Different Roles of Pitch and Duration in Distinguishing Word Stress in English

Jiahong Yuan, Stephen Isard, Mark Liberman

University of Pennsylvania, U.S.A.

jiahong@ling.upenn.edu, stepheni@seas.upenn.edu, myl@cis.upenn.edu

Abstract

Our study investigated the pitch and duration properties of word stress in a large speech corpus. We concluded that pitch and duration play different roles in distinguishing word-level stress classes. In the case of pitch, primary-stress vowels were different from secondary-stress and reduced vowels; in the case of duration, reduced vowels were different from the other two types of vowels.

Index Terms: stress, acoustic correlates, pitch, duration

1. Introduction

Studies on the acoustic correlates of word-level stress have demonstrated contradictory results regarding the relevant importance of the acoustic correlates [1, 2]. Most of the studies are based on small amounts of laboratory speech. In contrast, the acoustics of secondary stress - especially the three-way distinction of primary-stress, secondary-stress, and reduced vowels - has not been widely studied. In this study, we investigated the pitch and duration of vowels from different lexical stress classes in a large corpus. Our goal is to explore the role of pitch and duration in distinguishing different stress classes.

2. Data, Methods, and Results

The SCOTUS corpus includes more than 50 years of oral arguments from the Supreme Court of the United States. We extracted and utilized the "clean" turns (based on the transcripts) of eight Justices from 78 hour-long recordings from 2001. The phone boundaries were automatically determined, using word pronunciations from the CMU pronouncing dictionary and a forced aligner trained on the same data with the HTK toolkit. We tested the same forced aligner on the TIMIT corpus, where the average difference between the forced aligned phone boundaries and the manually labeled phone boundaries in TIMIT was about 15 milliseconds.

Our dataset contained *157,138* primary-stress vowels (labeled '1'), *10,368* secondary-stress vowels (labeled '2'), and *116,229* reduced vowels (labeled '0'). The durations of the phones were calculated from the boundaries in the forced alignment. The F0 was extracted using Praat and converted to a semitone scale. The base frequency used for calculating semitones was Justice dependent, defined as the 10^{th} percentile of all F0 values for that Justice. A simple linear regression was applied to the pitch contour of each turn, and the regression residuals were used for the pitch analysis below. Using the regression residuals instead of the real pitch values normalized the global downtrend of the pitch contours and captured the local pitch behaviors of the vowels.

Figure 1 shows the F0 contours of the vowels. We segmented each vowel into four equal parts, and averaged the pitches within each quarter. The F0 contour of the primary-stress vowels stayed well above zero, which represented the pitch regression line. The contours of secondary-stress and reduced vowels were very similar; both were below the regression line.

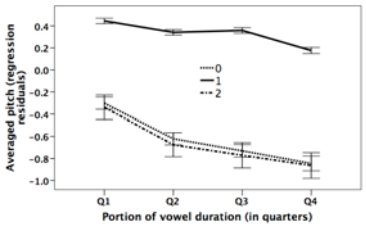

Figure 1. F0 contours of primary-stress ('1'), secondary-stress ('2') and reduced ('0') vowels.

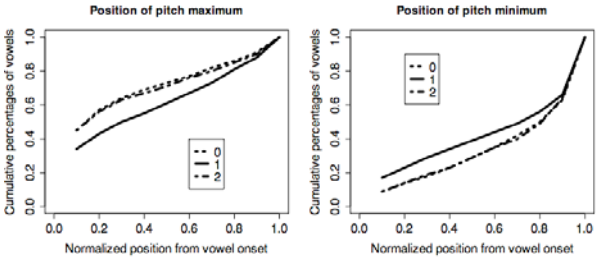

Figure 2. Cumulative percentages of the vowels whose pitch maximum or minimum is within different portions of the vowel.

Figure 2 shows the cumulative percentages of the vowels whose pitch maximum or minimum is within the first 10%, 20%, …, 100% of the vowel. Again, we noted similar patterns between secondary-stress and reduced vowels. Compared to these stress classes, fewer primary-stress vowels had an early pitch maximum and more primary-stress vowels had an early pitch minimum, relative to the regression line.

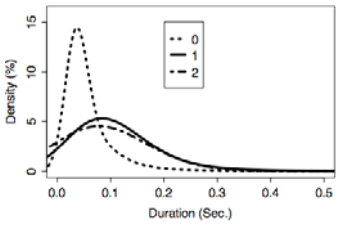

Figure 3. Density curves of duration

The histograms in Figure 3 show the frequency distributions of vowel duration for the three stress classes. Interestingly, the secondary-stress vowels were more similar to the primary-stress vowels in terms of duration. The reduced vowels were much shorter than these two types of vowels.

In conclusion, this study shows that pitch and duration play different roles in the three-way distinction between primary-stress, secondary-stress and reduced vowels.

3. References

[1] Okobi A. O., *Acoustic Correlates of Word Stress in American English*, PhD dissertation, Harvard-MIT HST, 2006.

[2] Wang, C. and Seneff, S., "Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the Jupiter domain", Proc. of EUROSPEECH, 2001.