



Automatic Sentential Vowel Stress Labelling

James L. Hieronymus

Centre for Speech Technology Research, University of Edinburgh.

ABSTRACT.

There is general agreement that sentential syllable vowel stress (called prominence by some authors) in American English is marked by pitch rise-falls, energy, and duration. None of these cues by themselves is sufficient, instead combinations of these cues are used by talkers to signal stress in continuous speech. After studying the stress marking strategies of 15 talkers of American English, an algorithm was devised which labels vowels with three levels of stress. The algorithm is based on combinations of pitch rise falls, relative energy and duration. The pitch is determined automatically in all voiced regions in the sentence. Then the regions are characterised as having rising pitch, falling pitch or steady pitch. Sequences of three regions are examined to find the pitch rise fall patterns which signal stress. The energy in the band 0-2500 Hz is determined throughout the utterance. All the energy measurements are made relative to the maximum energy in the sentence. If the energy of the vowel is within 11 db of the maximum it is considered energy stressed. The duration is determined from hand labels in the present implementation. Duration is corrected for prepausal effects. If two out of three cues are present, then the vowel is labelled stressed. If the vowel has the highest energy, longest duration, and highest pitch then it is labeled as highly stressed. If the vowel has very low energy relative to the loudest sound in the sentence, then it is labeled unstressed no matter what the other two cues indicate. The algorithm was tested on 125 sentences of American English and found to perform very well. The pitch stress was the most difficult. Detailed analysis of the results show that approximately 85 % of the syllables are correctly stress labelled.

INTRODUCTION.

The original goal of this research was to provide a way of automatically locating stressed (or pitch accented) syllables in continuous speech. Once the stressed syllables were located then various properties of these

syllables can be studied. In particular the vowel formant target frequencies were can be studied to determine if the formant targets are more stable for stressed vowels than the unstressed vowels.

In the following we will use "sentential stress" to denote marking of certain syllables within a sentence as receiving the emphasis. This marking is perceived with good accuracy for the most stressed syllable and secondary stress is less accurately perceived. This marking has syntactic significance, in that the stressed words (i.e.. words containing a stressed syllable in continuous speech) are often the subject, verb, important adjective or subject of a noun phrase.

Sentential stress or prominence has been studied by various researchers. Jones states that prominence is the general degree of distinctness of the syllable, this being the combined effect of timbre, length, loudness (which he called stress) and intonation. Fry in 1955 [2] showed that both duration and intensity were correlates of linguistic stress.

The speech of 15 talkers from the DARPA TIMIT acoustic-phonetic database was studied to determine the cues to stress marking. This database consists of read, phonetically balanced sentences. Some talkers used pitch extensively to mark sentential stress, while others seldom used pitch. Immediately it seemed that pitch was going to be the most problematic of the indicators of stress.

METHOD FOR MARKING PITCH ACCENT

The conventional wisdom is that pitch accent is marked by pitch rise-falls. The peak of the pitch is not necessarily the best point to label accented, because the pitch continues to rise somewhat to the middle of the next syllable. Ultimately a finite state system was devised to label pitch accent because of the many different pitch contours seen in practice.

In order to explain the pitch accent marking it is necessary to discuss the types of pitch contours seen in the speech data examined. The pitch was obtained from a trained knowledge based pitch tracker developed by Phillips at CMU [5]. The output from this tracker is

shown in Figure 1.

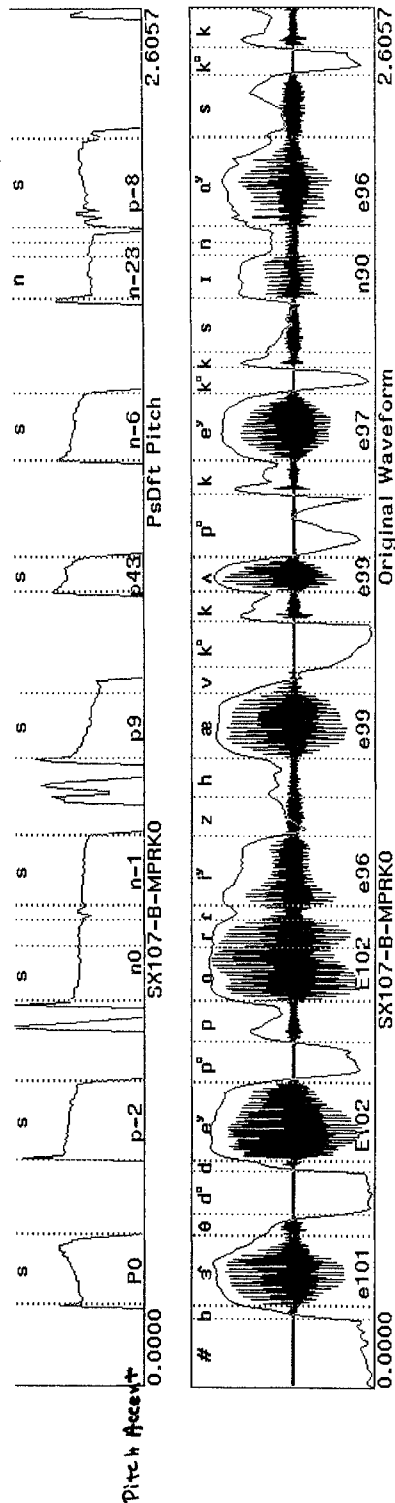


Figure 1: Pitch Track from Sentence sx107

The determination of pitch accents in sentences seems to require an estimate of the pitch and pitch slope in three adjacent voiced regions. Sometimes these regions correspond to one syllabic nucleus and sometimes to several nuclei. For each voiced region the pitch periods at the beginning and end are removed because of the common phenomena of anomalous pitch periods at the beginning and end of voicing. The pitch is gaussian smoothed with doubled or halved pitch corrected. The voiced region is examined for an overall pitch peak, and an average pitch, pitch at the beginning and end and an overall slope is computed. These parameters are input into the finite state machine.

The patterns in the finite state machine are shown in the following diagrams.

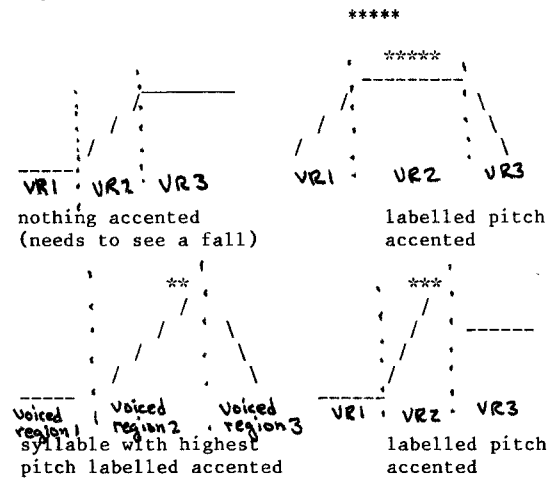


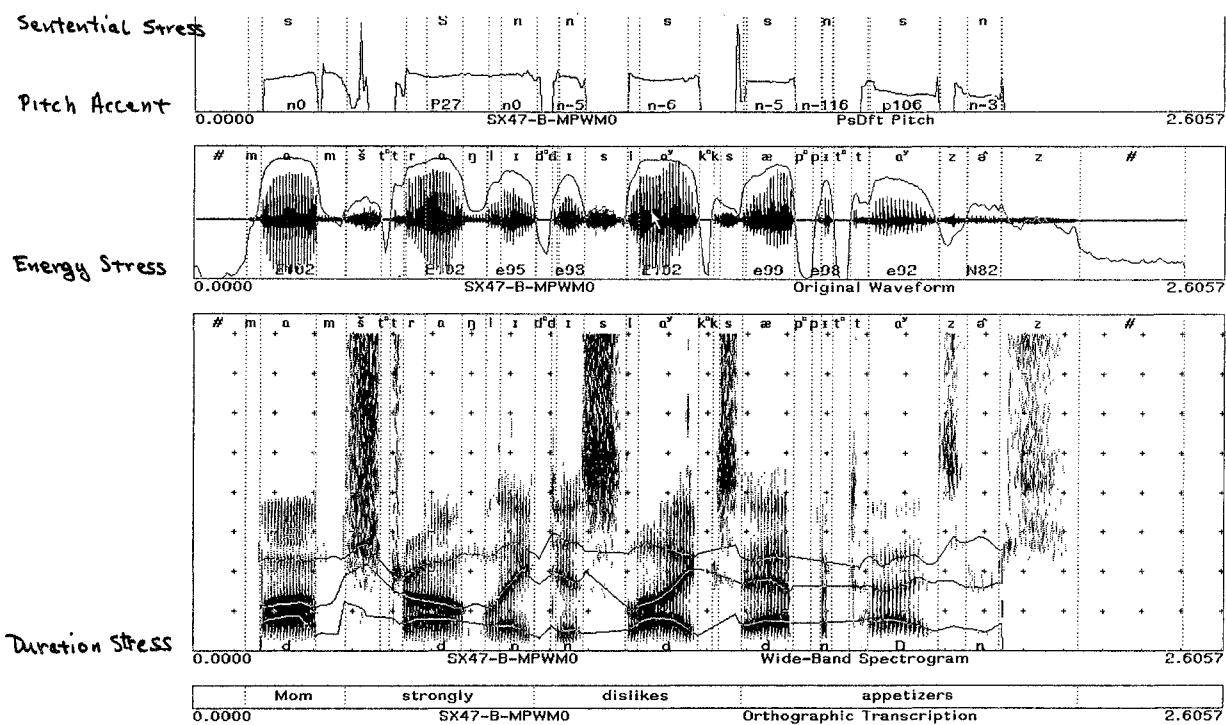
Figure 2: Some Pitch Patterns and Labels

The thresholds which had to be trained were the differences in pitch values at the edges of the voicing regions which would be significant, the amount of pitch rise within a region which would be required for a peak to be significant within the region.

METHOD FOR MARKING ENERGY STRESS

The energy stress is marked based on energy relative to the loudest portion of the sentence. Any vowel region (based on hand labels presently) which has energy within a threshold (11 db from the present training data) of the maximum energy for the sentence is energy stressed. If a vowel has energy greater than 20 db down from the maximum it can never be stressed.

The difficulty with this simple algorithm is that there are often very loud syllables early in the sentence.



Primary Stress = S unstressed = n Secondary Pitch accent = P
 Secondary Stress = S Primary Pitch Accent = P Primary Energy Stress = E
 Primary duration Stress = D Secondary Energy Stress = e Secondary duration Stress = d

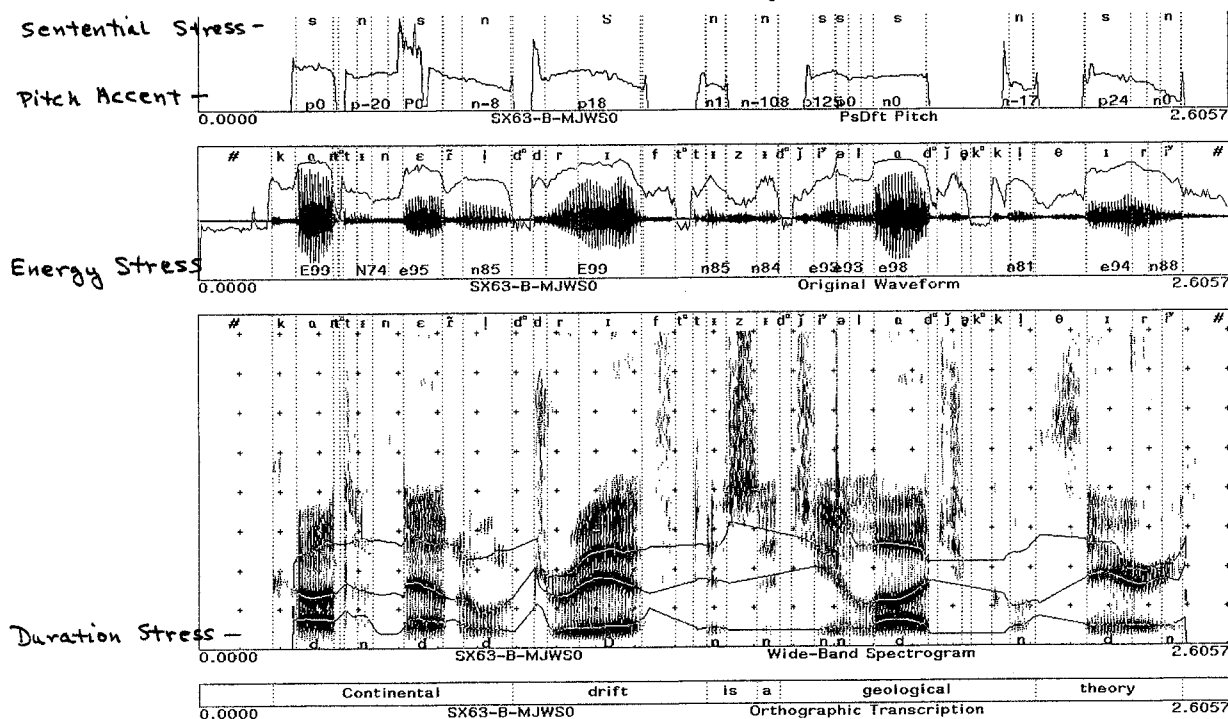


Figure 3: Spectrograms with Sentential Stress Labels

These are so much louder than the other syllables that some stressed vowels are lost due to the 11 db rule. An attempt has been made to remedy this somewhat by taking the average of the two highest energies in the sentence as the maximum for comparison purposes. This works somewhat better.

METHOD FOR DURATION STRESS

Duration is difficult to use as a parameter for stress determination because differing speaking rates and prepausal lengthening of vowels. For the problem of speaking rate, there should be some sort of speaking rate measurement to place the talkers in categories. Many of the speaking rate measures which were examined for this project were unreliable. Therefore a duration histogram technique was used to select the longest duration vowels which exceeded a threshold in absolute duration. The durations were compensated for prepausal lengthening before the histogram technique was applied. The absolute duration threshold was based on histograms of durations of unstressed vowels. Presently the value is 60 msec.

This simple algorithm works rather well. Part of its robustness is due to the fact that relative duration is always used for selecting the longest vowels. A fast talker will have a shifted histogram, which the technique will be able to use to find the longest vowels.

METHOD FOR SENTENTIAL STRESS LABELLING

The assignment of a stress label to a vocalic nucleus is based on a 2 out of 3 voting procedure. Since the study of the speech of several talkers revealed differing stress marking strategies between pitch, energy and duration, this procedure seemed to fit the data best. Presently the votes have equal weight, but some weighting based on training data might improve the performance. If two out of three stress marks are present in a vocalic nucleus, it is labelled stressed. If the nucleus has a sentence maximum energy, duration or pitch it is labelled stressed. If a vocalic nucleus has two out of three maxima, that is highest energy, duration or pitch then it is assigned primary stress. This often gives the right result. However the highest pitch in the sentence is often not indicative of stress, so the importance of this maximum needs to be diminished in the algorithm. Pitch peak alone should not determine a stress marking. This part of the algorithm is undergoing refinement. An example of the stress labelling is shown in Figure 3.

PERFORMANCE.

The algorithm was tested on 125 sentences from the DARPA TIMIT database. Since a perceptually labelled database is not available at CSTR, the test consisted of examining the sentences by hand and listening to the sentences to determine if the stress assignment was reasonable. A good test turned out to be measuring how often schwa and palatalised /ih/ were called stressed. In approximately 1360 vowels in the 125 sentences, 172 reduced vowels were called stressed. Most of these errors seemed to be caused by over emphasising the pitch contribution. The algorithm will be developed further to correct this overemphasis on pitch as a cue to sentential stress.

SUMMARY.

An algorithm has been developed which automatically marks sentential stress in continuous speech. While many of the sentences are correctly marked, the pitch accent assignment in the algorithm needs strengthening.

ACKNOWLEDGEMENT

This work was begun at the U.S. National Institute of Standards and Technology. A discussion with M. Y. Liberman at the beginning of this work was very helpful. Thanks to Michael D. Garris for programming and designing the finite state machine for pitch accent assignment. The spire and search speech research tools developed by Victor Zue's group at MIT were used extensively in this work.

REFERENCES.

- [1] D. Jones, *An Outline of English Phonetics* (Dutton, New York, 1940)
- [2] D. B. Fry, "Duration and Intensity as Physical Correlates of Linguistic Stress," *JASA* Vol.27 (1955), 765 - 768.
- [3] P. Lieberman, "Some correlates of word stress in American English," *JASA* Vol. 32 (1960), 451-454.
- [4] D. R. Ladd, "English Compound Stress." In D. Gibbon & H. Richter (eds), *Intonation, Accent & Rhythm: Studies in Discourse Phonology*, pp. 253-266: 1984.
- [5] M. Phillips, *DARPA Speech Rec. Workshop 1985*.
- [5] J. Pierrehumbert, *The Phonology and Phonetics of*