Non-Native English Speakers' Stress Patterns in Words and Sentences

J. Fokes and Z.S. Bond

Abstract

Non-native and native American English speakers produced tokens of three-syllable words such as confession *and four-syllable words such as* confirmation, *in isolation and in sentences. The acoustic-phonetic correlates of stress — amplitude, fundamental frequency, and duration — were measured for the first two syllables of each word. Both groups of speakers were highly variable. Non-native speakers differed most from native speakers in their control of amplitude and duration.*

Introduction

One of the more common reactions to English as produced by non-native speakers is that the rhythm or timing is inappropriate or incorrect. Even when non-native English speakers have control of the segmentals, their suprasegmental patterns appear to be distinctively non-native in character.

In a previous study investigating the acoustic-phonetic correlates of stress in non-native speech (Fokes, Bond, and Steinberg, 1983), we examined non-native productions of stressed vs. unstressed syllables in words of varying complexity. We found that non-native speakers used appropriate acoustic correlates of stress in relatively simple words such as *confess* and *complete*. When faced with longer words such as *compensation*, however, non-native speakers showed considerably more difficulty, not with the acoustic correlates of stress as such, but with their appropriate distribution. The most salient characteristic of non-native productions was a tendency to over-emphasize unstressed syllables, that is to produce them with distinct vowels and with fundamental frequency, amplitude, and duration values more appropriate to stressed than to unstressed syllables.

Although not directly comparable with our investigation of lexical stress, Adams and Munro (1978) report similar data concerning non-native speakers' production of sentence stress. Adams and Munro measured the fundamental frequency, amplitude, and duration of syllables in continuous text, selecting some which were judged to be stressed by a panel of listeners, and comparing these with identical unstressed syllables. Though the focus of their study was sentence stress rather than lexical stress, Adams and Munro noted that the "real difference between the stress production of the two groups lay not in the mechanisms they used to signal the feature (stress), but rather in their distribution of it" (p. 153), an observation suggestive of our findings concerning lexical stress.

J. Fokes, Ph.D.
School of Hearing and Speech Sciences
Ohio University, Athens, Ohio and

Z.S. Bond, Ph.D.
Department of Linguistics
Ohio University, Athens, Ohio

The purpose of the present study was to extend our understanding of non-native speakers' control of stress patterns by examining their productions of stressed vs. unstressed syllables in relatively long words both in isolation and in sentence context. We selected fundamental frequency, amplitude, and duration as the primary suprasegmental correlates of English stress (Lehiste, 1970).

Specifically, research questions were the following:
1) How do native speakers of English employ the acoustic correlates of stress, amplitude, duration and fundamental frequency, in two- and three-syllable words?
2) Do native speakers use these correlates differently when producing the same words in isolation and in sentence context?
3) How do non-native speakers of English differ from native speakers in their use of the acoustic correlates of stress?
4) Do non-native speakers use the correlates of stress differently from native speakers when producing words in isolation and in context?
5) Do non-native speakers have similar problems with the acoustic correlates of English stress?

Subjects

The subjects were three American graduate students and six non-native speakers of English enrolled in a class in English pronunciation designed primarily for graduate teaching assistants. All the non-native speakers had good knowledge of English, more than adequate to pursue academic work at the university level. Each subject spoke a different language natively: Korean, Persian, Japanese, Spanish, Hausa, and Chinese (from Brunei). All the speakers were male.

The number of different native languages of the subjects was advantageous, enabling us to compare English stress as produced by proficient speakers from quite different backgrounds.

The acoustic correlates of stress in these languages have received varying amounts of investigation. Beckman (1982, 1985) reports that Japanese speakers use changes in fundamental frequency rather than changes in either amplitude or duration to emphasize syllables. Dauer (1983) describes Japanese as a mora-timed rather

than a syllable-timed language. Spanish employs fixed stress, and according to Dauer, is traditionally classified as a syllable-timed language. Although the Chinese spoken in Brunei has, to our knowledge, received no investigation, all varieties of Chinese presumably employ contour tones on syllables. Hausa is a register tone language and contrasts long and short vowels. Korean also uses contrastively long and short vowels (see Maddieson, 1984).

The information about the suprasegmental structure of the native languages of our subjects is inadequate to make any clear predictions about specific difficulties they may encounter in English. Furthermore, there is some dispute about the role of transfer (interference) in the acquisition of the phonology of a second language (see Tarone, 1978). Although testing specific predictions concerning second language phonology would be valuable, we are interested in a slightly different goal: to describe one aspect of second language phonology, stress, as it is used in one language, English. If students of English as a second language experience similar difficulties with stress patterns, then these data may provide a basis for improved instructional procedures.

Materials

The materials were a subset of the test words used in our previous study: *conclusion, completion, confession, competition, confirmation, and combination.*

All the words begin with the same prefix. In the first three, the prefix is reduced and the second syllable receives main stress. In the second three, on the other hand, the prefix receives secondary stress while the second syllable is reduced. Thus, the words allow a comparison of different stress patterns using similar phonology and morphology.

The vowels in the stressed syllables of the three-syllable words are the tense /i, u/, common in the languages of the world, and the lax /ɛ/, occurring much less frequently. The unstressed vowels are reduced. All the syllables receiving secondary stress in the four-syllable words have the same vowel; the unstressed syllables have reduced vowels.

Procedure

The subjects were recorded producing three tokens of each test word in isolation, in three different orders. The subjects also produced the same words three times in a sentence, answering a question. For example, when the test word was *confession,* the subject was asked 'Was the confession accepted?' and responded 'The confession was accepted." All the responses were of the same form, requiring the speaker to use the test word as the subject of a sentence. In this position, the test word did not receive sentence emphasis.

The recordings were processed by a Voice Identification Inc. Pitch Trainer. Duration, amplitude, and fundamental frequency values of the first two syllables of each test word were obtained. The peak fundamental frequency was measured in each syllable of interest; peak amplitude was measured at the same point in the syllable. Syllable duration was measured from the release of the initial stop consonant, visible as 'noisiness' on the monitor screen, to the period of voicelessness or noisiness characterizing the intervocalic consonants /k, p, f/. The ending of the second syllable was determined by an amplitude drop for nasals and noisiness for fricatives. Whenever the segmentation of a particular word was dubious or the values could not be recovered from the monitor screen, the measurements were made from spectrograms.

For the three test words *conclusion, completion, confession,* the measured values allow a comparison between a reduced syllable and a syllable receiving primary stress; for the other three test words, *competition, confirmation, combination,* the comparison is between a syllable receiving secondary stress and a reduced syllable.

Results

It is possible to think of the words and their contexts as graded in complexity. The three words *conclusion, completion, confession,* when produced in isolation, probably represent the simplest speaking task. When serving as the subject of a simple sentence, these words present a little more difficulty. In that the stress pattern of the other three words, *competition, confirmation, combination,* involves secondary stress, these are probably a greater challenge for the speakers in isolation and particularly in sentence contexts.

On the basis of these considerations, we would expect that the non-native speakers' acoustic-phonetic realizations of stress would be most like the American patterns for the first three words in isolation, least like the American patterns for the second three words in sentence context.

Figures 1 to 3 show the relationship between stressed and unstressed syllables for the American speakers and the non-native speakers. Each data point is derived from the average values for one speaker for one word. The axes are ratios of the reduced to the stressed syllable, the y-axis representing fundamental frequency, the x-axis duration. A ratio of 1.0 means that the stressed and unstressed syllable are equal on a particular parameter; a ratio of .5 means that the unstressed syllable has half the value of the stressed syllable. Values greater than 1.0 mean that the unstressed syllable exceeds the stressed syllable on the parameter in question.

Amplitude is represented as the 'tail' on each data point, the longer the 'tail', the greater the amplitude difference. When the 'tail' is above a particular data point, the amplitude relationship is reversed: the reduced syllable is produced with a greater amplitude than the stressed syllable.

The data for the three American English speakers are given in Fig. 1. As can be seen from the figure, there is considerable variability in the fundamental frequency,

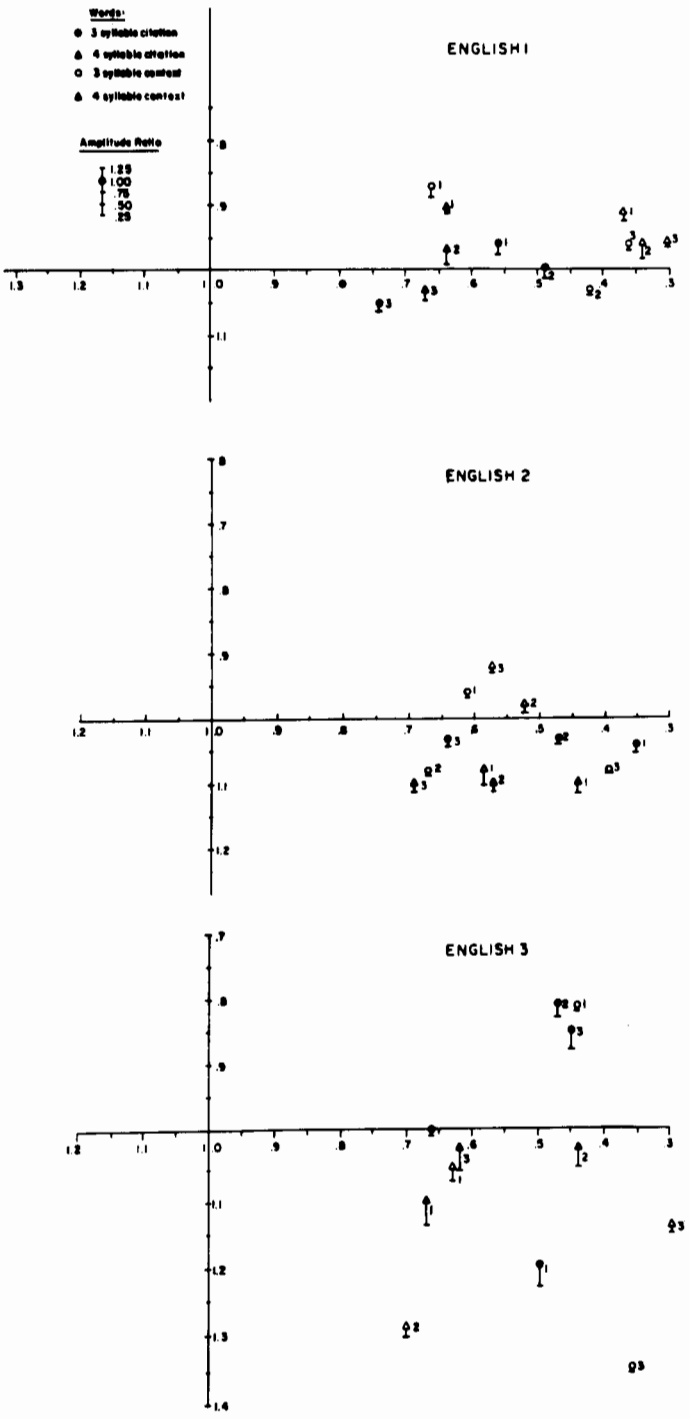

Figure 1.
Frequency, duration, and amplitude ratios of stressed/unstressed syllables in three and four syllable words when spoken in citation and context by three English speakers.

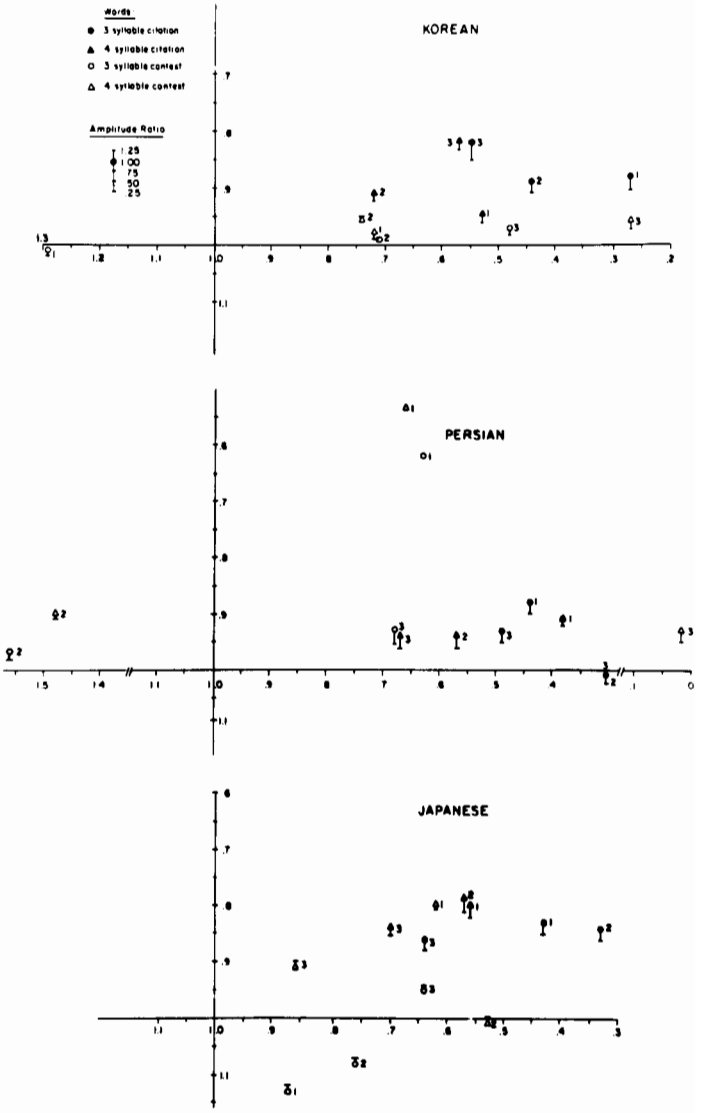

Figure 2.
Frequency, duration, and amplitude ratios of stressed/unstressed in three and four syllable words when spoken in citation and context by Korean, Persian, and Japanese speakers.

duration, and amplitude values which the Americans use to mark stress. For all productions of the three speakers, the duration and amplitude of stressed syllables are greater than of unstressed syllables. The duration ratios range from .30 (*combination* in sentences for two speakers) to .74 (*confession*). The durations of the syllables vary considerably, as would be expected for different speaker styles. Amplitude ratios range from .58 (*completion*) to .97 (*conclusion*). Fundamental frequency ratios, on the other hand, do not always lie in the expected direction. In fact, a majority of the productions do not

show the fundamental frequency changes expected of stressed vs. unstressed syllables. Instead, on occasion all three speakers use a slight FO fall with stressed syllables. Noteworthy is the consistency with which English speaker two pronounces the three syllable words in isolation and in context.

The productions of the Korean speaker are represented in Fig. 2, panel 1. The duration, fundamental frequency and amplitude values he uses in the three-syllable words in isolation correspond well to expected native speaker values. The four-syllable words, however, are more problematic. For the word *competition,* the fundamental frequency and duration values are reversed; for *confirmation,* there is no amplitude difference and a smaller duration difference than for the American speakers. These same words are also problematic in sentence context, in that *confirmation* is produced with the amplitude relationship reversed and both exhibit little fundamental frequency and duration differences. The word *completion* also shows little duration contrast when produced in sentence context, but its fundamental frequency and amplitude relationships are appropriate.

The data for the Persian speaker are given in Fig. 2, panel 2. He has little apparent difficulty producing appropriate values of duration, amplitude and probably fundamental frequency for the three-syllable words, in isolation and in sentence context. The durational relationship for the four-syllable word *confirmation* is reversed in both contexts and for the word *competition,* he has an exaggerated mean fundamental frequency difference, 141 Hz vs. 88 Hz.

The Japanese speakers's data are given in Fig. 2, panel 3. This speaker has some difficulty with amplitude as a correlate of stress in the four-syllable words, both in isolation and in context. He produces an amplituded difference in the expected direction only for *competition* within a sentence; for the other five productions, the unstressed syllable has a slightly higher amplitude than the stressed syllable. Two four-syllable words produced in isolation also show a reversed FO relationship (competition and confirmation) and little durational differences between stressed and unstressed syllables. Problems with duration characterize the word *combination* in sentence context. For this speaker, the lack of contrast in duration results from producing the stressed syllables at shorter durations (approximately 100 msec.) than is typical for Americans.

The data for the Spanish speaker are given in Fig. 3, panel 1. His productions also show difficulty with the four-syllable words. *Confirmation,* in both contexts, shows inadequate durational differences while *combination* has inadequate or reversed amplitude differences. With the exception of amplitude for *confession* in sentences, the remaining words appear to be quite appropriately produced. The durations which this speaker uses for the secondarily-stressed syllables are about 180 msec. while the reduced syllables are about 180 msec., suggesting that his difficulty with duration rests on inadequately

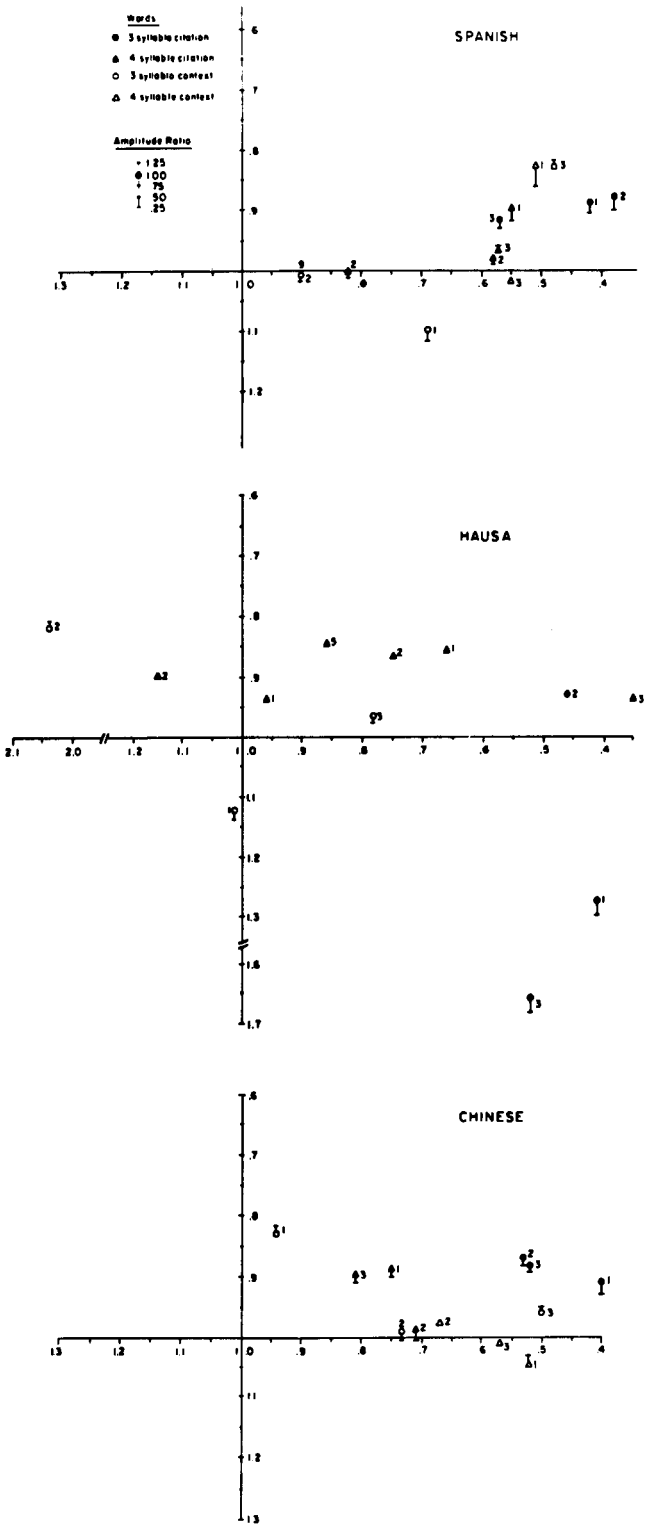

Figure 3.
Frequency, duration, and amplitude ratios of stressed/unstressed syllables in three and four syllable words when spoken in citation and context by Spanish, Hausa, and Chinese speakers.

reducing the unstressed syllables. Equivalent forms of most of these words occur in Spanish; whether this has an effect on the speaker is not clear from these data.

The amplitude of many of the productions of the Hausa speaker (Fig. 3, panel 2) could not be determined reliably. The duration relationships for *confirmation* and the fundamental frequency relationship for *conclusion* are reversed. The other productions employ appropriate FO values but tend towards equal duration for stressed and unstressed syllables, both by shortening stressed syllables and lengthening unstressed syllables.

The Chinese speaker (Fig. 3, panel 3) has some tendency to use inapropriate duration relationships, though less so than some of the other speakers. Whenever the duration relationship approaches 1.0, it does so because the unstressed syllable is excessively long. For most words, the FO values are relatively appropriate. The amplitude differences, however, are typically quite small or even reversed.

In the descriptions of stress patterns of words in isolation and in sentences, we have been concerned with ratios rather than with absolute values for amplitude, fundamental frequency, and duration. A characteristic of the speech of the non-native speakers which is not revealed by ratios is the comparative timing of syllables in isolation and in sentences. For the Americans, there is a very strong tendency to produce words in context at shorter durations than in isolation. The duration differences are most pronounced for syllables receiving main stress, that is for the three-syllable words. The differences in syllables receiving secondary stress and in reduced syllables are similar and much smaller.

The non-native speakers shorten the duration of syllables receiving main stress when they are produced in context, but somewhat less than the Americans do. However, they have a strong tendency to lengthen syllables receiving secondary stress in sentence context. Only the Spanish speaker produces these syllables at a shorter duration in context than in isolation while the remaining five speakers tend to produce longer syllables. Both the Chinese and Japanese speakers produce longer reduced syllables in context than in isolation. The remaining speakers shorten the reduced syllables somewhat, but less so than the Americans. These data are given in Table I.

Table 1. Average difference in msec; between syllables produced in isolation and in sentence context.

	Main Stress	Secondary Stress	Reduced Syllable
American (1)	30	14	18
American (2)	83	17	18
American (3)	65	22	3
Chinese	67	-15[a]	-9
Persian	51	-8	6
Japanese	48	-19	-6
Spanish	50	17	8
Korean	23	-19	11
Hausa	68	-39	6

[a] A negative number means that the syllable in question is longer in context than in isolation.

Table I.
Average differences, in msec., between syllables produced in isolation and in sentence context.

Conclusion

The American speakers consistently use duration as a correlate of stress and they vary amplitude though in rather different proportions. Their use of fundamental frequency is much less consistent. One of the speakers (1) varied fundamental frequency minimally; another (2) tended to raise FO for unstressed syllables; only one of the three speakers (3) produced most of the stressed syllables with a higher FO value than the unstressed syllables. These differences may be attributed to the speakers' style. The first speaker habitually used a monotone; the second speaker had considerable theatrical experience and training. The speaker who showed the expected FO pattern was untrained in public speaking and handled the speaking task unanalytically.

The acoustic correlates of stress as produced by the non-native speakers were variable and not particularly systematic. Though four-syllable words appeared to be more difficult than three-syllable words, no one word was invariably produced appropriately or inappropriately by the speakers. There was a considerable range in the productions for the three-syllable words, suggesting that the identity of the stressed vowel had minimal influence.

On the basis of our data, it appears that the chief difficulty of the non-native speakers was with duration; their stressed or secondarily stressed syllables may be too short. Above all, their reduced syllables tend to be too long. As might be expected, their productions of stress patterns in shorter words are more similar to native speaker patterns than in longer words in which they tend to control syllables with secondary stress and reduced syllables. Their control of stress patterns for words in isolation is better than for the same words in sentence context, though sentence context is not invariably a source of difficulty.

In addition, the non-native speakers tended not to shorten the absolute duration of syllables when words were produced in sentence context, particularly syllables receiving secondary stress. This tendency is also different from the native speaker pattern of shortening all syllables of a word in context.

At this point, we would like to introduce one caution. We have limited our analysis to peak fundamental frequency and amplitude values in stressed and reduced syllables. It is possible that not only peak values but fundamental frequency and amplitude contours contribute to native-like marking of stressed syllables, as suggested by Adams and Munro (1978). Whether this is so, and whether native and non-native speakers differ signifi-

cantly in their use of contours are questions which we have not investigated.

All three American speakers produced perfectly acceptable tokens of the test words yet the acoustic correlates of their productions were quite different. Many non-native speaker productions were not distinctively different from native speaker productions in duration, amplitude, and fundamental frequency, yet they sounded rhythmically inappropriate. Clearly, the rhythmic deficiencies of non-native speakers do not have simple acoustic correlates.

Some recent work of Rekart, Hoffman, and Danilof (1985) suggests an area for further exploration. They asked listeners to judge the 'accentedness' of speech synthesized with various characteristics found in Spanish-accented English. The only acoustic-phonetic property invariably associated with perceived Spanish accent was the formant pattern of vowels.

Perhaps inappropriate vowel quality, for both stressed and reduced vowels, is contributing to the perceived difficulties with stress of the non-native speakers described here. We intend to examine the role of vowel quality in futher work.

References

Adams, C., and R. R. Munro (1978), In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some non-native speakers of English, *Phonetica 35*, 125-156.

Beckman, M.E. (1982), Effects of accent on vowel duration in Japanese, *Journal of the Acoustical Society of America*, 71, S23.

Beckman, M.E. (1985), Perceptual cues to lexical accent contrasts in English and Japanese, *Journal of the Acoustical Society of America*, 78, S20-21.

Dauer, R.M. (1983), Stress-timing and syllable-timing reanalyzed, *Journal of Phonetics 11*, 51-62.

Fokes, J., Z.S. Bond, and M. Steinberg (1983), Patterns of English word stress by native and non-native speakers, in M.P.R. Van den Broecke and A. Cohen (eds.) *Procedings of the Tenth International Congress of Phonetic Sciences*, Foris: Dordrecht, pp. 682-686.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass: MIT Press.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Rekart, D., P. Hoffman, and R. Daniloff (1985), The evaluation of foreign accent using synthetic speech, paper presented at the Acoustical Society of America, Austin, Texas.

Tarone, E. E. (1978), The phonology of interlanguage, in J. C. Richards (ed.) *Understanding Second and Foreign Language Learning*, Rowley, Mass.: Newbury House.