An Acoustical Study of English Word Stress Produced by Americans and Koreans*

Byunggon Yang**

ABSTRACT

Acoustical correlates of stress can be classified as duration, intensity and fundamental frequency. This study examined the acoustical differences in the first two syllables of stressed English words produced by ten American and Korean speakers. The Korean subjects scored very high on the TOEFL. They read at a normal speed a fable from which the acoustical parameters of eight words were analyzed. In order to make the data comparison meaningful, each parameter was collected at 100 dynamic time points proportional to the total duration of the two syllables. Then the ratio of the parameter sum of the first rime to that of the second rime was calculated to determine the relative prominence of the syllables. Results showed that the durations of the first two syllables were almost comparable between the Americans and Koreans. However, statistically significant differences showed up in the diphthong pronunciations and in the words with the second syllable stressed. Also, remarkably high r-squared values were found between pairs of the three acoustical parameters, which suggests that either one or a combination of two or more parameters may account for the prominence of a syllable within a word.

Key words: English Word Stress, Prosody, Second Language Acquisition

1. Introduction

English is usually classified as a stress-timed language whereas Korean is categorized as a syllable-timed one. Thus, each English word has a unique stress pattern so that certain syllables are more prominent than others. A Korean learner will have difficulty making himself easily understood if he produces an English word without stressing certain syllables more strongly than others. This learner will also have problems listening to a native speaker's production. Stressed syllables can be defined in terms of vocal efforts from a speaker's point of view. They may be defined as sounding louder than unstressed ones from a listener's point of view. However, there is no one-to-one correspondence between stress and any single acoustical parameter. Acoustical correlates of stress can be divided into fundamental frequency (f0), duration,

* This research was supported by the Dongeui University Research Grant in 2001.
** English Department, Dongeui University

intensity, formant frequencies and spectral balance (Picket 1987; van Heuven and Sluijter 1996:243). Formant frequencies and spectral balance that should be measured after eliminating the influence of resonance can also be considered as an indicator of accented syllables. We will review the first three parameters to measure them appropriately.

F0 is the repetition rate of the vocal fold vibration. Normally the accented syllable is produced with a faster vibration, thus a higher f0. Since each person has a different mass of vocal cords, the comparison of absolute f0 values will not necessarily lead to a useful conclusion. In other words, some relative comparison or normalization within each word or subject will be necessary. Also, f0 varies according to vowel quality. Other factors being equal, higher vowels have higher f0 values (Lass 1997:233). Therefore, any comparison of syllables with different vowels should be done cautiously. Preceding consonants will have an influence on f0 of the following syllable nuclei (Hombert 1978). For example, the highest peak occurs immediately after a voiceless consonant, but after a voiced consonant f0 rises smoothly. Thus, any measurement of f0 at the beginning or in the middle of a vowel may not represent the whole syllabic segment.

The durational difference among accented or unaccented syllables is another important factor in the production or perception of English stress. Here again, the temporal parameter should be interpreted in terms of a relative ratio rather than an absolute value. The speaking rate will greatly affect the duration of each syllable. Generally, the accented syllable is produced "more elaborately, thus more slowly," than the unaccented syllable (van Heuven and Sluijter 1996:246). Also, the duration of vowels seems to be correlated with tongue height; a high vowel is shorter than a low vowel (Lass 1997). Peterson and Lehiste (1960) reported that the average duration of the syllable nucleus of the word 'bad' was 74 ms longer than that of the word 'bead.' They also found that the average duration of a vowel preceding a voiceless consonant was only two-thirds as long when it preceded a voiced consonant.

The intensity is also an essential factor in the production and perception of stressed syllables. Normally, it is expressed in Watt/cm^2 but we are more interested in the relative intensity expressed in decibels (dB). There is also an intrinsic intensity for each vowel. Lehiste and Peterson (1959) showed that low vowels have higher intensities than high vowels. For example, the average intensity of the vowel /a/ was 8.2 dB higher than that of the vowel /i/. The relative importance of these three parameters in the perception of stress have been listed in the order of duration, intensity and f0 (Lass 1997). Gay (1978) examined the American phonetician's production of vowels and reported that the unstressed vowels were produced shorter or lower in pitch than the stressed counterparts for both rapid and slow speech.

Since the English stress patterns determine the overall impression of foreign accent, we will examine how Korean learners with high English ability handle a long passage.

The research questions of this study are as follows:

1. What is the main acoustical difference in the first two syllables of a stressed word produced by American and Korean speakers?
2. What are the relationships among duration, amplitude and f0?

To answer those questions, we will analyze the acoustical values of English words in a long passage and compare them between ten Korean and American speakers. In order to form a homogeneous group, we chose Korean graduate students with very high English ability.

2. Method

Ten subjects (5 American males and 5 Korean males) participated in this experiment. They were graduate students of the University of Texas at Austin. None of them reported any speaking or hearing problems. The average age and height of the American subjects were 30 years old and 183 cm, whereas those for the Korean subjects were 33 years old and 176 cm. The average TOEFL score of the Korean subjects was 613, which is an extremely high English level, and their average length of stay in America was 26 months. The subjects read an English version of one of Aesop fables, "The North Wind and the Sun", in a sound-treated booth at a normal speed. The recording was made on an Apple G3 notebook computer at a sampling rate of 22 kHz per second. The text was chosen to examine the natural stress patterns of particular stressed words. From the passage, eight words were chosen for acoustical analyses of the first two syllables. They can be divided into the following two categories:

First syllable stressed	Second syllable stressed
closely	confess
making	considered
stronger	disputing
traveller	succeeded

Those words were extracted from the recorded spoken text and stored separately on a hard disk. The acoustical analyses were done by *Praat,* a speech analysis software program. Since each word consisted of more than one syllable, we focused on the first two syllables. We made two kinds of measurements: first, we measured the duration of the first two syllables manually from the waveform and spectrogram as shown in Figure

1; and second, the three acoustical parameters of the first two rimes were collected using an automatic *Praat* script.

We divided each word into two or more syllables by watching its spectrogram and intensity level at the same time. For example, the word 'confess' was divided into syllables as in Figure 1.

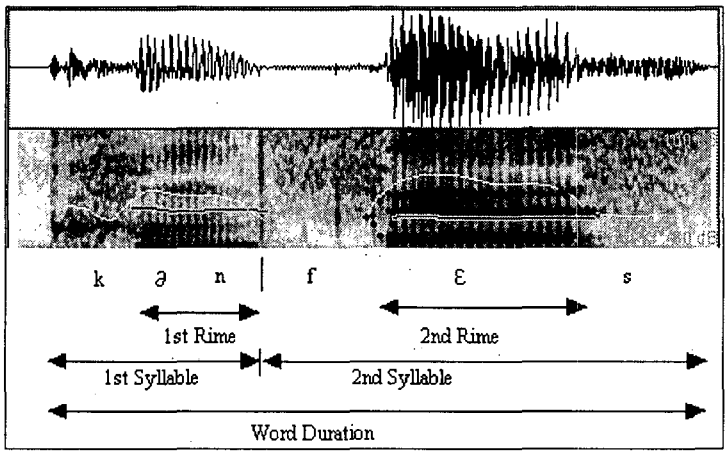

Figure 1. Division of the syllables of the word 'confess'

The waveform was zoomed in on to find the end of the periodic cycle of the schwa vowel and the frication onset of the consonant [f]. Because 'traveller' and 'stronger' were voiced between the vocalic segments, we applied a rule to divide them into two syllables at the point of the minimum intensity level as in Figure 2.

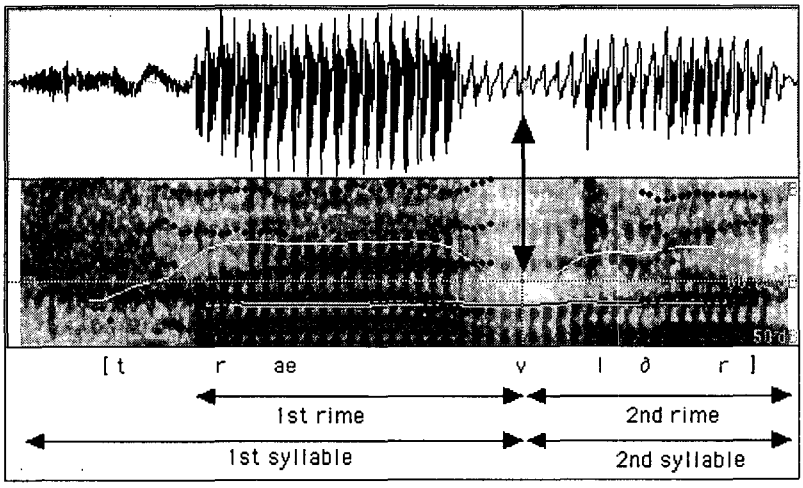

Figure 2. Division of the syllables of the word 'traveller'

The burst line of stops [k, p] in the words ('making' and 'dispute') was set to define each boundary. The word 'succeeded' had also a short silent gap followed by the frication of [s]. The end of the gap was set as the boundary. Secondly, the total duration of the first two syllables was obtained and divided by 100, at which time point the duration, amplitude and f0 were collected. Then the acoustical parameters were trimmed by removing those time points without f0. Rimes with vocalic segments usually had f0 values. F0 analyses were done every 5 ms with an autocorrelation method. Also, those values were discarded by the author if they were not within the possible range of each individual. There were two parts. The first part was set as the first rime and the second part, the second rime. Here again the point of minimum intensity level was set as the syllable boundary.

We assumed that the acoustical parameters of the word varied dynamically proportional to the total duration. Thus, measurements according to the speed of production would be comparable across speakers. Another assumption was that a syllable would be more prominent if the sum of each acoustical parameter within a rime was greater than that of the adjacent one. In other words, the sum of amplitude and f0 within each rime may better represent the energy of the given syllable. The rime length was also collected by finding the difference between its onset and offset points. Then the ratio was calculated to observe the relative prominence of the two syllables involved.

Though the validity of those absolute values collected from the linear scale still required perceptual testing, statistical analyses of the data between the native and non-native speakers were done by StatView+. Generally, objective acoustic scales do not always match subjective perceptual scales. Thus, some significant difference derived from any statistical analyses may not mean significant in the subjective perception.

3. Results and Discussion

3.1 Comparison of the duration of the first two syllables

Table 1 summarizes the average duration of the first two syllables of American and Korean subjects collected manually.

Table 1. Comparison of the average syllable duration. Ame indicates Americans, while Kor does Koreans. Diff means the difference between the two groups (5 subjects in each language group).

Syllable	clo	sly	ma	king	strong	er	tra	veler
Ame	244	307	159	190	412	158	291	214
Kor	259	339	184	188	433	137	286	231
Diff	-15	-32	-24	2	-21	21	5	-17
Ratio	clo/sly		ma/king		strong/er		tra/veler	
Ame	0.79		0.84		2.61		1.36	
Kor	0.76		0.98		3.16		1.24	
Diff	0.03		-0.14		-0.55		0.12	
Syllable	con	fes	con	si	dis	pu	suc	ce
Ame	146	441	122	167	264	161	226	273
Kor	182	457	146	245	260	144	236	253
Diff	-37	-16	-24	-78	4	16	-10	20
Ratio	con/fes		con/si		dis/pu		suc/ce	
Ame	0.33		0.73		1.64		0.83	
Kor	0.4		0.6		1.81		0.93	
Diff	-0.07		0.13		-0.17		-0.10	

From Table 1, one can note that the durational difference is quite minimal. The largest difference occurred in the syllable 'si' (78 ms) and the smallest difference in the syllable 'king.' The average difference of absolute values amounted to 21 ms (s.d.=18 ms). In addition, when we determined the ratio of the first syllable to the second one, the difference came out quite minimal (average ratio=0.16;s.d.=0.18). The largest ratio difference among the eight words occurred in the word 'stronger', where the Korean subjects produced an aspirated [t]. The others were roughly consistent with each other. Thus, one can say that there is not much difference in the production of the eight words in terms of duration of the two groups. These comparable durational results may be because we chose those subjects with a very high English level. If we had included any subjects with a lower level, the results might have been confounded. Next, we will examine the acoustical components of each syllable in detail.

3.2 Comparison of the acoustical ratio of the first two rimes

Table 2 summarizes the total average and standard deviation of the ratio of the first rime to the second rime.

Table 2. The total average and standard deviation of the ratio of the first rime to the second rime. GAve means the grand average of each stress group.

1st Syllable Stressed					2nd Syllable Stressed				
	American		Korean			American		Korean	
word	Ave	s. d.	Ave	s. d.	word	Ave	s. d.	Ave	s. d.
close	0.99	0.33	1.17	0.34	confess	0.94	0.65	0.72	0.21
making	0.78	0.13	1.48	0.52	consi	2.69	0.61	1.4	0.63
stronger	1.90	0.57	2.14	0.51	dispu	0.29	0.11	0.57	0.19
travel	1.31	0.32	1.21	0.42	succe	0.65	0.22	0.53	0.28
GAve	1.25	0.34	1.53	0.54	GAve	1.14	0.40	0.81	0.33

From Table 2, one can observe that there is not much difference in the average ratio of the group with the first syllable stressed except the word 'making'. The Korean group shows almost double the ratio of the American group. This may be because of a difference in producing the diphthong [eɪ]. The Korean group produced it with the duration of almost two different vowels, while the American group produced it as one vowel.

Table 3. Acoustical parameters of the first two rimes of the eight words produced by the Americans and Koreans.

	American						Korean					
	Duration		Amplitude		f0		Duration		Amplitude		f0	
Word	Ave	s.d.	Ave	s.d.	Ave	s.d.	Ave	s.d.	Ave	s.d.	Ave	s.d.
close	0.95	0.33	0.98	0.35	1.04	0.4	1.14	0.39	1.16	0.38	1.22	0.31
making	0.82	0.13	0.8	0.13	0.73	0.13	1.5	0.55	1.5	0.6	1.44	0.52
stronger	1.85	0.61	1.85	0.56	2	0.65	2.14	0.61	2.11	0.56	2.18	0.46
travel	1.3	0.37	1.29	0.28	1.34	0.37	1.15	0.35	1.2	0.39	1.12	0.62
Average	1.23	0.36	1.23	0.33	1.28	0.39	1.48	0.48	1.49	0.48	1.62	0.65
confes	0.84	0.49	0.83	0.51	1.15	0.96	0.7	0.24	0.66	0.22	0.8	0.21
consi	2.9	0.66	2.59	0.53	2.59	0.72	1.54	0.72	1.42	0.69	1.25	0.59
dispu	0.31	0.11	0.32	0.11	0.24	0.1	0.61	0.22	0.57	0.2	0.52	0.18
succe	0.67	0.28	0.62	0.22	0.67	0.2	0.63	0.32	0.5	0.29	0.46	0.26
Average	1.18	0.38	1.09	0.34	1.16	0.5	0.87	0.37	0.79	0.35	0.76	0.31

Since Table 2 reflects the overall picture of the two groups by examining only eight words, we will pursue the relationship in more detail. Table 3 lists average ratio values and standard deviations of the three acoustical parameters of each rime.

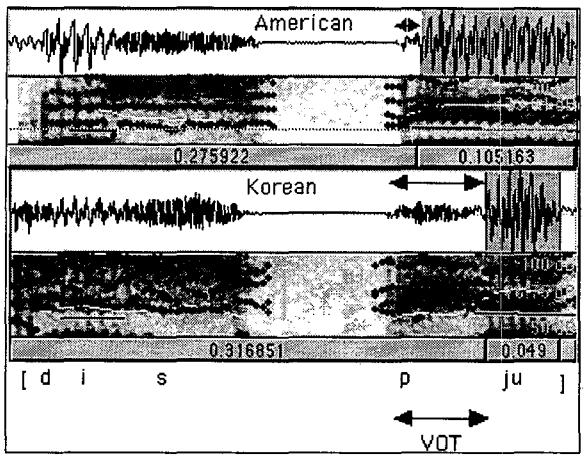

Figure 3. A comparison of Korean and American production of 'dis-pu.'

The acoustical parameters of the words with the first syllable stressed showed little difference in the average ratio except for the duration of the Koreans. In addition, the amplitude and duration of the second syllable stressed were almost half of the duration of the words with the first syllable stressed. The ratio results of the Americans appear quite similar. We will examine what causes these differences between syllables. The greatest difference occurred in the word 'considered'. The Americans produced the word with a first syllable (almost three times the length of the second syllable). However, the Koreans produced it with one and a half the length of the second syllable. This may be related to the sentence stress, which should be pursued in the future. The American subjects produced the second syllable within the sentence structure but the Korean subjects faithfully stressed the second syllable as in a separate word. On the other hand, the word 'dispute' showed the opposite result. The Americans produced it with a longer second syllable and the average ratio of the Koreans was double the ratio of the Americans. This difference seems to be related to the different Voice Onset Time (VOT) of the American and Korean stops. Figure 3 indicates that the Korean subjects produced the aspirated [p] with a longer VOT. Thus, the vowel length of the second syllable might have been reduced.

Figure 4 shows the average and standard deviation of the ratios of the group with the first syllable stressed. In Figure 4, the ratio difference seems quite great in the word 'making.' The other words are quite comparable with respect to the ratio of the first rime to the second rime.

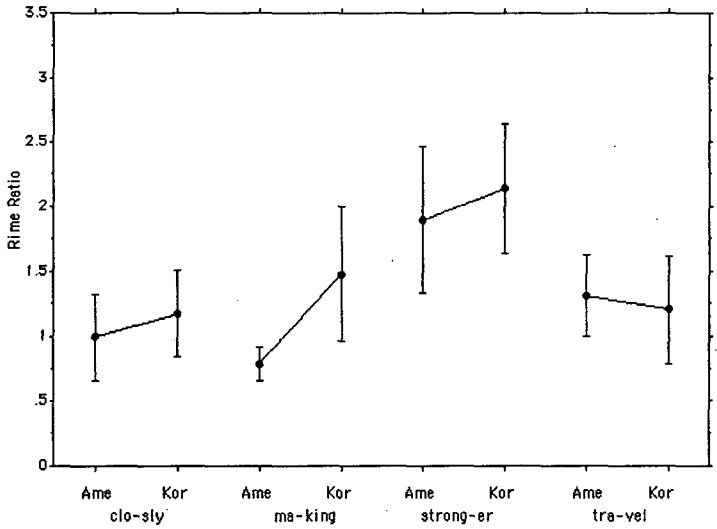

Figure 4. The average and standard deviation of the rime ratios of the group with the first syllable stressed. Ame indicates American whereas Kor does Korean.

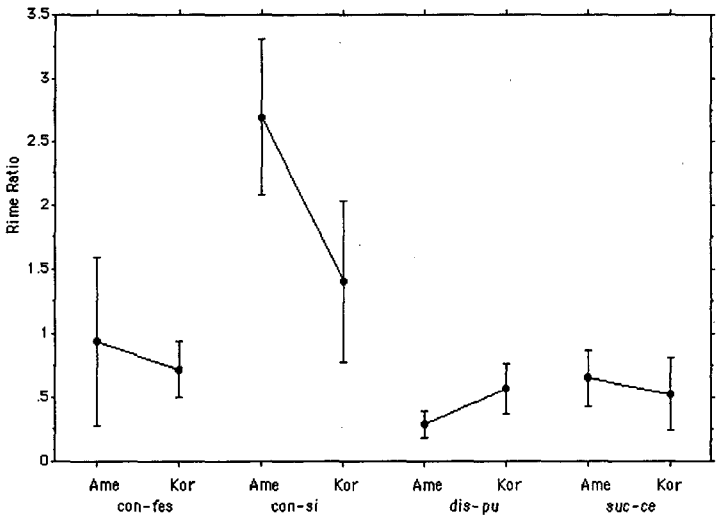

Figure 5. The average and standard deviation of the rime ratios of the group with the second syllable stressed. Ame indicates American whereas Kor does Korean.

Figure 5 shows the average and standard deviation of the ratios of the word group with the second syllable stressed. The ratio of 'con-si' and 'dis-pu' seemed considerably different, while the other two words in the group showed only a slight difference. The direction here seems that the American subjects showed higher ratios, which means that they produced the first syllable longer than the second one. However, the Korean production of the word 'dis-pu' showed a higher ratio than the Americans, which was contrary to the other three words in the group. Are those differences statistically significant? We conducted ANOVA analyses between the three acoustical parameters of

the first and second rimes using StatView+. Results showed significant differences in the words '(m)a-king' (df=29, F=25.4 p=0.0001); 'con-si' (df=29, F=32.2, p=0.0001); 'dis-pu' (df=29, F=0.59, p=0.0001). Individual statistical analyses per each parameter resulted in the same significant differences between the Korean and American groups. The other words turned out to be insignificant.

The second research question concerns the relationships among duration, amplitude and f0. When we plot the three parameters against each other, they appear as in Figure 6. Since these figures denoted lawful relationships, we conducted regressional analyses between the three parameters. Table 4 shows almost a perfect predictability with an average $r^2>0.9$. This suggests that the three acoustical parameters are positively correlated with each other. Thus, either one or a combination of two or more parameters may represent the same syllable prominence.

From the analyses we found that the three acoustical parameters showed strong predictability. Thus, we can claim that the prominence of one syllable over others can be expressed by the three factors altogether, or a combination of two or more parameters. Here, the subjects' English fluency must also be a strong factor in this experiment. If we included subjects with a lower level of English fluency, confounding factors might have led to an ambiguous conclusion.

Table 4. Slopes, intercepts and r-squared values between pairs of the three acoustical parameters by regressional analyses.

Word	comparison	slope	intercept	r^2
clo-sly	duration vs amplitude	1.00	0.03	0.99
ma-king		1.06	-0.08	0.99
strong-er		0.91	0.15	1.00
tra-vel		0.90	0.14	0.95
clo-sly	duration vs f0	0.86	0.23	0.78
ma-king		1.00	-0.07	0.99
strong-er		0.84	0.42	0.87
tra-vel		1.12	-0.07	0.74
clo-sly	amplitude vs f0	0.89	0.18	0.83
ma-king		0.94	0.01	0.98
strong-er		0.93	0.25	0.89
tra-vel		1.27	-0.27	0.82
average		0.98		0.90
con-fess	duration vs f0	1.76	-0.38	0.92
con-si		0.96	-0.21	0.98
dis-pu		0.87	-0.01	0.96
suc-ce		0.77	0.06	0.79
con-fess	duration vs amplitude	1.01	-0.03	0.99
con-si		0.87	0.08	0.99
dis-pu		0.90	0.03	0.99
suc-ce		0.85	0.01	0.89

Word	comparison	slope	intercept	r^2
con-fess	amplitude vs f0	1.74	-0.33	0.94
con-si		1.09	-0.27	0.96
dis-pu		0.96	-0.05	0.97
suc-ce		0.94	0.04	0.93
average		1.00		0.94

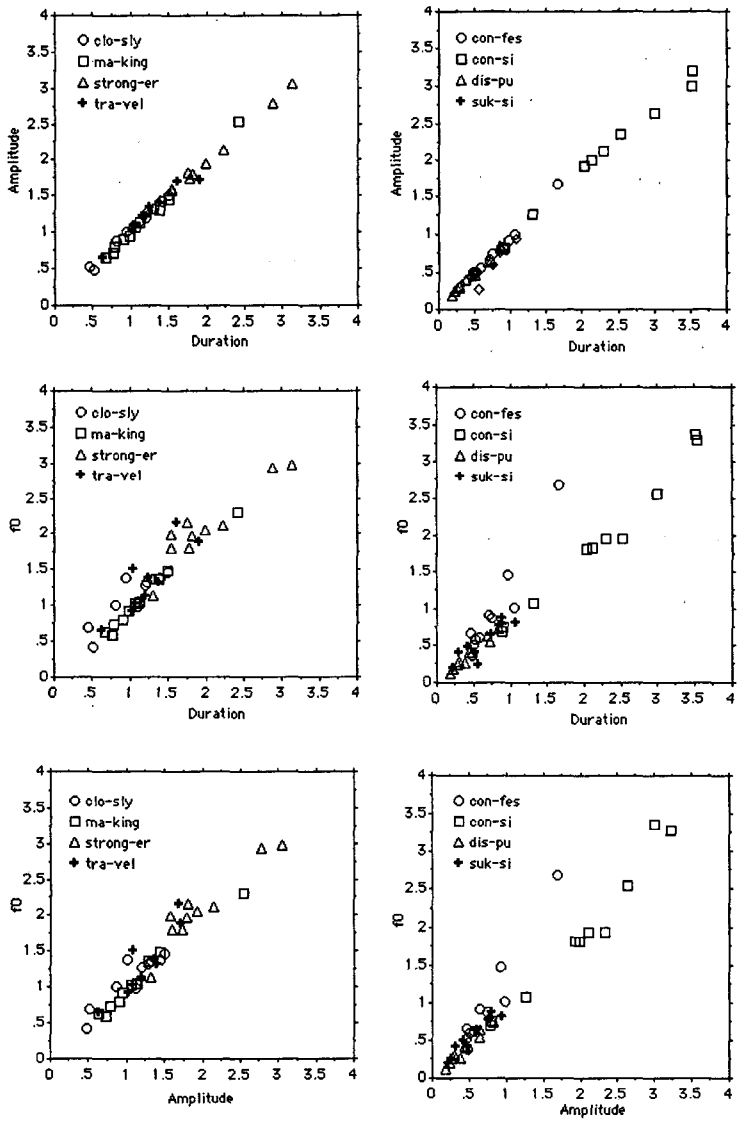

Figure 6. The three acoustical parameters plotted against each other.

4. Conclusion

This study examined the acoustical patterns of English word stress produced by a

group of ten Americans and Koreans. They read an English fable at a normal speed and the acoustical parameters of the first two syllables of eight words were analysed by *Praat*. The five Korean subjects had very high English scores on the TOEFL test and had stayed in America for more than a year. The three acoustical parameters of the first syllables and the second syllables were almost comparable with each other. The subjects' English level might have led to similar temporal management of the English words. However, the components of the parameters of the first rime and second rime came out significantly different with respect to pronouncing the diphthong [eɪ] and in the words with the second syllable stressed. This was attributed to the different methods of consonant and vowel production. The regression analyses among the three acoustical parameters yielded a remarkably high predictability. This suggests that the three parameters can represent the syllable prominence equally well. It will be desirable to study further Korean and American subjects' perception of the syllable stresses of synthesized words and how the Korean students with lower English ability will handle the stress patterns of English words.

References

Gay, T. 1978. "Effect of speaking rate on vowel formant movement." *Journal of the Acoustical Society of America* 63 (1), 223-230.

Hombert, J. 1978. "Consonant types, vowel quality, and tone." In Fromkin, V.A. (ed.), *Tone: A Linguistic Survey.* New York: Academic Press, 77-111.

Lass, N. J. 1996. *Principles of Experimental Phonetics.* New York: Mosby.

Lehiste, I. & G. E. Peterson. 1959. "Vowel amplitude and phonemic stress in American English." *Journal of the Acoustical Society of America* 31, 428-435.

Peterson, G. E. & I. Lehiste. 1960. "Duration of syllable nuclei in English." *Journal of the Acoustical Society of America* 32, 693-703.

Pickett, J. M. 1987. *The Sounds of Speech Communication.* Austin, Texas: Pro-ed.

Van Heuven, V. & A. Sluijter. 1996. *Notes on the phonetics of word prosody.* In Goedemans, R. van der Hulst H. & Visch, E. (eds.) *Stress Patterns of the World. Part 1: Background.* The Hague: Holland Academic Graphics.

Received: Jan. 11, 2002.
Accepted: Mar. 4, 2002.

▲ Byunggon Yang
English Department, Dongeui University
24 Kayadong, Pusanjingu, Pusan, 614-714, Korea
Tel: +82-51-890-1227 Fax: +82-51-890-1209
E-mail: bgyang@hyomin.dongeui.ac.kr