F0 Peak Lagging and Relative Timing in English Intonation

Sung-A Kim(Korea National University of Education)

In this paper, we examine f0 peak lagging phenomenon in English. F0 peak lagging refers to the fact that f0 peak corresponding to an accent is realized beyond the domain of the host syllable. We present experimental data of f0 peak lagging, which shows that f0 peak is heavily delayed when the duration of the accented syllable is relatively short. In addition, we show that f0 peak is also heavily delayed and realized in the following syllable in a focused word, even where the target vowel is not intrinsically short.

1. Introduction

Since Fry's pioneering work on English accent (Fry 1955,1958,1965), it has been observed that f0 peak corresponding to an accent in English is not often realized within an accented syllable itself. This is a partial reason that 'increased syllable duration' has long been considered to be a more direct correlate of accent in English since Fry's studies. Bolinger (1958) and many others, however, have observed that there is a strong association of an accent with some aspects of the fundamental frequency pattern.

Recently, the relationship between accent and aspects of fundamental frequency (f0 henceforth) is beginning to be uncovered by recent phonetic studies on suprasegmentals. They show that the realization of a pitch accent corresponds to f0 peak in the fundamental frequency contour. More crucially, they reveal that f0 peak corresponding to H* tends to lag behind the host syllable. This paper attempts to examine the f0 peak-syllable alignment in English and reveal some characteristics of f0 peak alignment patterns shown in native speakers' data.

The rest of the paper is organized as follows: In section 2, we will have a brief sketch of f0 peak–syllable alignment pattern in English, preceded by some accounts for the production factors inducing such patterns. In section 3, we will present experimental data of the f0 peak-syllable alignment patterns. Concluding remarks and implications will be discussed in section 4.

2. F0 Peak Lagging and Relative Timing in English

Since the early 1970s, it has been known that the mechanism processing f0 might take more time than the mechanism producing segmentals (Jevkin 1976). Fujisaki (1988) clearly states this point. The f0 value changes as a function of time and actually reflects the dynamic behaviors of some components of the laryngeal structure whose characteristics may be considered as second-order linear systems. In other words, f0 is a function of the strain of vocal folds, which in turn is a function of factors including the strain of muscles such as the cricothyroid and sternohyoid, mass of the thyroid cartilage, and stiffness of the cricothyroid joint. Even if the neural commands for producing a pitch target are issued simultaneously with those for producing the syllable that carries it, the pitch target would be attained more slowly than the CV targets, the control of the latter being approximately a first order linear system.

Ohala & Ewan (1973), Ohala (1978), and Sundberg (1979) show that speakers need longer time to produce a rising tone than a falling tone. When subjects are asked to change a pitch by 6 semitones in the shortest amount of time possible, a male speaker would need at least 90 msecs to complete 75% of the change when raising the pitch and 75 msecs when lowering the pitch and thus there was barely enough time for speakers to complete f0 peak within a single syllable.

In general, f0 production studies show that pitch change requires relatively longer time than formant changes because of the laryngeal inertia. Therefore, this physiological constraint imposed on the production of accent is concerned with consequent issues in f0-segment alignment.

A number of studies on the phonetic realization of accent and tone have found that phonetic prominence, specifically pitch prominence, may not align with the onset of the accented or tone-bearing syllable and is delayed (Steele 1986, Silverman and Pierrehumbert 1990 for English; Prieto et al. 1995 for Mexican Spanish; Liberman 1996 for Yoruba; Grimm 1997 for Oneida; Arvaniti et al 1995, 1998 for Modern Greek; Kim 1998a,b for English and Yao; Hayes and Lahiri 1990 preliminary results for Bengali; Hasegawa and Hata 1990 for Japanese; Barteles 1995 for Czech). Some of them are based on impressionistic observation (Grimm 1997). Among those, the following four studies are worth summarizing briefly because they present a more systematic analysis of the location of the f0 peak: Silverman & Pierrehumbert (1990), Pirieto et al (1995), Arvaniti et al (1995), and Kim (1998a, 1998b).

Silverman and Pierrehumbert (1990) examine the alignment of the f0 peak corresponding to the pre-nuclear high (H* following the terminology of Pierrehumbert 1980) tone in English under various conditions. They find the two factors of rhyme duration and upcoming prosodic context are the main sources of peak location variations in English. The f0 peak corresponding to H* is preferentially aligned 'past the end of the high tone-bearing rhyme, into the following unaccented syllable' (1990:87). Given an identical phrase position, there is a positive correlation between vowel duration and f0 peak lagging relative to the vowel onset. In contrast to this, the f0 peak is aligned early in the vowel when the vowel is close to a prosodic edge or when followed by another high tone in English.

Prieto, van Santen and Hirschberg (1995) report a similar result for Mexican Spanish. In addition, they observe that the same finding is obtained when f0 peak delay is measured relative to accented syllable onset as well as vowel/rhyme onset. Unlike Silverman and Pierrehumbert (1990), who propose that the location of the peaks are best expressed as a proportion of the accented syllable's rhyme length, they found that f0 peak can also be expressed by the duration of the syllable.

Arvaniti et al (1995,1998) report very similar f0 peak lagging patterns from a different language, Modern Greek. They also find that Greek pre-nuclear accents show a sharp rise that starts near the onset of the accented syllable, and peak is aligned just after the onset of the first post-accentual vowel (1998:22). Unlike Silverman and Pierrehumbert (1990), they could not find any evidence supporting the effect of the oncoming accent on the early alignment of the f0 peak. Arvaniti et al (1998) look at the f0 alignment separated by different number of unaccented syllables. They find that the number of the post-accentual syllables does not affect the f0 peak alignment, counter-evidencing the effect of the oncoming accent as previously assumed.

With regard to the f0 peak lagging in English, Kim (1998a, b) finds another factor that affects the the f0 peak lagging: intrinsic vowel duration. Unlike Silverman and Pierrehumbert (1990), she uses a high front vowel as a target rhyme and examines the f0 peak alignment patterns in English, in order to see how the f0 peak would align with the syllable when the duration of a host vowel is extremely short as in [i]. In the next section, we will have a close look at the experimental data.

3. Experiment and Results

3.1. The Purpose

We have already seen that pitch changes take longer than segmental changes because of the sluggish movement of laryngeal gestures. Therefore, f0 peak often lags behind a tone/accent-bearing syllable. The question that naturally arises is whether speakers have control over this f0 peak lagging where the rhyme duration is relatively short. Some might ask why the short rhyme or syllable duration matters to the f0 alignment. As mentioned in the previous section, the f0 pitch changes take more time than segmental changes (Ohala 1978). Therefore, the f0 peak alignment in the case where syllable or rhyme duration is considerably shortened provides an interesting issue with regard to f0 peak and segment alignment. When too little time was available to go from the consonant articulation to the vowel target, the articulators moved toward the target, but before reaching it they turned toward to next target. It is quite plausible that speakers undershoot a certain speech gesture where they do not have enough time to articulate it. In the f0 vertical dimension (i.e., f0 values), increase in speech rate induces a reduction in pitch range and in pitch displacements.

The experiment shown in this section was designed to examine whether there would be a similar undershooting in the horizontal dimension (i.e., the f0 peak alignment). Will the speakers rely on 'undershooting' to avoid a much delayed f0 peak aligned with the next vowel when they have little time to reach the target? The present experiment is different from those of Steel (1986) and Silverman and Pierrehumbert (1990) in that it uses real words. In addition, the present study focuses on the variations of f0 alignment when the syllable duration is relatively short. Unfortunately, previous studies including Silverman and Pierrehumbert (1990) use [a], a relatively long vowel, in their study. Thus they do not report any results in which the rhyme duration or syllable duration is under 0.1 sec. If speakers have a control over f0 alignment, they would try to avoid heavy f0 peak delay onto the next vowel. Otherwise, the next vowel will contain an f0 peak that is affiliated with the preceding syllable.

3.2. Stimuli and Speakers

Unlike Silverman and Pierrehumbert (1990), a real word with a short stressed vowel was used in the experiment, as shown in (1).

(1) The Corpus: Say '*minimize*' again.

Three native speakers of English uttered the target words inserted in a carrier sentence. They are in their twenties and thirties, and none of them reported a history of hearing difficulties. Notice that the target word indicated by italics is composed of sonorants except [z] in the word-final syllable in order to minimize segmentally induced perturbation on f0. Furthermore, it should be noted that a high front vowel that is intrinsically short is intentionally selected to be a target vowel.

Speakers were asked to speak the sentence in (1):"*say minimize again.*" To induce a broad range of f0 values and syllable duration, the speakers were asked to vary loudness and speech rate. There were three conditions with respect to loudness as used in Liberman et al. (1993): loud, normal, and soft. Loud speech tends to have a higher and broader pitch range (Liberman and Pierrehumbert 1984). A total of 144 tokens was obtained from each speaker. The utterances were all recorded on a Sony recorder, model MZ-R30, in a speech laboratory at the University of Texas at Austin.

3.3. Measurement Points

The target words are digitized and segmented by using of ESPS/*waves+* signals processing software package (Entropic Inc.). F0 will be measured from tracks obtained using the pitch-tracing facility of *Waves+*, which took measurements of f0 (Hz) every 10 msec over a 49 msec cos window. The following measurements were conducted in the experiment:

(2) Measurement Points:

a. T-syll duration: duration of the stressed syllable,

b. Peak lagging: the distance from the release of word initial [m] at the word initial position to f0 peak,

c. Ons-to-Next vowel: the distance from the release of word initial [m] at the word initial position to the onset of the following vowel.

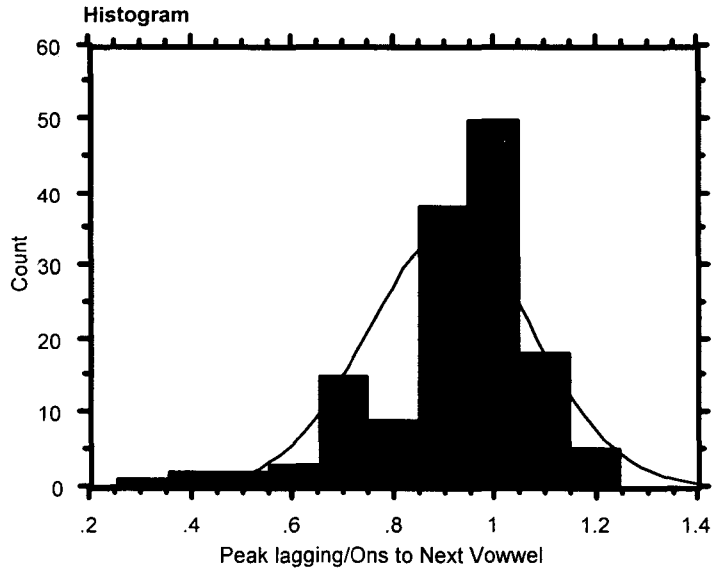
he mean of the duration of the stressed rhyme is tabulated in table 1 below. As the table shows, it is considerably smaller than 0.1 sec across the speakers.

Speaker	Mean	Std. Dev
KMJ	0.065	0.026
AJS	0.052	0.012

JPS	0.056	0.016

<Table 1> Mean Value of the Duration of the Stressed Rhyme

When the rhyme duration is relatively short, the f0 peak tends to be realized in the next syllable and even at the onset of the next rhyme, as shown in figure 1. Figure 1 is a result of the relative peak delay obtained from a male speaker JPS out of the three speakers who participated in the experiment. It provides a histogram of a relative peak delay where the value of peak lagging (or peak delay) is divided by value of the distance from onset of the stressed syllable to the onset of the next vowel, indicated by 'Ons-to-Next vowel.' If the outcome value of the division is higher than 1, it indicates that f0 peak is, in fact, realized beyond the onset of the next vowel.



<Figure 1> Distribution of Peak Lagging Divided by Ons-to Next Vowel

The histogram graphically illustrates the distribution of data. It is an especially good device for identifying the most frequently observed data. The peak at the curved line in figure 1 represents a mean value of the data. Figure 1 clearly shows that the majority of the data are centered around the value of 1. Specifically, values between 0.95 and 1.05 are the most frequently observed ones. This indicates that the f0 peak tends to be aligned with the beginning of the vowel next to the stressed one. The other speakers also show a similar

tendency, although there are minor differences among speakers. Due to space limit, let us look at a speaker's result.

In other words, if the syllable duration is short enough and the time required for pitch change is longer than the syllable duration, then the f0 peak affiliated with the first syllable is spilled over into the following sonorant portion. Although a neural command to reach a certain f0 value for a high tone may be sent at the same time as the command for the production of segments, speakers cannot vary f0 at an arbitrary rate (Fujisaki 1988), considering the short segmental duration. Therefore, an attempt to produce an abrupt change will be 'smoothed out' in a certain way.

Given this physiological constraint on the change of f0, it is obviously misleading to claim that the next vowel becomes accented because of the quite regular f0 peak aligned at the beginning of the vowel. The absolute temporal location of an f0 peak is not informative for the phonological representation of a given segment. Although the f0 peak is aligned with the beginning of the next syllable, it is a result of the presence of suprasegmental element in the preceding syllable.

Summing up, we saw in this section that the f0 peak corresponding to H* in English tends to lag behind the host syllable. When a focus is imposed on the H*, the focus falls on the H* tends to shift to the following syllable. We will sketch out this in the next section.

4. Discussions and Concluding Remarks

In the previous sections, we saw that f0 peak is delayed beyond the scope of the accented syllable. Given this phenomenon, the next question is how relevant it is to the issue of English intonation in general.

Suppose we have the following sentences in (3). The second word '*mom*' is the focused one in (3). We should recall that the duration of the accented syllable is significantly lengthened, when a word is focused.

(3) Does MOM have an examination?

The F0 contour for English speakers shows that the F0 rises gradually after focus, and the high f0 plateau is sustained in an interrogative sentence. The schematic f0 contour is in (4).

(4) Does MOM have an examination?

We can easily notice that the f0 peak of the focused word is realized on the following syllable. The delayed f0 peak is not surprising from the perspective of f0 peak lagging phenomenon. The f0 peak lagging phenomenon presented in the present study may provide an account for why English speakers have such intonation pattern in (4).

<References>

Arvaniti, A.; R. Ladd; and Ineke Mennen. 1995. Tonal alignment and the representation of accentual targets. Proceedings of the 13th International Congress of Phonetic Sciences. Stockholm, Sweden. 220-223.

Barteles, Christine. 1995. Pitch and non-pitch cues to word stress in Czech. Proceedings of the 13th International Congress of Phonetic Sciences. Stockholm,Sweden. 332-335.

Bolinger, D. 1958. A theory of pitch accent in English. Word 7. 199-210.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America 27. 765-768.

_____. 1958. Experiments in the perception of stress. Language and Speech 1. 126-152.

_____. 1965. The dependance of stress judgements on vowel formand structure. Proceedings of the 5th International Congress of Phonetic Sciences.

Fujisaki, H. 1988. A Note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. Vocal Physiology: Voice Production. 347-355.

Grim, C. 1997. The phonetic realization of pitch accent in the Ontario dialect of Oneida. MA project report. SUNY at Buffalo.

Hayes, Bruce and Aditi Lahiri. 1991. Bengali intonational phonology. Natural Language and Linguistic Theory 9. Dordrecht: Kluwer.47-96.

Jevkin, H. R. 1976. Auditory basis of progressive tone spreading. Journal of the Acoustical Society of America 60. S45.

Kim, Sung-A. 1998a. Phonetic assessment of tone spreading. Proceedings of the 24th Berkeley Linguistics Society. 129-138.

_____. 1998b. Tone Spread Decomposed. Proceedings of the Northeastern Linguistics Society. 155-169.

_____. 1978. The production of tone. Victoria Fromkin (ed.) Tone: A linguistic survey. New York, NY: Academic Press. 5-39.

Ohala, J. and W. Ewan. 1973. Speed of pitch change. Journal of the Acoustical Society of America 53. 345.

Pierrehumbert, Janet. 1980. The phonetics and phonology of English intonation. Ph.D. dissertation, MIT.

Prieto, Pilar; Jan van Santen; and Julia Hirschberg. 1995. Tone Alignment patterns in Spanish. Journal of Phonetics 23. 429-451.

Silverman, K. and J. Pierrehumbert. 1990. The timing of prenuclear high accents in English. J. Kingston and M. Beckman (eds.) Papers in Laboratory Phonology 1: Between the Grammar and Physics and Speech. Cambridge, MA: Cambridge University Press. 72-106.

Steele, S. A. 1986. Nuclear accent f0 peak location: effects of vowel, rate, and the number of syllables following. Journal of the Acoustical Society of America. Supplement 1.80. s51.

Sundberg, J. 1973. Data on maximum speed of pitch changes. STL-QPSR 4. 39-47.