Journal of English Linguistics
http://eng.sagepub.com/

Semiotic Layering through Gesture and Intonation: A Case Study of Complementary and Supplementary Multimodality in Political Speech

Norma Mendoza-Denton and Stefanie Jannedy

Journal of English Linguistics published online 8 June 2011
DOI: 10.1177/0075424211405941

The online version of this article can be found at:
http://eng.sagepub.com/content/early/2011/06/08/0075424211405941

Published by:
SAGE
http://www.sagepublications.com

Journal of English Linguistics
XX(X) 1–35

DOI: 10.1177/0075424211405941
http://eng.sagepub.com
SAGE

Semiotic Layering through Gesture and Intonation: A Case Study of Complementary and Supplementary Multimodality in Political Speech

Norma Mendoza-Denton[1] and Stefanie Jannedy[2]

Abstract

Face-to-face communication is multimodal. In face-to-face interaction scholars can observe the interplay of several "semiotic layers," modalities of information such as syntax, discourse structure, gesture, and intonation. The authors explore the role of gesture in structuring and aligning information in spoken discourse through a study of (1) the complementary co-occurrence of gestural apices and intonational pitch accents and (2) the supplementary co-occurrence of metaphorical gestures and elements in discourse. In the naturally occurring political speech situation the authors examine, metaphorical spatialization through gesture is key in indexing contextual relationships among the speaker, the politicians or government, and other external forces. The use of gestures simultaneously aligns with intonation and metaphorically manipulates political entities in space. Discourse context and social meaning are thus constructed together through the spoken and gestural channels and are supported through fine-grained structural alignment between intonation and gesture.

Keywords

gesture, intonation, spoken discourse, embodiment, multimodality, political speech, semiotics, public sphere

[1]University of Arizona, Tucson, AZ, USA
[2]Center for General Linguistics, Berlin, Germany

Corresponding Author:
Norma Mendoza-Denton, University of Arizona, Department of Linguistics, Tucson, AZ 85721
Email: nmd@u.arizona.edu

It has been widely accepted that cospeech gestures strongly correlate with intonation, with both of these modalities aiding in the structuring of verbally rendered discourse (McNeill 1992; Cassell, McNeill, & McCollough 1999; Loehr 2004). Yet most of the studies on cospeech gesturing analyzed in the literature were conducted using experimental elicitation or staged conversations, which give rise to gestures that are narrative and/or descriptive in nature. For example, participants in laboratory settings are asked to describe something concrete such as fish-trapping constructions (Enfield 2003) or to narrate preselected cartoon strips or films (McNeill 1992), while Loehr (2004) filmed and studied participants conversing freely on topics of their choice. In this study we analyze a short excerpt from a larger video-recorded corpus of twenty hours of spontaneous, naturally occurring data gathered at public congressional town hall meetings (THMs) in Tucson, Arizona. Speakers in THMs engage in political discourse, a task more abstract than elicited narrative, in which a speaker is trying to express a complex political viewpoint (Bohman 1996). The speaker analyzed in our data (whom we call by the pseudonym Mary-Jane) is a woman from Arizona who appears to be in her early forties; her primary interlocutor (then–U.S. House of Representatives Congressman Jim Kolbe, Republican, Arizona fifth district, served 1985–2007) is a middle-aged man in his late fifties. At the time of taping in 2001, they stand in a constituent–representative relationship to each other in the U.S. political system. In this THM, the cospeech gesturing deployed by Mary-Jane is used not only to describe the political landscape as she sees it, but also to persuade, cajole, shame, provide evidence, and otherwise convince her politician interlocutor and the audience to adopt her political point of view. The main findings in this article are (1) that cospeech gesturing correlates with intonational grammar and supports the structuring of the information that is acoustically rendered (shown previously in experimental research, here confirmed in naturally occurring interaction, and (2) that this speaker's use of the visuospatial field conveys speech-signal-extrinsic information on the relationship she posits among constituents, the government, and outside forces.

We thus hold there to be a parallelism between gesture and speech, with both channels carrying meaning on at least three different planes: structure, content, and social meaning. This layered semiotic parallelism can be supplementary, with both gesture and speech simultaneously highlighting a single message, or complementary, with the gestural and spoken channels dividing the information load to highlight different aspects of the same referent. In fact they can point at different, even conflicting information; thus, these two channels need not necessarily work toward a common goal. In this case study, contextual information about the political situation is transmitted through gesture, while concrete assertions are conveyed through speech. Discourse context and social meaning are thus constructed together through the spoken and gestural channels and are supported through fine-grained structural alignment between intonation and gesture.

The acoustic signal is an enormously rich source of structured information. From a phonetic-phonological point of view, natural languages display positional and co-occurrence restrictions of phonemes; these patterned phoneme configurations constitute the phonotactics that are typical of a specific language variety. There is a specific prosodic or

rhythmic structure, such that some parts of a word are uttered louder and longer, lending a specific syllable more perceptual prominence to indicate either lexical stress or informational salience. The acoustic signal also contains other grammatical information. Lexico-semantic content is conveyed by the choice of words spoken, and pragmatic content is affected by the context in which the words are uttered. Furthermore, the acoustic signal indexes social parameters and intentionality, carrying social information pertaining to the speaker (age, gender, ethnicity, etc.) and his or her interlocutors.

Cospeech gestures correlate with grammatical structure by featuring beats (in languages such as English and German beats are aligned with higher level prosodic prominences such as pitch accents) within gestural phrases. Beats also carry semantic and pragmatic content by being deictic, emblematic, or just emphatic beats aligned to parts of the information highlighted by the speaker. Simultaneously, these gestures transmit positional stances in terms of how the speaker recounts and relates to the world or abstract universe surrounding him or her.

Kendon (1996) thinks of gesture as a separate and distinct mode of expression with its own properties, which can be brought into a cooperative relationship with spoken utterances and used in a complementary way. Bolinger (1986:199) proposes that "gesture and speech/intonation are a single form in two guises, one visible and the other audible" and argues that gesture and speech stem from the same semantic intent. McNeill et al. (2001:23) asserts that "the organization of discourse is inseparable from gesture and prosody" and that "[they are] different sides of a single mental communicative process" (see also Enfield 2003:45-46). The results of McNeill's (1992, 2000; McNeill et al. 2001) experimental studies indicate that motion, prosody, and discourse structure are integrated at each moment of speaking.

The purpose of our research is to analyze the gesture–intonation timelines of spontaneously rendered speech in a case study of complementary presentation of information. Part of the information is presented through the speaker's verbal stream, while the broader and complementary political setting and the assumptions about political life on which it rests (what we here describe as an idealized public sphere) are presented gesturally.

We wish to highlight one well-known fact about the relationship of gesture when it is ancillary to spoken discourse: while spoken discourse has a high referential resolution, that is, it is able to pick out referents with relatively little ambiguity, gesture has a low referential resolution, so most of the information presented gesturally is complementary to speech and not recoverable solely from the gestural channel (an obvious exception to this being gestural systems that are full-fledged languages, such as American Sign Language [ASL]). Gestures, when used in conjunction with spoken discourse, have different affordances and limits to their ability to present information. Harper, Loehr, and Bigbee (2000) found gesture to be used for much more than just simple deictic functions. They explain,

> [L]anguage facilitates complex queries with the ability to express quantification, attribute and object relations, negation, counterfactuals, categorization,

> ordering, and aggregate operations. Gesture is more natural for manipulating spatial properties of objects (size, shape, and placement). (Harper, Loehr, & Bigbee 2000:3)

The gestural information stream is able to work on-line and incorporate context to take advantage of available material objects and their properties and of the spatial location of interlocutors and audience. We argue that the communicative constraints of the sociolinguistic situation in our case study maximized the need for the simultaneous presentation of information since the speaker was only grudgingly given the floor in the first place and was constrained in the time she was allotted for her turn at talk. In addition to these constraints, Rep. Kolbe constantly threatened to interrupt Mary-Jane and attempted to shift his attention away from her and to cut her off. Thus, the sociolinguistic pressure was great for the speaker to pack as much information as possible into her short turn at talk. These factors result in a naturally occurring situation where time and interactional constraints push the limits of both intonational and gestural information packaging.

Background

The Sequencing of Gestures

Gestural theory provides a useful framework for understanding the points at which gestures might be aligned. Gestural movements are described as having obligatorily one and at most five phases (McNeill 1992). The preparation phase (optional) marks the beginning of the motion in which the parts of the body involved in the gesture leave the neutral starting position and move to the position necessary for the upcoming gesture. The prestroke hold (optional) is the position of the hand and arm at the end of the preparation phase and before the beginning of the stroke. The stroke (obligatory) is the climax or peak of effort of the gesture; it is one of the most recognizable components of a gesture, and it is synchronized with the linguistic forms, such as accented syllables it is coexpressive with. The poststroke hold (optional) is the position the hand and arm remain in when the coexpressive spoken utterance is delayed. The retraction (optional) is the end of the motion in which the parts of the body involved in the gesture return to neutral positions. A beat or batonic gesture is smaller, occurring primarily for rhythmic emphasis (think of a conductor with a baton keeping the orchestra in time) and is described as having two phases which are typically flicks with the wrists or fingers. In the case of Mary-Jane, we will see that her impassioned argument occasionally motivates batonic gestures executed with her entire upper torso (see Figure 6 below).

McNeill et al. (2001) call the recurring combination of the same gestures with prosody and discourse organization "catchments." These catchments are recognized from recurrences of gesture-form features over a stretch of discourse (two or more gestures with partially or fully recurring features of shape, movement, space, orientation,

dynamics, etc.) and serve to offer clues to discourse cohesion in the text in which they occur.

Rather than assuming that gesturing serves only the interlocutor in structuring acoustically rendered information, we take the position that gesturing aids both the speaker and the interlocutor. Beattie and Coughlan (1999) have shown that gestures facilitate information comprehension, and Harper, Loehr, and Bigbee (2000) report that speakers felt hampered when asked to refrain from using gestures in three-dimensional descriptions. It has also been observed that gestures occur no less frequently when talking over the telephone though the speaker cannot be seen by the listener (Cosnier 1982; Rimé 1982) and that gestures occur in conversations among congenitally blind interlocutors speaking to each other (Iverson & Goldin-Meadow 1998), suggesting that gestures form a part of the speaker's arsenal for speech production. Gesture researchers have further shown that participants can recall information that was selectively presented in the gestural but not in the verbal channel (Kelly et al. 1999; Cassell, McNeill, & McCollough 1999; McNeill 1992) and conceptually integrate this information so as to be unable to remember the channel through which the information was presented.

We assume that the very same gestures so tightly aligned with grammar serve as a device to mediate between the speaker and the world. What does this mean? By having gestures that are ego centered and placing herself in the middle of a depiction of the political landscape, Mary-Jane sketches through gestures her view of an idealized public sphere and of the relationship it entails between a constituent and her government.

Intonational Grammar

Many descriptions of intonational systems and of the prosodic structure of typologically different languages are based on the Autosegmental Metrical (AM) framework of intonational phonology (Jun 2005). While the phonological descriptions are highly language and even dialect specific, they also share enough commonalities to make feasible the development of a single approach to prosodic modeling and intonational transcription. This common model allows for easier observation of universal and language-specific commonalities and differences among languages, greatly advancing our understanding of prosodic typology. The AM model takes a compositional approach and describes intonation as a sequence of distinct tonal events. Thus, an intonation contour (tune) is represented by a linear succession of tones that are aligned with specific syllables or with edges of phrases. In other words, in addition to the local prominences (accents), the prosodic structure of a language is also determined by the degree of juncture between two adjacent words, indicating prosodic groupings. Such phrases are often marked by tonal and/or durational events at their edges. The phonological representation of tones is mapped onto phonetic implementations, and both of these are language specific.

The grammar of intonation within the AM phonology that we are assuming for American English goes back to Pierrehumbert (1980) and was formalized in the Tones

and Breaks Indices (ToBI) intonation transcription manual for U.S. English, summarized in Beckman and Ayers (1994). According to this grammar of U.S. English intonation, there are pitch accents, intermediate phrases, and intonation phrases. Here, pitch accents are local prosodic prominences associated with a syllable carrying metrical prominence. Such prosodic prominences in English—often marked by a pitch movement—are not lexically specified in the grammar as they are for example in Japanese or Swedish. Pitch accents often serve to make some parts of an utterance acoustically and perceptually more prominent and highlight parts of the information against the stretch of unaccented material. They may, however, also serve only rhythmic purposes.

The grammar of English intonation can be summarized as follows: each larger intonational phrase (marked with "%" at the right edge) must contain at least one intermediate phrase (marked with a "-" at the right edge), which in turn must contain at least one pitch accent (marked with a "*"). Pitch accents are usually associated with the stressed syllable of a word. These accents mark local prominences above the level of the word in an utterance. There is a fixed inventory of pitch accents; they can have different tonal shapes to indicate high and low tone targets which are marked with labels such as H*, L*, or L+H* (to be read as "high-star," low-star," and "low-plus-high-star," respectively). The asterisk indicates that this tone is a pitch accent that is aligned with the prosodically strong syllable of a word. The tonal events between pitch targets are accounted for by interpolation between these pitch targets. For example, one should observe a falling fundamental frequency contour when a H* tone target is followed by a L* tone target. An intermediate phrase is a minor phrase and consists of one or more pitch accents plus a phrase tone associated with the right edge. This phrase tone can be either L- or H-. Evidence for this prosodic level comes from the phonetic interpretation of the phonology that applies to this domain (Beckman & Edwards 1994)—such as lengthening at the right edge. An intonational phrase is the largest intonational domain of U.S. English and consists of at least one or more intermediate phrases. The right edge of the intonational phrase has a phrase tone, taking the shape of L% or H%. These edge tones determine the shape of the F0 contour between the last pitch accent within the intermediate phrase and the end of this phrase. We use the ToBI transcription system to represent intonational events and to motivate our account of their alignment to gestural events.

The Public Sphere and Metaphorical Spatialization

We argue that Mary-Jane assumes a model of an idealized democratic public sphere, which is then exhibited at the level of gestures, specifically communicated by means of metaphorical spatialization. We first define the idealized public sphere, then discuss how such a model might be instantiated in gestures.

Philosopher and political scientist Jürgen Habermas (1989), in his influential work *The Structural Transformation of the Public Sphere*, argues that one of the crucial features of the European Enlightenment was the establishment of discursive democracies, crucially involving spaces where talk could be exchanged between the government

and the governed. These unprecedented fora allowed citizens to debate the news of the day and to participate publicly in political life. In Habermas's conception, the French and English coffeehouses of the eighteenth century, with their political debates, were the very first arenas in which citizens could, through talk, address matters of common concern (a matter also taken up by Gaudio 2003). According to Warren (1995:171), an idealized public sphere is "an arena in which individuals participate in discussions of matters of common concern, in an atmosphere free of coercion or dependencies (inequalities) that would incline individuals toward acquiescence or silence." Habermas argues that the idealized public sphere has long been lost as a result of the development of groups that exert an influence on public opinion but that features of the public sphere are commonly assumed in participatory democracy. Note that we are not suggesting that our respondent Mary-Jane has been reading Habermas; we refer to his account of the underlying assumptions of participatory democracy because we believe its features are reflected in her gestural modeling of government–constituent relationships.

For our purposes, one of the main features of the idealized public sphere is that it provides for direct, face-to-face dialogue between the government and its constituents. It is this space of argumentation, this dialogic form between the government and the people, that is gesturally invoked by Mary-Jane in her interaction with Congressman Kolbe. In her monologue (see the appendix for the transcript), Mary-Jane invokes her rights as a citizen and the government's responsibilities toward its people, but it is only in her gestures that she places these actors in metaphorical space and manipulates them in ways that make clear her assumptions about an idealized public sphere. We show that in her idealized public sphere, the people (represented by herself at ego–point of origin) and the government are in a dyadic discourse relationship, with shared attributes, properties, and responsibilities. Other countries and some institutions (such as prisons) stand outside of the people–government dyadic relationship and are represented as being spatially on the outer edge of the metaphorical space invoked.

Data and Method

We utilize video data collected in 2000–2001 during fieldwork for an ethnographic study of congressional THMs in Tucson, Arizona. The data form part of a larger project on political discourse, language, and power. A total of approximately twenty hours of THM data were collected at locations around southern Arizona. As THMs are public fora for discussion with government representatives, they are announced through the public media, on websites, and via fliers mailed to homes and posted in the relevant neighborhoods, in this case represented by Congressman (Rep.) Kolbe. A THM in this district normally lasted one and a half hours to two hours and in 2000–2001 was led by Rep. Kolbe, typically taking the form of an initial period of question writing by the audience on slips of paper circulated by congressional staffers. This was followed by a rehearsed monologue from Rep. Kolbe in which he stated as his aim the updating of his Arizona constituents on important happenings in Washington, D.C.

Rep. Kolbe would then select some of the questions that constituents had written out as questions to be answered without calling on anybody specifically (we may already note the divergences from Habermas's idealized public sphere). After this he often took a couple of spontaneous questions, ran to the end of the allotted time, and then invited those who wanted to talk to him further to stay and discuss matters after the official THM ended.

The video data of Rep. Kolbe and Tucson citizens that we analyze for this article were recorded simultaneously with two digital video cameras, both located to the left of the audience (from the audience perspective), one pointed toward Rep. Kolbe and the other aimed more generally toward the audience. Rep. Kolbe was outfitted with a wireless lavaliere microphone plugged into camera 1, and the audience sound was captured by a microphone mounted on camera 2.

The video was recorded in NTSC format, with roughly thirty frames per second corresponding to approximately one picture every 33 ms. The audio signal was obtained with a sampling rate of 32 KHz directly through the cameras. By aligning the sound tracks and the two videos, we were able to synchronize the videos exactly so as to gain accurate descriptions of the hand and arm movements from two different angles.

On this particular occasion, on a Saturday morning in February 2001, the THM was held in the cafeteria of a midtown Tucson school. The congressman introduced the researchers and advised participants in the THM that they were being videotaped for a research project and that their participation was strictly voluntary. If any of the constituents objected to being taped, they were encouraged to approach the researchers after the THM. None did.

We selected a particularly animated speaker for this analysis, as one of the larger interests in the research project has been the dynamics of power as expressed in the occurrence of face-threatening speech from constituents directed at the congressman. Thus, Mary-Jane is but one example in our "irate constituent series." The particular issue that Mary-Jane was arguing was the legalization of marijuana, which she supported. It is quite possible that the speech that she delivered was preplanned, since she had brought props with her; it is equally likely that Rep. Kolbe's answers were in some way stock answers. We do not know whether Mary-Jane had any prior encounters with Rep. Kolbe. However, this does not affect our conclusions on the minute complementary alignment of gesture and intonation, or on the supplementary information being presented by Mary-Jane's verbal and gestural channels. We hypothesize that any rehearsal that may have taken place would be at the level of general content, or possibly of specific phrasing. Although we cannot be certain, it seems unlikely that a promarijuana group seeking to legalize the substance would spend substantial time coaching one of its members on the microdynamics of gestural-intonational alignment.

The data we selected for microanalysis for this study last exactly 130 seconds (a reasonable amount of data within the gestural analysis literature (see Loehr 2004 for a discussion), and were transcribed according to the ToBI intonational transcription framework in addition to being subjected to a modified McNeill-style gestural transcription. Eleven tiers of body movement and gestural transcription were coded. For each

minute of verbal interaction, the linguistic and gestural annotation, transcription, and coding took about thirty to forty minutes. Transcriptions of the gestural tiers were done in teams of three as to ensure intertranscriber agreement, and the linguistic transcriptions were done separately by five transcribers and then compared and negotiated.

ToBI and Gesture Transcriptions

An audio file was extracted from the video for analysis in Praat (Boersma 2001). The researchers, assisted by trained linguistics and linguistic anthropology students at the University of Arizona, worked on transcribing the data on several different levels. Prosodic transcriptions were collaboratively made by a team of five coders who were trained to transcribe intonational events within the ToBI framework based on the Pierrehumbert (1980) system of English intonation. In cases of uncertainty or disagreement, they discussed the issue until a majority consensus was reached, yielding always at least 80 percent agreement across ratings. All prosodic transcriptions were then checked by an experienced labeler of U.S. English (one of the authors). It must be noted that our data were very difficult to annotate because of the fact that the recordings were made in a naturalistic setting.

In addition to room noises (random background noise) which show up in spectrograms as energy in all frequency regions, the audio data contain applause, coughs, and boos from members of the audience, partial interruptions by the congressman, and other noises that are unidentifiable. In Figure 1, created from our Praat display, we show a spectrogram (individual frequency bands) overlain with the fundamental frequency trace (F0) necessary for the tonal analysis and the sound pressure wave (waveform) in the middle display. This is how we coded the sequence "when did I lose the right" (full transcription included in the appendix). Note the transcription of the apices (apex) in the first tier, time aligned with the transcription of the pitch accents (tones) in the third tier.

While Praat allows a resolution to the millisecond, the video data allowed for a frame rate of only 33 ms between pictures. The video was evaluated in iMovie HD, a movie editing program that allows for simultaneous inspection of the sound pressure wave. The points and intervals of interest were evaluated from the video and the point in time was then marked in the Praat transcription tier. We recognize this small discrepancy in resolution between audio and video, but we believe the discrepancy does not change our claims about the speech–gesture synchrony.

Though there is a growing body of scholarly literature on gesture, no standard transcription system for cospeech gestures has been agreed on. There are two systems in relatively wide use; within language studies, the most prominent is McNeill's (1992) system, developed expressly for gestures. Based on gestural primes proposed by McNeill (1992), we transcribed movement of the left and right arm and the left and right hands on six tiers (range of movement, direction of movement, and palm configuration for each side). In addition, head and torso movements were transcribed as well as whether or not there was symmetry in the movements of both hands and arms.

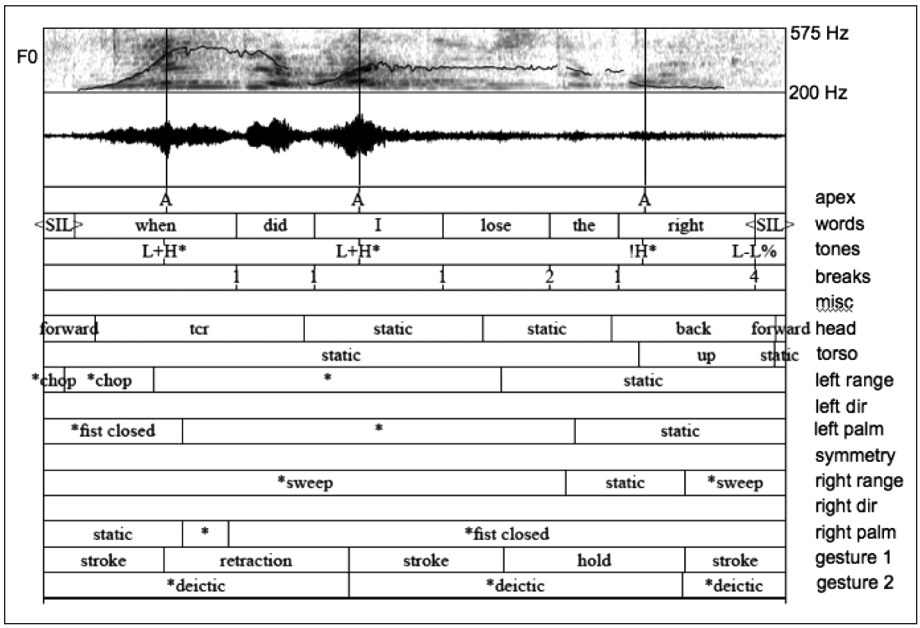

Figure 1. Praat display of multitier transcription of intonation and gestureIn this transcript bolded letters mark the syllables that are aligned with the pitch accents.

We also annotated the gestural phases that have kinetic properties (e.g., preparation vs. stroke) as well as the gestural phrases that are more semantic or pragmatic in nature (e.g., deictic vs. metaphoric). For the purpose of our study, it became crucial to mark not only intervals during which gestural movements were performed but also point events. We were immediately faced with the issue of data reduction, as twelve tiers of interval coding proved difficult to manage. As an initial step, we decided to investigate if, when, and how often pitch accents would co-occur with gestural movements. Since we had mainly coded intervals, and pitch accents are essentially point events in time as they associate with the metrically strong syllable of a word, we added another transcription level: we coded what Loehr (2004) has called the *apex*.

The apex is the point (event) in the hand or arm gesture during which the equilibrium of a particular gestural movement is reached (Browman & Goldstein 1990, 1992). In terms of our specification of the apex, we adapt Loehr's (2004:89) definition. He describes the apex as the "peak of the peak" or as the "kinetic goal of the stroke." This layer of transcription was added by the two authors: each author annotated the apices from video, and transcriptions (time points) were compared and transferred to Praat. The authors discussed cases of uncertainty and disagreement until a consensus was reached.

We follow Browman and Goldstein's (1990) Gestural Phonology account to consider apices the targets of the gesture. We noticed that in more rapid, highly emotional

speech, the cospeech gestures were not always fully carried out (this is called gestural undershoot): there often is just a succession of apices, tightly coordinated with pitch accents but without the hands or arms returning to a resting position. The gestures overlap, leaving only the peaks of the gestures to be observed. Coding apices allowed us to focus on the co-occurrence of pitch accents with apices, both point phenomena, and led to a more streamlined analysis. We recount this process here to alert researchers to the enormous data reduction problems inherent in exhaustive coding of video and its alignment with speech; for us, conversion of interval data to point data by using Loehr's conceptualization of apices was a turning point for the analysis.

Analysis

From the 130 seconds of Mary-Jane's monologue (see the appendix), we selected for microanalysis four segments that most clearly illustrate the various gestural, spatial-metaphorical, and intonational effects we aim to illustrate. All frame grabs are equally spaced in time, roughly 30 ms apart, the highest resolution possible in iMovie HD on a Mac. These artificial subdivisions of the data are not meant to be compared against each other, but to illustrate sequences that we felt had a unity of topic, sequencing, and purpose within the total data set. They are presented in the order in which they occurred, as clusters of frame grabs in Figures 3, 6, 9, and 13. Each panel has been divided into individual frame grabs to show the exact gestural sequences. Thus, P2:12-14 means panel 2, frames 12 through 14. In presenting the data, we first provide the transcript along with associated annotations, then the corresponding panels, and finally our discussion and any illustrative figures related to them. This will necessitate some back and forth for the reader between the figures representing the transcription, the panels, and analysis in the text. We thus begin at the first sequence, with the transcript shown in Figure 2 and the visual sequence (panels) shown in Figure 3.

We can summarize the gestures in Figure 3 as "metaphorical spatialization, plus box" to highlight the fact that two different kinds of gestural constructions are occurring. Through metaphorical spatialization, the speaker creates a rendering of the United States and other countries in space. Through the simultaneously interleaved depiction of a box with batonic gestures, she creates discourse and syntactic parallelism that support the ongoing narrative. Metaphorical spatializations can be viewed as extensions of abstract deixis (McNeill, Cassell, & Levy 1993), where the speaker can situate referents in space through a transposition of the concrete object space to an abstract space. The difference between abstract deixis and metaphorical spatialization in this case is that in the latter speakers not only point at overtly concrete targets but also to referents in a metaphorical space, to referents within what we have outlined as an idealized construction of the public sphere. Liddell (2003:96) explains the mapping of semantic structures into elements of real and conceptual space in ASL, where "contextual clues [in the environment] assist addressees in making appropriate mappings."

To begin the analysis, consider P1:5-11 in Figure 3, where Mary-Jane begins her utterance accompanied by a deictic gesture with the point of origin close to her chest.

M-J	the drug war will be over. (P1:1-3)
(0:20)	we won't have to ship guns and military to Colombia, (P1:4-11)
	it will be over. (P1:12-15)
	Bush won't have to talk to Fox about the drug war, (P1:16-24)
	it will be (P1:25-27)
	over. (P1:28-30)
Audience	((clapping))
	[there will not BE a drug war] (P1:31-36)
	if you LEGALIZE DRUGS (P1:37-42)

Figure 2. Transcript of panel 1 (shown in Figure 3): "It will be over. There will not be a drug war if you legalize drugs!" In this transcript bolded letters mark the syllables that are aligned with the pitch accents.

As she utters the phrase "we won't have to ship guns and military to Colombia," she extends her right arm and points out to the right, metaphorically placing Colombia to one side and away from the origin at ego. This metaphorical spatialization continues in P1:16-24, where Mary-Jane executes a bimanual pointing gesture immediately in front of her and then extends both arms to the left while saying, "Bush won't have to talk to Fox about the drug war," producing gestural apices and intonational pitch accents aligned with "Bush" (P1:18) and "Fox" (P1:20). Though both Colombia and then–Mexican president Vicente Fox (and, by extension, Mexico) were located to the south of Mary-Jane in real space, she has placed them in metaphorical space to the right and left, respectively, while Bush, who was northeast of her, was placed directly in front of her (Mary-Jane herself was facing absolute south). Although we would not expect a speaker of English to exhibit absolute cardinal directionality (as documented by Haviland 1993), people in Tucson often do, since the city is two hours north of the border with Mexico and has dramatic mountain ranges that orient residents to absolute directions. This apparent disparity in the directions of well-known places was our first clue that something beyond simple directional deixis was being represented through Mary-Jane's gestures. Figure 4 is a bird's-eye representation of her metaphorical-space

Figure 3. Panel 1: "It will be over. There will not be a drug war if you legalize drugs!" In this transcript bolded letters mark the syllables that are aligned with the pitch accents.

gestures, showing Bush, guns, and the military directly facing Mary-Jane, while Fox and Mexico are to her left side and Colombia to her right.

A crucial point about Mary-Jane's emergent orientation in space is that it involves the presentation of supplementary information about the speech situation. The conceptualization of the government and its accoutrements (guns and military) being directly in front of the speaker begins the creation of what we are calling a separate, metaphorical spatialization layer. This layer gets built up over time and exhibits a different (longer) time resolution than the batonic alignment of gestural apices and intonational peaks that we discuss below. We argue that the directional aspects of Mary-Jane's gestures and their lack of correspondence with real space provide the first clue that she

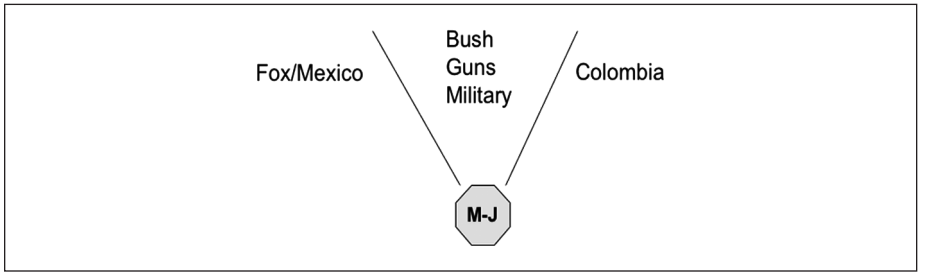

Figure 4. Bird's-eye representation of Mary-Jane's metaphorical spatialization of political actorsIn this transcript bolded letters mark the syllables that are aligned with the pitch accents.

is building up the metaphorical space of the public sphere. Note that we are not merely providing a description of locations chosen for these metaphorical mappings, but rather we advance the claim that Mary-Jane gradually builds and presents parallel content through her gestures. The placement of these political actors persists, as if suspended in air, through time, and the speaker continues to build against this depicted backdrop as her impassioned monologue goes on.

A catchment of gradually exaggerated gestures also occurs in this panel, interleaved with the above metaphorical spatialization. This catchment is not part of the larger U.S.–other countries metaphorical spatialization outlined above but is a separate subroutine interspersed across that narrative. The catchment builds a box that is not iconic or representative of the political sphere; it serves a complementary function, that is, to highlight through alignment and parallelism the message that is contained in the text. Mary-Jane repeats the phrase "it will be over" three times, each time with successively downstepped intonation and more emphatic gesturing. The first time, she brings her left hand (which is in ASL handshape G/X1; McNeill 1992:88) in one downward stroke from eye level to shoulder level (this is visible in Figure 3, P1:1-4). For the second iteration of the phrase "it will be over," she draws a three-sided, open-bottom box with both hands, starting at the center top (P1:12-15; P1:14 shows both palms facing each other as she sweeps them down to make the sides of the box). In the final iteration of the "it will be over" catchment (P1:25-30), she draws the complete box and slows down her speech, exhibiting an intonational phrase boundary, a long pause, and a gestural hold in the exact spot where she prepares to bring her hands in to close the box (P1:29), right before the last word, "over."

We have been making the claim that the metaphorical spatialization and the catchment outlining the box are "interleaved" through this segment. Example 1 shows this interleaved system of metaphors quite clearly:

(1) A The drug war will be over [gesture: left hand downward stroke: first side of box]
B We won't have to send guns and military to Colombia [gesture: metaphorical spatialization of the public sphere]

C It will be over [gesture: open bottom box]
D Bush won't have to talk to Fox about the drug war [gesture: metaphorical spatialization of the public sphere]
E It will be over. [gesture: redraw entire box, bottom closed to align with last word]

Example 1 yields nested structural parallelisms alternating between the metaphorical spatialization and the box-catchment routine: (1) lexical, intonational and syntactic parallelisms appear with coindexed pronouns at lines ACE, (2) intonational, gestural, and syntactic parallelisms appear at BD, and (3) catchment repetition along with gestural parallelism and expansion appear at CE.

Tannen (1989) has argued that repetition in discourse has a cohesive and focusing function, serving to highlight information and structure listener expectations. We believe the communicative constraints that we mentioned above motivate the use of four different semiotic levels of repetition (lexical, syntactic, phonological [intonation], and paralinguistic gesture), all within a space of nine seconds of spontaneous speech. The fine-grained coordination of moment-to-moment speech with modalities such as pointing has been discussed in the conversation-analytic literature (Goodwin & Goodwin 2000). Our description of the incorporation into emergent structure of affect-laden gesture and intonational phenomena contributes to this literature. We track the operation of two alternating, interleaved levels of communication, within each of which there are cohesive gestural, intonational, and discourse parallelisms, clearly indicating that speakers and listeners track and process simultaneous layers of information structure at different planes.

Figure 5 shows the transcript of the segment we have called "Excuse me, hello?" The panel stills are displayed in Figure 6. This segment is annotated with the ToBI tones marking pitch accents (*) as well as phrase tones (-) and boundary tones (%). The bolded words indicate that there is a simultaneous occurrence of a pitch accent with a gestural apex. It is particularly noteworthy that Mary-Jane deliberately separates the phrase "day one" into two intonational phrases. She gives both words prominence by accenting them and by making pointing gestures where her right index finger lands on the upward-open palm of her left hand ("day" in P2:10-11 and "one" in P2:12-14, both within Figure 6). Note also that Mary-Jane produces the last gestural apex in this segment with her entire upper body as the articulator (P2:27-36), which she abruptly stops midmotion at a precarious forty-five-degree angle as she says to the congressman, "Excuse me, hello."

The intonation contour L*+H L-H% occurring on "excuse me" has been described as carrying a holistic pragmatic meaning ranging from uncertainty to incredulity (Ward & Hirschberg 1985; Hirschberg & Ward 1992) depending on the pitch range of the phrase. Hirschberg and Ward (1992:243) argue that "when speakers use the contour to express incredulity, they generally express that incredulity about a value already evoked in the discourse." A striking fact about this use of the uncertainty–incredulity contour is that it is not aligned with the utterance that one might expect, given the partially ordered set relationship (poset) framework described in Ward and Hirschberg (1985).

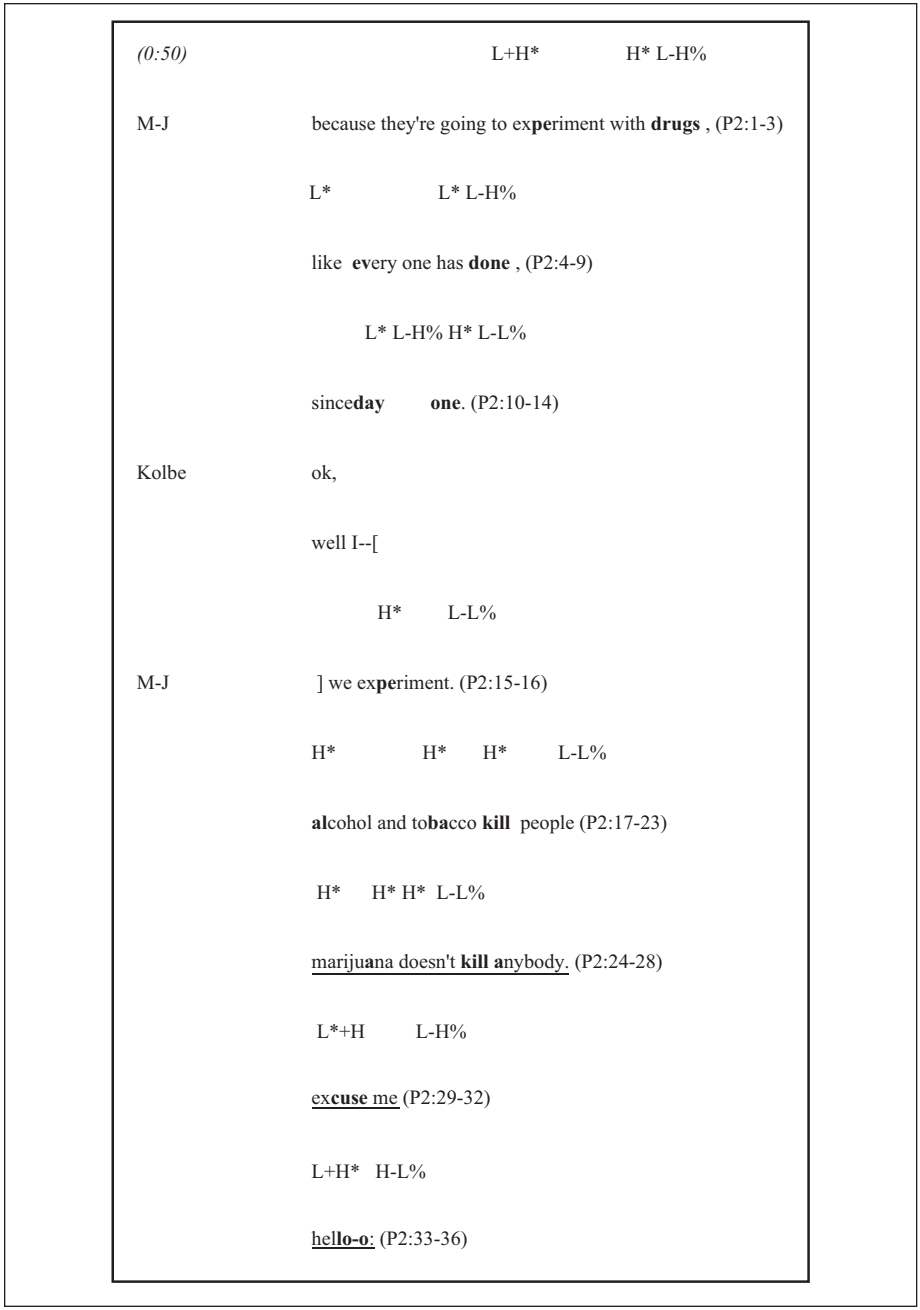

Figure 5. Panel 2 Transcript: "Excuse me, hello?" In this transcript bolded letters mark the syllables that are aligned with the pitch accents.

Figure 6. Panel 2: "Excuse me, hello?"

H* H* H* L-L%

alcohol and to**ba**cco **kill** people (P2:17-23)

H* H* H* L-L%

marijuana doesn't **kill a**nybody (P2:24-28)

L+H* L-H%

ex**cuse** me (P2:29-32)

Figure 7. Intonational (ToBI) transcription for "Marijuana doesn't kill anybody"

Consider the relevant part of the utterance once more (see Figure 7). In denying that marijuana kills people, Mary-Jane invokes the possible set of substances that do kill people and discursively selects for the interlocutor (Kolbe) and overhearers (other

constituents) deadly substances that are legal. The incredulity being expressed here, signaled by the wide pitch range that the contour carries and that upgrades it from uncertainty (Hirschberg & Ward 1992), is in the assertion that it is the noxious substances that are legal while the innocuous substance is illegal. These discourse facts all line up with the licensed uses of this contour according to Ward and Hirschberg (1985), with one exception: the tonal contour does not fall on the expected phrase ("marijuana . . ."), but instead on the following one ("excuse me . . ."). We believe that this finding has two possible interpretations with important implications as discussed in the paragraphs below.

On a first look, it is possible that this is the point where a speaker under strong communicative demands finally reaches the limits of the processing capacity for on-line information alignment. In this short segment, Mary-Jane simultaneously aligns pitch accents with syllables contained in informationally prominent words, gestural apices with those pitch accents, while making a complex argument and taking parts of her body to the limits of their physical space. Could the canonical alignment of the incredulity contour have been given up to meet the extra processing demands of this task? A second interpretation seems more plausible and may have more merit: the speaker, prefiguring that she was going to use a conventionalized expression with its own L*+H L-H% contour ("excuse me, hello?" is often uttered this way colloquially in the United States), chose to suppress the first contour (so as to avoid two identical contours together), following a kind of prosodic OCP (Obligatory Contour Principle; Leben 1973) presumably because the meaning is already accessible from a single rendition of the contour. This accessibility implies that the incredulity contour on "excuse me" (response to the rhetorical position that marijuana does kill people) can count for both "excuse me" and its preceding utterance "marijuana doesn't kill anybody."

On "hello," Mary-Jane produces a pitch contour that we described with the tonal sequence L+H* !H-L%. This contour has also been described as a "calling contour" (Beckman & Ayers 1994). It is the contour often used when shouting a name during the process of looking for somebody (e.g., to call somebody for dinner). It is our impression that Mary-Jane uses this contour to call Congressman Kolbe metaphorically to do his job and to reproach him.

Figures 8 and 9 provide the transcript and the panel stills for the segment we have titled "Alcohol and Tobacco." This segment superimposes a gestural layer on to the earlier metaphorical spatialization described in the first bird's-eye view in Figure 4. Prior to the frame sequences above, Mary-Jane has been referring to "kids you've been talking about today" who will go to prison because of their alcohol and drug use. The gesture in "kids" starts out directly in front of Mary-Jane, metaphorically in a space shared between the government and the people, and after drug use the "kids" get moved off to "prison," which also occupies the peripheral position earlier used for external entities (other countries) outside the relationship between the government and the people (see Figure 10).

Clearly visible from the built-up layers of metaphorical spatialization depicted in Figure 10 is the fact that another spatialized, gestural movement from "us" to "them"

M-J	why aren't you talking about alcohol and cigarettes? (P3.1:1-7)
	what you've done is you've taken that industry, (P3.1:8-12)
	and you are putting it on Third World . (P3.1:13-17)
Kolbe	what--
M-J	they are consuming the tobacco now. (P3.1:18-23)
	that's where all the tobacco money is coming from. (P3.1:24-30)
Kolbe	[we're spending]
M-J	[they're still making] a profit on cigarettes (P3.2:31-37)
	cause they're selling them to all the [people] (P3.2:38-40)
	that we're dumping our chemicals on (P3.2:41-48)

Figure 8. Panel 3 Transcript: "Alcohol and Tobacco"

takes place, and we will use it to illustrate the relationship between the gradually accruing metaphorical spatialization that Mary-Jane builds and the pitch–accent and apex alignment relationships in the gesture. As Mary-Jane discusses the practice of chemical dumping by the United States, she makes a gesture that starts out (with the word "taken") with both hands right behind her left shoulder (as though she were holding a ball and beginning to toss it; P3.1:10), and winds up gesturally in what is by now in her gesture system a stable place for the "Third World," to the far right (at the word "putting"). We have created a composite display, which we call a partiture display (Mendoza-Denton 2007), of the various semiotic layers in her sentence "You've taken that industry and you are putting it on the Third World." It is reproduced in Figure 11.

In Figure 11, we can finally see how the various components in semiotic layering come together. This type of transcription, from Mendoza-Denton (2007), is called a partiture display. Partiture displays are meant to show the time course of communicative organization in a multimodal framework. From the bottom of the display, we show the text layer, above it the tonal transcription layer with the pitch accents and phrase tones, then the apex layer, followed by the waveform, fundamental frequency

Figure 9. Panel 3: "Alcohol and Tobacco"

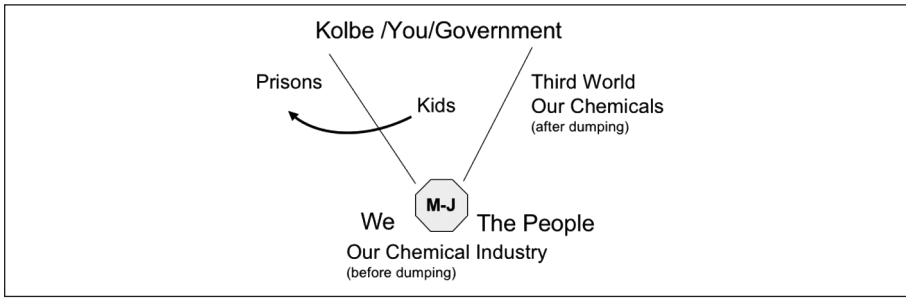

Figure 10. Bird's-eye view of spatialized entities from Figure 9

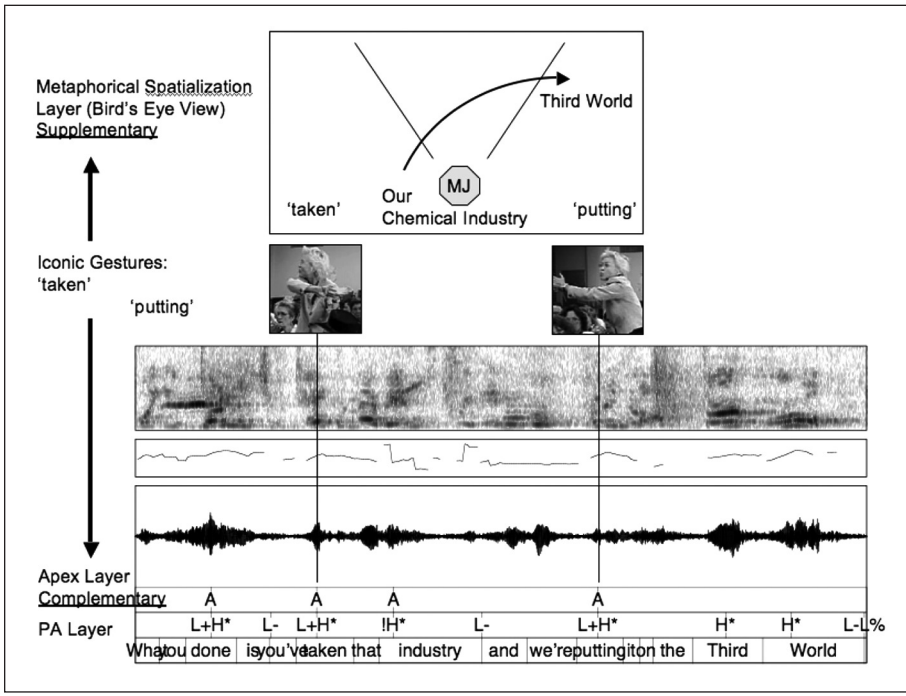

Figure 11. Partiture display showing multimodal alignment of text, pitch accents, apices, iconic gestures and metaphorical spatialization

track, and formant displays. In the center of the figure are two frame grabs from panel 3 in Figure 9 displaying the exact moment where the words "taken" and "putting" occur. Above this lies the metaphorical spatialization layer showing the movement in metaphorical space in real time, as it unfolds in this sequence. An important point to take away from the partiture display is that the dynamic iconic gestures in "taken" and

"putting" can be coded into gestural apices that align exactly with intonational pitch accents. We discuss the quantitative generalization on their co-occurrence in the concluding discussion. Gesture and intonation at this level stand in a supplementary relationship in that the information (on acoustic prominence, expressing pragmatic focus) that is found in the intonation grammar is supported or supplemented by the same information co-occurring in the apex layer. At the same time, those very gestures form part of an entirely different, slowly unfolding, metaphorical spatialization process that stands in a complementary relationship with the information in the text, pitch accent, and apex layers. By using "complementary" here we call attention to the fact that the information added in this way through metaphorical spatialization is new and non-redundant. That iconic gestures are able to participate in meaning making at two different levels (one in tandem with the intonational grammar and one that adds new information to the discourse) raises the possibility that cospeech gesturing is essentially (but not always, as in the case of purely emphatic batonic gestures) bivalent, belonging to two systems simultaneously (Woolard 1998). In this conceptualization, cospeech gesturing is the "bivalent hinge that integrates co-occurrence and contrast" (Woolard 1998:10).

Although a detailed discourse analytic account of the shifting pronominal referents in these data is beyond the aims and scope of this article, we will note that the pronoun "we" variously refers to (a) the people of the United States as a collective that produces chemicals that need to be dumped someplace ("people that we're dumping our chemicals on"), (b) the United States as an entity that ships guns and military overseas ("we won't have to ship guns and military"), (c) constituents that engage in collective behavior and do not share in the government's definition of what is legal and illegal ("we experiment"), and (d) individual citizens who have rights to control their own bodies ("yes we do"). The pronoun "you" likewise shifts in reference from a generic addressee, to representative Kolbe, to then-President G. W. Bush, and to the government in general. This variability in the instantiation of pronominal reference has been observed in political discourse cross-linguistically, as public figures act on the political stage (Van Dijk 2003; Wodak 1989). In this case, tracking the shifts in pronominal reference helps us to anchor and interpret the gestural data and to make explicit the implicit political model that Mary-Jane sketches with her gestures.

The final panel for our gestural/intonational microanalysis is panel 4 (Figures 12 and 13). In this segment Mary-Jane has taken advantage of a prop that she brought to the THM: a hemp-seed chocolate chip cookie in a ziplock bag. She holds up the cookie and addresses the audience, showing them the offending item which is made with hemp seed (the seed of the marijuana plant). She attempts to highlight the absurdity of the government's decree that marijuana is illegal by claiming that she does not have the right to eat a chocolate chip cookie. She ends her performance with a dramatic flourish, by uttering two parallel constructions, with the same syntactic structure, in a slow and dramatically delivered style. They are presented in Figure 14, in column format to highlight the parallelism. In conversation analytic transcription conventions, the periods in parentheses mean that there were significant pauses in the speech stream.

M-J they want to make (P4:1)

(.)

hemp seed, (P4:2-4)

a schedule one narcotic (P4:5-11)

when did I lose the right (P4:12-17)

(.)

to eat (P4:18-22)

(.)

the food (P4:23-24)

(.)

I want (P4:25-27)

(.)

to eat. (P4:28-30)

(1:40) when did I lose my rights

(.)

as a citizen of this country (P4:31-41)

(.)

to put into my body (P4:42-45)

(.)

and listen to whatever (P4:46-48)

(.)

or watch whatever I want. (P4:49-54)

Figure 12. Transcript of panel 4: "My rights as a citizen"

In this segment, not only is the syntactic–constructional parallelism evident, but we further note that the sequences of pitch accents in these utterances are in a downstepping relationship to each other, despite the fact that the standard ToBI transcription system assumes that intonational downstep cannot happen across intonational or

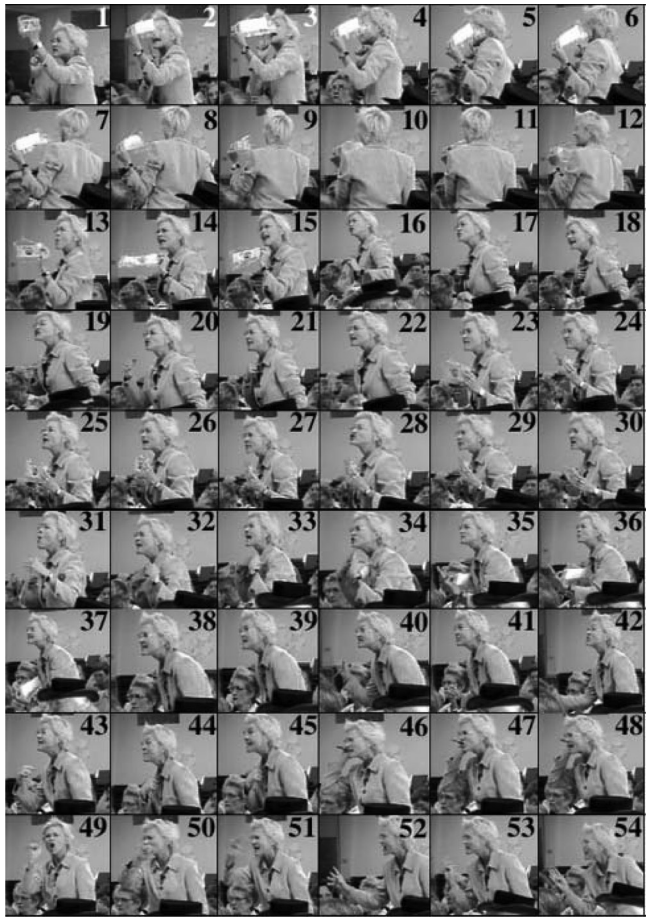

Figure 13. Panel 4: "My rights as a citizen"

intermediate phrase boundaries (Beckman & Ayers 1994). Within ToBI, one would assume a change in pitch range from one phrase to the next. The sequence "to eat, the food, I want, to eat" is composed of four separate intonational phrases, separated from each other by a small audible and visible pause. Each of these intonational phrases contains one intermediate phrase that in turn contains one pitch accent. There are accents on "eat," "food," "want," and "eat." We have observed that perceptually these accents appear to occur in a downstepping relationship to each other (Liberman & Pierrehumbert 1984) whereby downstep is defined to be a stepwise successive lowering of the pitch of H-tones that occur in a sequence. According to the tonal labeling conventions (ToBI) though, downstep across phrase boundaries is infelicitous.

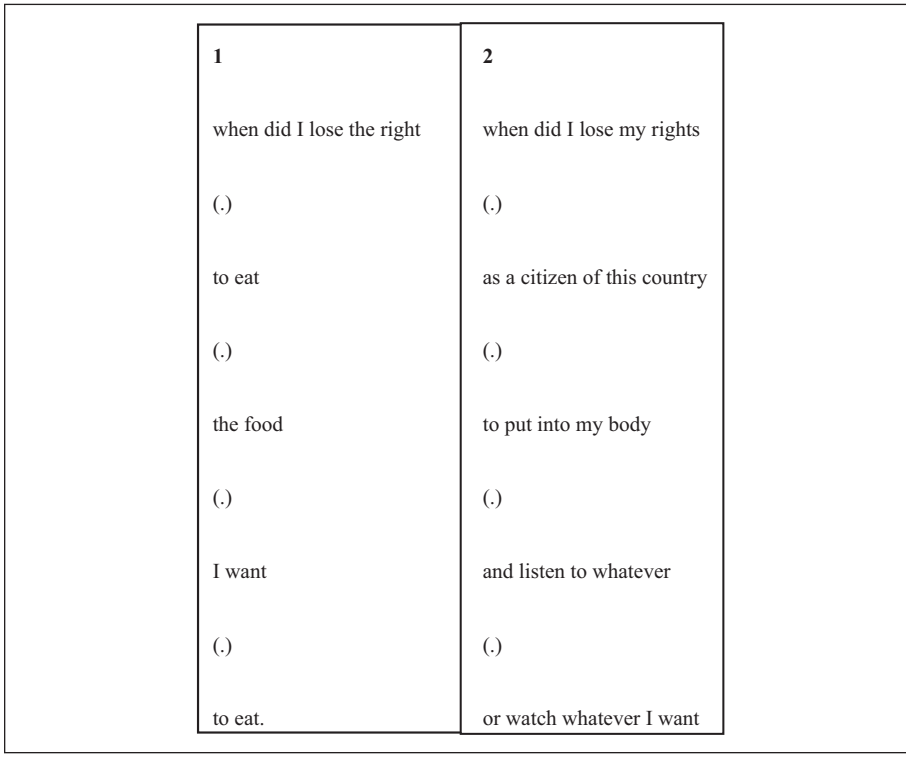

Figure 14. Parallelisms in "When did I lose my rights"

Rather, ToBI would assume a pitch range compression whereby the pitch range of each successive phrase is smaller and more compressed and thus accents are realized at a lower level.

Unfortunately measurements of the fundamental frequency contour during the time point of the F0 maximum did not generate any insights because the signal was too perturbed, but the quality of the recording allows us to understand the message and hear the pitch and the relationship of the accents to each other. In terms of interpreting the meaning of downstepped accents, Ladd (1996:90) suggests that the usage of a downstepped accent involves a choice rather than a phonological trigger and that "downstep" has a meaning, indicating "finality" or "completeness." Extending Ladd's downstep interpretation to our data, we stipulate that Mary-Jane uses the successive lowering in pitch of the accents (and the entire intonational phrase) to lend emphasis to her statements and give them some quality of "indisputability" and non-negotiability. We also suggest that for reasons of parallelism and emphasis, she employs downstepped H* (notation: !H*) accents. These !H* accents coincide in our data with the apices of the gestures as well, since Mary-Jane rocks her entire body back and forth

Table 1. Co-occurrence of Pitch Accents and Gestural Apices

Segment	Duration in seconds	Apices		PA		Co-occurring apices and PA		PA but no apices		Apices but no PA	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%
1. over	14.23	16	100.0			16	100.0			0	100.0
				24	100.0	16	66.7	8	33.3		
2. excuse me, hello	11.72	11	100.0			11	100.0			0	100.0
				15	100.0	11	73.3	4	26.7		
3. tobacco	14.26	16	100.0			15	93.8			1	6.3
				27	100.0	15	55.6	12	44.4		
4. lose the right	19.15	28	100.0			26	92.8			2	7.1
				32	100.0	26	81.3	6	18.8		
Total	59.36	71	100.0			68	95.7			3	4.2
				98	100.0	68	69.4	30	30.6		

approximately thirty degrees to land at the apex of these movements on the !H* accents in both of these utterances.

This brings us to a more general discussion of the correlation between pitch accents (PA) and apices. To establish some kind of a measure of what the relation is between the occurrence of an accent and the occurrence of an apex, we counted the number of times Mary-Jane produced accents, that is, prosodic prominences, as well as the number of times where her gesturing included an apex, the peak of the gesture. The numbers are shown in Table 1.

The general tendency apparent from this table comparing the "apices" and "PA" columns is that there are always more pitch accents than apices in any given segment. Note that we are not comparing the segments with each other; we have merely stated the numbers separately as to make the counting more transparent and easier. As we can see from the "PA but no apices" column, in verbally rendered speech, information is highlighted by acoustic prominences that need not have visually co-occurring gestural apices. In contrast, it only rarely happens that there is an apex occurring without a pitch accent (see the "apices but no PA" column). This suggests to us either that some prosodic prominences are just rhythmic in nature or that not all prominences are semantically "worthy" of being marked by cospeech gestures. It is possible also that the information given on these two different planes is complementary in nature, highlighting different parts of the message.

In the "co-occurring apices and PA" column, the values in the unshaded boxes (e.g., 16, 100.0) indicate that whenever there is a gestural apex, there also is a pitch accent. The number in the shaded box indicates that only 66.7 percent of all pitch accents are accompanied by an apex, a beat gesture. In terms of the total numbers, we find that

95.7 percent of all apices are accompanied by a pitch accent, whereas only 69.4 percent of all pitch accents are also marked by a gestural apex. The reason for this may be quite simple: people can speak without gesturing but rarely gesture without speaking. Thus, when there are gestural apices, it is quite likely that they are aligned with pitch accents.

We have looked for the co-occurrence of accent location and gestural apex. However, it may be promising also to differentiate between the accent types, which we have not done so far. It is possible that certain (less prominent) accent types such as a downstepped !H* or a L* may be reinforced by gestural means. However, we can make no claims regarding this hypothesis.

Discussion

We found that cospeech gesturing plays both a supplementary, reinforcing role in its co-occurrence with intonation and a complementary role (adding new content) in its relationship to discourse. Cospeech gesturing simultaneously participates in two processes at different levels of time resolution. Similarly, intonation is layered on and synchronized with the speech-content layer, providing supplementary information (which may be holistic as in tunes that have specified meanings) or directing a listener's attention to the speech material highlighted. This is done with pitch accents synchronized with the lexically stressed syllables in words that more often than not carry an extra communicative load. Just like gesture, intonation can also create a complementary layer of pragmatic information—for example, irony or sarcasm—not necessarily explicitly expressed in the speech-content layer or even contradicting the information rendered verbally. Thus, gesture is not privileged in its ability to exist on two semiotic layers simultaneously. But it does have an advantage over intonation (and other speech-accompanying layers) in its holistic expressivity (using McNeill's 1992 discussion of holistic expressivity, i.e., using space as well as time). That is, gesture is able to provide a wide range of complementary information free from the limits of a grammar, within which speech and intonation must exist.

If we assume the validity of Bolinger's (1986) claim that gesture and speech stem from the same semantic intent, then we commit ourselves to the notion that some degree of preplanning is involved in generating not only speech output but also gestural output to convey information on different planes. How information is structured and divided up across the two channels is not understood at this point. From our data it appears that complementary and contextual information can be transmitted via gestures while concrete assertions are made explicit via speech. We also do not know what constraints exist on (pre)planning complex gestures that we know are time aligned with linguistic structure in the final output.

Gestures are not just involuntary movements but finely coordinated structures of motion, aligned with semantic content. Since speech can be understood over the telephone or in the dark (with a complete absence of visual cues), we must assume that gestures facilitate the information transfer from the speaker to the listener–viewer but are not necessary for successful transmittal of content. Gestures play an important role

for the naturalness of speech and for cuing speaker stance. That said, it is also known that speakers gesture while speaking when nobody is there to see them. Therefore, it appears that gesturing is not just a facilitation device for the listener–viewer but also a mode of self expression for the speaker.

While cospeech gesturing may be innate behavior as strongly suggested by the use of gesturing by the congenitally blind (Iverson & Goldin-Meadow 1998), the type of gesture produced exhibits cross-cultural differences and is, just as other communicative actions, learned behavior. The mechanisms of acquiring cospeech gesturing consist of coordinating the individual tasks of the complex gestures with each other (e.g., lift right arm, rotate palm upward, release arm in this constellation) and then coordinating these motions with speech so that points of informational prominence in speech are accompanied for example by apices in gesture.

Infants as young as a few hours display the ability to mimic the sticking out of one's tongue when prompted (Meltzoff & Moore 1977:78). This suggests that our cognitive system provides for learning by example and imitation. The task is a formidable one, as inverse mapping has to take place: the infant observes tongue movement via the visual channel and then has to map the observed movements to his or her own motor patterns of the articulators (opening lips, lower jaw, extending tongue) to perform the task of sticking out the tongue. Advances in cognitive science and neurophysiology seem to suggest that mirror neurons "form a cortical system matching observation and execution of [goal-related] motor actions" (Gallese & Goldman 1998:495).

It appears that the same type of mechanisms should be involved to learn other motor skills such as finely coordinating hand, arm, and body movements to be timed with speech: based on our casual observation, often the offset of a complex gesture co-occurs with the end of a prosodic phrase (intermediate or intonational phrase) or, as we have shown in this article, apices, the peaks of the gestures, co-occur with pitch accents.

If the interpretation of intonation contours, that is, if the interpretation of prosodic focus is partly determined by context, and gesturing can provide some of this context, we have to take into account that this context is not just provided verbally or relates back to knowledge the interlocutors possess already. Rather, in face-to-face communication the gestural channel is able to provide the speaker's stance or part of the context in which to interpret the utterance. It appears that the verbal and gestural channels are interpreted simultaneously and holistically so that the semantic content can be recalled but the specific presentation and structure of information are more fleeting in nature and thus cannot easily be teased out by the receiver. That is, we tend to remember meaning rather than form but also make inferences that let listeners arrive at an interpretation (Bransford, Barclay, & Franks 1972).

Conclusion

This case study of the coordination of spontaneous speech with gesture, focusing on intonational alignment with pitch peaks (after Loehr 2004), has shown that wherever there is a gestural apex in our data, there is also a pitch accent. The reverse, however,

is not true because pitch accents often occur without gestures or the apices are phase-shifted from the gestural peaks. Our results concur with those of Loehr (2004) for staged but natural conversations, and they indicate that gestural phenomena are in robust co-occurrence with pitch accents in both laboratory and spontaneous speech. Our findings also include the discovery of a downstep relationship across intonational phrase boundaries as well as the pervasive use of lexical, syntactic, gestural, and intonational parallelism in the performance of a speaker under high pressure in an affect-laden, spontaneous communicative situation. Gestures also enhance the expressiveness by maximizing the simultaneous presentation of information on different channels. While the intonational and gestural alignments were observed within and across intonational phrases and semiotic layers, metaphorical gestural spatialization—unlike abstract deixis—had a larger domain, took longer to unfold, and sketched out complementary information on this speaker's notion of the relationship among a political representative, the government, and outside entities.

Appendix

Transcription of Town Hall Meeting and Transcription Conventions

Setting: Town hall meeting at St. Cyril's school, midtown Tucson, Arizona, February 2001. Filmed with two Sony DVTR8 mini-DV cameras in NTSC format (29 fps). This transcription was generated from the audio track corresponding to the audience microphone.

Transciption conventions include the following:

[ ]	Square brackets indicate overlap.
=	Equal signs indicate latching.
(.)	Periods in parentheses indicate pauses.
,	A comma at the end of a line indicates nonfinal intonation.
.	A period at the end of a line indicates falling intonation, though not necessarily the end of a clause or sentence
-	A hyphen marks an interruption, either a self-interruption or a cutoff.
(( ))	Double parentheses mark action descriptions.
<u>stress</u>	Underlining indicates extra loudness and emphasis
	Lines are divided by breath groups.

More detailed intonation transcriptions, when shown, appear above the line and are according to the ToBI framework, explained in the main body of the article.

Kolbe	and since you wanna comment on [this,
M-J	[I'm sorry]

(continued)

Appendix (continued)

Kolbe	we're gonna] get your comment=
M-J	=my-
	my only problem with-
(0:04)	n- n- n- not legalizing all drugs?
Green	right.
M-J	you take the criminal element out of it,
	when you end prohibition,
	(.)
	this is what we saw in the thirties,
	the drug war will be over.
(0:20)	we won't have to ship guns and military to Colombia,
	it will be over.
	Bush won't have to talk to Fox about the drug war,
	it will be
	(.)
	over.
Aud	((clapping))
	[there will not be a drug war]
	if you legalize drugs
Kolbe	ok [calm down,
(0:30)	calm down,
	(.)
	calm down,]
M-J	[you are making it a criminal] enterprise,
Aud	[boo:::::]
M-J	[you have got-]
	the government has made it a criminal enterprise,
	the government is making money,
	you are making money hand over fist,
(0:40)	you are building prisons so fast it's disgusting,
	you're not putting any of that money into education,
	you're locking those-
	you're building prisons to lock those kids up that you're talking about today,
	they're going to go to prison.
(0:50)	because they're going to experiment with drugs,
	like every one has done
	since day one.

(continued)

Appendix (continued)

Kolbe	ok,
	well I-=
M-J	=we experiment.
	alcohol and tobacco kill people.
	marijuana doesn't kill anybody.
	excuse me,
	hello-o:
	(.)
Kolbe	ok.
	(.)
	well-=
M-J	=why aren't you talking about alcohol and cigarettes?
	what you've done is you've taken that industry,
	and you are putting it on Third World.
Kolbe	what-=
M-J	=they are consuming the tobacco now.
	that's where all the tobacco money is coming from.
Kolbe	[we're spending]
M-J	[they're still making] a profit on cigarettes
	cause they're selling them to all the [people]
	that we're dumping our chemicals on
Kolbe	[we're-]
	we're spending a lot of money on a-
	(.)
	on education,
	which is what I think-
	(.)
	on tobacco,
	which is what I think we need to be doing,
	(.)
	[spending a lot on that].
M-J	[proper name] wants to make this
	a schedule one narcotic.
	a chocolate chip cookie,
	ok,
	a chocolate chip cookie,
	this is what they tried to run through on October the thirtieth,
	they want to make

(continued)

Appendix (continued)

	(.)
	hemp seed,
	a schedule one narcotic.
	when did I lose the right
	(.)
	to eat
	(.)
	the food
	(.)
	I want
	(.)
	to eat.
(1:40)	when did I lose my rights
	(.)
	as a citizen of this country
	(.)
	to put in my body
	(.)
	and listen to whatever
	(.)
	or watch whatever I want.
Kolbe	well,
	(.)
	uh,
	there-
	there always have been limits on doing-
	on some things.
	you do-
	we do-
	you do not have=
M-J	=I limit myself,
	[you don't limit me thank you].
Kolbe	[y-
	y-
	you] don't have absolute rights to do everything.
M-J	yes we do.
Kolbe	and we do -
	(.)
	we never have had.

Acknowledgments

We thank Congressman Jim Kolbe, his staff in Tucson, Arizona, and in Washington, D.C., and anonymous constituents at various town hall meetings in 2000–2001 for their consent and cooperation. We gratefully acknowledge the invaluable comments provided by the anonymous reviewers of our article and the editors of this journal, whom we wish to thank especially for their patience during the review process. Other comments came from audiences at conferences and talks, and we are especially grateful to Laada Bilianiuk, Aomar Boum, Mary Bucholtz, John Haviland, George Lakoff, Bettina Migge, and Mourad Mjahed. This research would not have been possible without the help of the many undergraduate and graduate students in the linguistic anthropology lab at the University of Arizona: we thank J. C. Baker, Natasha Gibson, Hannah Jones, Emily Kidder, Kerri Murray, Ashley Stinnett, and Landon Yamaoka. All errors and shortcomings of the article are ours.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a University of Arizona Vice Provost Faculty Research Grant, the University of Arizona Department of Anthropology Riecker Grant Program (both to Mendoza-Denton), and a University of Arizona Undergraduate Honors Program Grant (to Ashley Stinnett for fieldwork assistance). This work was supported (in part) by the German Federal Ministry for Education and Research (BMBF) (Grant Nr. 01UG0711). Further support came from SFB632-D3 (Humboldt University of Berlin and German Research Foundation) and the ZAS Berlin/BMBF (to Jannedy).

References

Beattie, Geoffrey & Jane Coughlan. 1999. An experimental investigation of the role of iconic gestures in lexical access using the tip of the tongue phenomenon. *British Journal of Social Psychology* 90(1). 35-56.

Beckman, Mary & Gail Ayers. 1994. *Guidelines to ToBI Labelling (Version 2.0)*. Columbus: Ohio State University.

Beckman, Mary & Jan Edwards. 1994. Articulatory evidence for differentiating stress categories. In Patricia Keating (ed.), *Phonological structure and phonetic form: Papers in laboratory phonology*, 7-33. Cambridge, UK: Cambridge University Press.

Boersma, Paul. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10). 341-345.

Bohman, James. 1996. *Public deliberation*. Cambridge, MA: MIT Press.

Bolinger, Dwight Le Merton. 1986. *Intonation and its parts: Melody in spoken English*. Palo Alto, CA: Stanford University Press.

Bransford, John, J. Richard Barclay, & Jeffrey Franks. 1972. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology* 3. 193-209.

Browman, Catherine & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary Beckman (eds.), *Papers in*

laboratory phonology I: Between the grammar and physics of speech, 341-376. Cambridge, UK: Cambridge University Press.

Browman, Catherine & Louis Goldstein. 1992. Articulatory phonology: An overview. *Phonetica* 49. 155-180.

Cassell, Justine, David McNeill, & Karl Erik McCollough. 1999. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition* 7(1). 1-33.

Cosnier, Jacques. 1982. Communications et langages gestuels. In Alain Berrendonner, Catherine Kerbrat-Orecchioni, Jacques Coulon, & Jacques Cosnier (eds.), *Les voies du langage: Communications verbales, gestuelles, et animales*, 255-304. Paris: Dunod.

Enfield, Nick. 2003. Producing and editing diagrams using co-speech gesture: Spatializing non-spatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology* 13. 7-50.

Gallese, Vittorio & Alvin Goldman. 1998. Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Sciences* 2(12). 493-501.

Gaudio, Rudolf. 2003. Coffeetalk: Starbucks™ and the commercialization of casual conversation. *Language in Society* 32. 659-691.

Goodwin, Charles & Marjorie Harness Goodwin. 2000. Emotion within situated activity. In Nancy Budwig, Ina C. Uzgris, & James V. Wertsch (eds.), *Communication: An arena of development*, 239-257. Stamford, CT: Ablex.

Habermas, Jürgen. 1989 [1971]. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Trans. Thomas Burger with the association of Frederick Lawrence. Cambridge, MA: MIT Press.

Harper, Lisa, Daniel Loehr, & Anthony Bigbee. 2000. *Gesture is not just pointing*. Mitzpe Ramon, Israel: International Conference on Natural Language Generation paper.

Haviland, John B. 1993. Anchoring, iconicity, and orientation in Guugu Yimidhirr pointing gestures. *Journal of Linguistic Anthropology* 3(1). 3-45.

Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20. 241-251.

Iverson, John & Susan Goldin-Meadow. 1998. Why people gesture when they speak. *Nature* 396. 228.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford, UK: Oxford University Press.

Kelly, Spencer, Dale Barr, R. Breckinridge Church, & Katheryn Lynch. 1999. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language* 40(4). 577-592.

Kendon, Adam. 1996. Gesture in language acquisition. *Multilingua* 15. 201-214.

Ladd, D. Robert. 1996. Intonational phonology. Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press.

Leben, William. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT dissertation.

Liberman, Mark & Janet Pierrehumbert. 1984. Intonational invariants under changes in pitch range and length. In Mark Aronoff & Richard Oehrle (eds.), *Language sound structure*, 157-233. Cambridge, MA: MIT Press.

Liddell, Scott. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge, UK: Cambridge University Press.

Loehr, Daniel. 2004. *Gesture and intonation*. Washington, DC: Georgetown University dissertation,

McNeill, David. 1992. *Hand and mind*. Chicago, IL: University of Chicago Press.

McNeill, David. 2000. *Language and gesture*. Cambridge, UK: Cambridge University Press.

McNeill, David, Justine Cassell, & Elena Levy. 1993. Abstract deixis. *Semiotica* 95(1-2). 5-20.

McNeill, David, Francis Quek, Karl-Erik McCullough, Susan Duncan, Nobuhiro Furuyama, Robert Bryll, Xin-Feng Ma, & Rashid Ansari. 2001. Catchments, prosody and discourse. *Gesture* 1(1). 9-33.

Meltzoff, Andrew & M. Keith Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198. 75-78.

Mendoza-Denton, Norma. 2007. *Modelling synchrony and entrainment in sociolinguistic variation*. Philadelphia: New Ways of Analyzing Variation Conference 36 poster.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: MIT dissertation.

Rimé, Bernard. 1982. The elimination of visible behavior from social interactions: Effects on verbal, nonverbal, and interpersonal variables. *European Journal of Social Psychology* 12. 113-129.

Tannen, Deborah. 1989. Talking voices: Repetition, dialogue, and imagery in conversational discourse. Cambridge, UK: Cambridge University Press.

Van Dijk, Teun A. 2003. Text and context of parliamentary debates. In Paul Bayley (ed.), *Cross-cultural perspectives on parliamentary discourse*, 339-372. Amsterdam: Benjamins.

Ward, Gregory & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61. 747-776.

Warren, Mark E. 1995. The self in discursive democracy. In Steven K. White (ed.), *The Cambridge companion to Habermas*, 167-200. Cambridge, UK: Cambridge University Press.

Wodak, Ruth (ed.). 1989. *Language, power, and ideology: Studies in political discourse*. Amsterdam: Benjamins.

Woolard, Kathryn. 1998. Simultaneity and bivalency as strategies in bilingualism. *Journal of Linguistic Anthropology* 8(1). 3-29.

Bios

Norma Mendoza-Denton (Ph.D. Stanford 1997, Linguistics) is associate professor of Linguistic Anthropology at the University of Arizona-Tucson. Her areas of specialization are linguistic anthropology and multimedia ethnography, with an emphasis on youth, bilingualism, and style in language. She is the founder and director of the Linguistic Anthropology Teaching Laboratory at the University of Arizona.

Stefanie Jannedy (Ph.D. The Ohio State University 2002, Linguistics) is the coordinator of the phonetics/phonology group at the ZAS (Center for General Linguistics) in Berlin, Germany. Her research interests include prosody, corpus linguistics, laboratory phonology, sociolinguistics and sociophonetics with an emphasis on spontaneous speech and youth varieties of German.