A corpus-based study of downstep in English*

Tae-Jin Yoon
(Cheongju University)

Yoon, Tae-Jin. 2013. A corpus-based study of downstep in English. *Studies in Phonetics, Phonology and Morphology* 19.2. 275-294. This paper presents evidence from acoustic analysis for a categorical distinction between downstepped and non-downstepped high-toned pitch accents (H* vs. !H*) in American English. The present study offers an explanation for the contradictory findings from prior acoustic studies of downstep, which call into question the status of the downstepped accent in American English as a legitimate prosodic category. It is shown in the paper that the experimental findings from naturally occurring speech corpus provide evidence for !H* as a distinct prosodic category. **(Cheongju University)**

Keywords: high-toned pitch accent, downstep, acoustic analysis

1. Introduction

The paper presents evidence for a categorical distinction between downstepped and non-downstepped high-toned pitch accents (H* vs. !H*) in American English on the basis of acoustic analysis. In particular, it offers an explanation for the contradictory findings from prior acoustic studies of downstep (Liberman and Pierrehumbert 1984 vs. Dainora 2001) which call into question the status of the downstepped accent in American English as a legitimate prosodic category (see Ladd 2008). In this paper, it is shown that the experimental findings from naturally occurring speech corpus provide evidence for !H* as a distinct prosodic category.

Downstepped high-tone is referred to as high tones that occur in compressed pitch range either triggered by a bi-tonal pitch accent (Pierrehumbert 1980, Beckman 1986), or a register feature on a high tone (Grice 1995, Beckman 1996, Ladd 1996, Beckman and Elam 1997, Snider 1998, Truckenbrodt 1999, Truckenbrodt 2002). Downstep has been claimed to be a central part of the theory of intonation that provides a crucial argument against multiple levels of tonal representation, and in favor of more restricted two-levels of tonal representation of high (H) and low (L) (Pierrehumbert 1980, Féry 1993, Grice 1995, Grabe 1998, Terken and Hermes 2000, Gussenhoven 2002).

As one of the hallmarks of autosegmental-metrical theory of intonation,

* This work was supported by the research grant of Cheongju University in 2013. Earlier version of the work has been presented at a previous LabPhon conference and ICEL (International Conference on English Linguistics). I thank Jennifer Cole and Mark Hasegawa-Johnson for their advice for constructive comments. Many thanks also go to the editorial chief and three anonymous reviewers for their insightful comments and suggestions. All errors are mine.

downstep has caught attentions of empirical studies cross-linguistically. Detailed empirical studies of downstep and related effects of tonal scaling are reported in Liberman and Pierrehumbert (1984) and in Ladd (1998) for English, Beckman and Pierrehumbert (1986) for Tokyo Japanese, Grabe (1998) for British English and German, Féry (1993) and Truckenbrodt (2002) for German, Snider (1998) for Bimoba (a Gur language spoken in the northern region of Ghana), and Connell and Ladd (1990) as well as Laniran and Clements (2003) for the tone language Yoruba.

In addition to the cross-linguistic studies of phonetic correlates of downstep, the correlation between the presence of downstepped accent and the domain of focus projection has also been discussed in Bartels (1999) for American English and is empirically tested in O'Rourke (2006) for Peruvian Spanish and Baumann et al. (2007) for German.

The accumulated evidence for the stepped-down accent from the previous accent has led to the claim that downstep is present in most languages of the world. For example, Beckman (1993) notes that the work on prosody during the 80s and 90s has let us "say with a fair degree of confidence" that "coherence among words or phrases can be signaled when each following F0 peak is systematically reduced relative to preceding peaks. (p. 259)" Therefore, the existence of downstep which serves to make words or phrases coherent appears to be too hard to deny.

As is acknowledged by Pierrehumbert (2000), however, none of the previous studies have been substantiated by a large-scale study of naturally occurring speech. The earlier studies cited above are based on carefully controlled and/or induced speech materials in a laboratory setting, which may or may not be attested in naturally occurring speech. For example, speakers in a laboratory setting may produce downstepping contours that are scaled in a regular fashion, but that may not be representative of speech in natural setting. In any given experimental situation, subjects confine their behavior to a small subset of their full range of capabilities. Besides, whereas many studies exist that distinguish final lowering effect and downstepped pitch accent, few studies exist that compare downstepped versus non-downstepped pitch accents. Therefore, a full inventory of naturally occurring variation is required for the substantiation of the findings by the earlier studies.

More recently, the categorical status of !H* is called into question by Dainora (2001). Based on the corpus analysis of naturally occurring speech, Dainora investigates the status of !H* by comparing the pitch drop in the tonal sequences (H* !H*) and (H* H*). She shows that !H* does not form a distinct tonal category. Based on her analysis, she claims that there is a single phonological High tone that can be used in the specification of pitch accent melody, and the so-called "downstepped" pitch accents are illusory, being no more than a subset of variants taken from the normal distribution of H* peak values. The failure to substantiate the status of the downstepped accents in American English as a legitimate prosodic category, thus, implies the wholesale question of the central part of the standard theory of intonation of American English in particular, and the theory of intonation in general.

The present paper offers an alternative approach to test whether the downstepped pitch accent possesses legitimate categorical status, and shows that the downstepped pitch accent in American English is indeed a legitimate prosodic category that can be predictable. Different analytic methods are applied to the same set of data from the Boston Radio News corpus (Ostendorf et al. 1995) as the one used by Dainora.

The paper is organized as follows: In section 2, theoretical and empirical studies are presented of the downstep phenomenon. In section 3, after discussing the drawbacks of hypothesis testing of categorical status based on the measurement of pitch drop, I illustrate an alternative approach to the test of downstep with respect to categorical status and predictability. Section 4 concludes the paper with discussion and conclusions.

2. Debate on the existence of !H* in Ameircan English

Pitch downtrend over the course of an utterance has been a research area on pitch, on both tone and non-tonal (or intonational) languages. Downtrend is nonspecific and refers to any phonetic or phonological phenomenon that involves lowering of pitch, including downdrift or automatic downstep, nonautomatic downstep or simply downstep (Snider 1998), final lowering, and declination. Declination refers to the 'gradual modification of the phonetic backdrop of F0 over the course of an utterance' (Connell and Ladd 1990). The term automatic downstep is sometimes used to refer to downstep when the low tone is non-floating and non-automatic downstep to refer to downstep when the low tone is floating (Snider 1998). Automatic downstep is the lowering of F0 in sequences of alternating Hs and Ls (sometimes referred to as downdrift). Non-automatic downstep is F0 lowering in tonal sequences with no obvious conditioning environment (sometimes referred to simply as downstep).

With regard to declination, Pierrehumbert (1980) advanced the hypothesis that much of declination can be accounted for as the result of downstep. Downstep is referred to as the stepwise lowering of pitch (or of the tonal space) at specific pitch accents. Beckman and Pierrehumbert (1986) replace Pierrehumbert's term downstep with the coinage catathesis. But the term downstep now appears to be well established. The neologism was originally motivated by a desire to steer clear of disagreements within the Africanist literature over distinctions like downstep vs. downdrift and automatic vs. non-automatic downstep (Ladd 2008).

Crucial to the characterization of non-automatic downstep is the setting of a new lower height for subsequent tones within the same phonological phrase, i.e., a 'lowering of the ceiling.' (Goldsmith 1976). According to Liberman and Pierrehumbert (1984), a downstepped pitch accent is preceded by a downstep trigger H*+L or L+H*, and it is easily observed when an utterance is made to list items. For example, in an utterance such as "There are blueberry, cranberry, mulberry, and gooseberry." the downstepped pitch contour is clearly seen as a characteristic of list intonation in English. Figure 1 shows the results of the berry experiment that is reported in Liberman and Pierrehumbert (1984) and that demonstrates stepped down pitch accent sequences.

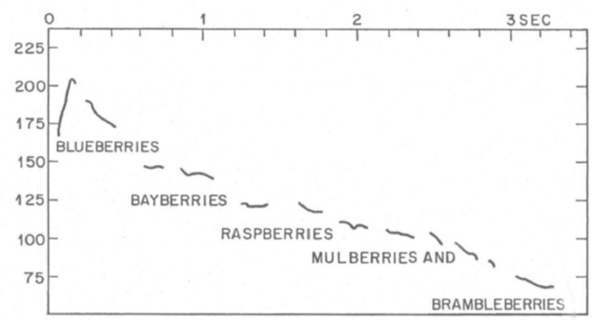


Figure 1. An F0 contour for the berry list *Blueberries, bayberries, raspberries, mulberries, and brambleberries*, produced with a sequence of step accents. Each step is smaller than the one before (taken from Liberman and Pierrehumbert 1984: 171).

Final lowering is a more abrupt lowering restricted to the ends of utterances. Liberman and Pierrehumbert (1984) and Beckman and Pierrehumbert (1986) report that final lowering in English and Japanese respectively, and suggest that it is restricted to the last 250ms or so of an utterance. In Figure 2 is an illustration of a listing-type utterance "The green candies, blue candies and orange candies." spoken by an adult when she spoke to a child. In the figure, the instances of downstepped pitch accent and final lowering are observed. The pitch peak of the second high tone on the word "candies" is relatively lower than the pitch peak of the first high tone on the same word "candies." The pitch peak of the final high tone on, again, the same word "candies" is much lower than the pitch peaks of the preceding two instances of "candies." Thus, if not indicated in the figure, the second and third instances of "candies" are to be marked with "!H*".

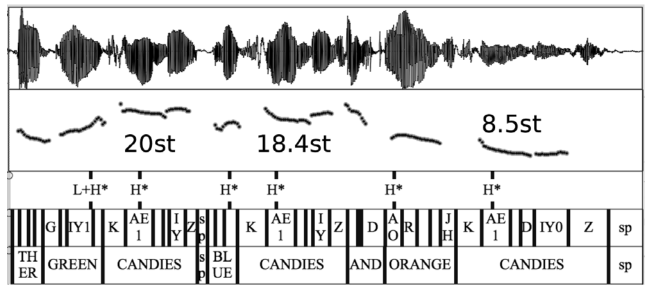


Figure 2. A waveform (in the first tier), a pitch contour (on the second tier), prosodic pitch contour (the third tier), the phone sequence (on the fourth tier), and the word sequence (on the 5th tier). Measurements are taken at the stressed vowel interval at each candies in an utterance “There’s green candies, blue candies, and orange candies.” The 1st [æ] is 313.10Hz(19.76 semitone), the 2nd [æ] is 288.74Hz (18.36st), and the third [æ] is 161.90Hz (8.34st).

A primary question is whether there are any downtrend phenomena such as non-automatic downstep in American English, and if so what is the specific phenomenon? If we observe the configuration as in Figure 1, we need to explain the reason why the height of the second peak is lower than the height of the first peak.

First, the lower height of the second peak than that of the first peak may be due to declination. The term ‘declination’ is widely used to refer to the tendency of fundamental frequency (F0) to gradually decline over the course of an utterance. And it is traditionally viewed to have a physiological basis. The continuous F0 decay observed in many languages is believed to be triggered by an automatic physiological mechanism. Previous studies on the effect of declination include Fujisaki (1983) for Japanese, Thorsen (1980) for Danish, Bruce (1977) for Swedish, Cooper and Sorensen (1981) for English, among others.

Second, the lowering of the second pitch peak may be due to downstep. With this respect, the works by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984) represents a major breakthrough in the study of declination. They discovered the stability of F0 peaks in descending contours under varied prominence conditions. They found that time-dependent lowering, i.e., declination, was almost absent in their data, and that the pitch descent could be modeled by an accent-by-accent decay, i.e., downstep. They proposed the continuous downtrend observed in certain contours in English (i.e., downstepping contours) could be explained as the deliberate use of the step accents by the speaker. The effect of downstep, but not declination, is also reported to be present in Mexican Spanish (Prieto 1998).

An ensuing question remains whether the downstep in American English is automatic or non-automatic. The issue is at most not satisfactorily resolved yet. Automatically downstepped high tones are referred to as high tones that occur in compressed pitch range triggered by a bitonal pitch accent

(Pierrehumbert 1980, Beckman 1996). Pierrehumbert (1980) proposed a downstep rule which was highly analogous to that found in some African languages. Some African languages such as Yoruba (Laniran and Clements 2003) have a downstep rule in which the second H in a H L H configuration is downstepped relative to the first (i.e., H L H → H L !H, where ! stands for downstep). Specifically, Pierrehumbert proposed that the second H in H+L H or H L+H configuration is downstepped (i.e., H+L H → H+L !H, H+L H → H+L !H). According to Beckman (1986), any two-tone accent (or bitonal accents, e.g., H+L and L+H) in English triggers downstep.

Non-automatically downstepped high tones are referred to as high tones that occur as a register feature on a high tone (Grice 1995, Ladd 1996, Snider 1998, Ladd et al. 1999, Truckenbrodt 1999). The alternative to a rule predicting downstep is a phonologically contrastive downstep feature and this is the solution adopted in the ToBI transcription system of the English prosody (Beckman and Elam 1997). The reason is based on some problems that the Beckman and Pierrehumbert approach encounters. In particular, experience to date with ToBI suggests that a L+H accent configuration can be followed either by a downstepped accent, or by one of essentially equal height. Note that according to Beckman and Pierrehumbert, in a sequence of H L+H, the second H will be down-stepped, as in H L+!H. On the other hand, a drawback of non-automatic downstep solution is that a downstep feature is never contrastive in initial position. Downsteps are stepped down in relation to what came before. Thus, it makes no sense to posit a downstep if nothing came before (Ladd 1996).

3. Analyses of downstep

I argue that the study by Dainora (2001) fails to consider the effects of peak height on the pitch measure. Specifically, the F0 peak of the first H* in the sequence might condition the magnitude of the pitch drop to a following peak. In an alternative analysis developed here, I analyze the peak of the second pitch accent (both H* and !H*) in relation to the peak of the preceding H* in the target sequences. I apply this analysis to the same set of data from the Boston Radio News Speech corpus used by Dainora.

3.1 The Boston Radio News corpus

The corpus used for this work was drawn from a subset of recorded FM public radio news broadcasts produced in Boston. The corpus is distributed by the Linguistic Data Consortium, and is the same corpus used in Dainora (2001). The corpus contains recordings spoken by seven radio announcers (Ostendorf et al. 1995). The subset of this radio news corpus, called the 'labnews portion' contains multiple renditions of four news stories, originally written for broadcast but recorded by six professional radio news speakers in a laboratory setting. Note that the speech style is not unscripted spontaneous speech, but is still considered to be naturally-occurring, because professional announcers are trained to read scripts as naturally as possible. The corpus is considered to be naturally occurring reading styles of speech. Example of the script is in (1)

(1) Wanted: Chief Justice of the Massachusetts Supreme Court. In April, the S.J.C.'s current leader Edward Hennessy reaches a mandatory retirement age of seventy, and a successor is expected to be named in March. It may be the most important appointment Governor Michael Dukkakis makes during the remainder of his administration and one of the toughest. As WBUR's Margo Melnicov reports, Hennessy will be a hard act to follow. (taken from the file f1ajrl1.txt)

The recordings are digitized in paragraph-sized units and labeled with the orthographic transcription. While the corpus as a whole contains recordings of seven speakers, not all of the data are prosodically labeled. So, a subset of the corpus was used for the work reported in this paper. The work reported in this paper used speech data from 3 female and 3 male radio announcers that are prosodically transcribed. The speech is labeled with prosodic markers using a version of the ToBI transcription system. For example, Figure 3 illustrates the F0 contours of the phrase "*Massachusetts may now…*" produced by a female speaker in the radio speech corpus.

Each announcer read almost the same script which consisted of about 114 sentences, whose average word count is 18. The sound files are accompanied by a word transcription that is time-aligned to the speech signal. Based on listening and visual inspection, I manually corrected sporadic regions of misalignment. The total number of sentences used for the experiment is 583, the number of word tokens is 10,548, and the duration of the speech corpus is about an hour (or 60 minutes).

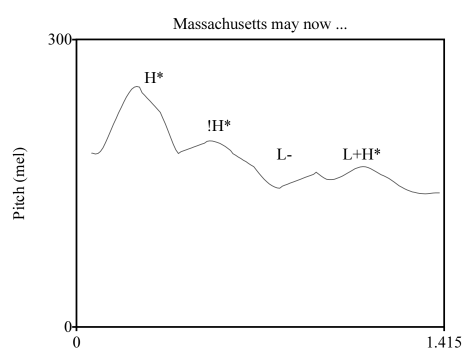


Figure 3. An F0 contour of the phrase "*Massachusetts may now…*" rendered by a female speaker. L+H* is lower in its pitch than that of the preceding !H* due to the effect of pitch reset induced by the intermediate boundary L-.

3.2 Pitch drop analysis

Dainora (2001) discusses the methodologies and data used in previous studies on downstep in English. She sums up the previous studies by stating that the authors of some previous papers aim to establish the existence of downstep or to document its phonetic implementation, while the authors of other papers assume that downstep exists and work on how to incorporate it into a theoretical phonological model. Since her concern is the empirical foundation of downstep in English, she made an extensive comment on a paper by Liberman and Pierrehumbert (1984), which is about a phonetic analysis of downstep in English. She raises an objection to the view of phonological encoding of downstepped pitch accent, stating that “speakers in a laboratory setting may produce downstepping contours that are scaled in a regular fashion, but that may not be representative of speech in natural setting.” (p. 46). The reasons for her concern include: (1) sample utterances have clear meaning differences or that occur in a long list of items. (2) to elicit multiple degrees of pitch range, the researchers instructed the speakers to speak with varying degrees of overall emphasis of excitement, and (3) two of the four speakers in the experiment were the authors themselves, who were aware of the nature of the research.

In order to provide an alternative methodology to the empirical study of downstep, Dainora adopted a corpus study of downstep. The corpus study of downstep has several advantages such as (1) the sheer number of data points for analysis, (2) the acquisition of data from speakers who are not aware of the nature of research, and (3) the availability of variety of sentences found in languages than the specialized examples found in the experimental literature.

She applies a pitch drop analysis to the Boston University Radio Speech Corpus. She argues that if downstepped high-toned pitch accent (e.g., !H*) belongs to a distinct category from non-downstepped high-toned pitch accent (e.g., H*), then we expect to see a bimodal distribution of F0 values, as in Figure 4. In the figure, E_A is the frequency of observing values of a category A, and E_B is the frequency of observing values of a category B. For example, if the x- axis represents the values of pitch drop, E_A will correspond to the pitch drop between H* and H* and E_B will correspond to the pitch drop between H* and !H*.

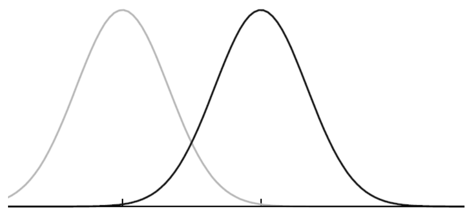


Figure 4. Hypothetical bimodal distribution. If downstepped high pitch accents belong to a different category from non-downstepped high pitch accents, then it is expected that a bimodal distribution will be revealed out of F0 values, such that the expected value of H* (E_{H*}) is higher than the expected value of !H*($E_{!H*}$)

Instead of the expected bimodal distribution, as in Figure 4, Dainora obtained a unimodal distribution through the pitch drop analysis, as illustrated in Figure 5. That is, the pitch drop measure defines a unimodal distribution, where H* and !H* belong to opposite ends of a single distribution. According to her, while some high-toned pitch accents exhibit pitch drops when they occur after preceding high-toned pitch accents, these downstepped tones do not form a distinct category. What she asserts is that there exists a possibility that the phenomenon of downstepping might be just a perceptual illusion that only exists in a laboratory setting or a very carefully produced speech. Thus, the difference between downstepped and non-downstepped tones is a “superficial one that does not belong in a model of intonation in English.” (p. 41).

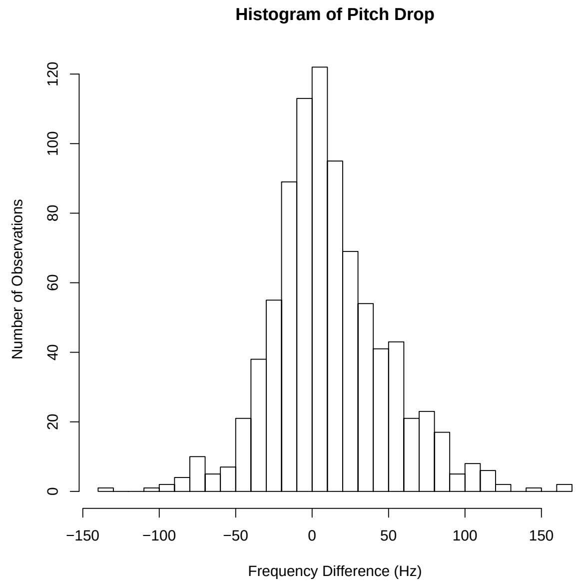


Figure 5. The pitch drop measure defines a unimodal distribution, where H* and !H* belong to opposite ends of a single distribution in the F0 dimension. The figure is obtained through the same measurement as taken by Dainora and the result replicates Dainora's result.

3.3 Linear regression analysis

The reason why the pitch drop analysis shows not bimodal distribution but uni-modal distribution may be due to the failure of considering the magnitude of the first pitch peak. That is, the F0 peak of the first H* in the sequence might condition the magnitude of the pitch drop to a following peak. The peak of the second pitch accent (both H* and !H*) needs to be taken into consideration in relation to the peak of the preceding H* in the target sequences. A regression analysis method is a way of modeling such an effect and helps us to see that H* and !H* form two distinct distributions when the F0 peak is plotted against the peak height of a preceding H*. In regression analysis for peaks in the sequence H* H* in the Boston Radio News corpus, the slope and intercept is 1.0 and 15.93, respectively (Y=1.0X + 15.93). For peaks in the sequence H* !H* in the Boston Radio News Corpus, the slope and intercept is 0.5 and 63.95, respectively (Y=0.5X+63.95) (see Figure 6). In the figure, density plots are also shown to illustrate two clusters.

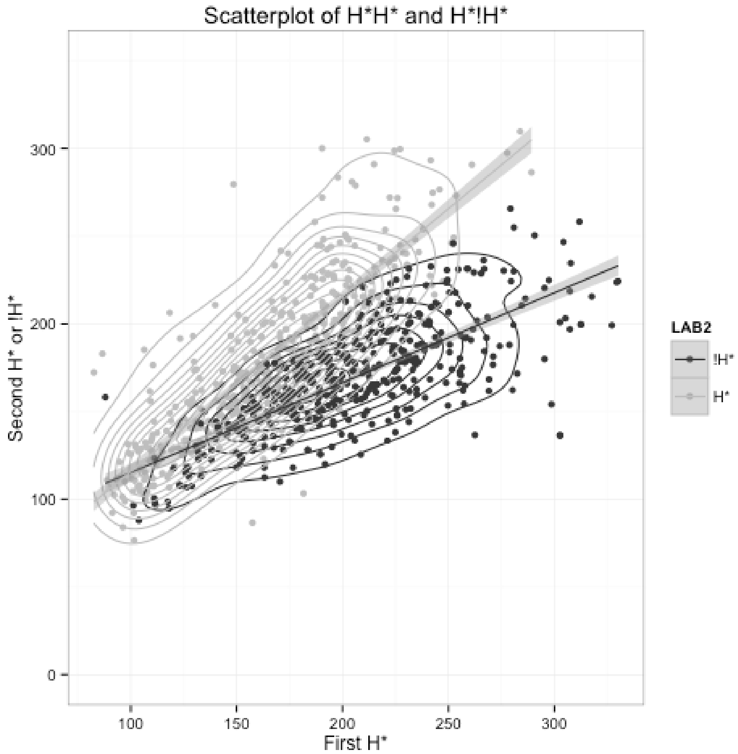


Figure 6. Scatterplot with overlaid density contours and regression lines of H* H* versus H* !H* sequences in the labnews portion of the Boston Radio News Speech corpus (BRNSC). As for H* H* sequence, the slope and intercept of linear regression is 1.0 and 15.93, respectively (i.e., Y = 1.0X + 15.93, where Y is the pitch value of the second H*, and X is the pitch value of the first H*). As for H* !H* sequence, the slope and intercept of linear regression is 0.5 and 63.95, respectively (i.e., Y = 0.5X + 63.95, where Y is the pitch value of !H*, and X is the pitch value of H*).

It is noted that there are cases in which the pitch values of the second H* is higher than those of the preceding H*. This may be due to the effects of nucleus accent or focal accent. These seemingly upstep-like phenomenon is not considered in this paper, because it is different from downstep in that downstep can occur repeatedly, whereas it is not the case for the higher second pitch peak than the first pitch peak.

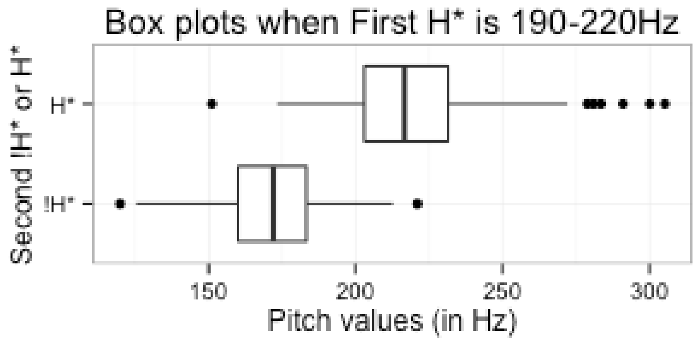
The regression analysis is limited in providing evidence for bimodality. It is easier to show bimodal distribution using numerical values from one dimension. One way of doing so is partitioning one dimension into a number of bins and treating those values in the partitioned bins as stationary. In statistics and machine learning, this process of partitioning continuous variables to nominal or discretized intervals is referred to as discretization. In classification experiment, this discretization is known to produce better results or better learning models. For more concrete evidence that the two clouds in Figure 6 form a bimodal distribution, the pitch peak values of the conditioning H* on the x-axis are arbitrarily partitioned into 7 bins, as shown in the first column in Table 1. The second and third column in Table 1 show the number of tokens in each bin and the mean and standard deviation of the pitch peak values of the following pitch accents (H* or !H*) in each line.

Table 1. Partitioning of the pitch peak values of the first pitch accent. The second and third columns and the fourth and fifth columns show the number of tokens of the second pitch accent of H* and !H*, respectively, and the mean and standard deviation of the pitch peak values of the second pitch accent. Note that two bins (i.e., 250~280 and > 280) have too few tokens to be valid comparison of the second pitch values between H* H* and H* !H*.

		H * H*		H* !H*
	tokens	Mean (Std)	tokens	Mean (Std)
< 130	73	133.16 (22.83)	12	111.50 (6.84)
130 ~ 160	93	160.11 (22.48)	45	138.82 (11.41)
160 ~ 190	102	189.23 (24.51)	95	155.32 (15.92)
190 ~ 220	87	220.79 (28.10)	110	173.23 (20.25)
220 ~ 250	28	250.59 (26.71)	89	186.35 (22.78)
250 ~ 280	6	267.51 (22.96)	40	199.61 (24.70)
> 280	2	298.02 (16.44)	15	233.87 (19.03)

Whether the distribution of categories is bimodal or not can be visualized using box-plots. In the box plot, the box contains 75% of the distribution, and a bar in the box indicates the median value of a variable. If non-downstepped H* and downstepped !H* in each bin in Table 1 does not form different distributions, but a unimodal distribution (as in Figure 5), then we will observe that the boxes of each category overlap quite significantly with each other, and the median value indicates the box of each category gravitates toward each other. If they form different distributions, then we will observe that the boxes of each category do not overlap significantly.

A box plot consists of a box and whiskers. The box goes from the 25th percentile to the 75th percentile of the data, known as the inter-quartile range (IQR). There's a line within a box indicating the median of the data. The whisker starts from the edge of the box and extend to the furthest data point that is within 1.5 times the IQR. Thus, a box plot is related to a density curve. Two box plots of different categories may reveal bimodal distribution if they have different means and deviations from each other. Figure 7 illustrates how box plots can be used for the visualization of bimodal distribution.



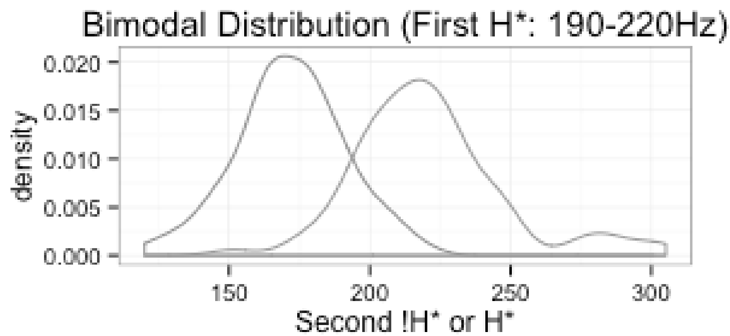


Figure 7. The boxplots (at the top) and density curves (at the bottom) of the pitch values of the second peak (H* or !H*) when the pitch values of the first peak (H*) occur with the range of 190Hz to 220Hz. Note the bimodal distribution of the density curves at the bottom of the graph.

Figures in Figure 8a and Figure 8b show the distribution of the second H* and !H* for each frequency range of the first H*. It shows that the boxes of each category in a given frequency range do not overlap significantly, suggesting that downstepped !H* is categorically different from non-downstepped H*. It is noted in the figures that the two categories diverge more as the frequency values of each increase.

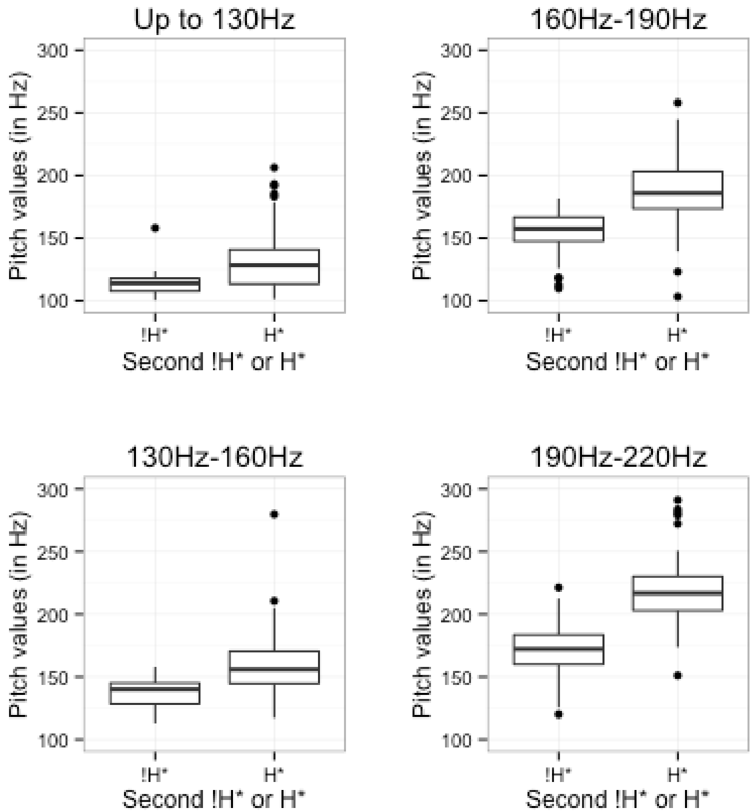


Figure 8a. Box and Whisker plots that illustrate the difference between the pitch peak values of the second H* and !H* in the sequences of (H* H*) and (H* !H*) for each frequency range in Table 1.

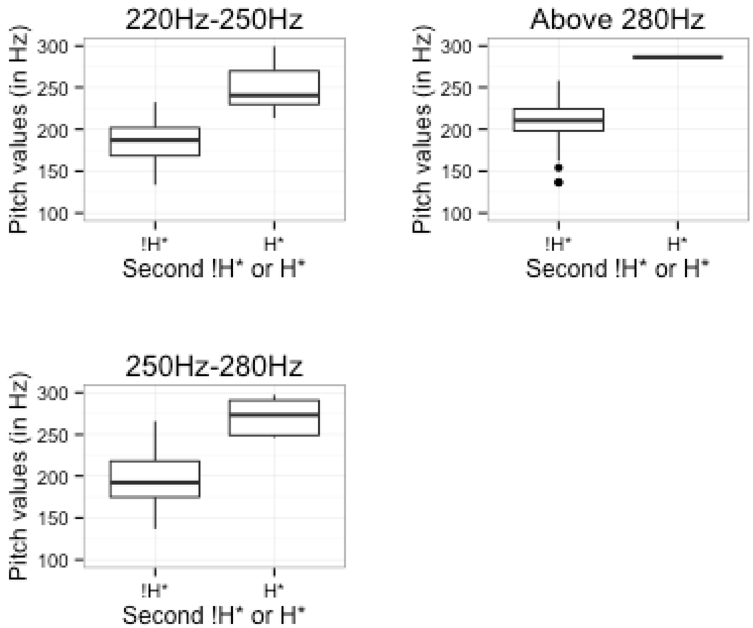


Figure 8b. Box and Whisker plots that illustrate the difference between the pitch peak values of the second H* and !H* in the sequences of (H* H*) and (H* !H*) for each frequency range in Table 1.

Two sample *t*-test can be used to test whether the mean values in each bin is statistically different from each other. Specifically, Welch two sample t-test is used to test a null hypothesis that the distribution of !H* does not differ from that of H* in each bin. The Welch t-test, which is a variant of t-test, is designed to test the null hypothesis even when sample sizes are unequal and variances are heterogeneous (Glass and Hopkins 1996). As shown in Table 2, the pitch peak values of the !H* are significantly different from those of the H* in each bin, leading to the conclusion that the null hypothesis cannot hold.

Table 2 Welch two sample t-test. The results show that the pitch peak values of the !H* is significantly different from those of the H* in each bin.

	Frequency range of the first H*	t	df	p-value
1	< 130	-6.51	58.29	< .001
2	130 ~ 160	-7.37	135.64	< .001
3	160 ~ 190	-11.58	174.73	< .001
4	190 ~ 220	-13.28	151.08	< .001
5	220 ~ 250	-11.47	40.12	< .001
6	250 ~ 280	-6.68	6.86	< .001
7	> 280	-5.08	1.38	< .05

4. Discussion and conclusions

The paper revisited the downstep in American English using the same data as in Dainora, but different analytic approaches. It is shown through a linear analysis and a classification experiment that the downstepped pitch accent (!H*) indeed constitutes a category different from non-downstepped, or normal high-toned pitch accent (H*).

Dainora has used the results of statistical tests to argue that difference between downstepped and non-downstepped high tones is a superficial one that does not belong to a model of intonation in English. She argues based on the data that “speakers produce tones that fluctuate around a given target point.”(p. 46) and indicates “the amount that the frequency of a tone falls between the initial tone and the following tone is a random amount.” She attributes our belief in downstepped high tone in English to the so-called ‘hot hand’ phenomenon (p. 46).

According to the ‘hot-hand’ phenomenon (Gilovich et al. 1985), people often believe that a basketball player who has made several shots in a row is more likely to make his or her next shot than a basketball player who is equally able but who has missed his or her last few shots. Basketball fans, players, and coaches who have observed thousands of games insist that the hot hand is real, even though evidence for hot hand phenomenon has yet to be provided.

A concern may still arise as to the scatterplot in Figure 6. In the figure, at a given point in the first H on the X-axis, it appears that the peak pitch values of the second pitch accent form a continuum. Given the seemingly continuous distribution, one might speculate that the transcriber simply chose the downstep label whenever the value of the second pitch peak value is lower than that of the first pitch peak value. In order to judge this predictability, independent perceptual evidence is needed that demonstrates that downstepped pitch accents are perceptually different from non-downstepped pitch accents.

Baumann et al. (2007) shows clearly that, at least in German, downstepped accents are in some sense pragmatically intermediate between non-downstepped H* accent and absence of accent. Specifically, downstepped accents are often used on expressions that refer to entities that are inferable or otherwise not completely new to discourse, whereas non-downstepped accents are more often used with expressions referring to completely new entities, and deaccenting tends to be reserved for entities that have been mentioned in the immediately preceding context (taken from Ladd 2008: 28). In a similar vein, we can find the distinction between downstepped and non-downstepped accents in English through the investigation of short utterances which have only two accents. Ladd (2008) provides a phrase like *my mother’s diaries* as an example of the short two-accented utterance (p. 77). The phrase *mother’s diaries* can have two possible prominence patterns: the first pattern is strong-weak (as in *my MOTHER’S diaries*) and the second pattern is weak-strong (as in *my mother's DIARIES*). In the strong-weak pattern, if the speaker wants to refer to my mother's diaries, not my father's, the speaker can put the focus on *mother's*.

```
         H*
         |
         S         W
(2)  My MOTHER's diaries
     (narrow focus: not my father's diaries)
```

In the weak-strong pattern, the focus is either in *diaries* or on the whole phrase. In the case of my mother's diaries, not her scrapbooks, the focus falls on *diaries*. In this case, the accentual peak on *diaries* is equal to or higher than that on *mother's.* If the speaker wants to single out my mother's diaries but not any of the family furniture, then the whole phrase '*my mother's DIARIES*' serves as a focus domain. In this second case of broad focus domain, we can have a downstepped accent on *diaries.*

```
          H*       H*
          |        |
          W        S
(3)  My mother's DIARIES
     (broad focus: not my mother's keys)

          H*      !H*
          |        |
          W        S
(4)  My mother's DIARIES
     (broad focus: not my father's baseballs)
```

Through these observations of the relationship between the occurrence of downstep and the focus domain, we may infer that the downstep in English may result from a pragmatic function of focus interpretation in addition to a phonological rule that dictates downstepping after a bitonal pitch accent.

To conclude, it has been shown in this paper that the downstepped pitch accent (!H*) indeed constitutes a category different from non-downstepped, or normal high-toned pitch accent (H*). Various unknown factors may influence the speech patterns found in natural speech, obscuring the comparison with speech obtained in a laboratory setting. Statistical methods can in some cases be applied to compensate for uncontrolled factors. The experimental findings from naturally occurring speech corpora provide evidence for !H* as a distinctive prosodic category, contrary to Dainora (2001) and in support of the findings of Liberman and Pierrehumbert (1984). The current study also resolves the uncomfortable state of affairs as expressed in Baumann et al. (2007). They acknowledged Dainora's finding by stating that it has been argued for English that "downstep cannot be reliably transcribed, implying that H* and !H* should not be treated as separate categories," but they had to assert their finding by saying that "a one-way ANOVA with the second peak minus the first peak as independent variable and accent type as factor revealed a highly significant difference between H* and !H* ($p<.001$)." They conclude with their findings that "for English, !H* was perceived as significantly less prominent than L+H* or H*." Given my analysis of the same corpus as in Dainora's previous studies, the uncomfortable state of affairs is now made clear.

REFERENCES

BARTELS, CHRISTINE. 1999. *The Intonation of English Statements and Questions: A Compositional Interpretation*. New York and London: Garland.

BAUMANN, STEFAN, MARTINE GRICE and SUSANNE STEINDAMM. 2007. *Prosodic Marking of Focus Domains - Categorical or Gradient?* Paper presented at the Speech Prosody. Dresden, Germany

BECKMAN, MARY. 1986. *Stress and Non-stress Accent*. Dordrecht, Foris.

____________. 1993. Modeling the production of prosody. In David House and Paul Touati (eds.). *Proceedings of an ESCA Workshop on Prosody*, 258-263. Department of Linguistics and Phonetics: Lund University.

____________. 1996. The parsing of prosody. *Language and Cognitive Processes* 11, 17-67.

BECKMAN, MARY and GAYLE AYERS ELAM. 1997. *Guidelines for ToBI Labelling*, 1-43. The Ohio State University Research Foundation.

BECKMAN, MARY and JANET PIERREHUMBERT. 1986. Intonation structure in English and Japanese. *Phonology Yearbook* 3, 255-310.

BRUCE, GÖSTA. 1977. Swedish word accents in sentence perspective. *Department of Linguistics and Phonetics Working Papers*, University of Lund.

CONNELL, BRUCE and D. ROBERT LADD. 1990. Aspects of pitch realization in Yoruba. *Phonology* 7, 1-30.

COOPER, WILLIAM and JOHN SORENSEN. 1981. *Fundamental frequency in stentence production*. Heidelberg: Springer.

DAINORA, AUDRA. 2001. *An Empirically Based Probabilistic Model of Intonation in American English*. PhD Dissertation. University of Chicago.

FÉRY, CAROLINE. 1993. *German Intonational Patterns*. Tübingen: Niemeyer.

FUJISAKI, HIROSE. 1983. Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing. In Peter F. MacNeilage (ed.). *The Production of Speech*, 39-47. New York Heidelberg and Berlin: Springer-Verlag.

GILOVICH, THOMAS, ROBERT VALLONE and AMOS TVERSKY. 1985. The hot hand in basketball: on the misperception of sequences. *Cognitive Psychology* 17, 295-314.

GLASS, GENE and KENNETH HOPKINS. 1996. *Statistical Methods in Education and Psychology*. Boston: Allyn and Bacon.

GOLDSMITH, JOHN. 1976. *Autosegmental Phonology*. Cambridge, MA, MIT.

GRABE, ESTHER. 1998. Pitch accent realization in English and German. *Journal of Phonetics* 26.2, 129-143.

GRICE, MARTINE. 1995. Leading tones and downstep in English. *Phonology* 12, 183-233.

GUSSENHOVEN, CARLOS. 2002. Phonology of Intonation. *Glot International* 6(9/10), 271-284.

KOTSIANTIS, SOTIRIS and DIMITRIS KANELLOPOULOS. 2006. Discretization Techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering* 32.1, 47-58.

LADD, D. ROBERT. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

______________. 1998. A critical remark on Intonational Phonology: response to Nina Gronnum. *Journal of Phonetics* 26.1, 113-114.

______________. 2008. *Intonational Phonology* (2nd ed). Cambridge: Cambridge University Press.

LADD, D. ROBERT, DAN FAULKNER, HANNEKE FAULKNER and ASTRID SCHEPMAN. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of Acoustical Society of America* 106.3, 1543-1554.

LANIRAN, YETUNDE and GEORGE CLEMENTS. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31.2, 203-250.

LIBERMAN, MARK and JANET PIERREHUMBERT. 1984. Intonational invariance under changes in pitch range and length. In Mark Aronoff and Richard Oehrle (eds.). *Language Sound Structure,* 157-234. Cambridge, Mass., MIT Press.

O'ROURKE, ERIN. 2006. The direction of inflection: Downstrends and uptrends in Peruvian Spanish broad focus declaratives. Paper presented at the second conference on Laboratory Approaches to Spanish Phonology.

OSTENDORF, MARI, PATTI PRICE and STEFANIE SHATTUCK-HUFNAGEL. 1995. The Boston University Radio News Corpus. *The Boston University Technical Report ECS-95-001*.

PIERREHUMBERT, JANET. 1980. *The Phonetics and Phonology of English Intonation*. Cambridge, MA: MIT Press.

__________________. 2000. Tonal elements and their alignment. In MERLE HORNE (ed.). *Prosody: Theory and Experiment*, 11-36. Dordrecht: Kluwer Academic Publishers.

PRIETO, PILAR. 1998. The scaling of the L values in Spanish downstepping

contours. *Journal of Phonetics* 26.3, 261-282.
SNIDER, KEITH. L. 1998. Phonetic realization of downstep in Bimoba. *Phonology* 15.1, 77-101.
TERKEN, JACQUES and DIK HERMES. 2000. The perception of prosodic prominence. In Merle Horne (ed.). *Prosody: Theory and Experiment*, 89-127. The Netherlands: Kluwer Academic Publishers.
THORSEN, NINA. 1980. Intonation patterns and stress group patterns in declarative sentences of varying length in ASC Danish. *University of Copenhagen Annual Report of the Institute of Phonetics* 14, 1-29.
TRUCKENBRODT, HUBERT. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30, 219-256.
TRUCKENBRODT, HUBERT. 2002. Upstep and embedded register levels. *Phonology* 19, 77-120.

Tae-Jin Yoon
Department of English Language and Literature
Cheongju University
298 Daeseongro, Sangdang-gu, Cheongju, Chungbuk
Korea 360-764
e-mail: tyoon@cju.ac.kr

received: June 25, 2013
revised: August 2, 2013
accepted: August 6, 2013