Does Intonational Meaning Come From Tones or Tunes? Evidence Against a Compositional Approach

Audra Dainora

Department of Linguistics and Philosophy
Massachusetts Institute of Technology
dainora@mit.edu

Abstract

The interaction between tones in a corpus is analyzed using conditional probability and mutual information, and a probabilistic model of intonation in American English is presented. The last pitch accent in the final intermediate phrase is found to be a strong predictor of boundary tone; this is modeled as a second order Markov process. The implications of these results for the compositional theory of intonational meaning of Pierrehumbert and Hirschberg (1990) are discussed. The conclusion is reached that the tones in a tune are interrelated in a way that a model that assigns separate meaning to each tone cannot capture.

1. Introduction

One important question in prosodic research is whether each tone contributes individually to the interpretation of an utterance, or whether intonational meaning can be determined only by examining a tune in its entirety. This paper addresses this issue by focusing on the interaction between pitch accents and boundary tones. The data introduced will help evaluate the contribution to meaning of individual tones versus whole tunes. While the compositional approach of [7] is often cited in recent literature, there has been very little debate about whether it is adequately explains intonational meaning.

I begin with a brief overview of the compositional approach to intonational meaning. I then analyze the interaction between tones in my corpus using conditional probability and mutual information as my primary statistical tools. I introduce a probabilistic model of intonation and discuss why a second order Markov model of intonation best explains the phonological patterns in the data and also suggests that intonational meaning is not compositional.

2. The compositional approach to meaning

The compositional approach to intonational meaning is exemplified by [1] and [6]. It is proposed in [6] that pitch accents, phrasal tones, and boundary tones each contribute to the meaning of the intonational tune. Pitch accents signal how the speaker intends the hearer to interpret information about the referents, modifiers, and predicates that correspond to the accented lexical items in the discourse. Phrasal tones and boundary tones convey whether an intermediate or intonational phrase is interpreted as a related unit with respect to the preceding and following intermediate or intonational phrases. The authors in [6] assign a meaning to each of the six types of pitch accents given in [5]. For example, the H* accent indicates that the associated lexical item is new information to be included in the hearer's mutual beliefs. This approach contrasts with that of numerous papers in which the authors discuss the meaning of tonal contours in terms of whole tunes or rises and falls. (See [4] for a list of relevant papers.)

3. Interactions between tones

I begin my investigation of the compositional approach to meaning by determining whether the choice of pitch accents has any bearing on the choice of the subsequent phrasal tone or boundary tone. If tones are not chosen independently of each other by the speaker than we can infer that meaning does not arise from the sum of the meaning of individual tones. The data used in this study are taken from the Boston University Radio News Corpus. A description of the corpus and the methodology used in this study are given in [4].

Table 1 shows the probabilities of phrasal tones and boundary tones conditional on the nuclear pitch accent of the phrase. Since there are only two possible phrasal tones and two possible boundary tones, I display the probability of a high tone occurring subsequent to the given nuclear accent. The probability of the relevant low tone (L- for phrasal tones or L% for boundary tones) is simply one minus the number given in the table.

	Frequency Of Occurrence Of H- Given Stated Pitch Accent	Frequency Of Occurrence Of H% Given Stated Pitch Accent
H*	8%	39%
L*	4%	83%
L+H*	9%	54%
L*+H	17%	17%
H+!H*	1%	26%

Table 1: Conditional probabilities of phrasal and boundary tones.

The data indicate that the nuclear accent is a significant determinant of the boundary tone. The pitch accent may also play a role in the likelihood of a high or low phrasal tone, but here the evidence is mixed. For example, a high phrasal tone is twice as likely to occur subsequent to L+H* as it is subsequent to L* (.08 versus .04). However, this difference is not very large numerically or statistically. The chance of a high phrasal tone occurring after H* is 8%. Greater variation is seen in the probabilities of H- subsequent to the two least common nuclear accents: after an L*+H nuclear accent we see H- in 17% of all cases, but subsequent to H+!H* we see H- in only 1% of cases. H+!H* is sufficiently rare that no statistical significance can be attached to this finding. The result for

L*+H, while not overwhelming statistically, is sufficiently sample size: we have some evidence that L*+H is more likely to be followed by a high phrasal tone than are the other nuclear accents.

In the case of boundary tones, we see a great deal of variability in the conditional probabilities of a high tone. While 83% of boundary tones subsequent to L* pitch accents are high, only 17% of boundary tones subsequent to L*+H pitch accents are high. H+!H* is also associated with a preponderance of low boundary tones, with only 26% occurrence of H%. In the middle are H* (at 39% high boundary tones), and L+H* (at 54% high boundary tones). But even these two numbers are quite far apart numerically and statistically speaking, and because of the large number of H* and L+H* phrases, the difference between these two probabilities is statistically significant at the 1% level.

Overall it appears quite clear that a phrase's nuclear pitch accent is an important determinant of the boundary tone: high boundary tones are far more likely subsequent to certain pitch accents than they are to others. These results are confirmed by calculating the mutual information shared between the pitch accents and phrasal and boundary tones. The results are presented in Tables 2 and 3.

	Observed/ Expected Probability With H-	Pointwise Mutual Information With H-	Observed/ Expected Probability With L-	Pointwise Mutual Information With L-
H*	1.05	0.051	1.00	-0.004
L*	0.59	-0.528	1.03	0.032
L+H*	1.19	0.173	.99	-0.015
L*+H	2.29	0.827	.90	-0.107
H+!H*	0.18	-1.738	1.06	0.063
Average Mutual Information			0.003794	

Table 2: Pointwise and average mutual information values for pitch accents and phrasal tones.

	Observed / Expected Probability With H%	Pointwise Mutual Information With H%	Observed / Expected Probability With L%	Pointwise Mutual Information With L%
H*	0.87	-0.144	1.11	0.104
L*	1.84	0.608	0.04	-1.161
L+H*	1.19	0.175	0.35	-0.170
L*+H	0.37	-0.995	0.01	0.417
H+!H*	0.57	-0.564	0.14	0.303
Average Mutual Information			0.037691	

Table 3: Pointwise and average mutual information values for pitch accents and boundary tones.

The average mutual information between pitch accents and phrasal tones is low (0.0038); this indicates that knowing the identity of a given pitch accent gives little information about the identity of the phrasal tone, and vice versa. However the mutual information between pitch accents and boundary tones is high. The identity of a pitch accent is a significant determinant of the identity of a boundary tone, and vice versa.

In our tests of whether phrasal tones predict boundary tones, the results are quite stark: high boundary tones succeed low phrasal tones about half the time (46%), but succeed high large to be quite suggestive even with the relatively small phrasal tones only quite rarely (16%). These results are summarized in Table 4.

	Frequency Of Occurrence Before H%	Pointwise Mutual Information With H%	Frequency Of Occurrence Before H%	Pointwise Mutual Information With L%
H-	16%	-1.041	84%	0.426
L-	46%	0.050	54%	-0.043
Average Mutual Information			0.015012	

Table 4: Conditional probabilities of boundary tones.

If the phrasal tone is low, we can say very little about which boundary tone will occur, but if the phrasal tone is high we can state with some confidence that the boundary tone will be low. Thus phrasal tone can be an important predictor of boundary tone, but only in the less common case of a high phrasal tone. Consistent with this, the mutual information shared between them is moderate (0.015012).

The number of examples found for each combination of phrasal tones and boundary tones is given in Table 5.

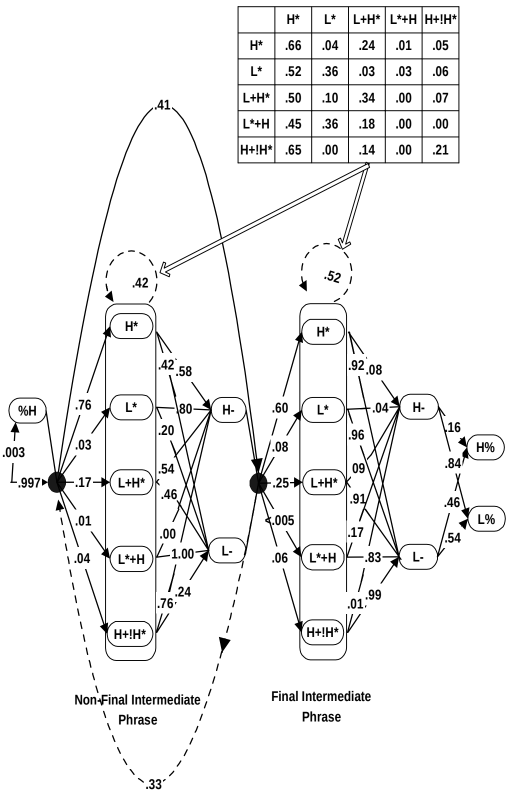
	Number Of Occurrences	Frequency Of Occurrence
L-L%	589	49%
L-H%	530	44%
H-L%	74	6%
H-H%	14	1%

Table 5: Frequency distributions of phrasal tone and boundary tone combinations.

Given the findings above, the results of Table 6 are no surprise. This table uses the nuclear accent and phrasal tone as conditioning information to predict the boundary tone. In other words, given one of the ten initial pairings of nuclear accent and phrasal tone, can we make a prediction about which boundary tone will occur?

	Number Of Occurrences	Frequency Of Occurrence In Sample Of 1207 Phrases	Frequency Of Occurrence Before H%
H*L-	674	56%	41%
L+H*L-	274	23%	58%
L*L-	89	7%	84%
H+!H*L-	77	6%	26%
H*H-	56	5%	16%
L+H*H-	26	2%	12%
L*+HL-	5	less than .5%	less than .5%
L*H-	4	less than .5%	50%
L*+HH-	1	less than .5%	0%
H+!H*H-	1	less than .5%	0%

Table 6: Frequency of distribution of nuclear accent and phrasal tone combinations.



	H*	L*	L+H*	L*+H	H+!H*
H*	.66	.04	.24	.01	.05
L*	.52	.36	.03	.03	.06
L+H*	.50	.10	.34	.00	.07
L*+H	.45	.36	.18	.00	.00
H+!H*	.65	.00	.14	.00	.21

Figure 1: A probabilistic model of intonation in American English.

Since both the nuclear accent and phrasal tone are independently strong predictors of the boundary tone, it is no surprise that in combination they provide substantial information about which boundary tone will be chosen. For example, looking at Table 6, if the first two tones in a sequence are high, the boundary tone is almost certain to be low. On the other hand, if the sequence begins L+H*L-, a high tone is not at all unexpected in the final position. Other similar predictions can be seen in the table. Overall the results reinforce the findings of Table 1 and Table 4 on the predictability of boundary tones by nuclear accents and phrasal tones. These data indicate that tones are not chosen independently of the tones that precede them. In certain phrases the identity of the boundary tone is almost predetermined by the preceding tones.

To summarize, the choice of pitch accent determines to a great extent the choice of boundary tone. The boundary tone after an L* nuclear accent is 83% likely to be high, while the boundary tone after L*+H is only 17% likely to be high. Significant results were also obtained for the other three pitch accents. These results are confirmed by tests of mutual information, which show a strong relation between the variables in question. There is also evidence that the choice of pitch accent determines the choice of phrasal tone. With regard to the interaction between phrasal tones and boundary tones, low phrasal tones are followed by high boundary tones about half the time, but high phrasal tones are followed by high boundary tones only 16% of the time. When we look at nuclear pitch accents and phrasal tones together, we see that they in combination act to predict certain boundary tones.

4. A probabilistic Markov model of intonation

Based on the results of the statistical analyses of my data set I present a formal model of intonation in American English, as shown in Figure 1. I adopt the assumptions of [2], [3], and [5] about the mappings between the frequency contour and the phonological level, as well as the finite-state grammar that produces tunes from combinations of pitch accents, phrasal tones and boundary tones.

The model in Figure 1 is hierarchical in nature and contains two stages. At the bottom level of the hierarchy are individual tones. In the case of pitch accents, anytime a pitch accent is spoken, the model allows for the possibility that it will be followed by another pitch accent (indicated by a backward curving dashed arrow), which returns us to the same place in recursive fashion. Phrasal and boundary tones lack this recursive feature (e.g. the model cannot have two phrasal tones in a row.) The second level of the hierarchy is the phrasal level. I define the last intermediate phrase in each intonational phrase to be the *final intermediate phrase* and all other intermediate phrases in the intonational phrase to be *non-final*. At the end of each non-final intermediate phrase, the model gives an opportunity for recursion: another intermediate phrase may follow. After a sequence of non-final intermediate phrases of some length (possibly length 0), the model moves on to the final intermediate phrase, after which the intonational phrase ends. The model I propose is considerably more complex than the model in [6] or a model consistent with the theory in [2]. But the true test of a model is how well it describes the data. Although parsimony is a desirable feature of models, a more complex model may be justified if it assigns higher likelihood to the observed data.

A central issue in understanding the nature of intonational tunes is whether they constitute *Markov chains*. A sequence of random variables is said be a Markov chain (or simply is Markov) if it has the following property: the future random variables of the sequence can be predicted by the value of the current random variable, without reference to any earlier elements of the sequence. Applying this definition to the problem at hand, a sequence of tones is Markov if the probability of a given tone occurring in a given position is conditional only on the previous tone, and not on any other earlier elements of the tune. Markov models are appealing because of their simplicity, and such models can accurately describe many phenomena.

The model in Figure 1, like a model consistent with the theory in [2], treats sequences of tones as being *first order Markov* processes. That is, the model assumes that, conditional on the most recent tone being of a certain type, information about earlier tones in the tune is not helpful in predicting future tones in the sequence. The term first order indicates that the model is only referencing one state; an *nth order Markov* would describe a model that referenced information going back *n* number of states (or in the problem at hand, *n* number of tones).

However, the data in Section 3 make it abundantly clear that tone sequences in spoken American English are not first order Markov chains. To give only one example, Table 1 makes clear that the boundary tone is far more likely to be H% if the nuclear accent is an L+H* than if it is an H*. Simply knowing the phrasal tone is not sufficient to give an accurate prediction of the boundary tone; we need to know the nuclear accent as well. Fortunately we can preserve many of the benefits of Markov modeling while still accurately fitting

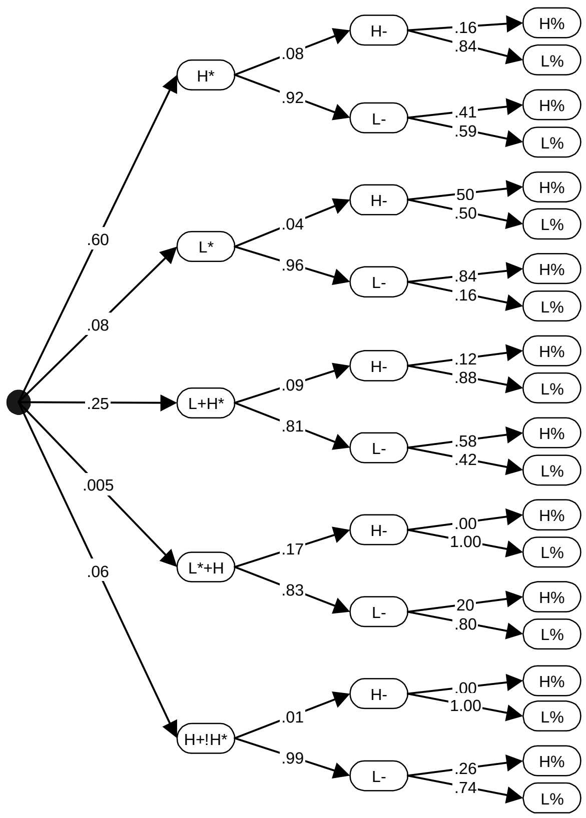


Figure 2: Second order Markov model showing interactions between pitch accents and boundary tones.

the data. The technique is of course to simply create a second order Markov process.

With this modeling approach in hand we can now proceed to use the figures computed from data in Section 3 to flesh out the model in more detail. In Figure 2, I explode these displays in order to illustrate the effects of the second order Markov approach. Note that Figure 2 only represents the model of the nuclear phrase (the last three tones of the intonational phrase). This is equivalent to the right-most section of Figure 1. Each of the twenty possible three-tone tunes that can form a nuclear phrase is represented by an individual branch of the figure. By multiplying the three probabilities associated with each tone in a given tune we obtain the overall probability of that tune, as given in the far right of the figure. For example, each nuclear phrase is 60% likely to begin with H*. Conditional on the H* nuclear accent, the next tone is 92% likely to be L-. Conditional on the nuclear phrase beginning H*L-, the tune is 59% likely to end with L%. By multiplying we get .60 X .92 X .59 = .33. Thus in this model the tune H*L-L% is 33% likely to occur.

Compare this to Figure 1. The H* and L- probabilities are unchanged. (.60 and .92 respectively.) However the probability of L% is .54 because in Figure 1 the model ‘does not know’ that the model began with H* when it is figuring out the probability of L%. Because the first order Markov model in Figure 1 does not take full advantage of the nuclear accent information, it obtains a less accurate probability estimate for this tune. It calculates the likelihood of H*L-L% to be .60 X .92 X .54., or .30, instead of the more accurate likelihood of .33 given in the second order Markov model in Figure 2. The model in Figure 2 gets an answer that exactly matches the data, whereas the model in Figure 1 fits less well.

5. Implications for a theory of meaning

The compositional approach to intonational meaning assumes that “the overall meaning of a tune is built up from the meanings of its smallest meaning-bearing constituents, that is, tonal morphemes” [1]. One example of sequences with clear compositional interpretations is numbers. The sequence 8,432 is composed of the individual meanings 8000 plus 400 plus 30 plus 2. The meaning of the number is exactly equal to the sum of the parts. Looking at the number as a whole does not give us additional information above and beyond the sum of the parts. Because of this strict compositionality one position in the sequence is not related to the other positions in the sequence. Knowing that the first three positions are filled by 843 does not reveal any information about the content of the last position in the sequence. It is just as likely to be 1 or 5 or 8 as it is to be 2.

The data illustrated in Figure 2 constitute a strong rejection of any theory that suggests that tones are chosen independently of the tones that precede them. In some cases the nature of the boundary tone is almost predetermined by other parts of the phrase. The strong interrelations among tones suggest that the meaning of a tune is more than the sum of its tones. Tones are behaving not like numbers but like phonemes in that they combine in predictable patterns. Knowing the identity of certain phonemes in a word allows for more accurate prediction of subsequent phonemes. For example, if a three phoneme English word begins with [ti] it is very likely to end with [m] or [n], as in *team* and *teen*, and less likely to end in [k], as in *teak*. Similarly in sequences of tones certain combinations are more common and others are more rare, as opposed to numbers where all strings are equally likely and meaningful. This argues for an important role for tunes in the intonational structure of English and against the view that the tones that make up intonational phrases are chosen independently of one another, each having its own meaning. My results suggest that a tunal approach might better account for the range of intonational meaning than a compositional approach. By applying the probabilistic approach initiated here we can further our understanding of how intonational meaning is conveyed.

6. References

[1] Bartels, C., 1997. *Towards a compositional interpretation of English statement and question intonation.* Ph.D. dissertation, University of Massachusetts.

[2] Beckman M.; Ayers G., 1994. Guidelines for ToBI labelling, version 2.0, Ohio State University, ms.

[3] Beckman M.; Pierrehumbert, J., 1986. Intonational structure in Japanese and English, *Phonology Yearbook*, 3, 255-310.

[4] Dainora, A., 2001. *An empirically based probabilistic model of intonation in American English.* Ph.D. dissertation, University of Chicago.

[5] Pierrehumbert, J.B., 1980. *The phonology and phonetics of English intonation.* Ph.D. dissertation, MIT.

[6] Pierrehumbert, J.; Hirschberg, J, 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*, Cohen, P.R.; Morgan, J.; Pollack, M.E., eds. Cambridge: MIT Press, 271-311.