INTONATIONAL MEANING: WHY *MOM* CAN BE BOTH EMOTIONAL AND RATIONAL[1]

Erin Good
University of Arizona

This paper seeks to evaluate and quantify the acoustic cues utilized in the production and perception of intonational meanings in English. Although much previous work has started from observed contours and looked for their meaning based on the context where they were found (Ward and Hirschberg 1985, Gussenhoven 1984, Crystal 1969), the opposite approach is taken here. Arbitrary contours generated in a systematic way (detailed below) were presented to study participants, who were then asked to rate these items on a series of semantic scales (scales with antonyms on either end).

The items were word-contour pairs consisting of one of four words in combination with one of twenty-seven contours. Contours were created on a grid which had three time points and three pitch levels. The words were chosen to have a variety of types of semantic/dictionary meanings. Participants' ratings were used to determine the emotive meaning of the contours, and thus the degree to which the contours interrelate. As defined by Osgood (1957), the emotive meaning of an utterance or word does not relate directly to its dictionary meaning, but instead relates to the location of that utterance or word in the semantic space. By looking at what items cluster together in semantic space and comparing the acoustic and lexical/semantic characteristics of these items it is possible to understand which dimensions play a role in the assignment of meaning to intonational contours.

Results show that the judgments on the semantic scales are influenced by both the intonational contours and the lexical items. Multivariate Analysis of Variance tests were performed to determine what acoustic characteristics contributed the most to the formation of clusters of items. It was found that the presence versus absence of High pitch points in the contour as well as the word used to convey the contour were the most influential factors for the participants. The approach taken here is able to uncover new shades of intonational meaning, as well as pinpoint the acoustic cues used to assess these meanings.

[1] I would like to thank all of people who helped to make this paper what is it today; Natasha Warner, Diane Ohala, Andrew Carnie, Scott Jackson, Adam Ussishkin, Lynnika Butler, and all of the students in 697a. All mistakes are my own.

Coyote Papers 14, 56-95
Linguistic Theory at the University of Arizona

1. Introduction

Ladd (1996) discussed a distinction between two types of intonational research. One type, which he terms *impressionistic* (or proto-phonological), seeks to identify pitch morphemes, and/or nuclear tones, and to explain how these smaller components of intonation relate to one another. This was the approach most used by linguists and language teachers who sought to explain how intonation is used by speakers to convey different sorts of meanings. This type of approach tends to lack a direct correlation between the labeling of the morphemes and phonetic/acoustic cues, but provides dynamic insights into how contours relate to one another and to the context in which they are used. An example of this approach can be seen in Bolinger (1972), who represented intonational contours by writing each letter of a word on a different line: letters went up in a rising contour and went down in falling contour (Figure 1).

Figure 1: Impressionistic representation of an intonational contour taken from Bolinger (1972).

Don't
be y
r
g
an

While this form of representation is highly intuitive it does not provide many details about what exact type of pitch change is present, nor are claims within this framework easily testable in an experimental setting. The other branch of intonation research was termed *instrumental* (or phonetic) by Ladd and it focuses on the acoustic cues to meaning, viewing intonation as a largely paralinguistic phenomenon. This method was mainly used by "experimental psychologists and phoneticians interested in speech perception and in identifying the acoustic cues to intonational phenomena" (Ladd 1996, p 12). This view claims that intonation is gradient and, for example, that certain scalar acoustic cues have direct syntactic correspondents. One example of this is the idea that the duration of a pause indicates the strength of a syntactic boundary (Cooper and Paccia-Cooper 1980).

The work presented here utilizes a method that is highly instrumental in that it manipulates fine phonetic details and examines the effects of these changes on the interpretation of the contour. As well, it is not just the acoustic characteristics which are being manipulated: the semantic content of the utterance is also controlled. Instrumental approaches like those of Cooper and Paccia-Cooper found direct links between syntactic boundaries and phonetic realizations of those

boundaries. The current work follows from the instrumental tradition and continues to view intonation as something that acts on more than just a phonological level.

At the same time, the method used here has been influenced by the impressionistic tradition. The intent here is to determine which of these acoustic cues are used to locate discrete intonational contours in the semantic space. The idea that the contours may have sub-parts which are combined to create intonational meanings comes directly from the impressionistic tradition and work by Pierrehumbert (1980). Ratings of the items by the participants on semantic scales can be seen as distinctly impressionistic, since these scales do not relate to semantic meaning but instead are measurements of emotive meaning.

Emotive meaning is a measure of the emotional content and/or impact of concepts. Osgood (1957) originally proposed the use of semantic scales to assess the emotive meaning of words within the field of psychology. Semantic scales have been adopted for use in anthropology as well as linguistics as a method for understanding how different concepts relate to each other. Responses to these impressionistic scales locate concepts from the lexicon in semantic space. In the context of the current study, items that have similar scores on the scales, and therefore similar emotive meanings, are predicted to be similar acoustically as well. The question asked in this paper is, "What counts as acoustically similar?".

There are a number of methods given in the literature for dividing up the acoustic space, and a variety of proposals have been made that seek to explain what aspects of phonetic variation in intonation are salient to listeners. Aspects of many of these proposals were used in the analysis of the data presented here. For example, the importance of the presence of a rise – any rise at all at any point in the utterance – has been highlighted by researchers of English intonation (i.e. Gussenhoven 1984, Crystal 1969, Pierrehumbert 1980). In the results section (6.5) it will be shown that having a rise anywhere in the utterance has a greater impact on how the utterance is interpreted than having a rise as the only pitch change.

Instrumental methodology is joined here with impressionistic questions in hopes of reaching a deeper level of understanding of how a number of factors influence the interpretation and assignment of meaning to intonational contours. By joining together these two approaches to the study of intonation, the current work provides a new perspective on what intonation research is about, and establishes a more unified approach to the field.

2. Background

2.1 Impressionistic Versus Instrumental Approaches

One of the first impressionistic approaches that also provided definite predictions about the phonological status of intonational contours was the IPO system (Institute of Perception Research) developed in Eindhoven, Holland to examine intonation in Dutch. This system gave a general theory of intonational structure that construed contours as "idealized sequences of pitch movements" (Ladd 1996, 14). Three acoustic characteristics were focused on in this system: 1) relative pitch height (high or low), 2) prominence-lending pitch movements (pitch changes on stressed syllables), and 3) non-prominence-lending pitch movements (pitch changes on unstressed syllables). These characteristics were combined in a linear fashion and the description of the contour types was directly linked to fundamental frequency (f0). No meaning was directly attributed to the contours thus described; instead, after the contour had been identified in context, a contextually dependent meaning was assigned to it.

An example of a contour described by this system is the colloquially named 'hat pattern', referred to as a Type 1 Rise within IPO. This contour has two variations: the 'pointed hat' is formed via an early rise from low to high followed by an early fall from high to low (Figure 2a), while the 'flat hat' starts with an early rise from low to high which is followed by a delayed fall from high to low (Figure 2b).

Figure 2: 'hat contour'

a. 'pointed hat' b. 'flat hat'

The IPO system is an offshoot of the British tradition which is exemplified by work with two distinct sub-approaches. One branch of the British tradition views contours as holistic non-decomposable units (e.g. Lieberman & Sag 1974), while the other views contours as subject to segmentation (e.g. Trager & Smith 1951).

2.1.1: Impressionistic approach – Intonation as non-decomposable

Those who view intonation as non-decomposable have generated a number of descriptive (or impressionistic) names for the intonational contours of English, including 'compound falling' and 'more emphatic' (Sweet 1878), but have

provided fewer details about what separates one contour from another at the phonetic level (Crystal 1969, Bolinger 1972). These descriptive names were assigned by the researcher based on the perceived general shape of the contours in question, and often had highly specific semantic meanings attributed to them (Pakosz 1984). For example, Halliday (1963) proposes five basic contours for British English, the second of which he describes as 'rising to mid pitch'. Affirmative clauses said with this contour are interpreted as 'contradictory', 'challenging', 'aggressive' and/or 'defensive'. Decisions about when a phrase had which of the above meanings were to be based on the context of the utterance and no account was given of how these meanings came to be associated with this contour.

An example of an objection to the impressionistic approach to non-decomposable contours comes from Pakosz (1982), who argued against such ungeneralizable terms as those given by Halliday. Pakosz proposed that discussion of intonational contours be based on the relative strength of the contours, rather than on the shape of the contour and contexts in which it was generally found. He set out several, still fairly impressionistic, criteria for assessing the "strength" of a contour. These criteria relate to: a) the presence of large changes in pitch, b) changes in the direction of the pitch track (rising to falling, and falling to rising), and c) general height of pitch. According to Pakosz, his criteria allow for intonation to be discussed independently of both semantic meaning and the context of the utterance. The meanings that contours have is assigned to them by the context, with the limitation that "stronger" meanings are assigned to "stronger" contours.

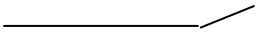
An example of a "strong" contour according to Pakosz is a contour that rises at start of the utterance and has a sharp fall at the end of the utterance. This type of contour is able to convey any sort of intense emotional meaning, ranging from admiration to indignation (Figure 3). A "weak" contour starts low and has a slight rise at the end of the utterance and can be used to express less intense emotions such as indifference (Figure 4). He argues that it is the context of the utterance that determines which emotion is attributed to a given contour; all that the contour determines is the relative strength of the emotion that will be assigned to it.

Figures 3 and 4: Examples of "strong" (3) and "weak" (4) contours according Pakosz (1982, 1984). A sample utterance is provided along with a schematized version of the contour located below the utterance. The emotion conveyed by the contour is given below the contour.

(3) That's really disgusting of him.

indignation or admiration

(4) Answer the door



indifference

Pakosz's approach falls somewhere in between the impressionistic approach and instrumental one, in that it calls for a direct relationship between the types of meanings and phonetic cues (instrumental), but does not go so far as to measure those cues (impressionistic).

2.1.2 Impressionistic approach – Intonation as decomposable

Those who favored the segmentation of intonational contours, rather than the treatment of contours as non-decomposable units, focused on the idea that English intonation consisted of five basic nuclear tones – Fall, Rise, Fall-Rise, Rise-Fall and Level - which could be modified in a variety of specific ways. These five basic tones would be composed with various types of rises or falls at the beginning and/or end of the nuclear tone to create entire intonational contours. Within this segmentation-based approach, Gussenhoven (1984) argued that experiments were necessary to uncover of the exact linguistic information carried by intonational contours. This argument in favor of an experimental approach places Gussenhoven more on the instrumental side of the impressionistic / instrumental divide.

Gussenhoven further argued that the type of meaning assigned to intonational contours should be similar to the type meaning present in words like *this*. The referent of *this* is fixed by the context of the utterance, and the word *this* itself is used referentially to refer to an object which is present and located close to the speaker. Gussenhoven contrasts this type of context-dependent referential meaning with the type of meaning carried by words like *dog*. In the case of *dog*, the meaning of the word is grounded in the real world and can be understood without any context.

The kind of meaning attributed contours in Pierrehumbert's 1980 dissertation is similar to that argued for by Gussenhoven. Intonational meaning in this system provides information about the relationship between the utterance and the beliefs that are shared between the speaker and hearer, as well as the

relationship between the utterance and the anticipated contributions of future utterances to the discourse (Pierrehumbert and Hirschberg 1990). Pierrehumbert's system for describing intonational contours is based on the idea that intonational contours consist of distinct compositional units, each of which contributes its own meaning to the utterance. Her work not only provides a theory within which meaning differences in contours are predicted based on their shape, but it also provides a system whereby differences between contours can be detailed, regardless of their meaning. This system is strongly impressionistic in that it seeks to create phonological categories for intonational contours. Yet Pierrehumbert did use definite phonetic/acoustic characteristics to define the contours in her system, which aligned her with the instrumental approach as well.

The descriptive system developed by Pierrehumbert is widely used in the current literature on intonation. Briefly, in the original version of this system, pitches are either high (H) or low (L). There are three different types of pitches : 1) Pitch accents, which are similar to nuclear tones, are marked with an asterisk (*); 2) Phrase accents, which control the pitch at the start/end of the phrase, are unmarked; and 3) Boundary tones, which appear at the end of the phonological phrase, are marked with a percent sign (%). Examples of this system, along with the traditional British descriptions of the contours and the descriptions used in the current study are given in Table (1). The correspondence between Pierrehumbert's contours, the traditional British contours, and contours used here was established by examining pitch tracks and comparing the pattern of the fundamental frequencies in each contour.

Table 1: Corresponding contour descriptions from three different systems: Pierrehumbert (1980), the British Tradition (as reported by Levis 2002) and the system employed in the current paper.

	Pierrehumbert	Traditional	Current study
a	L*LH%	Narrow low-rise	LLM
b	H*LH%	Fall-Rise	HLM
c	H*LL%	Fall (high-fall)	HML
d	L*HH%	Wide low-rise	LMH
e	H*HH%	High-rise	MHH

In an attempt to decide between various conflicting accounts of the contours given in Table (1), Levis (2002) looked closely at differences among these five contours in English. He used a Paraphrased Meaning Test, a standard method in psychology which asks participants to choose which of four paraphrases best fits what the speaker was trying to convey, to assess whether or not certain pairs of contours were in fact distinct.

In Levis's study it was found that only three of the five contours were distinct: there was one falling contour (1c), one general contour that had a rise to mid pitch (1a and/or 1b), and one general contour that had a rise to high pitch (1d and/or 1e). Thus of the five contours examined by Levis, two pairs (1a& b) and (1d&e) collapsed into a single contour. Levis's approach is strongly instrumental: he was not looking to locate these contours in a larger phonological framework, he was just evaluating whether or not the contours were distinct.

Another largely instrumental study was performed by Ward and Hirschberg (1985). While Levis was focusing on a way of distinguishing between contours, Ward and Hirschberg sought to learn more about the meaning of one contour in particular. They examined the Fall-Rise contour (L*+HLH%) in English and claimed that it was used by speakers to indicate uncertainty. They performed a corpora study which included 100 natural productions of this contour collected from television and radio shows. Phonetic cues, including a pitch peak late in the syllable and an abrupt drop in pitch for the syllables following the pitch peak, were used to distinguish instances of the contour in the study. By looking at natural productions of the contour and comparing the context of the production with the meaning that the researchers perceived to be conveyed, a meaning for the contour was determined. Ward and Hirschberg found that the presence of this contour on an item indicated speaker uncertainty about the relationship of the item to the topic under discussion.

Figure 5: Sample dialog from Ward and Hirschberg (1985): the word *Venus* is said with a Fall-Rise contour.

B: Did you see the morning star?

A: I saw \Ve/nus.

An example they give uses the dialog above (Figure 5), where the contour in question falls on the first syllable of the word *Venus*. In this dialog the speaker is uncertain whether or not *Venus* is the same as the *morning star*.

2.2 Intonation – Semantic interface

Ward and Hirschberg claim that the presence of the Fall-Rise contour on the word *Venus* in the example above indicates uncertainty about the appropriateness of the word in the context of the utterance. They make no claim that this meaning of uncertainty carried by the contour interacts directly with the meaning of the word. Rather, there are two levels of 'meaning' at work, a pragmatic one that is expressed by the contour and a lexical one that is contained in the word. Yet Caspers (2000) claimed that "it is notoriously difficult to test the contribution of

intonation to the meaning of an utterance" (134). She goes on to say that intonational meaning is influenced by both speaker intention and the literal meaning of the utterance.

Contrary to Ward and Hirschberg's view and more in line with Caspers's claim, Cruttenden (1981) claims that lexical meaning can override intonational meaning, and even force the choice of certain contours. One example relates to contrastive noun phrases. Cruttenden argues that contrastive noun phrases are generally said with a rising or 'open' intonation, but when preceded by a 'closed' lexical item like the quantifier *every*, the phrase will be said with a 'closed' or falling intonation. In this case the lexical item *every* forces the use of the falling intonation, showing a direct interaction between lexical meaning and intonation.

In contrast to Cruttenden's claims, yet still distinct from Ward and Hirschberg's, Mehrabian and Wiener (1967) present evidence that listeners, when given instructions to attend to both intonational and lexical meaning, gave more weight to the intonational meaning conveyed, even when it contradicted the lexical meaning. Mehrabian and Wiener crossed words with three semantic attitudes (positive, negative, neutral) with contours that conveyed the same three attitudes as cued by tone of voice.[2] For example, a 'positive' word like *love* would be said with a 'negative' tone of voice and participants would be asked to rate it as either positive or negative on a seven point scale. The results indicate that the participants placed more weight on how the word was said (the 'tone of voice' used) than on the semantic attitude conveyed by the word. So in the example given, *love* would be rated as negative despite its positive connotations.

A point to bear in mind throughout the remainder of this paper is that the work to date on the interface between semantics and intonation indicates that the nature and degree of interaction between intonational and semantic meanings is still highly debatable. Ward and Hirschberg claim that intonational meaning and lexical meaning function on different planes, while Cruttenden claims that lexical meaning can override intonational meaning. Linguists who work in the fields of syntax and semantics frequently makes claims about intonation being used to indicate scope or focus in certain constructions, without taking the time to explain how this accomplished. Intonation no doubt plays a role in the syntax and semantics of utterances, but this role is as of yet very ill defined, and more work investigating the interplay between semantic/lexical meanings and intonational ones needs to be undertaken.

[2] No information was provided in the paper about exactly what is meant by tone of voice. All that was said was that the speakers used were trained actors.

2.3 Summary

The line between what is instrumental and what is impressionistic is somewhat vague, as can be seen when the 'instrumental' approaches to intonation research taken by Levis (2002) and Ward and Hirschberg (1985) are compared with the 'impressionistic' approaches taken by Pierrehumbert (1980) and Gussenhoven (1984). All four researchers are interested in the same sorts of questions about the meaning of intonational contours, but differ in how narrowly they choose to focus their investigations.

The division between instrumental and impressionistic can thus be viewed, in part, as a division based on the size of the question asked; questions about the entire system of intonation are more impressionistic, and questions about particular contours are more instrumental. The current study asks questions about how changes to particular contours influence the meaning of those contours (instrumental), but it then steps back to ask how the types of acoustic differences that affect meaning changes can be viewed in the larger context of the interaction between intonational meaning and lexical/semantic meaning (impressionistic).

3. Research Questions

There are two main research questions addressed in this paper. This paper seeks to further explicate and understand: 1) what acoustic cues are used to differentiate between intonational meanings, and 2) how the meanings associated with those cues interact with the lexical/semantic meanings of the words over which the contours appear.

These questions are addressed experimentally in the study; twenty-seven distinct contours were crossed with four English words. The words were chosen first for their acoustic characteristics, and second for the variety of types/degrees of lexical meaning which they convey. The contours were arbitrarily generated to completely fill a set phonetic space and do not necessarily match known English contours. These word-contour pairs were rated on six different semantic scales in order to determine how similar the items were.

With these ratings it is possible to examine the characteristics used by listeners to distinguish between, and assign meanings to, the various contours. The contours differ from each other in distinct ways, making it possible to identify what types and degrees of change in the intonational contour are necessary to affect a change in the location of that contour in semantic space. Yet there is still a large degree of similarity between the contours. The contours exist in a rather small acoustic space, and each has many near neighbors in that space, making it possible to investigate a number of possible acoustic dimensions simultaneously and look for interactions between them.

For example, consider a contour which has a rise from mid pitch to high pitch at the end of the utterance (MMH) and a second contour which has a rise from low pitch to mid pitch at the end of the utterance (LLM). These contours have a number of similarities and a number of differences. If they are located near each other in the semantic space, then the similarities must be more important than the differences. The general pattern of the contour, flat at the start with a rise at the end, is the same for both. As well, both rise by the same relative amount. If these contours appear close to each other, then differences such as the starting pitch and the average pitch of the contours must be unimportant as compared to the similarities in the shape of the contours and/or degree of rise which they contain.

The first research question will be addressed via comparisons like that give above. The second research question goes one step further and asks if there is a difference in the location of a contour based not only on the acoustic composition of the contour but also on the word used to convey it. For example, if the word-contour pairs *no*MMH and *mom*MMH are located near each other in semantic space, then the fact that they have the same contour is more important than the fact that the contour was said over different words.

4.0 Methods[3]

4.1 Participants

Thirty-one native English speakers participated in this study. All participants were undergraduate linguistic students at the University of Arizona. None of the participants had any prior knowledge of the topic being investigated. Participants' gender[4], race and age, were not controlled.

4.2 Materials

4.2.1 Words

Four mono-morphemic English words were selected for use in the study. The words were chosen based on their acoustic characteristics as well as their semantic content. All of the words were CV(C) in shape, and contained only sonorant consonants and vowels. This was done to insure the presence of a definite pitch track for the entire course of the utterance, and to avoid micro-prosodic influences.

Within the constraints set by the acoustic parameters mentioned above, the words were chosen to have varying degrees and types of semantic content. *Yeah*

[3] The idea of creating arbitrary contours based on a three-by-three grid used in this study comes from unpublished work done in John Ohala's lab at the University of California Berkeley.
[4] Of the 31 participants, 8 were male and 23 were female.

[jæ] and *no* [no^{w}] were selected based on their having opposing dictionary-based meanings. *No* has a definite meaning of a negative response, but lacks a physical referent in the world. The meaning of *yeah*, in contrast, is less fixed than that of *no*. *Yeah* can be an affirmative response, but it can also be used as a reply to being called, with a meaning closer to 'I hear you/ I am here' than 'yes'. It can also be used to back-channel[5] and show support for what the other speaker is currently saying (Yngve 1970).

The proper name *Lynn* was selected because it has no dictionary meaning of its own. However, *Lynn* can have a physical referent in the world. Finally, the word *mom* was selected to have a particular lexical meaning that should not vary by context. After the fact it was realized that *Mom* is a bit of a special case, in that it can act like a proper name while at the same time it has definite dictionary meaning.[6]

Table 2: Type and nature of the lexical content of the words used in the study. Words with dictionary-style meaning have fixed content that is context independent. Words with context dependent meaning are assigned meaning based on the context of the utterance. Non-referential words do not refer to entities in the world, while referential words do.

Word	Type of meaning	Nature of meaning
No	Dictionary	Non-referential
Yeah	Dictionary or Context dependent	Non-referential
Lynn	Context dependent	Referential
Mom	Dictionary	Referential

As mentioned above the words were chosen to represent a variety of different semantic types, but were also selected based on the fact that they overlap with each other in a number of ways (Table 2). *Yeah* and *no* have opposing dictionary meanings on one use of *yeah* (when *yeah* means *yes*) and opposing degrees of content on another use of *yeah* (when *yeah* means 'I hear you'). *Lynn* and *mom* can both be used referentially and are thus similar, but *mom* has a definite dictionary meaning that *Lynn* lacks. *No* and *mom* both have definite dictionary meanings, yet at the same time *mom* can be much more vague in its meaning than

[5] An example of back-channeling is when the speaker is telling a story, and the listener says "yeah" whenever the speaker pauses. This done by the hearer to indicate that they are still listening and are interested in the speaker's story.

[6] Nonsense words were not used in this study, though they would have presented an ideal example of a 'lexical' item without any semantic content. This omission will be redressed in future research.

no depending on how *mom* is being used. Finally, *yeah* can lack dictionary style meaning as can *Lynn*, yet *Lynn* is referential and *yeah* is not.

Several different pronunciations of the words were digitally recorded with a 44.1 kHz sampling rate in a sound-attenuated booth by a female native speaker of English. The version with the flattest contour, produced at the middle of the speaker's normal speaking range, was selected for use in each case. Each word was approximately 500 milliseconds long. PSOLA (Pitch-Synchronous Overlap and Add) speech synthesis was used to further flatten the contours of each word. A declination of about .5 semitones over the course of the word, the same as was present in the original utterance, was maintained to more closely mimic natural speech. These flattened versions were then used to generate the contours used in the study.

4.2.2 Contours

The contours were created based on a three-by-three grid (Figure 6). Each contour had three time points at which the different target pitch levels for that contour were reached (represented in the columns). These time points were located at the start of the word, at the mid-point, and at the end of voicing.

Figure 6 gives examples of three of the contours used in the study. Contour (a) starts with high pitch (H), has a low pitch in the middle (L) and ends with a mid pitch (M). The name of this contour, for the purposes of this current study, is thus HLM. Similarly, contour (b) will be referred to as MHL and contour (c) as LMM. All of the twenty-seven possible contours that can be made on this grid were used in the study (Figure 7).

Figure 6: This three-by-three grid was used to create the contours in the study. Each row represents a different pitch level and each pitch level is separated by two semitones. The columns were aligned with the start, middle and end of the utterance. Three example contours are shown; contours were created by selecting a pitch level for each time point and synthesizing the pitch track to connect each of these time points.

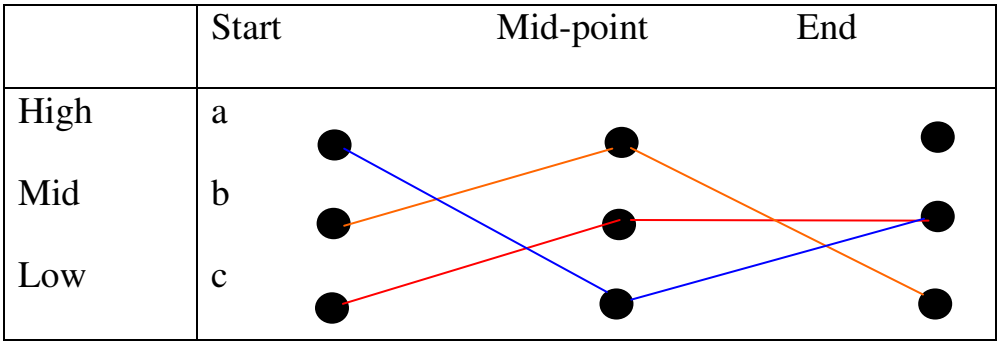
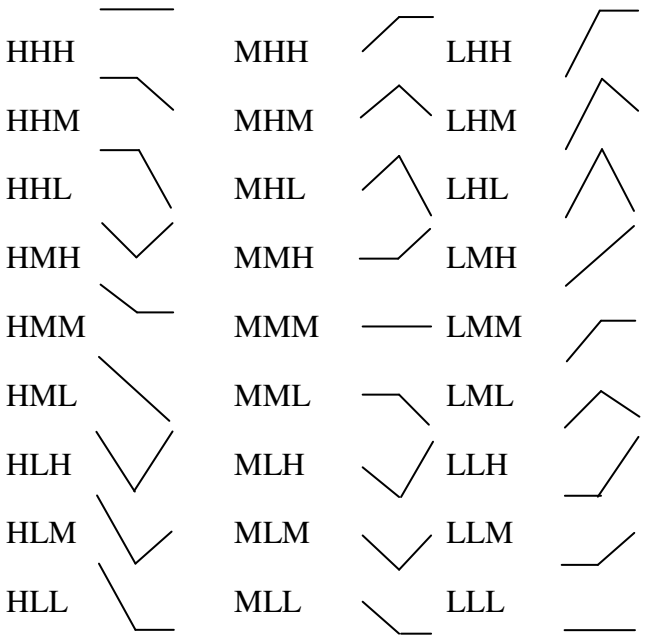


Figure 7: The 27 intonational contours used in the study. The lines represent the pitch tracks for each of the contours. The first column has the nine contours that start high, the second column the nine contours that start with a mid pitch, and the third column the nine contours that start with a low pitch.



The contours were generated using PSOLA techniques. The flattened contour generated from the original sound file was used as the midpoint in the scale and as the MMM contour in the study. Starting from the MMM contour, the pitch track of each word was manipulated so that it matched the target contour. Pitch was measured in semitones rather than in Hz, as semitones more accurately represent the way that pitch is perceived by the human ear (Ladefoged 1996). The starting semitone values for each word are given in Table 2.[7] The exact semitone values vary for the different words because the pitches of the original words varied slightly.

[7] A complete list of the pitch levels at each point for each word is given in the appendix.

Table 2: The exact starting pitch (in semitones) for each word at each pitch level is provided here.

Word	High starting pitch	Mid starting pitch	Low starting pitch
No	13.18	11.00	9.06
Yeah	13.72	11.53	9.35
Lynn	13.57	11.39	9.45
Mom	12.90	10.72	8.54

The changes in pitch were kept as close to the same as possible across words so that the contours created on the different words would be comparable. There was an average distance of 2.18 semitones between the high and mid points on the grid, and an average distance of 2.06 semitones between the mid and low points of the grid. These differences in pitch came from a compromise between the need to have as much difference between the points on the grid as possible, and a desire to have the stimuli sound as natural as possible.[8]

4.3 The scales

The study utilized the semantic differential technique originated by Osgood in 1957. This technique was first used in psychology to ascertain how words cluster together in mental/emotional space. Within this methodology, semantic scales consist of a pair of antonyms with the numbers one through seven written between them. In this study participants were told that, for example, a response of 'one' meant that the item they heard was highly weak, and a response of 'seven' that the item was highly strong, an answer of 'four' indicated that the item was neither weak nor strong. Research with this method has found that in the majority of cases, cross-linguistically, only three basic types of scales are necessary to locate every word in this space. The three scale types that Osgood found to be consistently important were: Evaluative (positive/negative), Potency (strong/weak), and Activity (active/passive). Forms of these scales appear in this study in addition to several other scales which were selected for used based on pilot work.

Figure (8) gives the scales used in the study in the order they were presented to the participants. Osgood's (1957) Potency scale (8f) was used directly in the

[8] If these values were converted into Hz they would be comparable to those used by van Bezooijen (1996) in her research on the correlation between the pitch of a person's voice and their perceived social status / gender. Van Bezooijen used 150 Hz for her low pitch level and 212 Hz for her high pitch level. The low pitch level used here ranged from 162–170 Hz and the high pitch level ranged from 211–219 Hz.

current study (strong/weak). The Evaluative scale, however, was split into two scales in this study: agree/disagree (8b) and optimistic/pessimistic (8c). In an earlier iteration of this project, only one version of the Evaluative scale was used (positive/negative), and participants reported having trouble using this scale. Thus Osgood's scale was split into two scales representing the two meanings that the participants in the previous study reported thinking of when they were presented with the positive/negative scale. The Activity scale (8a) surfaces in the current study as the calm/agitated scale.

Figure 8: Semantic scales used in the current study in the order of presentation. Each scale appeared on the screen singly.

a	Calm	1	2	3	4	5	6	7	Agitated
b	Agree	1	2	3	4	5	6	7	Disagree
c	Optimistic	1	2	3	4	5	6	7	Pessimistic
d	Questioning	1	2	3	4	5	6	7	Stating
e	Emotional	1	2	3	4	5	6	7	Rational
f	Weak	1	2	3	4	5	6	7	Strong

Two additional scales were also used here that do not relate directly the main semantic differential scales. The stating/questioning (8d) and rational/emotional (8e) scales were drawn from the results of the pilot study and reflect common dimensions over which intonation is claimed to vary.

Importantly, the mental space defined by these scales does not directly correspond with the lexical meanings of the words nor with the intonational meanings. Instead, the space represents a level of meaning where the lexical content of the word and the intonational content can interact. The scales measure emotive meaning. Emotive meaning reflects how the participant relates to the sign in question (in this case a word-contour pair), and how they perceive the relationship between that sign and other linguistic signs (Osgood 1957). Lexical meaning, in contrast, deals with how signs relate to the real world. Exactly what intonational meaning relates to is open to debate: Pierrehumbert and Hirschberg (1990) claimed that intonation carries information about how an utterance relates to the context and perceived shared knowledge of the speaker and hearer, Cooper and Paccia-Cooper (1980) claimed that intonational boundaries reflect syntactic structure, and Pakosz (1982) claimed that intonation was directly linked the emotional content of the utterance. This final claim is intuitively the most approachable, though the implementation of it by Pakosz is not universally accepted. Osgood focused on the emotive content of words and was able to show that meanings of this type are highly consistent across subjects for the same

population. Thus Osgood has shown that individual words have consistent emotive content, and as few would argue that intonational contours carry emotive content (to one degree or another), these scales provide a direct way to assess the meaning of word-contour pairs at a level of meaning where the content of both is active.

4.4 Procedure

E-Prime software was used to present the experiment and to collect responses. Participants were seated in a sound-attenuated booth with a computer monitor located outside of the booth and visible through a window. A keyboard with a mouse was located in front of them. The items were presented binaurally over headphones. Items were presented in a different random order to each subject.

First an instructional screen was presented, though instructions were given orally as well. Then a screen was shown with the numbers one through seven as they appeared on the scales, but without any words. While the fixation was displayed, the target item was played repeatedly with a 500ms delay between repetitions, for either 5 seconds (about 5 repetitions) or until the space bar was pressed. This allowed the subjects to move on when they were comfortable with the word, but established a maximum number of times that the stimulus would be heard by each subject. After the repetitions, a new screen was displayed with the first scale, and the item was repeated one more time. Once a response was recorded, the next scale was displayed and the item was repeated again. Each item was rated on all six scales in succession, before a new item was heard. After each of the scales had been responded to, there was a one-second pause before the beginning of the next trial with a new item.

Strict definitions of the scales were not given; the participants were encouraged to work out their own system for the precise meaning of the scales. Participants were told to interpret the meaning of the items as they would if they heard them in a conversation. Finally, participants were told to trust their first response and were encouraged not to think overly much about each item. Participants were allowed a break midway through the test. The majority of participants took approximately 45 minutes to complete the entire task, with a range of 25-60 minutes.

5. Data Analysis

Responses to all 108 items (four words crossed with twenty-seven contours) were collected for all six scales from 31 participants. This much data provided for a large degree of power in the experiment but also made the analysis complicated. In this section the data analysis procedures and the rationale behind them are detailed, while the results are reported in section 6.

The general goal of this paper is to determine what items (word-contour pairs) were given similar ratings, both on the individual scales and on the scales as a collective unit of measurement. There are several possible statistical approaches to this kind of data and many were attempted in the course of the data analysis process. Two methods were settled on as the most interpretable: 1) Multivariate Analysis of Variance, which allows for the effects of multiple dependent measures to be examined simultaneously, and 2) single mean t-tests, which provide a way to test for significant results on an item-by-item basis.

5.1 T-tests

Single mean t–tests were used to determine the rating of each item on each scale. A single mean t-test compares the mean response to a given item against the null hypothesis, in this case a score of four which is the midpoint on the scales. Recall that participants were told that a rating of four on the Weak/Strong scale for example, meant that the item was neither strong nor weak. A significant result on a single item t-test shows that the mean response to that item was consistently different from the null hypothesis, and shows the direction of that difference.

In Table (3) items containing the HLL contour are shown. All of the scales for which the mean response had a significance value greater than $p < .05$ are indicated with the direction of the response. For example, on the Questioning/Stating scale all of the items with a HLL contour were judged to be stating.

Table 3: T-test results for all items containing the HLL contour. The end of the scale on which the mean response to each item fell is provided. No response is given for items that did not reach significance ($p < .05$) on that scale.

Item	Questioning/ Stating	Calm/ Agitated	Agree/ Disagree	Optimistic/ Pessimistic	Weak/ Strong	Emotional/ Rational
no HLL	stating		disagree	pessimistic	strong	
yeah HLL	stating	Calm	agree			emotional
Lynn HLL	stating					rational
mom HLL	stating					

It would be simple to attribute differences found across words to differences in the lexical meaning of the words, and in some cases this is most likely the case. The difference between *no*HLL and *yeah*HLL on the Agree/Disagree scale is the best example. Yet the results are not so simple as that: lexically speaking, what would make *yeah* more emotional than *Lynn*, especially considering that fact that *yeah* is not always rated as emotional, the item *yeah*LML was judged to be rational. Word alone cannot be the only dimensional difference between these

items; rather the contour must be influencing the items differently on each of the scales, indicating an interaction between the words and the contours.

T-test results like those reported above are valuable for examining specific comparisons, but do little to explain larger trends in the data. Multivariate Analysis of Variance (MANOVA) can be used to investigate these larger trends. Single mean t-tests are then used to follow up the MANOVAs or examine particular questions about portions of the data.

5.2 Multivariate Analysis of Variance

A series of planned comparisons were performed that forced the twenty-seven contours into groups along a variety of dimensions. The groups were then tested via Multivariate Analysis of Variance to see how accurately they reflected patterns in the data.

5.2.1: Why Planned Comparisons

The twenty-seven contours were divided into twenty-three groups. These groupings were based on various potentially salient aspects of the contours, as suggested by previous work on intonation. This approach differs from past work in that the contours studied here are not natural: they vary in specific, restricted ways which allow differences in meaning to be attributed to specific differences in the configuration of the pitch track of the contour. Each contour differs from some other contour in only one dimension. The contours are grouped together based on similarity in these dimensions, and tested to see which dimensions were most consistently used by the participants to distinguish the meaning of the contours. The prediction is that if a grouping was formed along a dimension which is present in the data, there will be a significant result. Such a result would indicate that the dimension along which the grouping was made, for example the presence versus absence of a fall at the end of the contour, was used by the participants in the study to attribute meaning to the contours. A non-significant result would indicate that the dimension used to create that grouping was not utilized by the participants when they were assessing the meaning of the item. Examples of several of the dimensions used to form groupings for planned comparisons follow.

5.2.2 Examples of Planned Comparisons

One grouping which was used to performed a planned comparison consisted of the nuclear tones proposed in the British tradition, and another grouping was based on the presence versus absence of a change in the direction of the pitch track. The importance of a change in the direction of pitch track was argued for in Pakosz

(1982). Tables (4) and (5) show these two groupings of the contours. In Table (4) all twenty-seven contours have been divided into five different conditions as defined by the shape of the contour when all three pitch points are connected. These conditions represent the five nuclear tones used in the British tradition.

Table 4: Sample Grouping – A sample grouping of the twenty-seven contours based on the five Nuclear Tones of the British tradition.

Name	Condition Description	Contours in each Condition
Nuclear Tones	Flat	HHH, MMM, LLL
	Falling	HHM, HHL, HMM, HML, HLL, MML, MLL
	Rising	MHH, MMH, LHH, LMH, LLH, LMM, LLM
	Fall-Rise	HMH, HLH, HLM, MLM, MLH
	Rise-Fall	MHM, MHL, LHL, LHM, LML

Table 5: Sample Grouping – A sample grouping of the twenty-seven contours based the presence of a change in direction of the pitch track.

Name	Condition Description	Contours in each Condition
Change in Direction	a) Contains a change in direction (Fall-Rise and Rise-Fall)	HMH, HLH, HLM, MLM, MLH, MHM, MHL, LHL, LHM, LML
	b) Does not contain a change in direction (Flat, Fall, Rise)	HHH, MMM, LLL, HHM, HHL, HMM, HML, HLL, MML, MLL, MHH, MMH, LHH, LMH, LLH, LMM, LLM

In (5) the contours have been divided into two groups as defined by whether or not the contour changes from either a fall to a rise or a rise to a fall over the course of the utterance. If the contour does contain a change in direction it is in the first condition (5a), and if it does not change direction it is in the second condition (5b).

Pakosz (1982) and Pierrehumbert and Hirschberg (1990), among others, comment on the importance of high relative pitch in the analysis of intonational meaning. Higher-pitched utterances tend to be judged as more feminine, weak and emotional, while generally lower-pitched utterances are judged to be masculine, strong and rational (van Bezooijen 1996). Thus it is not only the changes in pitch that are of interest but also the distribution of different pitch levels relative to the other pitches used in the experiment. Another example of the type of groupings which were created is given in Table (6). Here the average pitch of the contour, rather than the pattern made by connecting the pitch points, was used to establish the conditions.

Table 6: Sample Grouping – A sample grouping based on the schematized average pitch of the entire contour.

Name	Condition Description	Contours in each Condition
Average Pitch	The average pitch of each contour was determined[9]	LLL
		MLL, LML, LLM
		HLL, MML, MLM, LHL, LMM, LLH
		HML, HLM, MHL, MMM, MLH, LMH, LHM
		HHL, HMM, HLH, MHM, MMH, LHH
		HHM, HMH, MHH
		HHH

All twenty-three of the groupings were coded as fixed factors and were tested in a series of Multivariate Analyses of Variance. Criteria for membership in every grouping are provided in the appendix. It should be noted that every contour appeared in every grouping, so the same data was present in each case, what differs and what is of interest here is how the data was divided and how that division affects the results of the statistical tests.

6. Results

As stated, a series of Multivariate ANOVAs (MANOVAs) were run using SPSS®, one for each of the twenty-three groupings created. Each MANOVA had Word (*no, yeah, Lynn* and *mom*) and one of the Groupings as fixed factors, with all six of the semantic scales as dependent variables.

As is normally the case when a large number of planned comparisons are performed, a Bonferroni correction[10] for family-wise error was applied. With this correction, a p-level of $p < 0.002$ is required for the results to reach significance. Even given the correction, nearly all of the groupings tested were significant. This indicates that many of the dimensions present in the contours were used by the participants to consistently assign meanings. In order to differentiate between the groupings and determine which of the groupings most closely resembled the primary dimensions used by participants in the study to distinguish meaning, an index of the effect size (eta-squared) was obtained for each MANOVA. Effect size provides information about what percentage of the variability in the data is

[9] The average pitch of each utterance was defined based on schematized pitch values. A low pitch point was given a value of 1, a Mid pitch point a value of 2 and a High pitch point a value of 3. The values for each contour were summed and contours were grouped based on their values. For example, MLH had a value of (2+1+3 = 6), while LLM had a value of (l+1+2 = 4).

[10] This correction takes the standard p-level of $p < .05$ and divides it by the number of comparisons tested. In this case that is .05/23 = .002.

accounted for by the fixed factors in the MANOVA (Keppel 2004). An eta-squared value of .30 indicates that thirty percent of the variance in the data can be accounted for by the factors used. Thus, the larger the effect size the more variability in the data is accounted for, and the less noise or random variation there is in the data. Less noise means that a clearer picture of what is happening can be seen.

There was no interaction between the Word factor and any of the Grouping factors. This means that each factor is acting independently on the data, so the lexical meaning of the words influences the words in a different manner than the intonational meaning. A lack of interaction suggests that the lexical meaning and the intonational meanings are acting on different aspects of utterance as a whole. This point will be returned to in the conclusion of the paper.

The main effects of the MANOVAs are discussed in the following sections. Results for the factor Word are given in (6.1), and results for different Groupings are given in (6.2–6.6). A summary of the results is presented in section (6.7). All of the results reported in the remainder of the paper, unless otherwise noted, have significance values of $p < .002$, and are thus significant even with the Bonferroni correction.

6.1 'Word' results

Unsurprisingly, there was a significant effect of Word in every analysis run. The effect size for this factor ranged from an eta-squared value of .214 for the Average Pitch grouping to an eta-squared value of .397 for the Minimum One Mid grouping. Nearly all of these effect sizes were larger than the effect sizes seen for the various grouping factors.

The large effect of Word is partially a result of the different ratings of *no* and *yeah* on many of the scales. The words *no* and *yeah* were chosen, in part, because they had opposite dictionary meanings. However, the dramatic effect of *no* as compared to the other words was not expected. In single mean t-tests, every single *no* item went in the same direction on the scales. The items containing *no* were consistently rated as Disagreeing, Stating, Strong, Pessimistic, Rational, and Agitated. Apparently *no* means 'no' regardless of how it is said.

It is interesting that *no* was rated so consistently even on scales that lack any apparent relation to its dictionary meaning. Given the proper context the word *no* can be said in isolation as a question, but it was never consistently judged to be questioning in this experiment, even when every other word said with a particular contour was so judged (Table 7). Apparently the highly definite and consistent nature of the lexical meaning of *no* made it less susceptible to intonationally based meaning changes.

Table 7: T-test results for items containing the MLH contour. Notice that even when the other three words are judged to be questioning, the item containing *no* is not. The end of the scale on which the mean response to each item fell is provided. No response is given for items that did not reach significance ($p < .05$) on that scale.

Item	Calm/ Agitated	Agree/ Disagree	Optimistic / Pessimistic	Questioning/ Stating	Weak/ Strong	Emotional/ rational
no MLH		disagree				
yeah MLH		agree	optimistic	questioning		
mom MLH	calm		optimistic	questioning		
Lynn MLH			optimistic	questioning		

No was used in the study because it had a strong dictionary meaning and this strength was borne out in the results reported above. In contrast, the other three words had more variable ratings on the scales. Though *yeah* was never rated as disagreeing, it did have every other meaning possible (Table 8).

Table 8: T-test results for several items containing *yeah*. Notice the large variability in meaning which *yeah* can have. The end of the scale on which the mean response to each item fell is provided. Responses in boxes have different meanings on the same scale. No response is given for items that did not reach significance ($p < .05$) on that scale.

Item	Calm/ Agitated	Agree/ Disagree	Optimistic/ Pessimistic	Questioning/ Stating	Weak/ Strong	Emotional/ Rational
yeah MLH		agree	optimistic	questioning		
yeah LHM	agitated		pessimistic	stating	strong	
yeah MLL	calm	agree		stating	weak	
yeah LLM		agree				rational
yeah MHH		agree				emotional

Table 9: T-test results for several items containing *Lynn*. Notice that *Lynn* items show less variability than *yeah* items. The end of the scale on which the mean response to each item fell is provided. Responses in boxes have different meanings on the same scale. No response is given for items that did not reach significance ($p < .05$) on that scale.

Item	Calm/ Agitated	Agree/ Disagree	Optimistic/ Pessimistic	Questioning/ Stating	Weak/ Strong	Emotional/ Rational
Lynn MLH			optimistic	questioning		
Lynn LMM			pessimistic	stating		rational
Lynn MHH	agitated	disagree				
Lynn HMM	calm	agree		stating		

The word that was predicted to have the most variability in meaning, *Lynn*, was variable, though less so than *yeah* (Table 9). Of all of the words, *Lynn* had the

fewest number of items to reach significance on any scale. Participants were seemingly unable to get much of a meaning for most of the *Lynn* items, perhaps because its lexical meaning was relatively unfixed in comparison to other words. Finally, *mom* has a definite dictionary meaning like *no*, yet at the same time can be used referentially like *Lynn*. Items containing *mom* were rated similarly to those containing *Lynn*, indicating that it was viewed more referentially by the participants (Table 10).

Table 10: T-test results for items containing *mom*. These items show a similar amount of variability to the *Lynn* items. The end of the scale on which the mean response to each item fell is provided. Responses in boxes have different meanings on the same scale. No response is given for items that did not reach significance ($p < .05$) on that scale.

Contour	Calm/ Agitated	Agree/ Disagree	Optimistic/ Pessimistic	Questioning/ Stating	Weak/ Strong	Emotional/ Rational
mom HHH	agitated					emotional
mom LLH	calm		optimistic	questioning		
mom LLL			pessimistic	stating		
mom MML					strong	rational
mom HHL		disagree		questioning		emotional

The main effect of Word in all of the MANOVAs originates in the consistency of the ratings of words on certain scales (i.e. *no* was rated the same on every scale, *yeah* was always rated as agreeing). The main effect of the groupings, discussed in section 6.2, originates in the changeable ratings of words relative to the contour with which the words were said. Thus the fact that items containing *yeah* were rated as agreeing, and items containing *Lynn* were rated as rational contributes to main effect of Word, while the overall the high variability of *yeah*, *Lynn* and *mom* contributes to the main effect of Grouping.

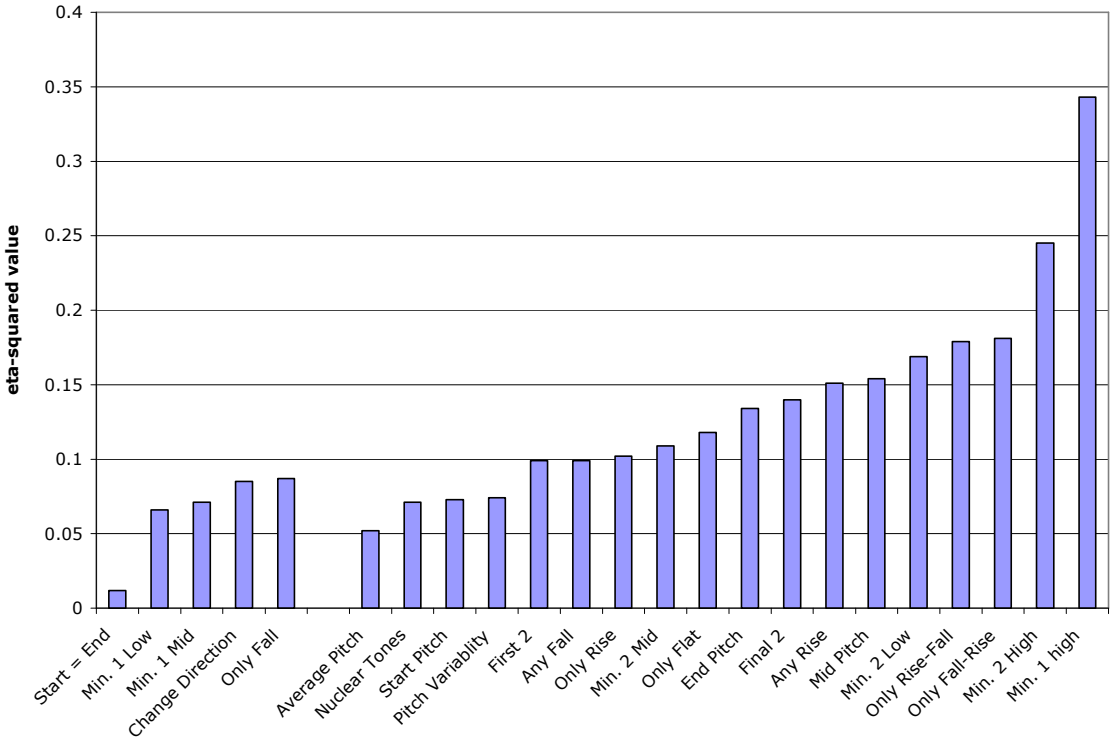
6.2 General Grouping Results

In the results for the Word factor, neither significance values nor eta-squared values were useful in determining which Groupings were of greater importance than the others. The Word factor was significant in every MANOVA run and had comparable eta-squared values across the board.[11] The results for the Grouping factor were more variable. The majority (18 out of 23) of the groupings tested had a high degree of significance, but there was substantial variation in the eta-squared values for the significant groupings. Figure (9) shows the eta-squared values for

[11] For significance levels and eta-squared values for both factors on each grouping, see the appendix.

all of the Groupings. Notice that the largest eta-squared value for the results of the Grouping factors (eta-squared = .343) is smaller than the majority of the values seen on the Word factor.

Figure 9: Eta-Squared values for all 23 Grouping factors are arranged with non-significant findings on the left and significant findings on the right. Within each group scores are arranged from smallest to largest.



6.3 Results for Minimum One High and Minimum Two High Groupings

The two groupings with the largest effect sizes are related to each other. Both of these groupings, Minimum One High and Minimum Two High, are defined based on the presence of high pitch points in the contours. A list of how the contours were divided into conditions for these two groupings is given in Table (11). Notice that the contours present in the first condition of the Minimum Two High grouping are a subset of the contours present in the first condition of the Minimum One High grouping. This subset-superset relation may indicate that the large effect size of the Minimum Two High grouping is linked to the large effect size of the Minimum One High grouping.

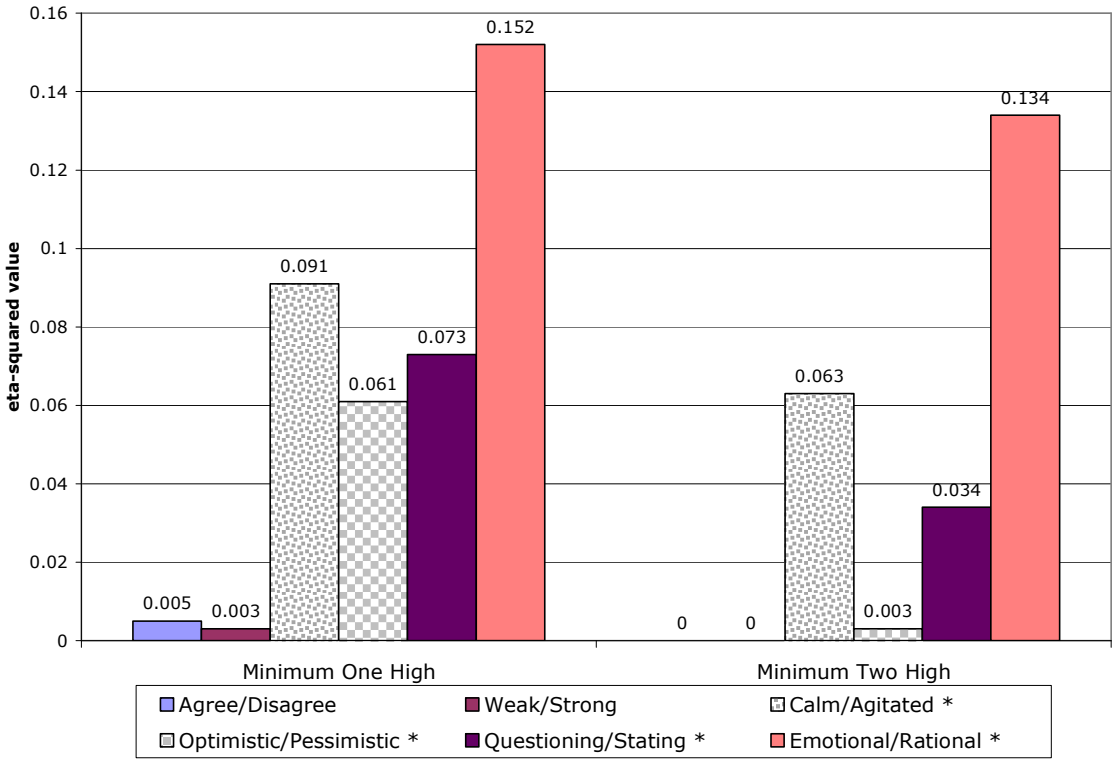
Table 11: Membership of the Minimum One High and Minimum Two High groupings.

Name	Condition Description	Contours in each Condition
Minimum 1 High	Contains at least one High pitch	HHH, HHM, HHL, HMH, HMM, HML, HLH, HLM, HLL, MHH, MHM, MHL, MMH, MLH, LHH, LHM, LHL, LMH, LLH
	Contains no High pitches	MMM, MML, MLM, MLL, LMM, LML, LLM, LLM
Minimum 2 High	Contains at least two High pitches	HHH, HHM, HHL, HMH, HLH, MHH, LHH,
	Contains no more than one High pitch	HMM, HML, HLM, HLL, MHM, MHL, MMH, MLH, MMM, MML, MLM, MLL, LMM, LML, LLM, LLM, LHM, LHL, LMH, LLH

As was shown in Figure (9), the amount of the variance accounted for (the eta-squared value) decreases from the Minimum One High grouping (.343) to the Minimum Two High (.245) grouping, indicating that the presence versus absence of any number of High pitch points was more attended to by the listeners than the presence of a large number of High pitch points. As both of these groupings have larger eta-squared values than any of the other groupings, the greater importance of high pitch, as compared to mid or low pitch, is also highlighted.

In Figure (10) the eta-squared values for each of the semantic scales for the groupings in question are given. The Emotional/Rational scale had the largest effect size in both groupings. From this it can be seen that the large effect of High pitches is linked most strongly to the Emotional/Rational scale, which corroborates the claims of previous researchers that emotion and high pitch are tied to one another (van Bezooijen 1996, Pakosz 1982, Pierrehumbert and Hirschberg 1990).

Figure 10: Eta-squared values for each semantic scale are arranged with the Minimum One High grouping on the left and the Minimum Two High grouping on the right. Within the responses to each grouping, non-significant values are on the left and significant values (marked with an asterisk in the legend) are on the right. Exact eta-squared values for each scale are provided in the figure.



On every scale the Minimum One High grouping has a larger effect size than the Minimum Two High grouping, showing that contours with at least one high pitch point have more consistent ratings, when compared to the remaining contours, than contours containing at least two high pitch points. Also notice that there is almost no effect on the Agree/Disagree or the Weak/Strong scales for these groupings. As will be seen throughout the results section, the importance of the scales varies from one grouping to another.

6.4 Rise-Fall and Fall-Rise groupings

The Rise-Fall grouping and the Fall-Rise groupings have the next two highest effect sizes of the all of the groupings tested. There are only five contours that have a Fall-Rise pattern. Thus when all twenty-seven contours are divided into two groups based on the presence versus absence of a Fall-Rise pattern, one group has five highly similar members and other has twenty-two highly variable members. This also true of the Rise-Fall grouping (Table 12).

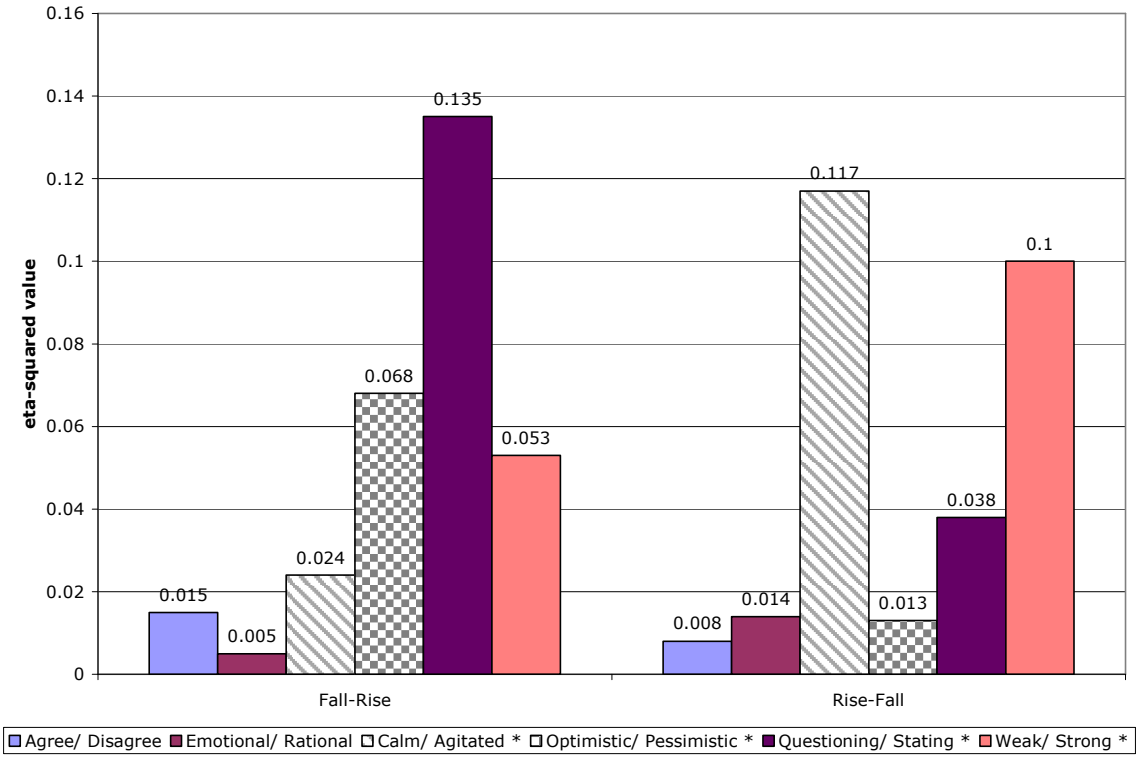
Table 12: Membership of the Fall-Rise and Rise-Fall groupings. Notice, for both of these groupings, the high degree of similarity within the contours in the first condition compared with the diversity of contour types in the second condition.

Name	Conditions	Contours in each condition
Fall-Rise	A fall-rise is the only contour present	HMH, HLH, HLM, MLM, MLH
	Any other contour is present (Flat, Fall, Rise, Rise-Fall)	HHH, MMM, LLL, HHM, HHL, HMM, HML, HLL, MML, MLL, MHH, MMH, LHH, LMH, LLH, LMM, LLM, MHM, MHL, LHL, LHM, LML
Rise-Fall	A rise-fall is the only contour present	MHM, MHL, LHL, LHM, LML
	Any other contour is present (Flat, Fall, Rise, Fall-Rise)	HHH, MMM, LLL, HHM, HHL, HMM, HML, HLL, MML, MLL, MHH, MMH, LHH, LMH, LLH, LMM, LLM, HMH, HLH, HLM, MLM, MLH

Recall that eta-squared values tell how much of the variability in the data is accounted for by a particular factor. Given the heterozygous nature of one condition of the group, the other condition must have highly consistent ratings for such a strong effect to emerge. Figure (11) shows the eta-squared values for each semantic scale.

The two scales with the largest eta-squared values for the Fall-Rise grouping are the Questioning/Stating and the Optimistic/Pessimistic scales. According to Gussenhoven (1984) there are only three main contour types in English, one of which is the Fall-Rise. Ward and Hirschberg (1985) worked on the Fall-Rise contour, and decided that it related to speaker uncertainty; given the scales used in the current study, the Questioning/Stating scale seems the most likely to relate to uncertainty. It is not much of a stretch to say that the other scale which plays an important role for this grouping, Optimistic/Pessimistic, also relates to speaker uncertainty; an uncertain speaker could easily be either optimistic or pessimistic about the correctness of their response.

Figure 11: Eta-squared values by semantic scale - Fall-Rise and Rise-Fall groupings. Within the responses to each grouping, non-significant values are on the left and significant values (marked with an asterisk in the legend) are on the right. Exact eta-squared values for each scales are provided in the figure.



In contrast, the two largest eta-squared values for the Rise-Fall grouping are found on the Calm/Agitated and Weak/Strong scales. This contour has been less discussed in the literature and has been grouped with other contours that end in a fall (Gussenhoven 1984). In the next section (6.5), differences between rising and falling contours are discussed further.

As a final note, the Emotional/Rational scale, which was so important for the groupings that focused on the presence of high pitch points, does not reach significance for either these groupings. This emphasizes the importance of having a variety of scales available for the measurement of intonational meaning. The influence of intonation on the meaning of utterances is quite complicated and cannot be captured by one or two scales.

6.5 Rises and Falls

To one degree or another, everyone who has done research on intonation in English has focused on the distinction between rising and falling contours. In this

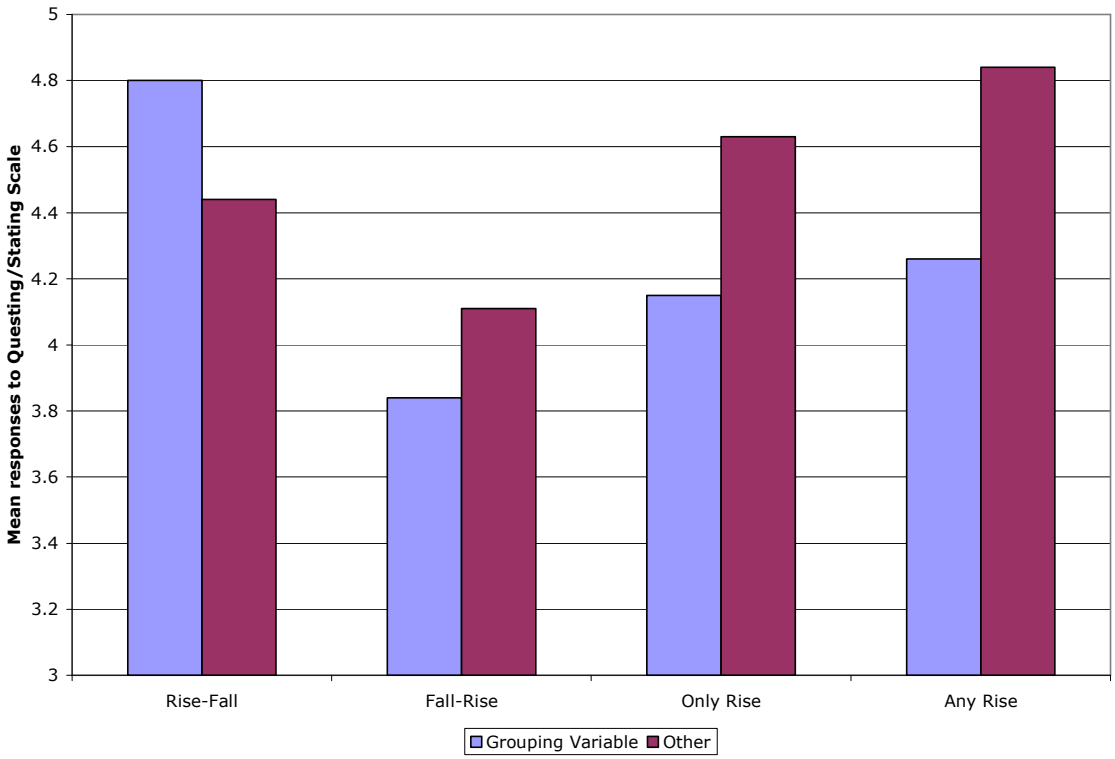
section two Rise-based groupings, Any Rise and Only Rise, and two Fall-based groupings, Any Fall and Only Fall, are examined. The difference between the two Rise-based groupings lies in where the Fall-Rise and Rise-Fall contours are located. Table (13) gives the membership of each condition of these groupings. In the Any Rise grouping, all contours that have a rise in pitch are in the first condition (13a), and those contours that do not contain a rise are in the second condition (13b). In contrast, in the Only Rise grouping only those contours that have a rise as their only pitch change are in the first condition (13c), while the remaining contours are in the second condition (13d). The Fall-Rise and Rise-Fall contours (shown in italics) switch from being located with the rises in the Any Rise grouping, to being located with the falls in the Only Rise grouping.

Table 13: Membership of the Any Rise and Only Rise Groupings (the Fall-Rise and Rise-Fall contours are shown in italics).

Name	Condition	Contours in each condition
Any Rise	(a) Contains a rise in pitch anywhere	MHH, MMH, LHH, LMH, LLH, LMM, LLM, *HMH, HLH, HLM, MLM, MLH, MHM, MHL, LHL, LHM, LML*
	(b) Does not contain a rise	HHH, MMM, LLL, HHM, HHL, HMM, HML, HLL, MML, MLL
Only Rise	(c) Contains only a rise in pitch	MHH, MMH, LHH, LMH, LLH, LMM, LLM
	(d) Does not contain a rise or has both a rise and a fall	HHH, MMM, LLL, HHM, HHL, HMM, HML, HLL, MML, MLL, *HMH, HLH, HLM, MLM, MLH, MHM, MHL, LHL, LHM, LML*

Both the Any Rise and the Only Rise groupings have relatively large eta-squared values, .151 and .102 respectively, yet only the Questioning/Stating scale reaches significance for either of these groupings. Rising contours have long been tied to the questioning intonation patterns, so the importance of this scale for these grouping is not surprising. What is surprising is that when the mean responses to the Any Rise and Only Rise groupings are examined, none of them have a mean that locates them on the questioning end of the Questioning/Stating scale. In fact, of the previous four groupings discussed, only the Fall-Rise condition of the Fall-Rise grouping is rated as questioning, and it is rated only very marginally so (Figure 12). According to the literature, rises are supposed be questioning, and when individual items from the current study are examined only those contours that end in a rise are judged to be questioning, yet neither of the groupings designed to evaluate the rising dimension have questioning mean responses.

Figure 12: Mean responses to the Questioning/Stating scale for the Fall-Rise, Rise-Fall, Any Rise & Only Rise groupings. Each of these groupings had significant values on this scale when tested. The Fall-Rise Grouping is the only one which was judged to be questioning.



The data is somewhat more interpretable when the falling groupings are considered (Table 14). Once again the difference between these two groupings lies in the location of the Fall-Rise and Rise-Fall contours.

Table 14: Membership of the Any Fall and Only Fall groupings (the Fall-Rise and Rise-Fall contours are shown in italics).

Name	Condition	Contours in each condition
Any Fall	A fall is present anywhere in the contour (Fall, Rise-Fall, Fall-Rise)	HHM, HHL, HMM, HML, HLL, MML, MLL, *HMH, HLH, HLM, MLM, MLH, MHM, MHL, LHL, LHM, LML*
	No fall is present (Flat, Rise)	HHH, MMM, LLL, MHH, MMH, LHH, LMH, LLH, LMM, LLM
Only Fall	A fall is the only pitch change (Fall)	HHM, HHL, HMM, HML, HLL, MML, MLL
	A fall is not the only pitch change (Flat, Rise, Fall-Rise, Rise-Fall)	HHH, MMM, LLL, MHH, MMH, LHH, LMH, LLH, LMM, LLM, *HMH, HLH, HLM, MLM, MLH, MHM, MHL, LHL, LHM, LML*

Both of these grouping have means that go in the direction predicted by the literature: they are both judged to be stating, but they have a much lower overall degree of interpretability to the participants. The Any Fall grouping has significant results for its MANOVA, but its eta-squared value (.099) is lower than any grouping discussed so far. The Only Fall grouping does not even reach significance for the initial MANOVA. So while the results go in the expected direction, the groupings themselves are apparently less important.

6.6 Groupings that relate to the Start versus the End of an utterance

There are two sets of groupings which relate to differences in the start or end of the utterance. The first set of groupings contains the Final Two and the First Two groupings. The Final Two grouping focuses on the pitch pattern created by connecting the final two pitch points of the item. This grouping has three conditions: Flat, Rising, and Falling. Similarly, the First Two grouping focuses on the pitch pattern formed by connecting the first two pitch points of the utterance, it also has three conditions: Flat, Rising, and Falling. The second set of groupings focuses on the pitch at either the end of the word (the End Pitch grouping) or the start of the word (the Start Pitch grouping). When both sets of groupings are viewed simultaneously a distinct pattern emerges. The Final Two grouping and the End Pitch grouping both have larger eta-squared values than the First Two and the Start Pitch groupings. As well, the Final Two and End Pitch groupings both have significant results on the Questioning/ Stating scale, yet neither the First Two grouping nor the Start Pitch grouping does. This indicates that what is happening at the end of the utterance, even an utterance as short as these (about 500ms), is more important than what happens at the beginning of the utterance, particularly in relation to whether or not the utterance is questioning or stating.

This finding corroborates work done by Pierrehumbert (1980) that caused her to establish boundary tones as one of the three primary components of intonational meaning. These boundary tones were created because of the influence of pitch changes at the end of an utterance on the meaning of the utterance as a whole. Pierrehumbert's system also includes phrase accents, which can appear at both the beginning and the end of a phrase, but the presence of a pitch accent at the start of a phrase is optional while the presence of a boundary tone at the end is not.

6.7 Summary of results

Tables (15–17) below provide a brief summary of the main findings discussed in section 6 of this paper. For reasons of space, only ten of twenty-three planned comparisons preformed were discussed here. By looking at a selection of the groupings which were significant and had a large effect size, it is possible to

understand some of the dimensions used by participants in the study to assess the meaning of the different contours. First, Table (15) provides a summary of the results for the Word factor. This factor had a significant effect in every MANOVA run, and a relatively large effect size across the board, with eta-squared values ranging from .294 to .393.

Table 15: Summary of results for the Word factor. The general responses to scales show only those scales which had consistent responses regardless of the contour with which the word was said.

Word	General responses on Scales	Observations
no	Agitated, Disagreeing, Rational, Stating, Strong, Pessimistic	Least influenced by intonation, never changed direction on scales
yeah	Agreeing	Most influenced by intonation, perhaps due to variety of possible uses in conversation.
Lynn	Rational	Fewest number of significant responses on scales, perhaps due to low amount of dictionary meaning
mom	Disagree, Strong	Larger number of significant responses than *Lynn*, treated as referential rather than as having definite dictionary meaning

Tables (16 and 17) summarize the results discussed for the various Grouping factors. The majority of Grouping factors had significant results in their respective Multivariate ANOVAs. However, unlike with the Word factor, the eta-squared values for the grouping factors had a wide range (.052 – .343).

Table 16: Summary of results for individual grouping factors discussed in the text. The general responses to scales show only those scales which had significant responses for that grouping.

Grouping	Response on Scales	Observations
Minimum One High	Rational	Largest eta-squared value for all of the grouping factors
Minimum Two High	Emotional	Supports findings in literature of correlation between high pitch and emotion
Fall-Rise	Questioning, Pessimistic	Agrees with previous research, indicates uncertainty
Rise-Fall	Stating, Strong, Agitated	Acts like a falling contour

The final table in this section (17) focuses on Grouping factors which were discussed in pairs.

Table 17: Summary of results for pairs of grouping factors discussed jointly in the text. The general responses to scales show only those scales which had significant responses for the groupings in question.

Grouping	Response on Scales	Observations
Any Rise & Only Rise	Stating	Goes against previous findings of rises being tied to questions, relatively important dimension to the participants.
Any Fall & Only Fall	Stating	Goes with previous findings of falls being linked with statements, relatively unimportant dimension to the participants.
End Pitch & Final Two versus Start Pitch and First Two	Questioning/Stating	Changes at end of an utterance are more attended to than changes at start of an utterance.

7. Discussion and Conclusions

This paper began with a discussion of two different approaches to the study of intonation. The *impressionistic* approach treats intonational contours as decomposable units and seeks to generate a treatment of these contours that is in line with current phonological theories. The *instrumental* approach treats contours as single units largely unrelated to one another yet subject to fine-grained phonetic analysis and experimentation for interpretation. In this study a fairly extreme instrumental method was used to answer a more impressionistic question. The method was instrumentally extreme in that the contours studied were not natural contours, but instead were created on a grid to completely explore a limited region of acoustic space. The rationale for this method was that the contours thus created were close enough to naturally occurring contours to provide insights into English intonational phenomenon, but by using these artificial contours a larger degree of control of the acoustic parameters could be obtained. This degree of control makes it possible to find direct links between pitch changes and changes in the interpretation of the contour, as pitch is the only thing which varied from one contour type to the next. The impressionistic question addressed asks what changes in sub-parts of the contour are required to affect a change in the meaning of the entire contour.

This question was addressed by dividing the twenty-seven contours into groupings defined along a number of acoustic dimensions. These dimensions were primarily ones previously suggested in the literature to be relevant to intonational meaning; some treated the entire contour as a unit (i.e. Nuclear Tone grouping) while others focused on smaller pieces of the contours (i.e. Final Two grouping). All twenty-seven contours were present in each grouping, so the same data was used in each of the tests run; what differed was how that data was divided.

This paper set out to not only uncover interactions between acoustic cues to intonational meaning in English, but also to investigate the interaction between these cues and semantic meaning. There was no significant interaction between any of the Grouping factors and the Word factor in any of the Multivariate Analyses of Variance which were run. This means that according to this test, there is no interaction between contours and words.

A main effect of Word was found in all of the analyses that were run. In section 6.1 this effect was attributed largely to differences in the response patterns seen for *no* as compared to the rest of the words. *No* was the word with the most fixed lexical meaning, and the most consistent meaning on the semantic scales. In contrast with *no*, the other three words all had variable ratings on the scales. The intonations with which these words were said, in combination with their lexical interpretations and uses, helped to determine how they were rated on each of the scales. This is a description of an interaction between lexical meaning and intonational meaning, for which no statistical evidence was found in the MANOVAs run. This lack of statistical evidence for an interaction in the face of intuitive evidence for one is confusing and necessitates further investigation.

Univariate ANOVAs can be used as a follow-up test to Multivariate Analysis of Variance tests. These ANOVAs are ongoing, but one series of univarate ANOVAs has been run on the Final Two grouping. This grouping, which had an eta-squared value of .140 in the MANOVA, showed a significant interaction between the Word factor and the Final Two Factor on three of the scales.

Table 18: Results from a series of follow-up ANOVAs which were run on the Final Two grouping. Significant values are marked with an asterisk. The Bonferroni correction applied previously to the Multivariate ANOVAs continues to apply. Notice that only three of the scales show a significant interaction between Word and the Final Two grouping factor.

Scale	Interaction	Final Two	Word
Questioning/Stating	$p < .001$ *	$p < .001$ *	$p < .001$ *
Agree/Disagree	$p < .01$	$p = .062$	$p < .001$ *
Emotional/Rational	$p = .580$	$p = .057$	$p < .001$ *
Calm/Agitated	$p < .001$ *	$p < .001$ *	$p < .001$ *
Strong/Weak	$p < .001$*	$p < .001$ *	$p < .001$ *
Optimistic/Pessimistic	$p < .005$	$p < .005$	$p < .001$ *

The lack of a significant interaction in the MANOVA run on this grouping probably relates to the presence of an interaction on only three of the six scales. As was noted in the results section, different scales are relevant to the interpretation of the various dimensions present in the data, so a non-significant interaction on a MANOVA which takes into consideration all of the scales as a collective should not be surprising, when upon further investigation an interaction on a subset of the scales can be seen.

Results from the single mean t-tests can also be used as a form of follow-up test. These results show that in some cases very small variations in the acoustics are able to affect dramatic changes in the meanings assigned to utterances. For example, when *mom* was said with one falling contour (HHL) it was interpreted as emotional, yet when it was said with another falling contour (MML) it was considered rational. The only difference between these contours is that one starts higher than the other. This change was only seen on one scale, and occurred between a pair of contours that would have been lumped together in many of the groupings. There are several levels at which the data collected for this study can be analyzed, and analysis is ongoing.

At the current level of analysis, several acoustic cues have been shown to influence the intonational meaning of the items tested. The single most important cue, as defined by having the largest effect size, is the presence of a High pitch point. This result is new: though others have observed that relative pitch influences the interpretation of intonational meaning, this is the first finding of the salience of High pitch in contrast with Mid or Low pitch in the same environment. Recall that the groupings with the two largest eta-squared values were the Minimum One High and Minimum Two High groupings. The Minimum One Low and Minimum One Mid groupings did not even reach significance.

Another acoustic cue that emerged as important in this data was the presence of a pitch change at the end of the utterance, particularly when compared with a

similar pitch change at the start of an utterance. The Final Two and End Pitch groupings both had Higher eta-squared values than the First Two and Start Pitch groupings. This finding is hinted at in the literature: Pierrehumbert's system requires more detailed coding of the pitch at the end versus the beginning of an intonational phrase, but the data presented here confirms the greater relevance of the ending pitch.

It is apparent from these results that there are a number of acoustic as well as lexical dimensions present in the data presented here. The majority of the previous claims about what factors influence intonational meaning - claims ranging from the importance of the average pitch of the contour, to the presence of a change in the direction, to the contrast between rises and falls - are supported by these results. What is most intriguing is that all of these factors can be attended to at once by listeners. Much previous work has focused on just one or possibly two of these factors and manipulated them, no one has attempted to manipulate so many different acoustic cues simultaneously. While approaches that focus on just one acoustic factor, like average pitch, have their purposes and provide insight into that single factor, intonation as a whole is a complex system that can only be understood when a wide variety of influences are taken into account.

Appendix

Semitone values used for each word for each point on the grid.

Yeah

High	13.72	13.56	13.39
Mid	11.53	11.38	11.21
Low	9.35	9.20	9.03

No

High	13.18	13.07	12.82
Mid	11.00	10.89	10.64
Low	9.06	8.95	8.70

Mom

High	12.90	12.83	12.49
Mid	10.72	10.65	10.31
Low	8.54	8.47	8.37

Lynn

High	13.57	13.32	13.05
Mid	11.39	11.13	10.87
Low	9.45	9.20	8.93

Eta-square values – These are the eta-squared vales for each of the twenty-three MANOVAs run. The results are arranged with the with the non-significant findings at the top of the table and significant findings beneath them. Within each set (non-significant and significant) the finding are ordered from smallest to largest eta-squared values, just as in Figure (9).

		Grouping	Word
Not significant ($p > .002$)	Start Pitch = End Pitch	0.012	0.393
	Minimum 1 Low	0.066	0.391
	Minimum 1 Mid	0.071	0.397
	Change of Direction	0.085	0.387
	Only Fall	0.087	0.367
Significant ($p < .002$)	Average Pitch	0.052	0.214
	Nuclear Tones	0.071	0.288
	Start Pitch	0.073	0.366
	Pitch Variability	0.074	0.317
	First 2	0.099	0.363
	Any Fall	0.099	0.381
	Only Rise	0.102	0.341
	Minimum 2 Mid	0.109	0.375
	Only Flat	0.118	0.318
	End Pitch	0.134	0.349
	Final 2	0.14	0.338
	Any Rise	0.151	0.364
	Mid Pitch	0.154	0.353
	Minimum 2 Low	0.169	0.372
	Only Rise-Fall	0.179	0.391
	Only Fall-Rise	0.181	0.281
	Minimum 2 High	0.245	0.386
	Minimum 1 High	0.343	0.351

References

Bollinger, Dwight. 1986. *Intonation and Its Parts*. Stanford University Press.
Bybee, Joan. 2002. "Phonological Evidence for Exemplar storage of Multiword Sequences". SSLA, 24, 215–221.
Caspers, Johanneke. 2000. "Experiments on the meaning of four types of single-accent intonation patterns in Dutch". *Language and Speech*. 43 (2). 127–161.
Cooper, William and Jeanne Paccia-Cooper. 1980. *Syntax and Speech*. Cambridge, MA: Harvard University Press.
Grice, H. P. 1975. "Logic and conversation". *Speech acts (Syntax and Semantics III).* Peter Cole &Jerry Morgan (eds.) New York: academic Press. 41–58.
Cruttenden. 1981. "Falls and rises: meanings and universals". *Journal of Linguistics*. 17, 77–91.
Crystal. 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press.
Gussenhoven, Carlos. 1991. *Tone segments in the intonation of Dutch.* In T.F. Shannon & J.P. Snapper (eds.) The Berkeley Conference on Dutch Linguistics, 1989. Lanham MD : University Press of America. 139–155.
Gussenhoven. 1984. *On the Grammar and Semantics of Sentence Accents*. Foris Publications.
Gussenhoven, Carlos. 2002. "Intonation and interpretation: Phonetics and phonology." *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, ProSig and Université de Provence Laboratoire Parle et Language. 47–57.
Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. The Hauge: Mouton.
Keppel, Geoffery and Thomas D. Wickens. 2004. *Design and Analysis : A researcher's handbook*. Pearson: Prentice Hall.
Ladd, Robert D. 1996. *Intonational Phonology*. Cambridge University Press.
Ladefoged, Peter. 1996. *Elements of Acoustic Phonetics: Second Edition*. University of Chicago Press, Chicago.
Lieberman, Mark and Ivan Sag. 1974. "Prosodic form and discourse function." *Proceedings of the Chicago Linguistics Society*: 10: 416–27.
Levis, John M. 2002. "Reconsidering Low-Rising Intonation in American English." *Applied Linguistics*. 23:1 56–82.
Mehrabian, Albert and Morton Wiener. 1967. "Decoding of inconsistent communications". *Journal of Personality and Social Psychology*. Vol. 6, No. 1, 109–114.
Nosofsky, Robert M. 1988. "Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality." *Journal of Experimental Psychology: Learning Memory and Cognition.* Vol. 14 700–708.
Ohala, John. 1984. "An Ethological Perspective on Common Cross-Language Utilization of f0 of Voice." *Phonetica*. 41. 1–16.
Osgood, Charles E., George J. Suci, Percy H. Tannenbaum. 1957. *The measurement of Meaning*. University of Illinois Press.
Pakosz, Maciej. 1982. "Intonation and Attitude". *Lingua*. 56. 153–178.
Pakosz, Maciej. 1984. "Contextual considerations in the semantic analysis of intonation". *Journal of Pragmatics*. 8. 169–181.
Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation.* PhD thesis, MIT, published 1988 by IULC.
Pierrehumbert, Janet and Julia Hirschberg. 1990. "The meaning of intonational Contours in the interpretation of Discourse". In P.R. Cohen, J. Morgan, & M.E. Pollack (eds.) *Intentions in communication*. Cambridge MA: MIT Press. 271–311.

Trager, George L. and H. L. Smith. 1951. *An outline of English Structure*. Norman, OK: Battenburg Press.

Van Bezooijen, Renée. 1996. "The effect of pitch on the attribution of gender-related personality traits." In "*Gender and Belief Systems"* eds. Natasha Warner, Jocelyn Ahlers, Loela Bilmes, Monica Olive, Suzanne Wertheim, and Melinda Chen. Proceedings of the 4th Berkeley Women and Language Conference. April.

Ward, Gregory and Julia Hirschberg. 1985. "Implicating uncertainty: the pragmatics of Fall-Rise intonation". *Language*, V 61, Number 4. 747–776.

Yngve, V. H. 1970. "On getting a word in edgewise" In *Papers from the Sixth Regional Meeting, Chicago Linguistic Society*. 567–577.

Erin Good
Dept. of Linguistics
University of Arizona
Douglass Building, Room 200-E
PO Box 210028
Tucson, AZ 85721
emgood@email.arizona.edu