

Grouping Intonational Phrases in Discourse:  
A Phonological and Phonetic Toolset

Hunter Hatfield

Ling 640F  
May 9, 2005

## **Abstract**

This paper examines grouping of Intonational Phrases in a speech by Malcolm X using the Autosegmental-Metrical model of intonation. Two types of grouping are identified – local and global, where local means that proximate Intonational Phrases (IPs) are associated with one another by specific local cues that happen in a linear string, and global means that large chunks of discourse are distinguished from one another in a stylistic, gradual way. The primary distinction is that a local group will be implemented with specific cues that occur at easily identified points in time, while a global group is implemented gradually and inconsistently. The local grouping cues to be discussed include the repeat of a tonal sequence, the use of nuclear accents that refer across IP boundaries, the “dephrasing” of an entire IP, a pattern of suspending and completing boundary tones, and the phonetic realization of pauses. Global grouping cues include the overall use of different phonological tones and patterns, the length of an IP, speech rate, pitch range, and pitch level. The paper remains exploratory in nature and no final phonological model of Group phrases is asserted.

## **Introduction**

Most studies of intonation using the Autosegmental-Metrical model perform no analysis higher than the level of the Intonation Phrase (IP). A great amount of research goes into discovering the internal structure of IPs. Do they have intermediate phrases (iP) inside the IP, such as in English, or is the grouping like an accentual phrase (AP) in Korean or Japanese? What types of phonological accents can be contained within these phrases and what cannot be? What prosodic categories mark iP, APs, and IPs? All of this moves toward describing the phonological structure of a single phrase in a given language.

However, an IP is a small piece of most spoken discourse. An IP might only be a fragment of an utterance. The question then arises: is there any structure to intonation above the individual IP? Are there Group Phrases of any sort, and if so, what can they contain, how are they marked, and what is their purpose? This paper looks at discourse intonation in the narrative of a single speaker, particularly in a political speech by Malcolm X.

Most work on intonation in discourse that uses the A-M model comes from the field of speech processing, and especially computational linguistics, where there is a need to find specific cues for marking discourse segments. Nakatani et al 1995 studied news and direction-giving corpora and found correlations between pitch range, amplitude, and timing and discourse segments. In a directions-giving task, both spoken and read, they found a relationship between discourse structure and maximum and average F0, speaking rate, and pause duration. Grosz and Hirschberg 1992 focused on discourse boundaries – segment beginning and segment ending cues – as well as items like direct quotation,

indirect reported speech, and parentheticals. They studied many of the same phonetic cues as this paper, as well as phonological contour type, and nuclear accent. They also found that pitch range and pauses were associated with segment boundaries. Hermann 2001, working in the same tradition, studied the phonetic marking of discourse final utterances, rigidly controlling for phonological factors, in order to test the hypothesis that discourse final phrases will have a greater final lowering than discourse medial phrases. She does find multiply realizable cues of frequency, amplitude, and final lengthening. A major strength of these studies is the use of a specific method of discourse analysis, while the present study has relied on native speaker intuition.

Ladd 1996 approaches this topic differently from the the above authors. His focus (section 6.3) is the nature of the prosodic hierarchy. Specifically, are there distinct levels of intonation which cannot be violated, what is called the strict layer hypothesis, or can a feature at level A be associated directly with a feature at level C without an intermediate level B? He argues for a limited recursiveness within the prosodic hierarchy, such that one phrase can join with another phrase to create a phrase of the same type at a higher level. This paper will largely stay neutral on this issue, though it does side more with Ladd in that there is no enforcement of the strict layer hypothesis in the trees presented. Instead we will let the phonetic and phonological cues group as they please and not attempt to impose/discern a systematic theory.

This paper will provide evidence for a small set of phonological and phonetic tools with which a speaker can group IPs into larger units:

- Duplication of Tonal Patterns
- Nuclear Pitch Accents across IPs
- Dephrasing of Final IP
- Pause Duration
- Suspended and Final Boundaries

There are two main purposes to such groupings. 1) The speaker can signal when a thought is complete such that a new thought or topic will emerge; and 2) The speaker can indicate to the listener how to evaluate or understand what is being said. For instance, by grouping the word personal in one IP with the word political in another IP, the listener understands that the later phrase's political stance is to be understood in complement to the former personal stance. Both purposes ultimately give instructions to the listener about how to

understand the speaker's intentions by grouping particular words and phrases in particular ways. There is also likely another reason for these groupings, which might be considered simply aesthetic. Certain things sound good. I will say little about this aspect of discourse intonation, though it should not be forgotten. This will come up again in the future research section in relation to music.

## **Materials and Methodology**

To look at the structure of intonation across large discourses, a speech given by Malcolm X in 1964 called "The Ballot or the Bullet" is used. There were several reasons for this choice of research materials. First, Malcolm X is generally considered a powerful speaker, and so the hope was that a "good" speaker would in fact use intonational devices when talking, if they existed at all in the language. Also, this particular speech clearly had a wide range of stylistic variation even upon first listening. Minute 1 and minute 20 do not sound stylistically the same. This implied that Malcolm X was indeed deliberately using intonation to get his message across and, moreover, there would be at least 2 distinct intonational patterns to compare. Finally, there is the simple fact that Malcolm X's speeches are readily available online for public use. This speech (X: 1964) was downloaded from American Rhetoric's Online Speech Bank.

This speech is available from the Speech Bank in two files, each approximately 20 minutes in length. For the current research, various pieces from the first 20 minutes are employed. This 20-minute file was then progressively divided into roughly 10-second units, which could then be transcribed using the ToBI transcription system and Praat acoustic analysis software. The files had to be divided manually in order to not divide intonation phrases. There are some limitations to this data. Since this is a publicly delivered speech, there is considerable background noise including the echo of the microphone and the vocal participation of the audience. Whenever the audience interrupts Malcolm X, usually through applause, that data is discarded. So a pause that contains substantial audience applause is not measured for duration, as it is not under Malcolm X's control. However, some pauses which end discourse segments are followed by lengthy applause.

All transcriptions and measurements were done by the author using Praat. Several items were measured:

- Pause – the duration in seconds between 2 IPs. Any IP whose boundary was realized in a final lengthening instead of in actual silence was marked with a Pause of 0 seconds. A way of measuring final lengthening should be included in future research. Audience-created pauses are discarded.
- Pitch Excursion – Pitch Excursion was calculated as simply the highest pitch value in the IP minus the lowest pitch value in the IP. Any pitch calculations that appeared to be software pitch tracking errors were not used for this calculation. Also, at this point, the pitch effects of microprosody are used. In particular, Malcolm X's voiced consonants are often very clear and can swing the pitch lower by large amounts – a “microprosodic” effect that swings the frequency 50 Hertz at times. It might be better in future proceedings to not use microprosodic pitch for this calculation, as it is often at odds with the general pitch of the vowels in the IP.
- Pitch Level – High Pitch +Low Pitch divided by 2. This should be enhanced in the future to indicate what portion of time is spent at a certain range or level.
- High and Low Pitch – Frequency of the highest and lowest points in the pitch track.
- Duration of IP – the number of seconds from beginning to end of an IP, excluding adjacent pauses.
- Number of Words – a count of the words in an IP. This should have a companion syllable count in future research.
- Seconds per Word – Duration of IP divided by the Number of Words. Future work should add a seconds per syllable speech rate calculation.
- Number of Pitch Accents – A count of the pitch accents in an IP. Phrase and Boundary tones are not included.

Apart from measuring such items, each IP's phonological tone sequence was listed and grouped. The grouping was based upon the cues presented in the Results section below. Since this was an exploratory project, often IPs were grouped based on native speaker's intuition. Then, the author looked to see if there was any basis for such groupings. A possible drawback to this method is that there may be times where a grouping cue, which is in fact present in one group, is present at some other time, but does not group. For example, let's say IP A and IP B sound grouped to a native listener. Upon examination it is noticed that these two IPs repeat the same tonal sequence. The hypothesis is then that the repetition of a tonal sequence is a grouping mechanism. This is later backed up by seeing the same pattern in, say, IPs C and D, and IPs E and F. However, it is possible there are two IPs G and H, which for other reasons do not sound

grouped, but in fact also repeat tonal patterns. Such a condition would likely be missed in the current methodology. The hope is that this project identifies likely grouping cues, which can later be evaluated systematically.

## **Results**

### **Local Grouping Tools**

Intonational phrases can be grouped locally using a set of tools:

1. Repetition of a tonal phrase pattern
2. Nuclear accents that refer across IP boundaries
3. “Dephrasing” of an entire IP
4. Suspending and completing boundary tones
5. Phonetic realization of pauses.

These tools segment a discourse and let the listener know how to interpret what is being said. They may be used separately or in conjunction. Naturally, the more cues used at once, the greater the juncture of IPs. However, using the tools separately allows the speaker to construct a rather elaborate hierarchy of IPs, where pairs of single IPs are grouped together, then those pairs are grouped once again at another level and so on. These tools can create at least 3 levels of grouping above the IP before larger global grouping cues, discussed below, are required.

Let’s first examine a brief section approximately 15 minutes into the speech. Malcolm X is discussing the civil disobedience strategy of sit-ins during this time. He argues against using sit-ins as a tactic, and instead standing and fighting. The discussion contains 13 IPs:

- (1) It’s not so good to refer to what you’re going to do as a sit-in.
- (2) That right there castrates you
- (3) Right there it brings you down
- (4) What- what goes with it
- (5) What- think of the image of someone sitting
- (6) An old woman can sit
- (7) An old man can sit
- (8) A chump can sit
- (9) A coward can sit
- (10) Anything can sit

- (11) Well you and I been sitting long enough
- (12) And its time for us to do some standing
- (13) And some fighting to back that up

IPs (6) - (10) form one group. This grouping is established with 2 cues. First, the tonal sequence is repeated each time. Each IP uses a simple H\* on the noun followed with an L-L% boundary. The only variation is that old is also marked with an H\* in the first phrase. So the tonal sequence is:

- (6a) H\* H\* L-L%
- (7a) H\* H\* L-L%
- (8a) H\* L-L%
- (9a) H\* L-L%
- (10a) H\* L-L%

The second cue is in the phonetic realization of the H\* pitch accents. The frequency for each H\*, in order, is 295Hz, 301Hz, 345Hz, 328Hz, and 373Hz. Notice that the highest accent comes at the end. Similar to a nuclear accent internal to an IP where nothing can have a greater realization after the nuclear accent, this nuclear accent indicates that no further phrase is coming that should be compared with these five. It would be quite unexpected after this nuclear phrase to have a 6<sup>th</sup> IP saying a dog can sit unless dog takes on a higher accent than anything.

A similar grouping occurs with IPs 2 and 3, though it is not as obvious. The tone sequence for IPs 2 and 3 are:

- (2a) H\* L- H\* L- L%
- (3a) H\* !H L-L%

These tone sequences are closer than they appear. In IP (2) the second H\* is also a step lower than the first H\*, just as the !H is in (3). However, the disjunction in (2) is slightly longer than in (3). In other words, if there was no stronger pause inside (2), it would have the same tonal sequence as (3). There is no additional grouping cue here, so the grouping is not as strong as in (6) – (10).

IPs (11) – (13) are also grouped in a parallel fashion, though not in an identical way. Their tone sequences are:

- (11a) H\* H\* L-L%
- (12a) H\* H\* L-L%
- (13a) H\* L-L%

IPs (11) and (12) again repeat a tonal sequence. (13) has a slight variation. Malcolm X does not use the phonetic realization of the H\*'s pitch height to group the IPs, but he bangs the podium on sitting, standing, and fighting making it clear these three actions are to be understood with reference to one another. Based on this, we could argue for either a single group of (11) – (13), or one group of (11) and (12), which is then grouped as a unit with IP (13). This latter interpretation is strengthened by a longer silence between (12) and (13) than between (11) and (12).

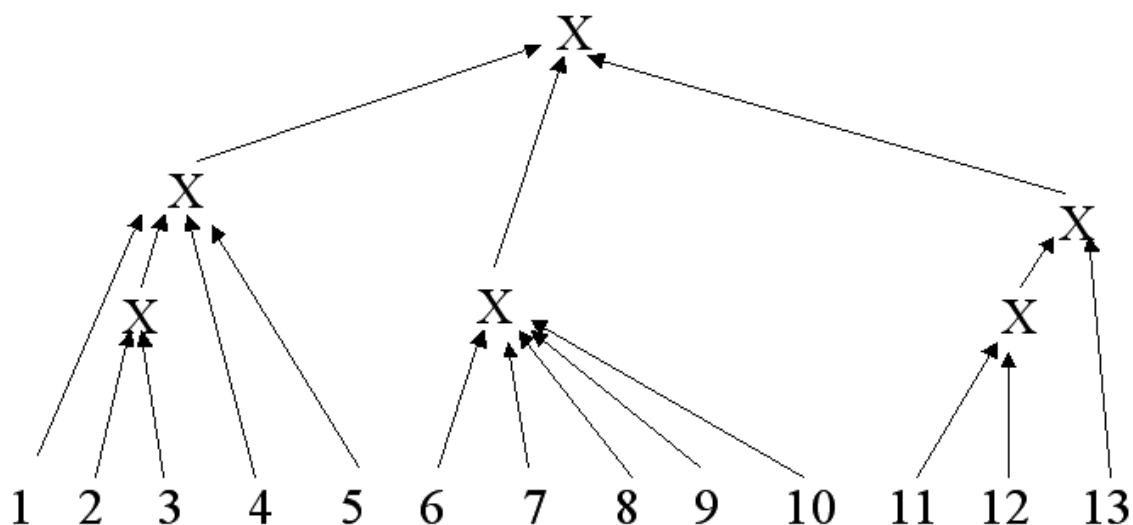
Using the tools of tonal repetition then, we have grouped several adjacent IPs. Using the other tools, however, we can do further work. For instance, IPs (1) – (5) form a group. This is signaled by two cues. First the pitch realization of the L% boundary at the end of IP (5) is significantly lower than the previous IPs. The L% realization of (1) – (5):

- (1b) 135
- (2b) 182
- (3b) 165
- (4b) 157
- (5b) 98

This boundary realization, in addition to the fact that (6) – (10) are strongly grouped as discussed, indicate (1) – (5) as a group.

Finally, (1) – (13) are marked off from adjacent IPs by the coherence of the preceding and following IPs with other groups. We end up with the following structure:

**Figure 1.**



The X simply indicates the present agnosticism about what these higher layers are and how they relate to one another. Despite this neutrality, it is worth noting that the tools of tonal repetition and nuclear pitch accents appear to be used more frequently in smaller groupings, while groupings done through pause and boundary realization occur in larger groupings. This is discussed further below.

Another example will display the use of other grouping tools, as well as show the complexity of groupings. Here we will look at just 5 IPs:

- (14) But when we come out here
- (15) we have a fight
- (16) that's common to all of us
- (17) against an enemy
- (18) who is common to all of us

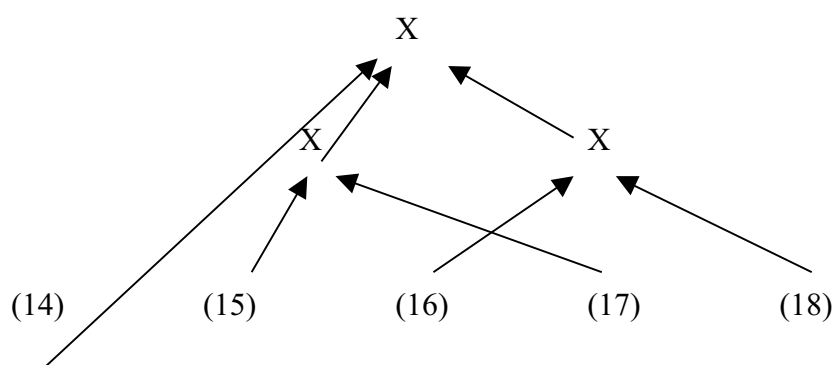
Interestingly, IPs (15) and (17) group together, while IPs (16) and (18) do. Then all 5 IPs group to make a larger unit. The tonal pattern for all 5 IPs is:

- (14a) H\* L-H%
- (15a) H\* L-H%
- (16a) H\* L- H\* L-H%
- (17a) H\* H\* L- H%
- (18a) H\* L- H L-L%

One can tell that (15) and (17) group due to a nuclear pitch accent in (17) that seems to refer back to (15), not (16). The H\* on enemy in (17) is realized at 224Hz, while that on (fight) is realized at 204Hz. Also, the tonal pattern is almost the same for an additional connection. Next, the tonal sequence repeats in (16) and (18), apart from the boundary tone discussed below. The result is two interleaved pairs in (15) – (18). It is not clear what to do with (14). The tonal pattern repeats between (14) and (15), but it's H\* is not clearly related to the (15), (17) group. The result is that it sounds linked, but just not as strongly.

Three tools are then used to group the 5 IPs together. First, Malcolm X repeats the L-H% at the end of the first 4 phrases, changing to a final L-L% at the final IP. There is also a substantial pause after the final IP, though we cannot measure its exact intentional duration due to the fact that the audience joins in for an extended time. Finally, after the nuclear accent in IP (17), which is the penultimate IP, the final IP is dephrased. The last IP is dephrased by reducing the pitch globally across this IP compared to the preceding IPs. The pitch levels for (14) – (17) are 146Hz, 167Hz, 154Hz, and 163Hz. The pitch level for (18) is 133Hz. The result is the following grouping:

Figure 2.



Again tonal repetition and nuclear accents seem to create smaller groups, while pause, dephrasing, and boundary tones create larger groups. This is especially true for the boundary tones. What we just looked at is actually the last 5 IPs of a large IP group containing 16 IPs. There are earlier cues closing groups within this, but the boundary tone never switches to the final L-L% fall until the last IP. As will be discussed in the Global

section, this fall actually closes out the first major topic of the speech, and Malcolm X will now switch to a new topic that will last over 5 minutes.

Despite these tendencies for tonal repetition to be at a small level, and pauses, dephrasing, and boundary tone at a higher level, the relationship between the cues is not a simple one. For instance, one of the reasons that the dephrased IP sounds so dephrased or final is because it follows on a nuclear accent. Also, one can find L-L% boundary tones inside an intonation group. In such circumstances, typically the final L-L% has a greater phonetic realization than the internal ones. See Hermann 2001 for further research. All of this is why these cues are being described as tools. It appears that the speaker has this set of tools to effect grouping, but there is no requirement to always use them. They can be present or not in any given discourse, but if an intonational cue is wanted, the identified set will be employed. This is no different than intonation inside the IP, which can realize a phonological structure in many ways. For instance, a large disjuncture inside an IP can signal a syntactic break, but a syntactic break does not require a large disjuncture. It is a tool the speaker can employ as needed.

To see evidence of groupings at play in further IPs, see Appendix 1, which contains a piece of the data analysis performed for this research project.

### **Global Intonational Parameters**

In this section we will look at parameters that effectively segment very large pieces of discourse, but are typically not implemented at any single point in time, such as the cross-IP pitch accents or final boundary tones are. This will include parameters, such as pitch range, pitch level, IP duration, speech rate, pause duration, and the type and number of tones used. Note that pause duration is included here, as well as in the local grouping section. The difference is that a relatively longer pause in a small group of IPs can create a sense of finality, while in this section we will be looking at pause duration almost as a stylistic parameter – does the speaker generally allow lots of time between phrases or just a little?

That Malcolm X is speaking differently in the first couple minutes of the speech compared to 15 minutes into the speech is immediately apparent to any casual listener. Some of the ways his speech changes can be seen in Table 1:

**Table 1.**

	Section 1	Section 2
IP Duration (seconds)	1.5	1.89
Word Count per IP	4.68	7.25
Speech Rate (seconds)	.35	.27
Pause Duration (seconds)	1.00	.52
Pitch Level (Hz)	128.47	234.25
Excursion (range in Hz)	67.16	167.5

So in section 2, Malcolm X is speaking in longer phrases with more words per phrase than in section 1. He has lengthened the IPs by speaking a little faster, reducing the pauses by 50%, and simply speaking for a longer time before breaking. Moreover, his pitch range has changed dramatically, rising over 100 Hertz. In fact his average low in pitch for section 2, 150Hz, is not much lower than section 1's high, 162Hz. However, this information by itself is partial at best. What we need to know is how these changes are implemented.

To examine this problem, we will look at the IPs before and after a major topic change that occurs about 5 minutes into the speech. The basic structure of the speech for the first 10 minutes is 1) brief greetings to the audience for less than a minute, 2) a discussion of religious beliefs, their role in his personal life, and their proper role in political discourse, which takes about 4 minutes, followed by 3) his political philosophy of black nationalism. We will look at the transition from religious philosophy, which Malcolm X regards as critical to his personal life, to political philosophy, which he argues people of all faiths can share. Table 2 looks at 43 IPs before this shift and 24 IPs after, comparing phonetic realization across the two topics. Numbers in boldface will be discussed.

**Table 2.**

	IP Dur	Word Count	Speech Rate	Pause Duration	Pitch High	Pitch Low	Pitch Level	Excursion
Ave 1	<b>1.17</b>	<b>3.91</b>	.36	.40	188	116	<b>152</b>	72

Ave 2	<b>1.97</b>	<b>6.88</b>	.31	.26	205	126	<b>165</b>	79
Max 1	<b>2.63</b>	<b>11</b>	1.07	1.22	267	156	196	141
Max 2	<b>4.91</b>	<b>16</b>	.48	.88	257	166	181	132
Min 1	<b>.50</b>	<b>1</b>	.13	0	<b>127</b>	93	126	17
Min 2	<b>.93</b>	<b>1</b>	0.19	0	<b>170</b>	91	121	12

Table 3 contains a comparison of tone sequences in each section of the discourse.

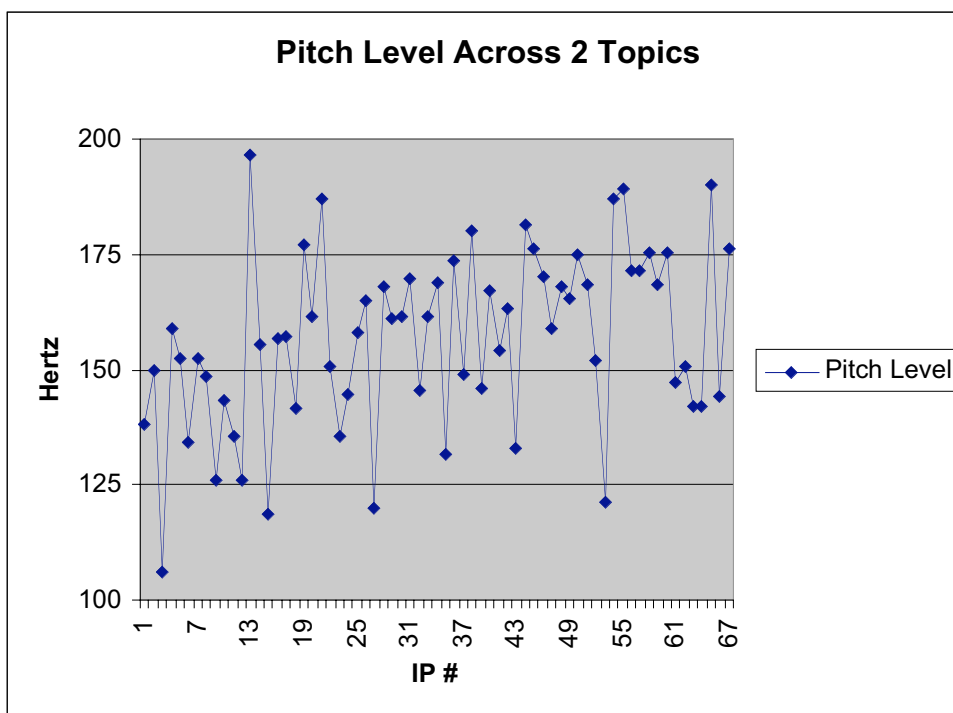
**Table 3.**

	Pitch Accent Count	# IPs with simple pitch accent	# IPs with complex pitch accent	% of IPs with simple pitch accent	% of IPs with complex pitch accent
Topic 1	1.58	43	4	97.73	9.09
Topic 2	2.92	24	13	100	54.17

Looking at Table 2 first, we can see the most substantial difference between the two sections is the length and duration of IPs. The average duration of an IP grows by 67%, and the average number of words grows by 76%. From Table 3, we can see that the number of pitch accents per IP also grows by 85%. The type of tone accents also changes. In section 1, less than 10% of the IPs have complex tones – anything other than H\* or L\*, while in section 2, a full 54% of the phrases have complex tones. So in section 2, Malcolm X is speaking longer in each IP and using different markers for prominence than in section 1.

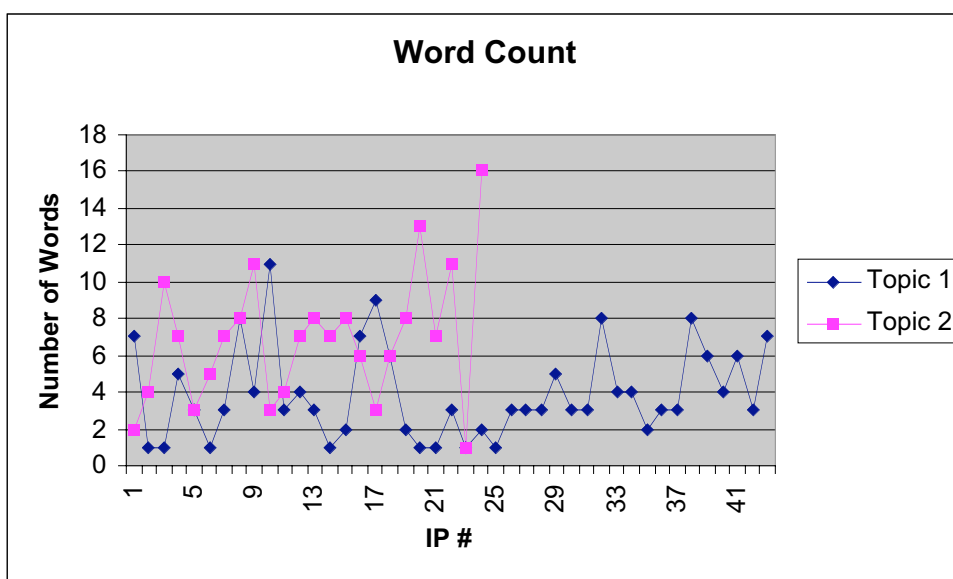
Surprisingly, pitch level does increase between the sections, but it increases by only 9%. Also, the highest high pitch for Section 1 is higher than section 2, and the lowest low pitch for section 1 is lower than section 2, yet the excursion size is about the same on average. What this all means is that these global pitch changes are not implemented on every IP in a topic, and there will be great fluctuations from one IP to the next. Figure 2 shows that, while there is a general trend upwards for pitch between the two sections, it is realized only gradually. (IP 44 is the topic change)

**Figure 3.**



Compare this graph to Figure 3, showing Word Count. While there is overlap between the two sections, section 2 is realized more consistently above section 1.

**Figure 4.**



So what we have is that when you look at large numbers of IPs on either side of a major topic boundary, the pitch changes, but only to a small degree and inconsistently, while IP duration and tonal sequences change significantly and on a much more consistent basis. However, to confuse the issue further, when you look at the exact topic boundary and just the 4 IPs on each side, the picture changes.

**Table 4.**

	IP Duration	Word Count	High Pitch	Low Pitch	# of Pitch Accents	IPs with complex accents	Pitch Level
IP 1.1	.8	4	204	130	1	0	167
IP 1.2	1.7	6	199	109	2	0	154
IP 1.3	1.21	3	224	102	2	0	163
IP 1.4	1.6	7	166	100	2	0	133
IP 2.1	.93	2	220	143	1	0	182
IP 2.2	1.93	4	205	147	3	0	176
IP 2.3	3.3	10	194	146	4	2	170
IP 2.4	1.37	3	169	148	3	0	159
Ave 1	1.33	5	198	110	1.75	0	154

Ave 2	1.88	4.75	197	146	2.75	25%	172
-------	------	------	-----	-----	------	-----	-----

Table 4 shows that at the very local level surrounding the topic shift, the IP duration is implemented immediately (1.22 versus 1.88), as is the # of pitch accents. However, there is no major implementation of complex tones at the boundary. Zero of four IPs have a complex tone before the boundary, and only one of four does so after it. However, while a pitch change was haphazard and small looking across 20 IPs, it is quite significant when only looking across four phrases. Particularly, Malcolm X raises the low pitch from 110 to 146 and keeps it consistently higher, before it later falls again in other parts of section 2.

## **Discussion and Conclusion**

This project has made good headway in identifying local grouping mechanisms. They can be identified, and they frequently occur with non-intonational similarities, such as repeated words or clausal structure, providing external evidence of grouping. The next major stage to enhance this work would simply be a more systematic look at the grouping tools. As for the global parameters, it still seems that there are major stylistic differences in intonation between major sections of the speech, but their implementation is largely a mystery. The number of words and number of pitch accents was implemented rather uniformly across a topic boundary, but most of the other parameters were not. One possibility is that such stylistic effects are largely emergent from local choices. For instance, let's say a nuclear pitch accent raises the pitch above previous IPs. If the next IP does not always reset as low or lower than the previous one, you would have a slow increase in pitch. Similarly, if repeated tonal sequences are used locally to group nearby IPs, then if you keep grouping IPs of similar structure again and again, you will end up with a stylistic shift.

## **Further Research**

There are several future directions for this work. First, this research should be integrated more fully with other research on intonation in discourse, particularly the work of Hirschberg, Grosz, Hermann, etc. With a greater model of how discourse in general

operates, the use of the local grouping tools should become clearer. Additionally, more systematic evidence is required. For instance, how often is the cross-IP nuclear accent used? Is it ever used where it is not intended to end a group of IPs? With evidence of the systematic use of tools, then we might have more to say on the Strict Layer Hypothesis as discussed in Ladd 1996. A further piece of earlier research that could be relevant for this work is the generative theory of tonal music of Lerdahl and Jackendoff. Their music theory uses various principles for grouping rhythm and melody in music and building hierarchical trees. As music also works in pitch and duration, there might be a useful connection there. Exactly how much this would be moving from discourse structure to the aesthetics of rhetorical performance is not known. Finally, part of the systematic analysis would ideally include generalizing away from one speech, one speaker, one rhetorical style, one dialect, etc.

## **Appendix 1**

This is an informal table showing analyzed IPs with groupings color-coded. A red grouping is the smallest grouping, gold next smallest, blue the next smallest, and gray the largest. If the colors are in a single column for a set of IPs, then they are grouped. If in different columns, they are not. Red, gold, and blue groupings are local in the sense described in the paper. The gray grouping is a global topic segment as discussed in the Global section. Note that some of the sections require all 4 levels and some do not. Once again, no assertion is made whether or not these are distinct levels that happen to start and end at the same time, like an IP with a single iP, or if one of the grouping levels simply does not exist for that section of the discourse. This appendix is merely to show that grouping occurs in more than the handful of phrases discussed here, though space does not allow for the justification of any of these groups at this time.







## **References**

Grosz, Barbara. 2002. "Discourse structure, intentions, and intonation." In *The Languages of the Brain*. Edited by Albert M. Galaburda, Stephen M. Kosslyn, and Yves Christen. Cambridge, MA: Harvard University Press.

Grosz, Barbara and Hirschberg, Julia. 1992. "Some intonational characteristics of discourse structure." In *Proceedings of the 2<sup>nd</sup> International Conference on Spoken Language Processing*. Banff, Canada, pp. 429-432.

Hermann, Rebecca. 2000. "Phonetic markers of global discourse structures in English" *Journal of Phonetics*, 28, 466-493.

Ladd, Robert D. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lerdahl, Fred and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press.

Nakatani, Christine, Hirschberg, Julia, & Grosz, Barbara. 1995. "Discourse structure in spoken language: studies on speech corpora." In *Working Notes of the AAAI-95 Spring Symposium on Empirical Methods in Discourse Interpretation*, pp. 106-112, Menlo Park, CA. American Association for Artificial Intelligence.

Wennerstrom, Ann. 2001. *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press.

X, Malcolm. 1964 "The Ballot or the Bullet". Delivered in Detroit, Michigan.  
Downloaded from <http://www.americanrhetoric.com/speeches/malcolmxballot.htm>.