specialties, the distances increased, which has been interpreted to be the result of the decreased focus on interpersonal relations and an increased focus on the technical aspects of the case at hand. The present results therefore 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 the subsequent speaker to “chime 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 may create a bridge between two turns that could not be achieved by verbal means alone. They also signal solidarity and a common mood (Damasio, 1999).

6. “Heating up and cooling down”: Conflict and resolution

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 leads to the convergence of speech patterns and prosodic displays; prosody, as other nonverbal communicative means—is a resource for interaction deployed for pragmatically dealing with issues at hand. If this is the case, then the divergence and convergence of prosody would be associated with practical action in more complex ways than along differences in institutional
positions. The episode analyzed here exhibits and exemplifies precisely this feature: increasing pitch levels when differences appear in the content of talk—rather than the decreases observable in section 4 and 5—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 trump the commitment made before. On the other hand, speakers calm the situation when their conversational contribution is produced with lower speech volume, pitch, and speech rates. These changes in prosody with respect to conflict and its resolution are articulated and exemplified in the following episode, one of numerous 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, check on homework, wanting to pass, and avoid detentions that would keep her off the basketball team—has an idea and attempts to explain it. Victoria 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 in the classroom at the time of this episode, his participation is not central. The interactions involving students and Victoria are most salient in this vignette, especially interactions between Victoria, Mirabelle, and three of her peers. Mirabelle sits near the back so that most students cannot see her without turning around (Figure 6).

Figure 6. Seating arrangement of some of the key players in the episode.
Just prior to the conflict beginning to articulate itself, Victoria has presented a “trick” for figuring out the number of valence electrons an atom of a specific element has given a periodic table of elements. Mirabelle then announces that she has figured out a systematic way that one can use 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 6.1 picks up at this point (the underlined speech elements are represented in terms of pitch and speech intensity over time in Figure 7, each alphabetical label referring to the panel of the same letter).

**Episode 6.1**

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 been speaking before, Mirabelle restates that there is another way to figure “it” (valence) out. After a considerable pause—nearing the one-second mark of the standard maximum silence in conversation (Jefferson, 1989)—Victoria utters in a determined fashion, “this IS the way to do this.” As Figure 7.b shows, the pitch rises be 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 (Virginia) way, the one she just explained is the only way to remember valence, that is, the topic at hand. In the background, Alex comments calmly and with low speech intensity, “I hope so,” thereby both sustaining Virginia’s claim but also providing a resource for avoiding a heating up of the possible conflict that announces 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.

In turning their bodies and heads to look squarely at Mirabelle, the three peers exhibit their expectation that something is to be forthcoming. 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 the students, the teacher perhaps has articulated a challenge, which Mirabelle, to preserve her social capital with the peers, has to take up. Until this point, however, Mirabelle has stayed calm despite the determination and irritation that Virginia displays in her voice. Contrasting Figure 7.a and Figure 7.c 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.

Tracy is the first to ask Mirabelle in a low pitched and low intensity voice, where the two was coming from the latter has been subtracting from the atomic numbers. Episode 6.2 begins as Victoria repeats the question but, as Figure 7.d. shows, her pitch is rising near 300 Hz and speech intensity is increasing 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. At this point, Mirabelle launches her body forward, raises her voice and her pitch repeatedly moves to 500 Hz and beyond as she utters, “I’m just saying, just do the number two” (Figure 7.e). Her frustration is apparent. Mirabelle continues in an accusatory way that her teacher did not want her to copy text, with the possible implication that what she (Mirabelle) searches for is understanding and having a surefire way to remember valence. The frustration is also apparent from the way she
Prosody as interactional resource

uses her body to direct others’ attention, for example, to a particular place in the periodic table (Figure 8). At this stage, her speech rate has increased from an average of 5 syllables per second to over 8 syllables per second.

**Episode 6.2**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>01 V:</td>
<td>Where is the ‘two coming from.</td>
</tr>
<tr>
<td></td>
<td>02</td>
<td>(0.52)</td>
</tr>
<tr>
<td></td>
<td>03 Ta:</td>
<td>&lt;&lt;p&gt;from.&gt;</td>
</tr>
<tr>
<td></td>
<td>04</td>
<td>(0.19)</td>
</tr>
<tr>
<td>b</td>
<td>05 M:</td>
<td>&lt;&lt;f&gt;I'M JUST SAYing just do the number ‘TWO: (0.14) ↓1:</td>
</tr>
<tr>
<td>c</td>
<td></td>
<td>if you=don=t=wan=me=da copy text from noWHERE, or you can take</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>t=the two come from the (0.16) the ‘ROW,&gt; I mean</td>
</tr>
<tr>
<td>e</td>
<td></td>
<td>(0.42)</td>
</tr>
<tr>
<td>f</td>
<td>06</td>
<td>(0.42)</td>
</tr>
<tr>
<td>g</td>
<td>07 Tr:</td>
<td>&lt;&lt;p&gt;that won't WORK f'r (0.57) for=ALL=of them.&gt;</td>
</tr>
<tr>
<td>h</td>
<td>08</td>
<td>(0.42)</td>
</tr>
<tr>
<td>i</td>
<td>09 V:</td>
<td>it=doesn't=work for=ALL of them. (max 238, ending 185)</td>
</tr>
<tr>
<td>j</td>
<td>10 M:</td>
<td>‘well `what. you &lt;&lt;f&gt;want&gt; then the &lt;&lt;f&gt;↑’TWO&gt;</td>
</tr>
<tr>
<td>k</td>
<td>11</td>
<td>(0.30)</td>
</tr>
</tbody>
</table>

Figure 8. 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.

Following the outburst, Tracy, who has turned around, to face Mirabelle, suggests in a very low intensity voice that this way does not work for “all of them”; her pitch was more than 200 Hz lower than the final syllable Mirabelle has uttered (Figure 7.f). Virginia’s restatement that Mirabelle’s method does not work for all instances, the pitch has returned 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), yet again moves beyond 500 Hz when she comes to the point of the number “two,” also associated with an increase in speech intensity to above 80 dB (Figure 7.g). Her speech rate,
however, has decreased to 3.5 syllables per second. The question “well, what?” already appears to be more conciliatory, and this possibility for a resolution can be heard in the prosody.

The voice intensity and sharply rising pitch at the end of the previous episode 6.2 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 have further calming effect. Thus, Mirabelle still begins turn 04 of episode 6.3 above 300 Hz but progressively descends, following the pitch of Gavin (moving considerably above his range), who overlaps hers and then drops progressively to a value below 200 Hz (Figure 7.h).

**Episode 6.3**

<table>
<thead>
<tr>
<th>Turn</th>
<th>Time</th>
<th>Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>0.30</td>
<td>(then? then? you're gonna be)</td>
</tr>
<tr>
<td>02</td>
<td></td>
<td>but once &lt;&lt;dim&gt; we start getting in here it doesn't work anymore&gt;</td>
</tr>
</tbody>
</table>
| 03   |      | i ain't talking about ((points to the left part of the periodic table))
| 04   |      | [this I'm talking about that] ((points to the right part of the periodic table)) |
| 05   |      | [no she talking about] row, |
| 06   |      | and according to rows it says, the (simple?) rows it checks (??) |
| 07   | 1.55 | if that's the case |
| 08   |      | [no,] no |
| 09   |      | i don't know (0.46) <<f>well [i] |
| 10   |      | 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. |
| 11   |      | alright, so n=a would be nine. |
| 12   |      | arright, so n=a would be nine. |
| 13   |      | right? so then you have nine for a long [time.] |
| 14   |      | [alright.] this is one |
| 15   | 1.51 | valence, one= |

Following two brief, partially overlapping turns, Victoria begins to explain again that she provided but 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 7.i). As if she has realized during the brief 0.40-pause
the potential of her prosody to heat up again the conflict, her prosody changes to take on lower values that are experienced 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 250–300 Hz and the speech intensity repeatedly goes to 80 dB and beyond (Figure 7.j).

Additional students “come to help,” as apparent from the final episode depicted here. Thus, sitting behind 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 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 7.k). In this turn, her speech rate has returned to normal 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 6.4**

01 S: ((oriented toward Mirabelle, who cannot see her))
you sk[ip the middle. ]
02 V:    [and that's 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 don’t care about them when we’re talking about this right now.
09 (0.48)
Prosody as interactional resource

Previous research suggests that there is prosodic alignment between speakers in face-to-face conversations and that the tendency is for the person(s) with less power (derived from institutional position) and status to align himself or herself with the person with relatively more power and status (Gregory, 1994). The previous sections already provide evidence that institutional position (power) and status do not determine the deployment of prosodic features. Quite to the contrary, the evidence provided 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 interactions and interaction rituals suggest that the preferred state in interactions is the alignment of rhythmic and periodic features, because such alignment produces and intensifies positive emotional energy for participants (Collins, 2004). Considerable changes in periodic features (e.g., pitch, beat) on the part of one participant with respect to others means breaking out of the repetitions at the microlevel, and indicates that the person is not responding to the cues of others, who are frustrated in the process.

The data in this section, featuring the emergence and resolution of a conflict, provide further support to the contention that prosody generally and pitch in particular constitute interactional resources that are available and used for pragmatic purposes. More so, 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. Similarly, when Daniel hesitates or disagrees with Anne, his pitch remains in his normal...
range. On the other hand, in conflict situations, pitch levels 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 pushing each other. Conversely, the repeated contributions at lower pitch levels and speech intensity are correlated with a cooling effect that led to the ultimate defusing of the conflict.

Both Victoria and Mirabelle exhibit emotional states. It is apparent from the inflection of the voice—in addition to pitch and speech intensity—that she is annoyed increasingly with what appears to be Mirabelle’s resistance to accepting the articulated trick as the only way of remembering valences. The annoyance is embodied in the articulation that there is only one correct solution, and the one Victoria has articulated is it. The unfolding interaction shows that Mirabelle takes the teacher’s stand as an outright refusal or rejection of her method, even before she has articulated it. When she hears her method being questioned, Mirabelle bursts out, her pitch tripling in value, the speech intensity quadrupling, her body movements vigorous 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 any experienced danger, whatever it might be. All of the productions require higher than normal levels of energy, so that one legitimately articulates the situation as one that is charged and high 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, anger is the explosive reaction against the frustrations experienced in seeing the attempt at presenting her method for remembering and recalling valences. Mirabelle therefore produces high levels of emotional energy with negative valence; and this negative emotional energy is expressed largely by vocal means.

In this situation charged with high levels of negative emotional energy, the quiet, low-intensity and low-pitched contributions several classmates produced 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 defusing the conflict. 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 settle.

7. **Prosody as tool for entrainment: At the source of solidarity**

A major question about how humans are able to make happen and bring off the specially human forms of interaction has been answered in terms of the emotional valence of jointly lived situations and the emotional valence tied to the setting 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 interaction (Turner, 2002) suggest that emotions—articulated and therefore communicated by prosodic and other observable means—constitute the essential feature that binds individuals into a collective. This “gluing” is 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. Sociologists, social psychologists, and anthropologists concerned with time and temporality of human interactions have borrowed this notion to describe the synchronization of human behavior (McGrath and Kelley, 1986). In the previous sections, we show that speakers (unconsciously) align their pitch levels and pitch contours with other speakers become aligned. During our ethnographic work we observe other rhythmic phenomena in classrooms that align themselves with rhythmic patterns articulate and exhibited by the speaker(s). This is the case even when a member of the audience does not or cannot see the speaker. 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 Virginia and Alex.

After Mirabelle and Virginia have had their initial exchange about the method for remembering the valence associated with the atoms of a particular element, Mirabelle launches
into articulating her method. (In terms of the overall event, the following episode 7.1 immediately follows episode 6.1.) Mirabelle orients toward the periodic chart of elements in the front of the classroom and moves her eyes from Victoria to the periodic table as she counts out the atomic numbers from three to six, from each of which she proposes to subtract the number two (Figure 9). The speech intensity for the entire episode is depicted in Figure 10, which also includes the words and numbers on which the intensity spikes occur.

Figure 9. Mirabelle produces a beat gesture ending in forward position precisely with the utterance of the result of each calculation.

**Episode 7.1**

01 M: U:m (0.45) You see thrEe right; you take two away from three; (0.32) j'st gonna be One (0.16) that's how many valence electrons you got. (0.29) Take two away through four (0.32) and it'd be two. (0.50) take two away of fI:ve and it be three. (0.37) take two away of=the=six; (0.36) <<dim>and it be (0.39) <<all>t=gotta=be> and so and so on> (0.55) I lost my numbers.

In this episode, Mirabelle presents her method of arriving at valence, orienting and staking out her own ground. She does so in a rhythmic way, scanted by the rises, an initially increasing involvement (production of intensity), and, decreasing involvement as she gets into the zone on the periodic table where her method no longer works, at which time she reduces the energy in the speech, the sound fades away, and the pitch drops. When the stressed and unstressed syllables are represented using a meter notion from poetry, the rhythmicity of the speech production clearly is evident:

```
/    ^ /  ^   /     ^ /   /   /   ^
Take two away from three and you got one

/    ^ /  ^   /     ^ /   /   /   ^
Take two away from four and it be two

/    ^ /  ^   /     ^ /   /   /   ^
Take two away from five and it be three

/    ^ /  ^   /     ^ /   /   /   ^
```

In this episode, Mirabelle presents her method of arriving at valence, orienting and staking out her own ground. She does so in a rhythmic way, scanted by the rises, an initially increasing involvement (production of intensity), and, decreasing involvement as she gets into the zone on the periodic table where her method no longer works, at which time she reduces the energy in the speech, the sound fades away, and the pitch drops. When the stressed and unstressed syllables are represented using a meter notion from poetry, the rhythmicity of the speech production clearly is evident:
Take two away from six and it g’na be and so on and so on

In this situation, the ingredients for a focused encounter and the production of solidarity: bodily copresence, barrier to the outside, mutual focus of attention (periodic table of elements), and shared mood (Collins, 2004). Mutual focus and shared mood are linked through a feedback if 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 observe—in addition to expressing the basic rhythmic 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 10). That is, the cyclical hand movement visible in Figure 9 is patterned such that the foremost position and the stressed syllables fall together.

Figure 10. Mirabelle vocally produces a rhythm that she also produces gesturally; Gavin, who cannot see her, precisely reproduces the same rhythm.

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 production is available to others. With each number in the series three to six, 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 6]) also move their regard to gaze at the table. 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 10 shows, even though Gavin cannot see Mirabelle (Figure 6), his rhythm perfectly reproduces the forward position of Mirabelle’s hand, which itself is aligned with the rhythm with which the account of her method is produced. He also produced 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 peaks. 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 the resources for the coordination, prosody is the likely candidate for allowing synchrony to emerge.

In the following episode, the production of patterns of synchrony is evident when the students are entrained into Victoria’s re-articulation of her 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 7.2, two students and Victoria suggest to Mirabelle that for the rows where there are metal elements in the center of the periodic table, 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 7.2**

01 V: the maximum amount is eight, that’s 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 skip the middle. ]
07 V: [and that’s as far] as (0.25) you
[skip ] the middle
08 X: ((oriented toward Stacy))
[I wasn’t] ? 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 7.2, moves
slightly backwards than forwards, her lips silently forming “one” as her head reaches the
foremost position (Figure 11). 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 the beat of the chalk against the chalkboard. 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 with the
words. Behind Tasha, Ivory and another student begin to count at “three” and 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
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 serves 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 of social relationship).

![Beat gesture in synchrony with the teacher's counting and action of hitting chalkboard with chalk.](image)

In these episodes, alignment and synchrony are observed beyond the pitch levels and contours presented in the earlier sections. In this section, rhythmic patterns of beat gestures with the hand, rocking movements of legs, rocking head movements, beat of chalk against the chalkboard, and stressed syllables are produced and reproduced in synchrony by members across the classroom. Not only members who had the speaker in view reproduce the synchrony—e.g., students counting aloud in synchrony with the teacher—but also members who do not have the current speaker in view. It is important to note that this synchrony therefore is not mere cognitive alignment—something that is processed consciously, interpreted, and applied—but in fact is an embodied phenomenon. Such rhythmic patterns can be interpreted as baseline patterns of interactional solidarity (Collins, 2004).

There are suggestions that such beats are produced as “leaders,” individuals with more power or status, set the pace and “followers” become entrained into the rhythm (Kemper and Collins,
Prosody as interactional resource

1990). To be able to reproduce synchrony, a person actually has to anticipate rather than react to the beat, and therefore already be in alignment with the other; but the production of a beat serves as a resource for further synchrony. Whereas this might lead some to think it terms of a dichotomy of mind and body, theorizing the present phenomena do not require dichotomies. Rather, both in phenomenology (e.g., Merleau-Ponty, 1945) and cultural-historical activity theory (e.g., Vygotsky, 1986), the irreducible unit for understanding societal phenomena such as “doing school” is the activity as a whole, involving all members both in terms of their material bodies and individual and collective forms of consciousness. This activity organizes the events at all 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 that are produced across a great variety of bodily productions, not only including speech and gesture (McNeill, 2002) but also, as shown here, other rhythmic patterns produced by different body parts.

In this corpus, 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 Virginia produced is consistent with her repeated head nods following Virginia’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 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 arriving at an altercation of the kind so prevalent in the inner-city schools, altercations that frequently lead to the suspension of students.

8. Emotion, prosody, and power/status relations

Emotions are central to way in which human beings orient to the situation; and they are modified by the articulation of emotions of other participants in a face-to-face interaction (Turner, 2002). Besides facial expression, prosody is a major resource for displaying and finding emotions in the setting. There have been suggestions that participants with less power and status converge in their prosodic parameters with those of more power and status; and differences and conflict over hopscotch games are expressed 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 an interactional resource that is deployed *pragmatically* 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 social 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: 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: 216).

Our study allows us to construct four major claims. 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. In the examples
provided here, the interviewer (Daniel) modified his pitch level to match that of the interviewees when he finally acceded to a request for making an evaluation or responding to a question on the part of the interviewees. Second, when two or more teachers coteach the same classes over time, the new teachers begin 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, the teachers tended 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, similar to the phenomenon observed among hopscotch playing children. Third, raised pitch levels and speech intensity when differences in the cognitive 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 (rocking head, hand beat, rocking leg, pen hitting desk, chalk hitting chalkboard).

These results suggest the need to theorize the production of prosody differently than some previous research suggested. Consistent with the claims of at least one other study (Yaeger-Dror 2002), our research shows that different social situations permit and sometimes require variation of strategies so that prosody varies (sometimes substantially) with the nature of the social situation. 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 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 corpus 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 (unconscious) move on the part of the teacher may find its explanation in the same concept of solidarity, here exhibited by the regular teachers with extended experience in the school, who tends to speak in a way that students are familiar from their home and street cultures.

We find solidarity produced and reproduced 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 and to see how she would respond. 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.

The phenomenon of entrainment is not one describable in causal terms, 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 forward, 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 synchronization by something on the order of a second or two. Even if there were non-conscious ways of being caused into being in synchronization, he would still be behind her, for he could only know when she peaks after
having heard it. This means, Gavin has to anticipate when she peaks, which is a production of his own, which therefore requires that he is already in tune and the coincidence of his movement with those of Mirabelle is a consequence of being in tune. Such an anticipation clearly is observable at the instance when Tasha utters “three,” a moment when Virginia hits the chalkboard 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 has anticipated the correct placement of the count with respect to the noise from the chalk, but Virginia, who produces the chalk noise, is out of alignment with her speech, which only follows the rhythm of the chalk beats.

References


Figure 2. Above. Anne and Dan during a stretch of conversation in their normal ranges. Below. When Dan hesitates in responding to queries, he stays in his normal range; when he does respond to the interviewee’s questions, his pitch moves up into her range and then drops until he reaches his normal range.
Prosody as interactional resource  

Figure 3. After working with Alex for a while, Chris has “picked up” some of the utterance features characteristic of Alex’s speech.
Figure 4. Virginia, a Hawaiian American new teacher, is in conflict with Alex (Cuban African American). Her pitch (□) does not align with his, nor does she “pick up” features of his speech. The differences are also observable in interactions with students, here Mirabelle (●), where pitch continuity observable in successful classes does not appear.
Prosody as interactional resource  52

Figure 5. a. Pitch discontinuity and continuity signal disagreement and agreement. Bottom. b. In the course of coteaching with Alex, Chris came to align pitch levels with students and Alex alike.
Figure 7. Classroom conflict and resolution are correlated with rising (“heating up”) and falling pitch levels (“cooling down”).