An Experimental Study of Prosodic Determinants of English Naturalness

Yoshiko FUJISAKI

Department of Humanities, Faculty of Humanities & Economics

Abstract: The purpose of this study is to discover a correlation of determinant factors for 'naturalness' judgement of the English language. Phonetic specimens produced by Japanese people are experimentally measured and statistically analyzed on the three major prosodic parameters –duration, pitch and intensity– to reveal that 1) the 'native-like' speakers, as well as the native speakers of the language, showed greater compression effects ; 2) narrow pitch range is one of the factors of 'unnaturalness' ; 3) intensity least contributes to the degree of 'naturalness'. It can be concluded that all judgements of the three prosodic features in natural speech depend on the complicated interaction of a number of factors.

Key word: prosodic features, psychological isochrony, F0

0. Introduction

Linguistic sounds are said to be multi-dimensionally perceived in the phonological system of a language. Exposed to varieties of different sounds, a native speaker of a language is to make an appropriate distinction between meaningful and meaningless sounds in that language. In other words, s/he is able to hear whether each sound uttered should be classified 'natural' to the language or not when the utterance contains any 'foreignness', either qualitative or quantitative, at any phonetic level. What s/he hears is evidently composite realizations of physical functors, some of which are accepted into the phonological system of his/her language as such, while the others are unconsciously ignored and discarded as 'noise'. What are the physical functors that enable a native speaker to discriminate and incorporate the selected natural members into his/her system ? And how do those functors multi-dimensionally correlate with each other ? The purpose of this study is to discover a correlation of determinant factors for 'naturalness' judgement of a given language, so that certain essential and universal 'naturalness' of linguistic sounds may be objectively observed on the phonetic level. Specifically, English sounds, uttered by Japanese people and evaluated by the native speakers of American English, are experimentally measured and statistically analyzed, so as to reveal some relative correlation of prosodic parameters of the language.

1.0 Aspects of a Language

1.1 Segmental versus Suprasegmental

What lyrics and melodies are to music, segments and suprasegments are to spoken languages. These two highest hierarchical levels of speech sounds, together with hierarchically lower phonetic levels of each aspect, are responsible for acceptability as the language. It may well be said that a language cannot exist as a language without mutual reflections and correlations between the two.

The segmental and suprasegmental components are the two comprising units of the hierarchically highest tier directly under 'speech sounds'. The segmental units are the string of vowel and consonant sounds uttered in sequence, forming the words and sentences of the spoken languages, whereas the suprasegmental units are the features of prominence and transition added to the string of vowels and consonants, to facilitate the interpretation of spoken language[1]. In the English language, the patterns of prominent and non-prominent syllables produce a particular rhythmical effect, and this pattern is said to be the backbone of intonation.

A series of precedent studies[2], conducted on determining which of the two would make a greater contribution for a language to be perceived as more native-like, shows that suprasegmental, or namely prosodic features, rather than segmental ones are far more important in judging how native-like the speech is. The suprasegmental features are, according to Lehiste(1970), defined to be features whose arrangement in contrastive patterns in the time dimension are not restricted to single segments defined by their phonetic qualities[3].

1.2 Prosodic Features of English

The prosody of continuous speech can be analyzed and described in terms of the variation of such prosodic features as length, pitch, loudness, tempo, pause, etc. which are perceived features. Length, pitch, and loudness[4] are said to be the most important of the prosodic features of English. Pitch concerns the varying height of the pitch of the voice over one syllable or over a number of successive syllables ; length concerns the relative durations of a number of syllables or the duration of a given syllable in one environment relative to the duration of the same syllable in another environment ; loudness concerns changes of loudness within one syllable or the relative loudness of a number of successive syllables[5]. These three features together are the components of prominence, or stress, in the language.

Precedent experimental studies have led to different definitions of stress in perception. More than half a century ago, Bloomfield(1933) remarks on the importance of loudness by saying that stress is produced when one syllable is uttered louder than the other or others. Fry(1955, 1958), however, finds that pitch is more important than intensity and duration, although both duration and intensity ratios are important for the judgement of stress, and concludes that the higher the pitch, the greater the perception of stress[6].

1.3 Isochrony in English

According to the differences in their rhythmic units, languages have been traditionally classified into three different types : stress-timed, syllable-timed and mora-timed[7]. In the case of stress-timed languages, one of which is English, the interval between strong stresses, generally called a 'foot', is said to have an equal duration regardless of the number of syllables contained in the interval, while syllable- or mora-timed languages, the latter to which Japanese belongs, keep equal duration of syllable or mora. This distinction of timing units puts the two languages, English and Japanese, into fundamentally different rhythmic categories. This rhythmic categorization, based on durational characteristics, presumes the nature of isochrony in the English language, which has long been disputed by a number of linguists[8]. It has been discovered through experimental examinations that each foot in English, which carries a different number of syllables and yet is supposed to keep an equal duration, does not have the same physical length of time at all, but the duration of the intervals is naturally prolonged as the number of syllables contained increases[9].

Lehiste(1970) supports the existence of 'isochrony' in English, although she admits that there is no such thing as 'absolute isochrony'. The basis of her assertion is a phenomenon called 'Compression Effect', which implies a tendency toward shortening the unstressed vowels, and even the stressed vowels, in a foot as the number of syllables contained increases, so that isochrony can be preserved.

What is intriguing is that listeners perceive each foot as isochronous even though the physical durations of actual utterances may be different; that is, there is physical dissonance between productive and perceptive compositions. If this is the case, the conception of isochrony may be something psychological, or rather merely theoretical, which follows that there must be some other norm for a listener to perceive 'isochrony'. In other words, there must exist certain cues within the sound signal by which the listener can discriminate the rhythm. It would be, therefore, naturally assumed that the norms in their senses inevitably, and mutually, affect each other in producing and perceiving languages of different rhythm categories, and some parameters, other than 'time' or 'duration', must be introduced in the process so that the barometer of 'naturalness' of a language as a whole may be finally determined. This would be the reason why synthesized speech sounds somewhat 'unnatural', if physically equal durations are preserved[10]. The case proves that some other parameters make greater contribution to 'naturalness' of languages, and it is left to be discovered what exactly these parameters are that make them sound or heard more like what they are.

1.4 P-Center (Perceptual Center)

Although a sequence of digits may be produced by a human speaker such that they are perceived as isochronous, they are perceived as occurring irregularly, if tokens of each naturally spoken digit are presented with isochronous acoustic onsets. Morton, Marcus, and Frankish(1976) first introduced 'P-center' and defined it as a neutral term to describe what is regular in a perceptually regular sequence of speech sounds. Ever since the term was introduced, researchers have been making attempts to spot the 'psychological norm' in perception, using different acoustic parameters.

Marcus(1981) has shown that P-center location is affected by both initial consonant duration and subsequent vowel and consonant duration, although he suggests that there is little value in attempting to determine any single acoustic or articulatory correlate of P-center location, or in attempting to define P-center location absolutely in time. Howell(1988a, 1988b), on the contrary, suggests that the amplitude envelope of a syllable significantly affects a subject's judgements of its P-center location. His experimental findings say that if the energy tends to occur early in a syllable, its P-center will be nearer the start of the syllable, and conversely, if the energy tends to occur late in a syllable, its P-center will be nearer the end of the syllable. Other researchers[11] are now attempting to find acoustic determinants of P-center location in the combinations of different parameters, and thus speech production and perception have to be treated and examined in the reality of its dynamic form.

2.0 Experiments

2.1 Recordings

A total of ten native speakers of Japanese, eight females and two males, serve as subjects (Subject A~J). The stay of the subjects in English-speaking countries range from zero to twenty-eight months at the time of recording. Provided with sufficient time for practice, each subject is instructed to produce selected sentences at a comfortable speaking rate and to read each sentence through without unnecessary pausing. S/he is also instructed to repeat reading the text over again from the beginning, in the case of having misread it, until there is no misreading involved. Table 2.1 gives the background of each subject.

Table 2.1

subject	A	B	C	D	E	F	G	H	I	J
sex	f	f	f	f	f	f	m	f	m	f
age	35	32	25	33	27	45	23	26	25	38
stay (mo.)	12	0	0	26	2	2	2	28	0	3

A specimen sentence is selected from the IPA booklet[12] as the actual utterance to be evaluated. It reads as follows:

'They agreed that the one who could make the traveler take his coat off would be considered stronger than the other one.'

The recordings are held in an anechoic room, using DAT(SONY TCD-D10) and DAT(SONY DT-120). The recorded utterances are A/D converted with a sampling frequency of 40kHz to be analized.

2.2 Evaluation

In order to rank each subject's utterance according to the degree of 'naturalness' of English, the evaluation test is given to nine native speakers of English from the U.S.A. These informants are from Kentucky, California, Texas, New Hampshire, and Vermont. The test is carried out through a questionnaire on which the informants are asked to give intuitive and subjective evaluations of the readings and put the ten subjects' utterances in a ranked order, from the 'most native-like' to the 'least native-like'. They are allowed to listen to the text as many times as necessary. Table 2.2 shows the results of the evaluation made by the nine native speakers of English. All the informants reported normal hearing in both ears. The numeral '1' in the Table indicates the subject's reading is evaluated the 'most native-like', and '10', the 'least native-like', on a scale of one to ten[13].

The evaluation result has been applied to 'Cluster Analysis' (nearest neighbor method) for the purpose of grouping the ten subjects' utterances, which is shown in Fig. 2.2.

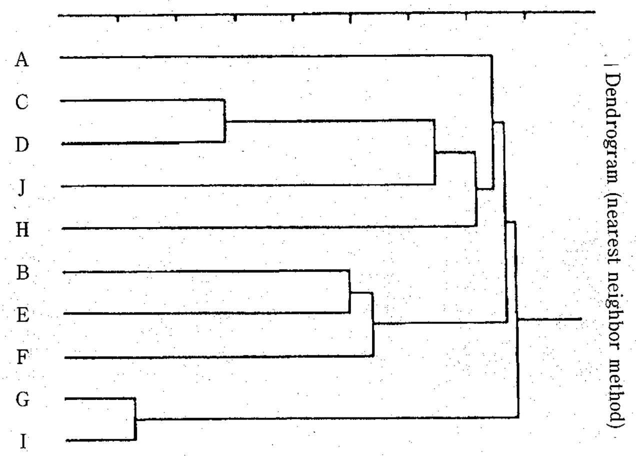
According to the results shown in Table 2.2, and Fig. 2.2, Subject H and A are chosen as the 'most native-like', and Subject I and G as the 'least native-like' speakers of the language to be analyzed. Informant JMf30's utterance is also analyzed to serve as a model utterance of a speaker of English, so that cross-proficiencial comparison can be made. Thus, the experimentally analytical and statistical treatment in the following chapter is applied to the utterances of the four, H, A, and I, G, out of the ten subjects as the representatives of the two contrastive groups, i.e. the 'most native-

Table 2.2
The Evaluation by Native Speakers of English

informants	A	B	C	D	E	F	G	H	I	J
GYf 30	5	6	1	2	8	7	9	3	10	4
JMf 30	5	6	4	3	7	8	10	1	9	2
BLm 20	1	7	4	6	5	9	10	2	8	3
WKm 40	5	8	6	3	2	7	9	1	10	4
WDm 20	1	8	6	5	7	3	10	2	9	4
LTm 20	1	6	3	4	7	8	10	2	9	4
CSm 20	7	6	2	3	5	4	10	1	9	8
HYm 30	1	5	4	7	6	8	9	2	10	3
HYf 20	1	7	5	4	7	8	10	2	9	7

The code for each informant indicates his/her initials, sex and age group.

Fig. 2.2
The Cluster-Analyzed Result of the Evaluation by the Americans



like' and the 'least native-like'.

3.0 Experimental Analysis

Based on the data obtained in the previous chapter, the utterances of the subjects H, A, I, G, and that of a selected native speaker, N, are experimentally examined through different parameters concerning the three major prosodic features.

3.1 Duration

Duration, one of the three major prosodic features of English, has been analyzed concerning the following parameters: total utterance duration (TUD), total unvocalized utterance duration

(UVD), interstress duration (ISD), mean syllable duration (SYD), stressed vowel duration (SVD), voice onset time of voiceless plosive consonants in initial position (VOT).

The stored texts with a sampling frequency of 40kHz are visually segmented with the reference to their wave forms, sound spectrograms and formant frequencies in a 25-msec. window. The onset of the vowels is determined to be as where a sharp rise appears in the first formant, and the offset as where the formant disappears.

3.1.1 TUD & UVD

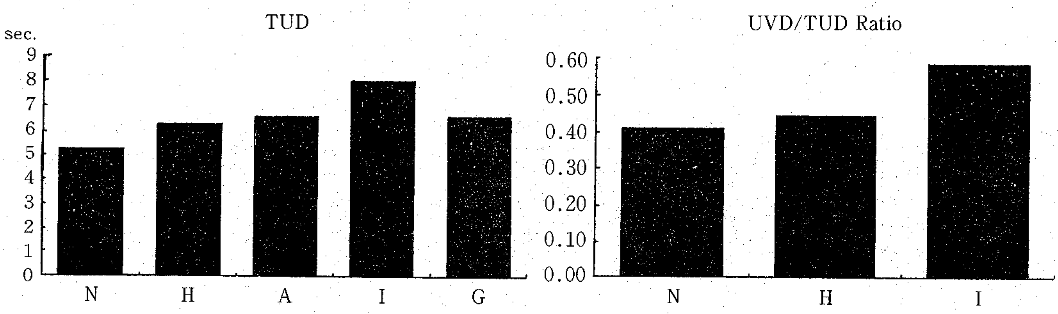
After measuring TUDs and UVDs, the ratio of the two is obtained in percentage and graphed out, and is shown in Table 3.1.1 and Fig. 3.1.1.

Table 3.1.1 : TUD & UVD

	N	H	A	I	G
TUD	5.29	6.33	6.64	8.07	6.55
UVD	2.19	2.86	*	4.76	*
UVD/TUD	41.5%	45.1%	*	58.9%	*

The numbers in TUD and UVD are in sec. The mark, *, indicates that it was unable to be obtained.

Fig. 3.1.1



3.1.2 ISD

The durations between stressed vowels are obtained by measuring the starting time of the beginning of each stressed vowel, and are numbered from 1 to 9 on Table 3.1.2. The content of each foot numbered in the issued sentence is described in the IPA phonetic symbols. The number in the parentheses shows how many syllables are in each ISD.

Table 3.1.2: ISD

foot number	N	H	A	I	G
1 (3) 'g r i d ð ə t ð ə	611.5	654.4	890.4	1200.3	966.8
2 (3) 'w ə n h u k ə d	462.8	701.4	707.3	880.4	606.2
3 (2) 'm e k ð ə	421.7	439.3	408.2	580.4	617.3
4 (3) 't r a v ə l ɹ	520.8	447.7	557.8	727.4	497.0
5 (2) 't e k ɪ z	399.5	470.4	428.4	589.7	396.5
6 (1) 'k o t	178.1	152.9	181.4	180.6	112.6
7 (4) 'ɔ f w ə d b i k ə n	831.6	1131.5	1184.4	1675.8	1389.7
8 (2) 's ɪ d ɹ d	421.7	747.6	584.7	432.6	256.7
9 (4) 's t r ɔ ŋ g ɹ ð ə n ð ə	650.1	665.3	672.0	856.8	910.1

The numbers are in msec.

3.1.3 SYD

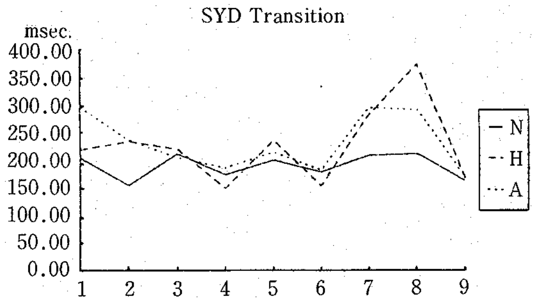
Based on the data in Table 3.1.2, the means of syllable durations are obtained for each ISD as is shown in Table 3.1.3a, and Fig. 3.1.3 shows the change of syllable duration from one foot to another. The data in Table 3.1.3a is then applied to F-test and t-test at the 5% level to see whether any significant difference can be observed in the five subjects' variables. The following can be concluded from the results of the tests shown in Table 3.1.3b. As far as syllabification is concerned, a significant difference can be observed between the native speaker and the non-native speakers of English—between N and H, A, I, G—whereas it cannot be said that there is a significant difference among the four Japanese subjects' data.

Table 3.1.3a: SYD

foot number	N	H	A	I	G
1	203.8	218.1	296.8	400.1	322.3
2	154.3	233.8	235.8	293.5	202.1
3	210.9	219.7	204.1	290.2	308.8
4	173.6	149.2	185.9	242.5	165.7
5	199.8	235.2	214.2	294.9	198.3
6	178.1	152.9	181.4	180.6	112.6
7	207.9	282.9	296.1	419.0	347.4
8	210.9	373.8	292.4	216.3	128.7
9	162.6	166.3	168.0	214.2	227.5

The numbers are in msec.

Fig. 3.1.3



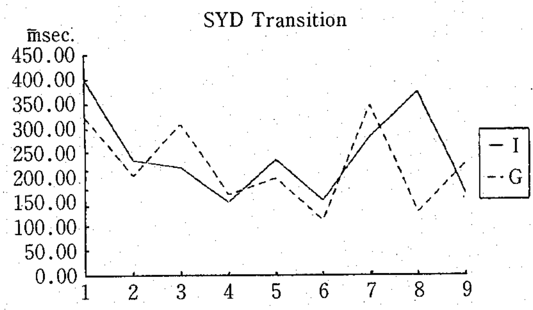


Table 3.1.3.b: Results of F-test

	N	H	A	I	G
Mean (mesc.)	189.1	225.8	230.8	283.5	223.7
Variable	434.4	4459.5	2425.3	5980.0	6464.5
S.D.	20.8	66.8	49.2	77.3	80.4

3.1.4 SVD

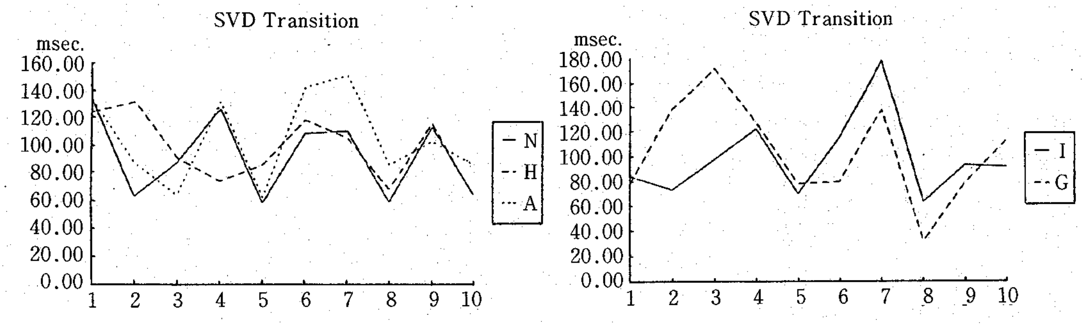
The issued sentence in this study includes a total of ten stressed vowels, and each of the vowels is extracted from each respective foot and its duration is measured. Table 3.1.4 gives the measured values and Fig. 3.1.4 visually depicts the durational change of stressed vowels.

Table 3.1.4: SVD

foot number	Vowel	N	H	A	I	G
1	i	132.7	123.5	132.7	84.0	77.3
2	ə	63.0	131.0	87.4	73.1	138.6
3	e	87.4	90.7	63.8	97.4	172.2
4	a	126.0	73.9	131.1	121.8	126.7
5	e	58.0	85.7	62.2	69.8	78.2
6	o	108.4	117.6	142.8	115.1	79.8
7	ɔ	110.0	105.0	151.2	178.1	137.7
8	ɩ	58.8	68.0	85.7	63.3	31.7
9	ɔ	112.5	115.1	102.5	93.3	78.8
10	ə	63.9	63.0	85.7	91.6	113.4

The numbers are in msec.

Fig. 3.1.4



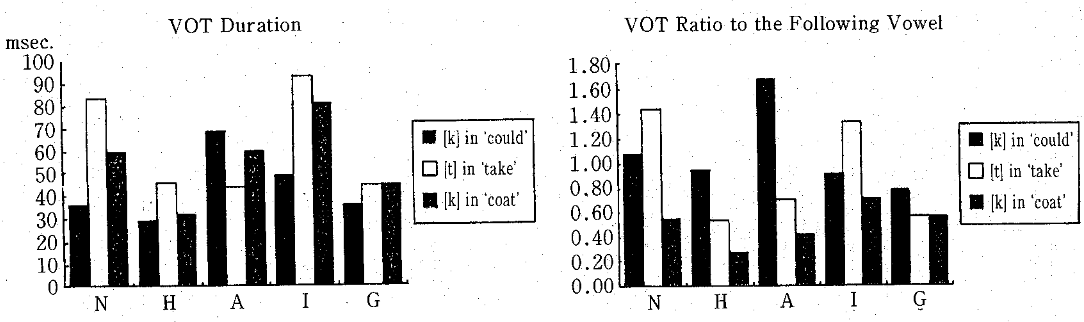
3.1.5 VOT

In order to observe not only vowel durations but also to see how consonants are articulated in each subject's utterance in its durational phase, the three words which hold stop consonants in their initial positions – 'could [kəd]', 'take [tek]', 'coat [kot]'– are examined. Actual VOT for each voiceless stop consonant in its initial position of the word is measured, and the ratio to the duration of its following vowel is obtained, so that some correlation of the two consecutive phonemes may be observed. Table 3.1.5 gives the values, which are graphed out in Fig. 3.1.5.

Table 3.1.5: VOT

VOT (msec.)	N	H	A	I	G
[k] in 'could'	36.1	29.3	68.8	48.7	35.8
[t] in 'take'	84.0	45.9	43.7	93.3	44.4
[k] in 'coat'	59.8	31.8	59.9	81.5	44.7
VOT/Folloing Vowel					
[k] in 'could'	1.07	0.94	1.67	0.91	0.77
[t] in 'take'	1.45	0.54	0.70	1.34	0.57
[k] in 'coat'	0.55	0.27	0.42	0.71	0.56

Fig. 3.1.5



3.2 Pitch

First, the overall F0 contour form of the entire utterance of the sentence is displayed together with its wave form for all the five subjects as in Fig. 3.2a[14]. They were obtained in a 10-sec. window, with a 0-500 Hz time domain for the F0 plot (100Hz divisions on the vertical axis), and with 400-msec. time spans (on the horizontal axis) for the wave forms. Fundamental frequencies, then, are examined with regard to the ten stressed vowels extracted in Chapter 3.1.4. Because of the differences of pitch ranges of the subjects, pitch-ratios are obtained based on the highest and lowest F0 values of each one of the ten vowels, so that the F0 transitions within the stressed vowels can be compared among the five subjects. Table 3.2 gives values for each measurement (which is graphed in Fig. 3.2b) as well as visual figures of F0 contours within each vowel.

Fig. 3.2a

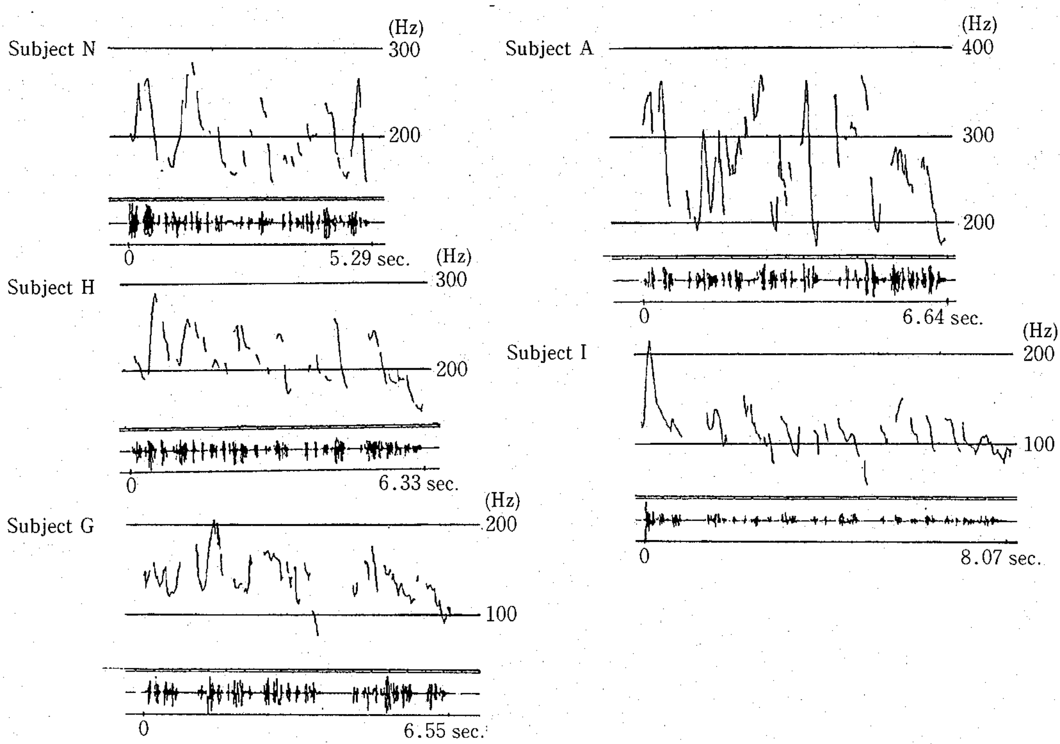


Fig. 3.2b

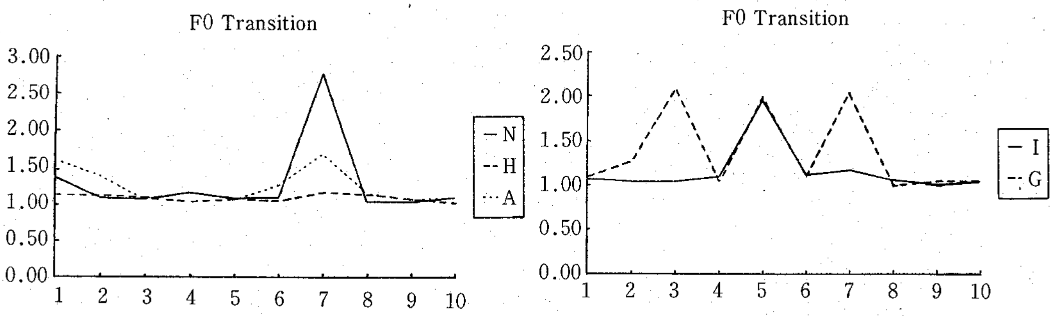
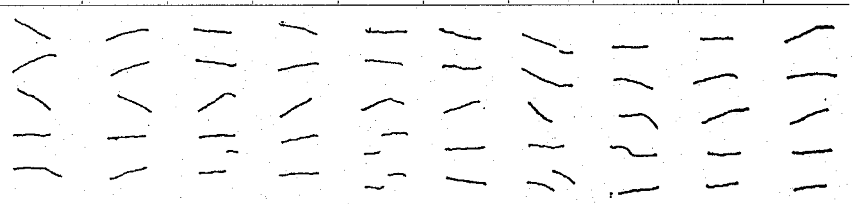


Table 3.2: Fundamental Frequency

stressed vowels		1	2	3	4	5	6	7	8	9	10
highest value (Hz)	N	248	210	221	189	178	239	157	199	227	265
	H	287	257	211	243	217	239	202	254	245	189
	A	330	295	297	365	248	362	291	359	277	274
	I	137	139	115	130	114	117	113	133	129	112
	G	147	206	267	170	150	150	106	157	149	134
lowest value (Hz)	N	181	193	207	163	166	220	57	194	224	246
	H	255	230	194	249	205	231	175	227	232	187
	A	206	214	281	318	236	287	174	315	263	254
	I	128	133	110	118	58	104	96	125	127	107
	G	134	162	128	164	75	137	52	156	142	127
FO-ratio	N	1.37	1.09	1.07	1.16	1.07	1.09	2.75	1.03	1.08	1.01
	H	1.13	1.12	1.09	1.02	1.06	1.03	1.15	1.12	1.06	1.01
	A	1.60	1.38	1.06	1.15	1.05	1.26	1.67	1.14	1.05	1.08
	I	1.07	1.05	1.05	1.10	1.97	1.13	1.18	1.06	1.02	1.05
	G	1.10	1.27	2.09	1.04	2.00	1.09	2.04	1.01	1.05	1.06

visual transition N H A I G



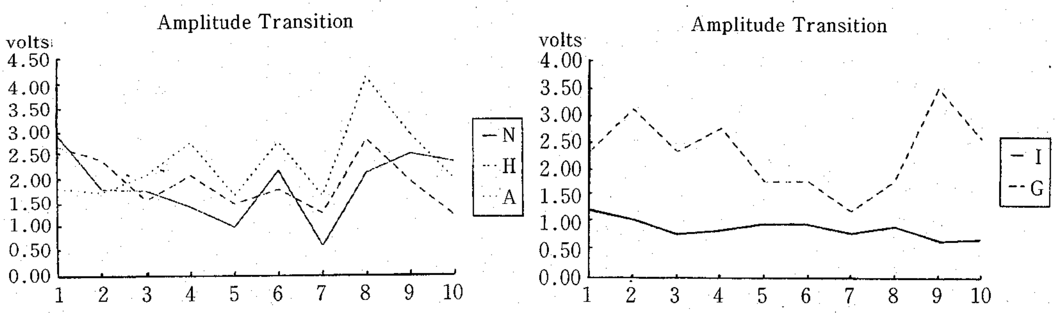
3.3 Intensity

Intensity is examined in terms of amplitude in volts for each of the ten stressed vowels as in §3.2, and is shown in Table 3.3 and Fig. 3.3.

Table 3.3: Amplitude

stressed vowel		1	2	3	4	5	6	7	8	9	10
amplitude (volts)	N	2.95	1.82	1.79	1.45	1.02	2.22	0.62	2.17	2.57	2.40
	H	2.72	2.42	1.59	2.12	1.52	1.83	1.32	2.88	1.99	1.26
	A	1.84	1.75	2.12	2.80	1.70	2.81	1.68	4.15	2.97	2.02
	I	1.26	1.09	0.81	0.88	1.01	1.00	0.83	0.97	0.70	0.74
	G	2.33	3.14	2.35	2.79	1.80	1.81	1.26	1.82	3.53	2.59

Fig. 3.3



4.0 Discussion

It has been generally noted that Japanese speakers of English find it rather difficult to pronounce unstressed syllables with shorter durations than stressed ones. It is this inability of controlling the durations of syllables, depending on whether they are to be stressed or not, that most likely contributes to the increase in the interstress durations, in addition to the tendency toward having longer pauses in the utterance. Although longer duration is not the sole factor for the native speaker of English to give 'non-native-like' judgement[15], listeners are influenced by the durational patterns of voicing and/or syllables.

The present study has revealed that the 'most native-like' speakers, as well as the native speaker of English referred to as the model utterance, showed compression effects in the feet with different numbers of syllables. This can be seen in the transitions of feet numbered 6-7-8 in Fig. 3.1.4, whose numbers of syllables are 1-4-2, respectively. As for the 'native-like' and native speakers, the stressed vowel duration numbered 7, which has four syllables, does not increase sharply when compared with the preceding ones, each of which consists of a single syllable. On the contrary, extremely sharp gradients can be most contrastive, in the transitions of the 'least native-like' speakers between the two consecutive vowels. The same is observed in the transitions between the stressed vowels numbered 7 and 8, although the direction of the gradients is opposite. The data can be interpreted such that the 'native-like' speakers attempt to reduce the duration of the foot numbered 7 by pronouncing the stressed vowel with much shorter duration, so that the total duration of the foot would not be too long.

What, then, differentiates the Japanese most 'native-like' speakers of English from the native speaker of the language is, as shown in the transitions of the feet numbered the same, the native shows a greater compression effect than the Japanese speakers by keeping more equal mean syllable durations so that 'psychological isochrony' is likely to be preserved. It may be said that the degree of this compensatory effect is, in some way, one of the factors to affect the differences of 'native-like' judgements between the native speakers of English and of Japanese[16].

Also, it has been reported that narrow pitch range is one of the factors of 'unnaturalness' of English. According to F0 contours of the subjects, Subject A maintained the widest pitch range from the beginning to the end, followed by Subjects N and H with moderate range, and Subject G and I, the narrowest. Since pitch range differs from person to person, however, pitch ratio within each individual was used for comparison. Fig. 3.2 appears to offer the greatest differentiation of the two groups, 'native/native-like' and 'non-native-like'. The limited data obtained in this study indicates that pitch may be the greatest contributor of all the prosodic parameters involved in this study. The three peaks in Subject G's utterance can be due to pitch accent of his native language, Japanese. Both Subject I and G's pronunciation of the stressed vowel numbered 5 with extremely high frequency may indicate some correlation with VOT of its preceding voiceless stop consonant, [t]. Further experiments in more detailed parameters are necessary to come to a definite conclusion on the matter.

In their experiments with synthesizing and parameter-exchanging methods, Suzuki, Ohyama and Kiritani(1989) report that intensity has the least effect of the three principal prosodic features on 'naturalness' of English. However, as far as the amplitudes of stressed vowels in the issued sentence are concerned, it seems that a certain common inclination can be observed in the utterances of the native and 'native-like', which are quite different from the ones of the 'non-native-like'. It would be worth continuing to search for more determinant parameters within 'intensity', from different phases and methods, even though it is true that intensity least contributes to the degree of how English-like an utterance can sound. All judgements of the parameters of the three prosodic features, in natural speech, or rather stress in English, depend on the complicated interaction of a number of cues.

5.0 Conclusion

The above-mentioned discussion remains in need of further empirical validation. A definitive conclusion on the data obtained in this study must also remain tentative until such time as these and other fundamentally interrelated issues are more clearly understood. At present, what can be concluded is that there appear to be language-dependent differences at multi-dimensional levels in sound cognition as function of a language, and that these differences, in turn, seem to be a function of the phonetic structure of the language per se. Further clarification on the relationships amongst phonetic features must await substantiation, and it would be of value to replicate the present study with different, narrower samples of more subjects and informants.

NOTES

1. Bowen (1975), xii-xiii
2. Suzuki, Ohyama, and Kiritani (1989, 1990)
3. Lehiste (1970), p3
4. The term 'loudness', used as one of the three prosodic features of English, should be clearly distinguished from other terms implying similar content, such as 'intensity' and 'amplitude'. "Loudness is the

subjective property of a sound that is most directly related to intensity, which is a physical characteristic of sound. Intensity differs from amplitude in the sense that the amplitude of a wave does not depend on the frequency of the wave whereas intensity does."(Lehiste, 1970, p112, and p115)

"The acoustic correlate of loudness is intensity or the amount of energy which is present in a sound or sequence of sounds, variations in the pressure of air coming from the lung. ...the relationship of absolute intensity to perceived loudness is by no means linear (a sound has to be much more than doubled in absolute intensity before it will be heard as twice as loud).(Cruttenden, 1986, p3)

5. Cruttenden, (1986), p2
6. He notes that what is important is not the magnitude of the frequency change, but the fact that a frequency change has taken place, or rather, not how much F0-change is made but whether F0 changed or not.
7. Pike (1945), Price (1980), Hoequist (1983b), Kiritani (1989)
8. Bolinger (1965), Lehiste (1977), Hoequist (1983b)
9. This has been experimentally proved in this study.
10. Lehiste (1970, p61) remarks on this matter as follows: The relative aperiodicity of sound waves produced at the glottis seems to contribute to the naturalness of many samples of synthesized speech may be due to the excessive regularity of the F0 generator.
11. Hoequist (1983a), Cooper (1986), Fowler et al. (1988), Howell (1988a, 1988b)
12. International Phonetic association (1949), p21
13. Informants were asked not to overlap the ratings for different subjects.
14. F0 contours listed here are shown in a 10-sec. window.
15. The English-like evaluation is also made by 138 Japanese college students. Unlike the native speakers of English, native Japanese speakers put much higher importance on TUD. This may have been one of the crucial reasons why the Japanese informants, freshmen/sophomore college students, rated Subject G as well as D, H and A as 'native-like'. The following gives the TUDs of all the subjects in the order of shorter durations.

Table 3.1.2: ISD

D	N	H	C	A	G	E	J	I	F	B
5.26	5.29	6.33	6.48	6.64	6.65	6.85	7.51	8.07	8.29	9.43 (sec.)

16. The Japanese informants rated the subjects, H and G, as approximately the same degree of 'native-likeness' while the American informants ranked Subjects H 'most native-like' and Subject G 'least native-like' without hesitation. This seems to indicates that the native speakers of English can auditorily recognize the presence of the compression effect in Subject H's utterance (but not in Subject G's) and give credit to its existence, whereas the Japanese informants either can not perceive such effect in utterances or do not perceive it as meaningful.

REFERENCES

Albrechtsen, Dorte., B. HENRIKSEN, and C. Færch., 1980 "Native Speaker Reactions to Learners' Spoken Interlanguage" *LL*30, pp.365-396

Bloomlield, L., 1933 *Language*, Henry Holt and Co., N. Y. (translated by Hattori, S., Miyake, K., and Hino, S)

Bolinger, D. L., 1965 "Pitch Accent and Sentence Rhythm", *Forms* of English: *Accent, Morpheme, Order*. Abe, I. & T. Kanekiyo, eds., Tokyo, Hokuou Publishing Co.

Bowen, J. Donald., 1975 *Patterns of English Pronunciation*, Univ. of California, Los Angels, Newbury House Publishers, Inc., Rowley, Massachusetts

Cooper, A. M., 1986 "P-centers are Unaffected by Phonetic Categorization", *Perception & Psychophysics* 39(3), pp. 187-196

Cooper, A. M., 1988 "The Syllable's Rhyme Affects Its P-center as a Unit", *Journal of Phonetics*, 16, 231-241

Crystal, David., 1987 *The Cambridge Encyclopedia of Language*, Cambridge Univ. Press

Cruttenden, Alan., 1986 *Intonation*, Cambridge Univ. Press

Fowler, C. A., D. H. Whalen, & Andre M. Cooper, 1988 Perceived Timing is Produced Timing: A Reply to Howell, *Perception & Psychophysics*, 43, 94-98

Fry, D.B., 1955 "Duration and Intensity as Physical Correlates of Linguistic Stress", *JASA* 27

Fry, D.B., 1958 "Experiments in the Perception of Stress", *Language and Speech* 1, pp.126-152

Fujimura, Osamu., 1981 "Temporal Organization of Articulatory Movements as a Multi-dimensional Phrasal Structure", *Phonetica* 38, pp.66-83

Hœquist, Charles, Jr., 1983a "The Perceptual Center and Rhythm Categories", *Language and Speech*, Vol. 26, Part 4, pp.367-376

Hœquist, Charles, Jr., 1983b "Syllable Duration in Stress-, Syllable- and Mora-Timed Language", *Phonetica* 40, pp.203-237

Honma, Yayoi., 1985 *Acoustic Phonetics in English & Japanese*, Yamaguchi Shoten.

Howell, Peter., 1988a "Prediction of P-center Location from the Distribution of Energy in the Amplitude Envelope: I", *Perception & Psychophysics* 43, pp.90-93

Howell, Peter., 1988b "Prediction of P-center Location from the Distribution of Energy in the Amplitude Envelope: II", *Perception & Psychophysics* 43, p.99

Kenyon, J.S., and T.A. Knott., 1953 *Pronouncing Dictionary of American English*, Springfield, Mass.: G. & C. Merriam

Ladefoged, P., 1975 *A Course in Phonetics*, New York, Harcourt Brace Jovanovich

Ladefoged, P. & M. Lindau., 1989 "Modeling Articulatory-Acoustics Relations: a Comment on Stevens' 'On the Quantal Nature of Speech'", *Journal of Phonetics* 17, pp.99-106

Leben, Willie Ronald., 1980 *Suprasegmental Phonology*, Garland Publishing Inc., New York & London

Lehiste, Ilse., 1970 *Suprasegmentals*, Cambridge, Mass., MIT Press

Lehiste, Ilse., 1974a "Interaction between Test Word Duration and Length of Utterances", *Ohio State Univ. Working Papers in Linguistics*, No. 17, pp.160-169

Lehiste, Ilse., 1974b "Variability in the Production of Suprasegmental Features", *Ohio State Univ. Working Papers in Linguistics*, No. 17, pp. 170-179

Lehiste, Ilse., 1977 "Isochrony Reconsidered", *Journal of Phonetics* 5, pp.253-263

Liberman, Mark., 1979 *The Intonational System of English*, Garland Publishing, Inc. New York & London

Luthy, Melvin J., 1983 "Non-native Speakers' Perception of English 'Nonlexical' Intonation Signals", *LL* 33, pp. 19-36

Marcus, Stephen Michael., 1981 "Acoustic Determinants of Perceptual Center Location", *Perception & Psychophysics* 30(3), pp.247-256

Mochizuki-Sudo, Michiko., and Kiritani, Shigeru., 1989 "Production and Perception of the Rhythmic Pattern of English by Japanese Learners", Ann. Bull. *RILP* No. 23, pp.75-84, Tokyo University

Morton, J., Marcus, S. M., & Frankish, C. R., 1976 "Perceptual Centers (P-centers)", *Psychological Review* 83, pp.405-408

Nakatani, Lloyd H., et al., 1981 "Prosodic Aspects of American English Speech Rhythm", *Phonetica* 38, pp.84-106

Ohala, John., 1981 "Speech Timing as a Tool in Phonology", *Phonetica* 38, pp.204-212

Oliver, Douglas., 1983 "Fundamental Frequency Studies as a Preliminary to the Literary Criticism of Poetry", *Journal of Phonetics* 11, pp. 1-35

Pike, Kenneth., 1945 General Characteristics of Intonation, *The Intonation of American English*, Univ. of Michigan Press, pp.20-41

Pompino-Marschall, B., 1989 "On the Psychoacoustic Nature of the P-center Phenomenon", *Journal of Phonetics*

17, pp.175-192

Price, P. J., 1980 "Sonority and Syllabicity: Acoustic Correlates of Perception", *Phonetica* 37, pp.327-343

Reenen, Pieter Van., 1982 *Phonetic Feature Definitions, Their Integration into Phonology and Their Relation to Speech, A Case Study of the Feature NASAL*, Foris Publications

Scovel, Tom., 1969 "Foreign Accents, Language Acquisition and Cerebral Domiance", *LL*, pp.245-254

Scovel,Tom., 1977 "The Ontogeny of Ability to Recognize Foreign Accents", *First Annual Second Language Acquisition*, Research Forum, UCLA 1977

Stevens, Kenneth N., 1989 "On the Quantal Nature of Speech", *Journal of Phonetics* 17, pp.3-45

Suzuki Hiroshi, Ghen Ohyama, Shigeru Kiritani., 1989 "A Comparison of Evaluations by American Japanese Listeners of English Spoken by Japanese Speakers", *Ann. Bull. RILP* No. 23, pp.51-57

Suzuki, Hiroshi, Ghen Ohyama, Shigeru Kiritani., 1990 "In Search of a Method to Improve the Prosodic Features of English Spoken by Japanese", Presentation at ICSLP 90, vol.2, pp.965-968

Takefuta, Yukio., 1982 *Nihonjin Eigo no Kagaku: Sono genjou to Asu he no Tenbou*, Kenkyusha

Tarone, Elaine., 1980 "Communication Strategies, Foreigner Talk, and Repair in Interlanguage", *LL* 30, pp.417-431

International Phonetic Association, 1949 *The Principle of the International Phonetic Association*, Department of Phonetics, Univ. College, London

Manuscript received: September 28, 1993

Published: December 27, 1993