

## PROSODIC VARIATION ACROSS DISCOURSE TYPES

*Cynthia A. McLemore*

Institute for Research in Cognitive Science  
University of Pennsylvania  
Philadelphia, PA 19104  
*cam@arch.ling.upenn.edu*

### 1. INTRODUCTION\*

In this study, I compare the frequency and distribution of a small set of prosodic features in two different types of discourses, or speech activities. The goals of this investigation are to refine methodologies for transcribing and characterizing intonational regularities in natural speech, and to uncover the ways in which intonational forms are used for particular, situated ends.<sup>1</sup>

The data considered here are "natural" in the sense that they weren't elicited for the purpose of study; however, each dataset consists of speech performed in fairly constrained situations, rather than arising spontaneously in the course of conversation. Constrained data of this nature are useful in that they simplify the problem to some extent; the constraints themselves yield clues to prosodic patterning. Nevertheless, hypotheses about intonational function that are formulated on datasets of this kind are necessarily preliminary, if conversation is taken as the fundamental paradigm of language use (Fillmore 1981), which it undoubtedly should be. The data examined in this study include:

- A portion of a second-grade mathematics lesson ("Lesson")
- One exchange from a call-in radio talk show interview with a politician ("Interview")

Both speech activities are recurrent ones for the primary speakers, who occupy social roles with which these activities are associated: the teacher conducts daily lessons in the classroom, the politician gives interviews (more broadly: answers questions and espouses policies) on a regular basis.

Each dataset may be considered a token of a discourse type in a very broad sense: i.e. members of the English-speaking community, including the participants, have names for these speech activities, indicating that they are recognized and evaluated as distinct (see Silverstein 1979, Swales 1990:58). That is, a "lesson" is not an "interview", and neither discourse is a television commercial (Gumperz 1982:102-105), a sports commentary (Ferguson 1983), a meeting announcement (McLemore 1991a), a traditional narrative (Woodbury 1987), or the opening of a telephone exchange (Lieberman & McLemore 1992).

While the two discourses examined here clearly belong to distinct genres, the distribution of prosodic features in them reflect more general characteristics that cross-cut genres. Biber (1988:170) distinguishes genre from text type: the latter represents groupings of texts based on their linguistic form, regardless of their genre. In order to arrive at such general principles, the discourse-internal correspondences between intonational form and function will be examined and related to previous findings, and the intonational features of the two discourses will be compared. The assumption motivating this approach is that, while the prosodic structure of a discourse may arise from the rational, more or less conscious, intentions of speakers, the meanings created by intonational choice ultimately can be understood only by an ordering of the facts of use.

### 2. INTONATIONAL DESCRIPTION

#### 2.1 Segmentation

The segmentation of natural speech into discrete intonational phrases is far from straightforward (Du Bois et. al. 1991:100-114; McLemore 1991a:28-44).<sup>2</sup> In this study, segmentation has been based on sound structure without regard to syntactic, semantic, or pragmatic constituency (as much as possible); i.e. criteria for intonational juncture include pauses, pitch excursions and

\* This research was supported by NSF STC grant number DIR89-20230, and OERI grant number R117610003-92.

<sup>1</sup>Two more long-term goals of this undertaking are to characterize sources of intonational variation in a principled way, and to shed light on the intonational phonology of English by examining recurrent patterns in phonetic data that correspond to apparent functions.

<sup>2</sup>Indeed, Lieberman, McLemore & Woodbury (1991) argued that evidence for independent hierarchic prosodic units, such as the intonational phrase, generally is lacking; rather, such 'units' are motivated by local, gradient phonetic cues or phonological processes and other constituent structures (pragmatic, syntactic, semantic). See also Woodbury (1992).

salient changes in scaling values. This is an extremely narrow view of segmentation; however, without a more fully articulated account of syntactic, semantic, and pragmatic constituency in discourse, it seems wise to avoid attributing their effects to prosodic structure.<sup>3</sup>

All self-interruptions (preceding repairs, restarts, and so-called hesitation phenomena) were coded as junctural markers, since they disrupt the speech stream; ruling this type of juncture out *a priori* could only be motivated by considerations of function, which properly follow a formal characterization.

## 2.2 Tune

The tonal description used in this, and previous, research has its roots in early generative treatments (see Liberman & Pierrehumbert 1984), in that discrete target tones are used to describe rises (LH) and falls (HL, HM). However, the notational conventions used here are largely intended as a pre-theoretical discovery procedure, or null hypothesis, as in phonology generally, where systematic description of phonetic data is a prerequisite for theorizing. The tonal transcription differs from the revision of Pierrehumbert (1980) in Silverman et. al. (1992) primarily in that a minimal phonology is assumed here, consisting of one tone type with three categorical values and variable text-tune alignment (T, with the values H, L, or M, aligned with stress, T\*, or not, T); i.e., no independent categories of phrase accents and boundary tones are postulated. For example, the different phonetic forms that would be described in those systems as H\* L\* L L% would be described here more specifically according to actual phonetic form: H\* L\*, with no implicit L L%; or H\* L\*— (where '—' indicates a simple temporal function corresponding to a sustained final L\*). In addition, no tone is designated as the "nuclear accent," although for the most part the stress immediately preceding the juncture is the only one considered for the purposes of this study (the complete transcripts are fully notated for stress).

Sustained tones are notated with a following dash, T— (H—, L—, M—), indicating that the current pitch value is held relatively constant until the next (notated) tone, pause, or turn change. Tonal interpolation is otherwise a relatively direct path from one tone to the next (e.g., H L indicates a straight downward movement; L H indicates a straight rise from L to H).

<sup>3</sup>In segmenting phrases based on vowel length alone, Wightman et. al. 1991 categorize gradient junctural strength into five levels of phrasing. Although this description has since been incorporated into the intonational transcription in e.g. Silverman et. al. 1992, it has not been used here, since the segmentation criteria include vowel length and textual relations.

In addition, a tone with unspecified value, T, has been used to notate contours set apart from surrounding speech by pauses or shifts in scaling, but which show no internal pitch change. Most such utterances are particles or 'discourse markers', e.g., *um, uh, and, well, so* (cf. Hockey 1991, 1992). While it is possible to specify a local tonal value based on the last value of the preceding phrase, this descriptive criteria often results in a counterintuitive specification. For example, if the preceding junctural tone is scaled very high, and the constant value on a following *and* is lower, it could be described as L, but might nevertheless sound quite high. Analysis in terms of scaling values rather than tonal category would be more useful at this stage of theorizing (cf. Shriberg 1992).

### 2.2.1 A Note on Mid

Mid is used as a descriptive category for reference to the set of values at endpoints of falls that sound non-Low. Phonetically, the criterion used for identifying junctural tones as Mid is primarily that the end value is scaled higher than a preceding L (usually in the same phrase, although in some cases the lower L was in the immediately preceding phrase). When a preceding L tone wasn't available (e.g., in turn-initial utterances), a tone was coded as Mid if it sounded Mid.<sup>4</sup> In the Lesson data, most of the analysis was performed auditorily rather than instrumentally, since many of the very final values for junctural tones were impossible to recover instrumentally.

It may be the case that the phonetic form of gradiently scaled Low tones conflate with that of target Mid tones. For the most part, the criterion used for identifying Mid excludes sequences (more than two) of non-low Low tones that are progressively, gradiently scaled (i.e. in a declining pitch range), since each Low in such a sequence would generally be *lower* than a preceding one, rather than higher. On the other hand, sequences of progressively declining Mid junctural tones in the tonal environment of LHM, where L is lower than M, would be identified as Mid.

In the Interview data, a comparison of L, M and H junctural tone values for the two individual speakers, "Caller" and "Mayor," shows that M values are consistently distinct from L and H junctural tone values when the immediately preceding H peak value is considered for each case (this H value was not used to code M):

<sup>4</sup>In addition to Mid tones identified in terms of relative scaling values, forms that *sound* Mid, in fast speech at least, include falls followed by a slight rise whose value is less than a preceding H\*, and falls to low in which the L is sustained (see Liberman & McLemore 1992 for examples). Several clear cases of the slight rises were notated as Mid. No cases of sustained L were notated as Mid, in order to allow investigation into the functional patterning of these different phonetic forms.

FIGURE 1:

Mayor: L o M m H x

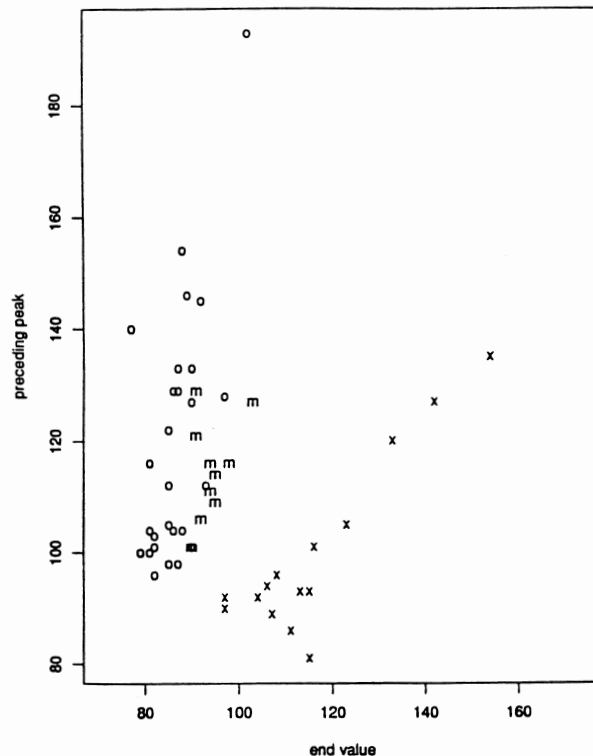
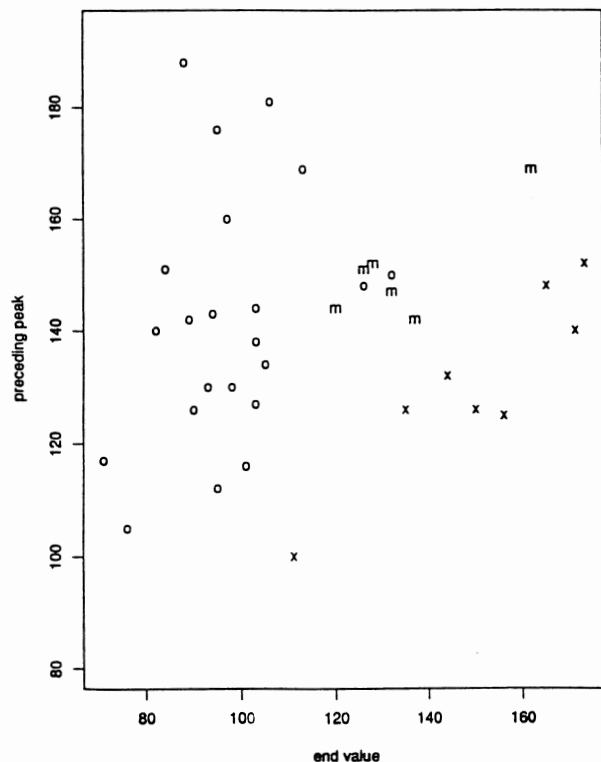


FIGURE 2:

Caller: L o M m H x



## 2.3 Text-Tune Alignment

Tones that align with a stressed syllable are marked with an asterisk, T\*, (H\*, L\*, M\*), and referred to as *accents*. One consequence of the close phonetic description applied here is that junctures can occur immediately following accents. In fact, there are a number of phrases in the Lesson data that are overall rising or overall falling contours, with clear accents at each end point. In the data analysis that follows, I have made a distinction between junctural tones (rises, falls, and levels; i.e. pitch movements that help create a juncture) and what will be called "final accents," stress-aligned tones that precede a juncture created by a pause or e.g., scaling shift.

Thus, the contour referred to as the "vocative chant," which Liberman (1975) characterized as (L) H\* M, has the following possible variants (ignoring the optional L):

H\*M H\*M- HM\* HM\*-

The analysis of intonational features in this study is limited to the two-tone sequence preceding a juncture, including also:

H\*L H\*L- HL\* HL\*-

L\*H L\*H- LH\* LH\*-

Sample pitchtracks, referred to in the analysis below, are shown in the Appendix.

## 3. INTONATIONAL FUNCTIONS

### 3.1 Rises and Falls

In McLemore (1991a, b) I argued that rising, falling, and level junctural tones have the fundamentally iconic<sup>5</sup> general functions of connecting, segmenting, and continuing, respectively. That is, as pitch excursions, both rises and falls segment the speech stream; but rises carry additional information: as the first part of an incomplete pitch peak that implicates a second part, they are used and interpreted as connecting to the second half of a dyad.

These abstract functions are essentially relational; the things related may fall primarily or simultaneously into the three general domains of interaction (e.g. turn-taking), text (textual relations or discourse structure), and information

<sup>5</sup>More specifically, junctural tones are diagrammatic icons arising from acoustic phonetic form, much like a map is a diagrammatic icon for relations between places. They aren't entirely "natural" or universal (cf. Bolinger 1980), but rather depend on the culture-specific evaluation of the signalling roles of intonational primitives in the system and cultural assumptions about which domains are relevant to interpretation.

structure (given/new, background/foreground). Rises, for example, function abstractly as connectives, and convey a broad range of more specific meanings depending on their textual, interactional, and discourse structural environments. They connect turns when the speaking floor is at issue, and textual units when the relation between such units is at issue; they connect participants, when address, participation, and attention are salient themes in context; and among at least some groups of speakers (e.g., the Texas sorority that I studied), they function much like text-aligned H accents to foreground new or exceptional information at the phrasal level when relative ranking of utterances in terms of shared knowledge is at issue (McLemore 1992c). (In the latter case, the relation is not so local and linear as the others, but rather paradigmatic, i.e. in the choice of contrastive tonal value). To summarize, junctural H tones:

- (segment and) connect
- implicate a second part, in associated text or interaction
- foreground associated information or action

In contrast, falls to low convey the least amount of relational information; in the sorority data, this form was found to generally segment textual phrases and turns, and under certain circumstances to co-occur with old, expected, or otherwise unexceptional discourse contributions. Unlike rises, falls alone do not elicit response (i.e. without additional conventions, textual information, or other cues coming into play). To summarize, junctural L tones:

- segment
- don't provide any information about what follows, in associated text or interaction
- background associated information or action

More specific interpretations of junctural tones, such as uncertainty, hesitation, conclusiveness, etc., arise from (more or less conventionalized) co-occurrence with text and aspects of context.

### 3.2 Levels

When junctural tones (H, L) are sustained (H–, L–), they become transparently iconic signs for continuation (as the tone is sustained, so is some aspect of the speech activity underway).

Tonal perseveration for H (H–) makes its general function continuative rather than connecting (i.e. current values for the speech activity underway are maintained, rather than changed as with H), and has the effect of changing its interactional value. That is, H– is more relevant to the interpretation of textual relations than to participant relations; unlike H, the intonation itself doesn't elicit response (i.e. but can if it co-occurs with text or other cues that do).

Perseveration of L (L–) also changes its general function from segmenting to continuing, (again, current values are maintained rather than changed, as with L). Like L, L– doesn't overtly cue interactional behaviors, although it can co-occur with them; like H–, L– cues a local continuative relation between textual units.

In the sorority corpus, both H– and L– junctural tones were found to co-occur primarily with old or expected information (consistent with Ladd's 1978 observations, as well as the data analyzed in Walker 1992), although within that functional space, H– still appears to mark information as foregrounded.

### 3.3 Falls to Mid

If falls to low segment the speech stream, and thereby text and interactional units, what do falls to mid do? As with an intonational rise, the form itself is incomplete when compared to a whole pitch peak; since function follows form closely in intonation, it isn't surprising that falls to mid seem to mark incompleteness. The theory outlined above would predict that since falls to mid are falls that don't completely segment, they should share some characteristics with both rises and falls (see also Liberman 1975). The actual functional correlates of HM junctural tones and levels will be examined in the following sections.

## 4. COMPARISON OF DATASETS

The classroom data ("Lesson") consists of an excerpt of approximately 10 minutes of speech from a second grade mathematics lesson conducted in an inner-city Parish school in Pittsburgh, Pennsylvania. The portion of the lesson examined is the initial portion, called "pre-team" by participants, in which the teacher sets up the problem to be solved and works out preliminary solutions with contributions from the students. (Twenty-one of 225 phrases in the transcript are students' utterances, or joint teacher-student utterances; they have not been included in the intonational analysis because they are largely inaudible. An additional 2 phrases spoken by the teacher were not included because they were inaudible, resulting in a total of 202 phrases.)

The radio talk show ("Interview") dataset consists of an excerpt of approximately 4 minutes of speech from a call-in interview talk show aired on the public radio station in Austin, Texas. This particular exchange is between a caller (C), and the Austin mayor (M); the host's introduction and closing have been removed (for a total of 103 phrases). It is one of several exchanges between the interviewee and callers during the one-hour program. The intonational characterization below is of the entire exchange, including the speech of both the caller and the mayor, which display striking similarities.

In both the radio talk show interview and classroom contexts, participants bring knowledge about appropriate interactional behaviors to the verbal exchange. Call-in radio talk shows have a recurrent basic structure: the host introduces callers who direct questions or comments to the guest (and if there are no callers, the host plays this role); the guest responds; the host has the option of limiting the duration of either speaker's turn (by e.g., introducing another caller or requesting clarification). The nature of the communicative medium also imposes constraints on interaction: participants know they have a limited amount of time for the exchange, and that silence is to be avoided (see Coles 1991, Goffman 1981:197-330). This means that if one has the speaking floor, there is an especially urgent obligation to keep it filled with sound; and on the other hand, participants must respond promptly when a response appears to be called for.

A second grade classroom also has a recurrent, basic interactional structure (see Resnick et. al. 1991): the teacher has primary obligation to maintain the speaking floor, and allocates it either by calling on individual students (verbally or gesturally) or by indicating that a contribution from the class is required. Indeed, part of the lesson taught in the classroom is appropriate interactional behaviors generally. Furthermore, when verbal or nonverbal contributions, or attention more generally, are elicited from students, they are obligated to respond, or otherwise be (implicitly or explicitly) reprimanded.

#### 4.1 Intonational Frequency Differences

The most significant (and obvious) difference between the two datasets is in the number of accents per intonational phrase. The pedagogical discourse shows fewer accents per phrase (1.5) than the interview discourse (2) – i.e. phrases are shorter and less complex intonationally. This difference is even more striking than the numbers suggest, since about half of the one-accent phrases in the Interview consist solely of discourse markers or so-called pause fillers, while only one third of the one-accent phrases in Lesson do.

A comparison of overall contour types (tonal sequence corresponding to the whole 'phrase') in the two discourses indicated very little difference in the frequency of phrase-internal tonal elements. However, junctural forms are different at the level of better than  $p=.999$ . The greater variability in junctural forms than in phrase-internal composition is undoubtedly due in part to the fact that a majority of the intonational phrases segmented in Lesson contain *only* the junctural tone sequence.

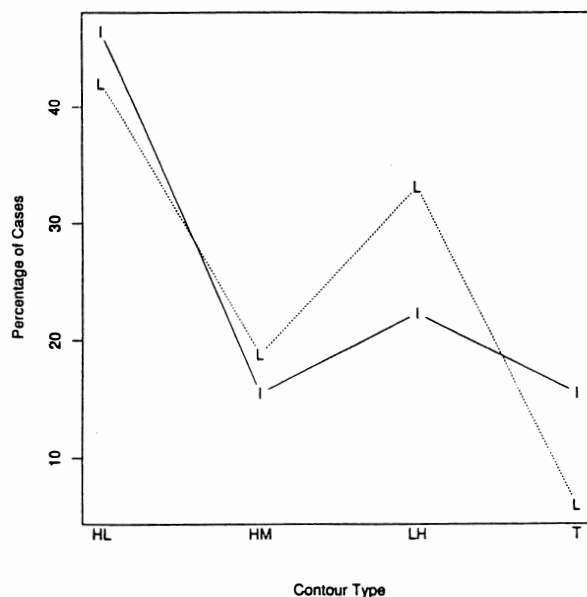
The most striking and significant difference in junctural tone frequency across the two datasets is that forms of LH (rises and high final accents) occur more in Lesson than in Interview:

Proportions of Junctural Form Types for each dataset:

	HL	HM	LH	T
Lesson	.42	.19	.33	.06
Interview	.47	.15	.22	.15

FIGURE 3:

Junctural Tone Types: LESSON and INTERVIEW



A higher proportion of phrases in Interview consist solely of level pitch, with no tonal value assigned (T above), than in Lesson. Since the T category itself is heterogeneous in pitch values relative to surrounding material (although constant in lack of pitch movement), it will be compared to level or sustained tones more generally (i.e. the class of H-, L-, M-, T).

The table below shows the proportion of junctural forms in each dataset according to stress alignment and tonal perseveration.

Proportions of Stress Alignment and Tonal Perseveration Variations:

	T*T	TT*	T*T-	TT*-	T
Lesson	.58	.20	.10	.06	.06
Interview	.58	.07	.05	.15	.15

A greater proportion of levels occur in Interview than in Lesson (i.e. the class of final level tones and phrases without tonal movement, T\*T-, TT\*-, and T below), and, as noted, there is a higher proportion of final accents in Lesson than in Interview.

These junctural form types are further distinguished by tonal value in the tables that follow; proportions shown are relative to the total of all three tables below (i.e. of 103 junctures in Interview, 202 junctures in Lesson).

Falls to mid occur more frequently in Interview than in Lesson, while rises occur more frequently in Lesson:

	FALLS		RISES
	H*L	H*M	L*H
Lesson	.30	.06	.21
Interview	.36	.15	.07

Among level junctural tones, sustained Mids occur more frequently in Lesson (especially H\*M-), while sustained Highs occur more frequently in Interview:

	LEVELS					
	H*L-	HL*-	H*M-	HM*-	L*H-	LH*-
Lesson	.02	.03	.07	.02	.01	.01
Interview	.04	.02	--	.01	.01	.12

Finally, the two datasets also show a difference in the frequency of final accents, which are more common in Lesson overall: High final accents occur more frequently in Lesson than in Interview:

	FINAL ACCENTS		
	HL*	HM*	LH*
Lesson	.07	.04	.10
Interview	.05	--	.02

To summarize, with respect to the general categories shown in Figure 3 above: In Lesson, most occurrences of LH are rises, L\*H (about two thirds), while in Interview, most are high levels, LH\*- (over half). The two datasets are similar in the proportion of falls to low; however, among variants of HM, falls to mid (H\*M) occur more frequently in Interview, while a greater proportion of mid levels (HM-) are evident in Lesson. Among the greater proportion of final accents in Lesson, High final accents are far more frequent in Lesson than in Interview.

## 4.2 Intonational Distributions

Rises and high levels, and falls to mid and mid levels, are examined more closely below for their distribution within and between the two datasets, with respect to interactional behaviors (e.g. turn-taking), shared knowledge (including e.g., repetitions), and aspects of discourse structure (e.g., local textual relations).<sup>6</sup>

### 4.2.1 Rises (L\*H)

In the Lesson dataset, rises pattern as cues to interactional behaviors in general, but underdifferentiate verbal interaction from nonverbal action or attention (i.e. connecting participants more generally rather than simply speaking turns). The majority of junctural tones preceding turn change are rises:

	LESSON:	
	H*L	L*H
Turn change	3	13
No turn change	57	30

There are 18 instances of student or whole-class response to the teacher: about two-thirds of the teacher's phrases (13 of 18) that elicit responses are spoken with L\*H; 3 are spoken with H\*L; one with LH\* and one with T- (neither of which is shown in the table above).<sup>7</sup>

Of the rises preceding turn change, 4 occur on vocatives (*Nicole?*), 5 begin counting sequences which the students join in, and 5 occur on imperatives, questions, or variations on those forms, such as truncated *be* statements (e.g., *three groups of four is?*). In three of the latter 5 cases, the teacher names a student prior to the utterance that finishes her turn.

The 3 falls that precede turn change also occur on vocatives, but interestingly, differ from vocatives carrying rises in that they occur on names of boys rather than names

<sup>6</sup>Intonation is represented in the text as follows:

Falls to low are marked with a period ("."), falls to mid are marked with an addition sign ("+"), rises are marked with a superscript question mark ("?"), and sustained tones are indicated with a dash ("-"). Pauses are shown in angled brackets ("< >"), measured in seconds or noted as "<.>". CAPS indicates a High pitch excursion, bold indicates a Low (not necessarily marked).

<sup>7</sup>In Lesson, the single instance of T- that is followed by response occurs on a truncated statement, which the student, named in an earlier utterance, finishes (after "///"): *seven rows- I sound like more than- // three rows.*



of girls (e.g., *Martin*).<sup>8</sup> The numbers are quite small and merely suggestive of gender socialization (cf. the finding that American women use intonational rises more frequently when talking with other women (Edelsky 1978), or report that they do (McLemore 1991a, b)). Some vocatives addressed to boys take rises when attention or other non-verbal action appears to be elicited (as in e.g., line 123 shown below).

In addition to cueing verbal interactional behavior, L\*H is used on utterances that elicit specified behaviors, or attention more generally (e.g. with students' names; see also Figure 4 in the Appendix):

- 123     AntoNIO?  
          L\*   H
- 124     there's another TEAM?  
          L\*        H
- 125     team four?  
          L\*H

It is important to note what is co-occurring in the classroom, as well as what follows use of the form: on videotape it is apparent that the teacher sometimes uses rising intonation on a current utterance, combined with gaze, to elicit attention from students who may have become distracted (cf. Keenan, Schieffelin & Platt, 1978).

Direction of attention or (verbal or nonverbal) action appears to be implicated in every occurrence of L\*H in Lesson. This suggests that the children must attend to other cues (e.g., gesture, text) or conventionalized routines to determine the more particular significance of a rise for interactional behavior.

In the Interview dataset, the majority of junctural tones preceding turn change are falls to low, H\*L, although 2 of the 5 rises that occur do precede turn change:

INTERVIEW:		
	H*L	L*H
Turn change	8	2
No turn change	31	5

<sup>8</sup>In one of the cases mentioned above, L\*H occurs on a boy's name, but is exceptional in being a vocative tag to an overall rising yes-no question rather than occupying its own phrase. In one of the instances of H\*L, the same boy's name carries a fall when it occurs as a vocative tag to an overall rising WH question.

There are 12 changes of speaker: 8 follow intonational falls, and 2 follow intonational rises; one follows HL\* and one HL\*- (not shown in the table above). The 8 falls that precede turn change are not significantly different in form from the 31 falls that don't; they are scaled to quite low values, but similarly low values occur within single-speaker turns.

Both of the rises (L\*H) are used by the Mayor, one in his initial single-phrase turn (*hello Rick?*), and the other in his second single-phrase turn, a yes-no question (*uh Rick is that area posted for no parking?*). In both cases, the rises redundantly cue the subsequent response.

No uniform pattern is evident in the additional 5 rises used in Interview within same-speaker turns, although all occur between clauses: in 2 cases L\*H appears to create additional cohesion between clauses (*if C<sub>1</sub> then C<sub>2</sub>; C<sub>1</sub> because C<sub>2</sub>*), and in 3 cases L\*H has the effect of foregrounding a new referent (see example below).

Why don't these rise cue a response? Because the particular domain to which an intonational form is interpreted as relevant, text or interaction, depends in part on the pragmatic and semantic import of the text with which the form is associated, as well as the interactional conventions at work in a given exchange. In short, when the 'connection' function is dropped into a textual location where turn-change is not plausible (either because there is no coherent meaning to respond to or because it's clear from textual structure that the speaker is not finished), that function is applied to textual units rather than turns or interactants.

#### 4.2.2 High Levels (LH-)

High levels of the form LH\*- constitute 12% of the total junctural forms in Interview, which makes them by far the most frequent form of level junctural tone in that dataset. In comparison, High levels constitute only 1% of junctural tones in Lesson. The following example from Interview is especially illustrative of the different uses of rises and high levels:

- 53     AUstin is a great climate?  
          H\*        L-        L\*    H
- 54     It's a good CLIMATE FOR BICYCLING-  
          L        H\*-
- 55     uh THAT community is growing?  
          H\*                    L    L\*    H
- 56     OUGHT TO BE GROWING-  
          T\*-

- 57 and it's a **pr**oDUCTIVE THING to DO 'N-  
L H\*- L H\*-
- 58 it's less intrusive to our **en**VIronMENT-  
L H\*-
- 59 uh more Energy efficient as **they** SAY-  
H\* L H\*-

(See Figure 5 in the Appendix.)

The rises in 53 and 55 occur on a new referent or predicate in the discourse, while the High levels occur on discourse old or inferrable information (see Prince 1991 for given/new distinctions). The use of High levels on text framed as a listing sequence (i.e. in which the theme is continuous across junctures that correspond to a parallel textual frame) is similar to the use of phrase-final levels in the sorority planning meeting discourse reported in McLemore (1991a, c), in which the discourse is structured by a written program under discussion; there, too, L\*H and LH- had distinct distributions, in which L\*H occurred on exceptional or unordered items. In this case, the aspect of text relatively foregrounded by L\*H happens to be newness (the material from which the subsequent list is formed), and the text relatively backgrounded by LH- is a continuation of it.

Most of the High levels in Interview are used in this way – i.e., in a series of similar utterances containing old or inferrable information. (Two of the 5 repetitions in Interview take High levels, and a third is categorized as T, but sounds quite high). Two exceptions occur early in the exchange; they occur on new referents, and have a foregrounding effect similar to L\*H (*I've noticed when I'm bicycling especially on SHOAL CREEK – I that the BIKE LANE-*).

High levels aren't generally used in Lesson. L\*H is used on the parallel counting sequences, which are usually begun by the teacher and joined by the students (so elicitation of participation is appropriate). There are 45 repetitions in Lesson; none take High levels. Except for the counting sequences, which take L\*H, the majority of repetitions occur with falls (to mid or low), or (mid or low) levels.

L\*H- doesn't precede a turn change in either Lesson or Interview.

#### 4.2.3 Falls to Mid (H\*M)

Like High levels, falls to Mid never precede a turn change in either dataset<sup>9</sup>; they also have a tendency to co-occur

with old information. Occurrences of H\*M constitute 15% of the total junctural forms in Interview compared to 6% of Lesson; in addition, their usage patterns are slightly different in the two datasets.

Uses involving self-interruptions and apparent disfluencies of various kinds (e.g. when the following utterance begins with or consists of *uh* or *um*) account for about half of the total number of H\*M in Interview, compared to only about 1/6 of the occurrences of H\*M in Lesson.

Consider the following general pattern of H\*M use. The phonetic form results from physical contingencies when a speaker stops talking abruptly after a High peak (e.g., from Lesson: *SOMEone's+ / SOMEone's TALKing while you're TALKing.*) This would seem to instantiate neither a target Mid tone nor a higher-scaled Low tone; indeed, the Mid value could hardly result from tonal *target* per se at all. However, this kind of occurrence is in principle difficult to distinguish from the use of H\*M in less mechanical cases of self-interruption (repair, disfluency, etc.) – which may also be used for deliberate communicative purposes, such as floor-holding. For example, consider lines 23 and 24 from Interview (which follow the Caller's report of a problem):

- 23 is there **ANY**thing you can do as **MA**yor to **HE**LP us+  
H\* L H\* L H\* M
- 24 **HE**LP us+ <.2>  
H\* M
- 25 **RI**ders who **are** **TRY**ing to **ge**t **OU**T there **and**. <1.0>  
LH\* L H\* L H\* L

This usage is suggestive of an oblique interactional function of Mid (e.g., floor-holding); such a function is all the more oblique because of its affinity to relatively unplanned 'accidents' of speech. (Overt interactional behaviors such as speaker change aren't the extent of interaction in speech communication; all speech forms have interactional consequences, including not changing speaker turns. Intonation is an important resource in avoiding turn changes at points where there might otherwise be opportunities for them. See Sacks et. al. 1974.)

Another general pattern of H\*M also has an instantiation that appears to result from physical contingencies; that is, when speech rate is accelerated through a juncture so that attaining a very low value for Low would be physically difficult (i.e. undershoot). This is especially apparent when the following phrase is shifted upward in pitch range. An

<sup>9</sup>Of course, given the appropriate textual material or local conventions, any junctural form can precede a turn change. In some dialects of American English, as well as Scottish and

British English, H\*M appears to be conventionally associated with certain types of interrogatives. (See e.g., Brown et. al. on Scottish English.)



example from Interview is shown in line 89 (see also Figure 6 in the Appendix):

88 AND uh-

89 Certainly we'll enCOUrage the poLICE to do  
everything we CAN+

90 we CAN'T reLY upon the poLICE to MAKE people  
oBEY the LAW in ALL Cases.

Again, however, the effect of this usage of H\*M on both local textual relations and turn-taking (or lack of it) is indistinguishable from slower, seemingly more deliberate uses of H\*M.

In both types of uses, whether deliberate or 'disfluent', H\*M functions somewhat like a fall and somewhat like a rise, indicating a cohesive relation between two utterances (and their associated texts and acts) not by overtly marking connection or continuation, but by not quite segmenting the two utterances. In both uses, the material on which H\*M occurs is in some sense treated by the speaker as less important than the material that follows it. There is a tendency for H\*M to background, rather than foreground, when the relative rank of information is relevant to the communicative event. Consider the following example from Lesson:

60 but SEven+ <.>  
H\* M

61 seVEN rows sound like MORE.  
L\* H L- H\* L

Here, the teacher responds to a students' answer by re-introducing the term *seven* into the discourse (*seven* ended a counting sequence several utterances prior, and other terms referring to the items counted have been used in the meantime, e.g., *number of cupcakes on this tray*). In line 60, *seven* is marked as salient by the High accent, but backgrounded by the M junctural tone.

In this usage, H\*M is similar to the so-called 'backgrounding' or B contour examined in Liberman & Pierrehumbert (1984) and Steedman (1991) (who describes it as L+H\* L H%). Based on the data analyzed here, as well as that in Liberman & McLemore (1992), it appears that in fact (...)H\*LH and H\*M are variations of the same form. (See McLemore 1992 for a more indepth discussion of formal and functional patterns of falls to mid and their variations.)

Finally, another general pattern of H\*M use is closely related to the backgrounding function, but differs from it in that rather than occurring on old information or a theme or topic about which more follows, it occurs on utterances out of the blue that set up an expectation for more speech

or action, e.g. in 2 instances from Lesson on *you know what*. Since this use of H\*M, in combination with such textual phrases, has the effect of creating suspense, I will refer to it here as the *anticipatory* function of H\*M. It exploits the locally cohesive function of H\*M, as well as the backgrounding usage. Most occurrences of H\*M in Lesson are backgrounding or anticipatory.

The anticipatory function of H\*M is even more apparent when the final M is sustained as a level, i.e. H\*M-, which is a frequent form of HM in Lesson, as discussed below.

#### 4.2.4 Mid Levels (H\*M-)

Mid levels of the form H\*M- occur at 7% of junctures in Lesson; they are the most common form of level junctural tones in the Lesson dataset. None occur in Interview.

Nearly one third of H\*M- occurrences are within a sentence, as in the following example (see Figure 7 in the Appendix):

33 LAUREN said+-  
H\*- M-

34 oKAY?  
L H\*

35 LAUREN said+-  
H\*- M-

36 "there are THREE rows. "  
L- H\* L

This excerpt is from part of the fictional narrative the teacher uses to set up the problem; line 36 contains the first mention of a number that will be crucial to the solution. The use of H\*M- in the quotative frame of lines 33 and 35 is both backgrounding and anticipatory with respect to what follows (line 36). The sustained final Mid tone enhances the anticipatory function by simple temporal duration. (Note the relation between the intermediate scaling of Mid and its tonal lengthening; each aspect suggests continuation.)

As the *okay* with LH\* in line 34 indicates, the teacher is concerned with holding the students' attention; indeed, in other occurrences of H\*M-, the teacher is simultaneously performing gestures, such as holding up a certain number of fingers, to which the students must attend.

Recall that H\*M and H\*M- are never used to elicit response; they are also never used to directly elicit students' attention. Rather, the teacher uses L\*H or LH\* to overtly elicit attention or response, and forms of Mid to actively *hold* students' attention. That is, the use of Mid in some (but not all) cases indicates that the speaking floor may be in question and the current speaker is claiming it. That is

also the interactional function of level junctural tones, as noted previously, which makes sustained M- all the more effective in this regard.

### 4.3 Summary

In Lesson, phrases are shorter and simpler intonationally: many phrases contain one continuous pitch movement, overall falling or overall rising, some with emphasis at either or both ends (TT\*). Although the teacher is the primary speaker, participation of the students is considered necessary to accomplishing the pedagogical purpose of the discourse, hence the frequent elicitation of students' attention, action or verbal response (L\*H). At the same time, the teacher must maintain interactional order so that students follow the main points of the lesson (H\*M, H\*M-). The discourse is pre-planned for the most part; i.e. the teacher has outlined the lesson, if not the particulars of her speech, and the discourse incorporates speech routines that are used in other datasets from this classroom. When intonational junctures occur within otherwise cohesive textual units, or on repetitions, they are more often accompanied by enhancing gestures and actions than by indications of disfluency, suggesting a careful, emphatic speech style (H\*M, H\*M-); indeed, there are numerous repetitions and rephrasings (see Resnick et. al. 1991 on the forms and pedagogical functions of *revoicing* in this classroom). Consistent with this, there are no occurrences of *uh* or *um* in the dataset, but rather discourse markers that, together with the whole range of intonational forms, appear to play a role in eliciting attention and structuring the discourse and the activity overall.

In Interview, phrases tend to be long and intonationally complex. With some frequency, junctural forms break up otherwise cohesive textual units (H\*M, T), often accompanied by self-interruptions and pause fillers. In part, this undoubtedly reflects the unplanned nature of the discourse, although these patterns also have useful interactional implications (i.e. preventing speaker change). Turn change is not cued explicitly with intonation, but apparently by other (semantic and pragmatic) sources of interactional information. Emergent parallelistic structure is evident in some same-speaker turns (LH\*-), and given/new relations are marked intonationally in some of these sequences (L\*H, LH\*-).

## 5. CONCLUSION

The different junctural forms clearly have a high degree of overlap in aspects of their functions; e.g., H\*M and LH\*- are similar in (generally) not cueing interactional behavior, creating local textual cohesion, and co-occurring with non-new information. The distinction between them is subtle, and has as much to do with the associations they accrue from both general and local patterns of use (i.e. their more

or less conventional indexical value) as with the fundamental differences in their (iconic) form.

Nonetheless, for individual junctural forms there are clear tendencies in usage patterns with respect to interaction, information, and textual relations. For the community of speakers from which the present data were drawn, it would be surprising to find uses (conventionalized or not) that aren't generally consistent with the patterns observed here; rather, local discourse variation is evident in the selection of specific forms, and the aspect of their function that is made salient by the contingencies of the communicative event itself (including, e.g., the domain a given use is primarily taken to comment upon, text or interaction).

Consider the case of H\*M. In both Interview and Lesson, the patterning of H\*M suggests that it is used, like levels, to continue across a juncture. However, like L\*H and H\*L, the phonetic form contains a pitch excursion which effectively segments the pitch stream; it differs from H\*L in not segmenting all the way to Low, and unlike LH\*-, it segments in part. Like L\*H, it is used to mark more specific relations between the contours it partially segments, as well as their associated text and acts. In Interview, this cohesive function is most salient; i.e. it occurs primarily in self-interruption, mid-clausally, and on particles like *um*. Its use has interactive implications (i.e. in terms of floor holding), which are less direct than e.g., the use of L\*H; it differs crucially from L\*H in not explicitly cueing interactional behaviors. In Lesson, the backgrounding and anticipatory functions of the final mid value are most salient, and the interactional consequences (of floor- or attention-holding) are more clearly evident, since all forms of attention and interaction are visible. The most frequently occurring subtype of HM is a mid level, in which the anticipatory function is enhanced, with respect to both text and interaction.

## REFERENCES

- Brown, G., K. Currie, & J. Kenworthy. 1980. *Questions of Intonation*. Baltimore: University Park Press.
- Bolinger, D. 1980. Intonation and "Nature." In M. Foster & S. Brandes, eds., *Symbol as Sense: New Approaches to the Analysis of Meaning*. New York: Academic Press.
- Coles, F. 1991. Pre-Sequences in a Radio Talk Show. In *Texas Linguistics Forum 32: Discourse*, C. McLemore ed. Austin: University of Texas Department of Linguistics and the Center for Cognitive Science. 21-42.
- Du Bois, J., S. Schuetze-Coburn, D. Paolino & S. Cumming. 1991. *Discourse Transcription*. University of California at Santa Barbara ms.

- Ferguson, C. 1983. Sports Announcer Talk: Syntactic Aspects of Register Variation. *Language in Society* 12: 153-172.
- Fillmore, C. 1981. Pragmatics and the Description of Discourse. *Radical Pragmatics*. Cole, P. (ed.). 143-166.
- Goffman, E. 1981. Radio Talk: A study of the ways of our errors. In *Forms of Talk*, pp. 197-330. Philadelphia: University of Pennsylvania Press.
- Gumperz, J.J. 1982. *Discourse Strategies*. London: Cambridge University Press.
- Hockey, B. A. 1991. Prosody and the Interpretation of *okay*. Paper presented at the AAAI Fall Symposium, November 1991.
- Hockey, B. A. 1992. Prosody and the Interpretation of Cue Phrases. Paper presented at the IRCS Workshop on Prosody in Natural Speech, Philadelphia, PA, August 5-12.
- Keenan, E. Ochs, B. Schieffelin & M. Platt, 1978. Questions of Immediate Concern. In E. Goody (ed.), *Questions and Politeness*. Cambridge: Cambridge University Press. 44-55.
- Ladd, D.R. 1978. Stylized Intonation. *Language* 54:3, 517-541.
- Liberman, M. 1975. *The Intonational System of English*. MIT Ph.D. Dissertation. (Published by Garland, New York, 1979.)
- Liberman, M. and C. McLemore. 1992. The Structure and Intonation of Business Telephone Openings. In *The Penn Review of Linguistics* 16: 68-83.
- Liberman, M., C. McLemore & A. Woodbury. 1991. On the Nature of Prosodic Phrasing. Paper presented at the LSA Summer Institute Workshop on Grammatical Foundations of Prosody, July 5 1991, University of California at Santa Cruz, Santa Cruz, CA.
- Liberman, M. and J. Pierrehumbert. 1984. Intonational Invariance Under Changes in Pitch Range and Length. In *Language Sound Structure*, M. Aronoff and R.T. Oehrle, eds. Cambridge, Mass: MIT Press.
- McLemore, C. 1992. Discourse Context and Prosodic Function. Paper presented at NWAWE XXI, Ann Arbor, Michigan, October 1992.
- McLemore, C. 1991a. *The Pragmatic Interpretation of English Intonation: Sorority Speech*. University of Texas Ph.D. dissertation.
- McLemore, C. 1991b. The Interpretation of L\*H in English. In *Texas Linguistics Forum* 32: *Discourse*, C. McLemore, ed. Austin: University of Texas Dept. of Linguistics and the Center for Cognitive Science. 175-196.
- McLemore, C. 1991c. Boundary Tone Values and Shared Knowledge. Paper presented at Grammatical Foundations of Discourse and Prosody, LSA Summer Institute workshop, Santa Cruz, California, July 1991.
- Pierrehumbert, J. 1980. *The Phonetics and Phonology of English Intonation*. MIT Ph.D. dissertation.
- Prince, E. 1991. The ZPG Letter: Subjects, Definiteness, and Information-Status. In S.A. Thompson and W. Mann eds., *Discourse Description: Diverse Analyses of a Fund-Raising Text*. Amsterdam/Philadelphia: John Benjamins.
- Resnick, L., M. Leer, V. Bill, and L. Reams. 1991. From Cupcakes to Equations: The Structure of Discourse in a Primary Mathematics Classroom. University of Pittsburgh Learning Research and Development Center ms.
- Sacks, H., E. Schegloff, and G. Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50: 696-735.
- Shriberg, E. 1992. Intonation of Clause-Internal Filled Pauses. Paper presented at the IRCS Workshop on Prosody in Natural Speech, Philadelphia, PA, August 5-12.
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, J. Hirschberg. 1992. TOBI: A Standard for Labeling English Prosody. *ICSLP Proceedings*.
- Silverstein, M. 1979. Language Structure and Linguistic Ideology. In P. Clyne, et. al. (eds), *The Elements: A Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society. 193-247.
- Steedman, M. 1991. Structure and Intonation. *Language* 67.2: 260-296.
- Swales, J. 1990. *Genre Analysis*. Cambridge, U.K.: Cambridge University Press.
- Walker, M. 1992. When Given Information is Accented: Repetition, Paraphrase and Inference in Dialogue. Paper presented at the IRCS Workshop on Prosody in Natural Speech, Philadelphia, PA, August 5-12.
- Wightman, C., S. Shattuck-Hufnagel, M. Ostendorf, and P. Price. 1991. Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. Unpublished ms.
- Woodbury, A. C. 1992. Prosodic Elements and Prosodic Structures in Natural Discourse. Paper presented at the IRCS Workshop on Prosody in Natural Speech, Philadelphia, PA, August 5-12.
- Woodbury, A. C. 1987a. Rhetorical Structure in a Central Alaskan Yupik Eskimo Traditional Narrative. In J. Sherzer & A. C. Woodbury, eds., *Native American Discourse: Poetics and Rhetoric*. Cambridge: Cambridge University Press.

# APPENDIX

Figure 4

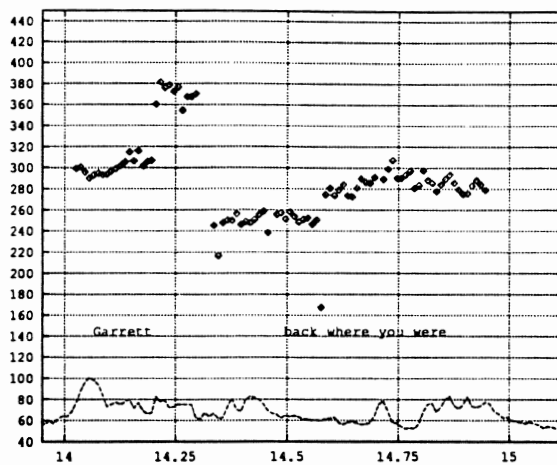


Figure 5

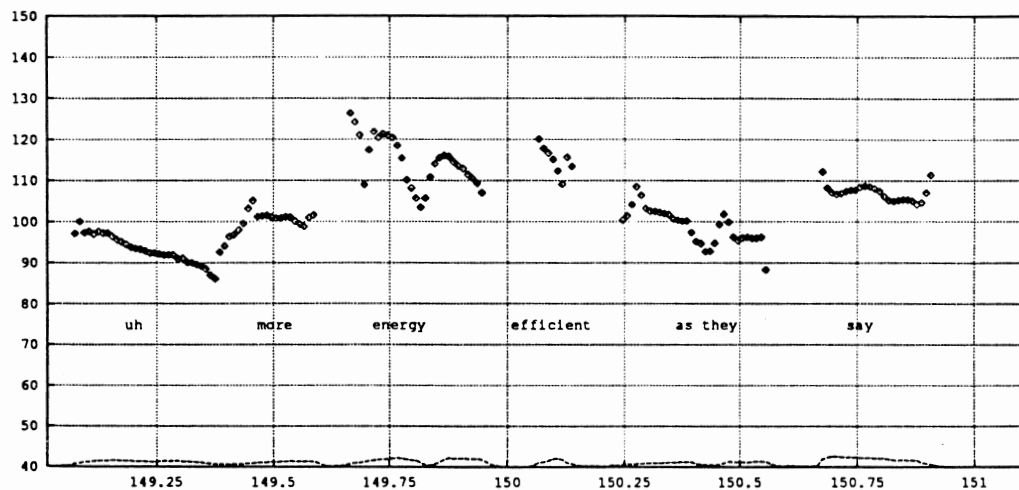


Figure 6

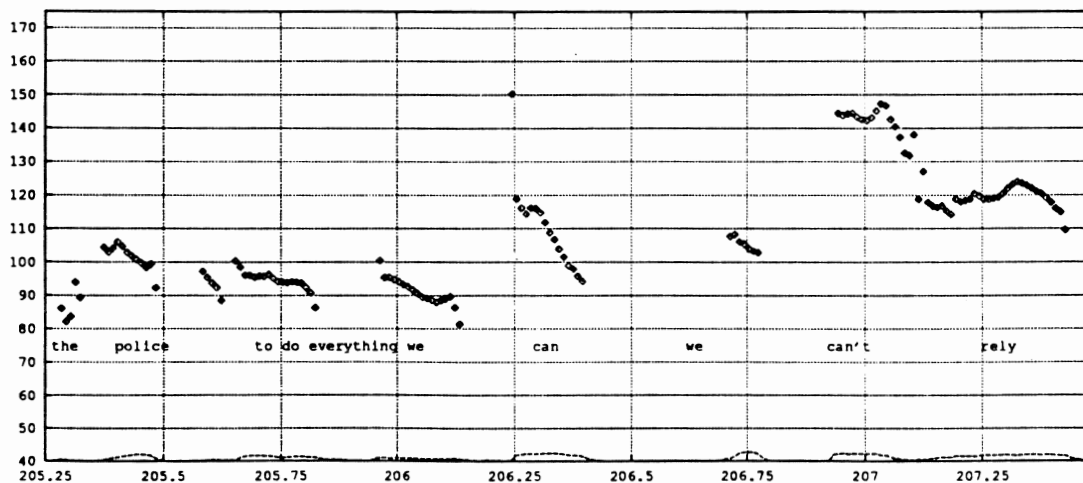


Figure 7

