A Suprasegmental Threshold for L2 Pronunciation

Tamami Katayama
(Graduate School of International Media, Communication, and Tourism Studies, Hokkaido University)

Abstract

This study has attempted to identify a threshold of intelligible pronunciation amongst L2 speakers, focusing on such suprasegmental features as rhythm, stress and intonation. In particular, it has analyzed examples of the best pronunciation and the poorest pronunciation in the utterances of sixteen Japanese learners of English, as evaluated by four native English instructors. The instructors evaluated pronunciation through such perspectives as the segment, intonation and rhythm, using a five-point scale with midpoints from 1 (poor) to 5 (excellent). After labeling each segment of the utterances, I examined the spectrograms, fundamental frequencies, and segmental duration with a speech analyzer, *praat*. The results show that L2 speakers with the highest

evaluations indicated stress by heightening the pitch, and producing several words as a unit within a phonological phrase, while emphasizing the stress of a nuclear word. On the other hand, L2 speakers with the poorest evaluations indicated word accents by lowering the pitch, and produced every word separately.

1. Introduction

One of an L2 learner's main goals is to communicate successfully through the effective use of the target language. While L2 pedagogy has focused on communicative skills rather than strict accuracy in the performance of the target language, it has sometimes forgotten that pronunciation, including the suprasegmental (or prosodic) features, plays a vital role in oral communication since without it communication becomes impossible. Even so, it is not realistic for learners to seek to acquire native-like pronunciation; rather, the practical goal should be to acquire comprehensible pronunciation (Celce-Murcia, Brinton, & Goodwin, 2003). Since the factors that make L2 pronunciation comprehensible to native speakers have not been adequately clarified, we need further empirical studies to bring about better understanding.

One promising approach to developing strategies for communicative pronunciation is to focus on suprasegmental features of pronunciation. The segmental features themselves are vowels and consonants, which "involve airstream mechanisms, states of the glottis,

primary and secondary articulations, and formant frequencies ... The principal suprasegmental features are stress, length, tone, and intonation" (Ladefoged, 2001, p.225). Researchers have investigated segmentals and suprasegmentals as important aspects of pronunciation (Anderson-Hsieh, Johnson, & Koehler, 1992; Riney, Takada & Ota, 2000). Some researchers (Anderson-Hsieh, Johnson, & Koehler, 1992) found that suprasegmental features affected native speakers' understandings. Others have argued that suprasegmentals are more important in the communication of meanings (McCrosite, 2004). It therefore appears rational to prioritize suprasegmental features, and this study focuses on the suprasegmental features of Japanese speakers who are learning English as an L2. The aim of this research is to investigate what features affect L2 speakers' comprehensible pronunciation.

2. Background

Earlier studies have indicated that a number of suprasegmental features are language-specific (Kubozono, 1998, 2001; Ladefoged, 2001), that suprasegmental features of the L1 can be transferred to production of the L2 (Celce-Murcia et al., 2003, Masaki, Takazawa, & Arai, 2000), and that L1 suprasegmental transfer can hinder native speakers' understanding of non-native speakers' L2 production (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro, 1995). These factors taken together suggest that a focus on the suprasegmental features of L2 speech is likely to provide a promising approach to the

investigation of communicative pronunciation. Because Japanese and English are typologically very different, it is generally assumed that negative transfer of first language occurs when Japanese L2 learners attempt to acquire competence in English as the target language.

When a particular syllable is more prominent relative to others within a word, it can be said that the syllable in question has an accent (Kubozono, 1998). The prominence within a word (a word accent) plays an important role in oral communication. Another function of word accent is to discriminate one word from another: this is called the distinctive function (Kubozono, 1998), and is used to discriminate two separate words by means of such prosodic features as strength or pitch, when their phonemic structures are the same with respect to the array of phonemes. For example, there are pairs of words, such as "ame" (rain and candy) in Japanese, and "increase" (a verb and a noun) in English. The Japanese pair can be distinguished in terms of the differences in pitch patterns, such as [High Low]—[Low High], while the English pair can be distinguished by the difference of intensity (stress) patterns such as [Strong Weak]—[Weak Strong]. This explains why the term "stress", signifying intensity of performance, is used for English accent (Kubozono, 1998).

Those languages that employ a word accent are categorized as either a pitch-accented language or a stress-accented language. The former recognizes the prominence of a word as the alteration of pitch, while the latter recognizes it as variation of strength (Kubozono,

2001). "In Japanese the accent is invariably realized as a high pitch, so that Japanese is often called a pitch accented language" (Ladefoged, 2001, p.240). Since English accent is represented as making use of intensity, it can be said that English has a stress accent (Kubozono, 2001).

In English, speakers place prominence on a specific word or a specific part in a word in the flow of speech to avoid a monotonous chain of words. This is called sentence stress, and is carried by the words with more information (Kubozono, 1998). The stresses that can appear on words are sometimes modified and dropped off when the words are part of sentences. "As a general rule, English tries to avoid having stresses too close together. Very often, stresses on alternate words are dropped in sentences where they would otherwise come too near one another" (Ladefoged, 2001, p.98). The Japanese language differs from English in this respect. In Japanese, stresses on words are not modified nor dropped off even when the words are part of a sentence, as an accent is fixed within a word. Furthermore, "word and sentence stress combine to create the rhythm of an English utterance – that is, the regular, patterned beat of stressed and unstressed syllables and pauses" (Celce-Murcia et al., 2003, p.152). As stress in English is closely associated with rhythm, both stressed and unstressed syllables have to be pronounced correctly in order to establish the rhythm of an utterance.

Rhythm is the recurrence of a regular linguistic structure that gives rise to a sense of comfortable diction.

Rhythm is roughly classified into two groups, that is, syllable-timed rhythm and stress-timed rhythm. While the recurring of syllables that have the same length produces a syllable-timed rhythm, a stress-timed rhythm is the result of the recurring of sentence stresses. Japanese is a typical example of the former and English that of the latter (Kubozono, 2001). Celce-Murcia et al. (2003) note that "the stress-timed nature of English means that the length of an utterance depends not on the number of syllables but rather on the number of stresses"(p.152). In a language where the length between the stresses is equal, it is impossible to pronounce every syllable with the same length because speakers adjust the lengths of syllables depending on the location of stress. In this way, stress-timed rhythm and syllable-timed rhythm are inherently incompatible (Kubozono, 2001).

Celce-Murcia et al. (2003) emphasize the significance of the stress-timed nature and rhythm in English when learning that language. They point out that learners whose first language is syllable-timed do not differentiate between stressed and unstressed syllables in English, and that this failure to use stress effectively interferes with native speakers' comprehension. In addition, she claims that learners, while being on the whole ignorant of the significance of contrast between stressed and unstressed syllables, are confident in their pronunciation if they believe that they have articulated each word and syllable distinctly. Nevertheless, her claim requires further empirical evidence to support it. As Japanese and English have different types of accents, pitch-accented versus

stress-accented, Japanese learners of English may experience interference from their first language when they speak English. Their manner of producing stress is very likely to hinder native speakers' comprehension.

Masaki, Takazawa and Arai (2000) investigated empirically the effects of the typological variation with respect to phonetic realization of word accents. They sought to find out whether there were any differences in the perception of accents between native speakers of pitch-accented languages like Japanese and stress-accented languages such as German and Spanish, when they encounter a language that in this respect differs from their own. First of all, the participants identified the placement of accents of foreign words that they did not know, in languages such as Norwegian, Hindi, Finnish, Hungarian, and Greek. Then they analyzed the acoustic parameters of the words that affected the judgments of accents. The results indicated that the German speakers who use stress accent perceived accents in syllables whose pitch rose significantly, while Japanese speakers and Spanish speakers identified accents in syllables whose pitch dropped. Indeed, in Japanese phonology, the rapid pitch fall in a word is called an accent (Kubozono, 1998). In addition, they reported that Japanese speakers tended to recognize accents in the syllable whose pitch did not fall, when the following syllable was produced at a low pitch. As for the Spanish speakers, the further investigation is required. Since English, like German, has a stress accent, it can be predicted that native English speakers will show the

same tendency as German speakers with regard to the perception of accents.

3. Research Purpose

The purpose of this study is to examine the suprasegmental features of L2 speech which affects pronunciation ratings by native English speakers, and to suggest the threshold level of L2 pronunciation. In particular, this study compares speeches of a native English speaker (NS), and two non-native speakers with the highest evaluation (SH) and the lowest evaluation (SL), and examines whether any acoustic differences are found between them. The following research question needs to be addressed:

> What is the threshold level of suprasegmentals in L2 pronunciation?

According to Masaki et al. (2000), there were differences in the perception of accented syllables between native speakers of a pitch-accented language (such as Japanese) and speakers of a stress-accented language (such as German and English) when they encounter a language that differs from their own. Speakers of a pitch-accented language perceived an accent in the syllable where the pitch falls, whereas speakers of a stress–accented language perceived an accent in syllables with a remarkable pitch increase. As these different perceptions of an accent may impede the production of stress for Japanese learners of English, the following prediction can

also be made:

> When Japanese learners speak English, they produce stresses with a pitch accent instead of a stress accent and fail to produce a stress-timed rhythm. This makes it hard for native speakers to comprehend the speech.

I hope to suggest which components of the suprasegmental features affect native English speakers' evaluation by comparing pronunciation, and offer some linguistic explanations as to why Japanese speakers continue to have difficulty in producing intelligible English.

4. Methodology

4.1 Participants

Sixteen Japanese speakers consisting of eleven graduate students and five undergraduate students, took part in the study. They ranged in age from 19 to 31 years. This study adopted the pronunciation of a native English speaker from California as a model, while four native English-speaking instructors at college level joined as raters.

4.2 Procedure

The sixteen Japanese participants were instructed to read aloud a text (see Appendix) without preparation and video-taped by the researcher. The four native English-speaking instructors then watched the

video-tapes. In order to avoid the order effect, another version of the tape was created, in which the order of presentation was reversed. Two of the four raters watched the first version, and the other two watched the reversed version. The raters evaluated the pronunciation of the participants from the perspectives of segmental features, intonation, and rhythm, using a five-point scale with midpoints. The criteria of each perspective are as follows:

1. Segmentals
   Each word is clearly distinguishable and understandable.
2. Rhythm
   "Stressed syllables are sufficiently prominent and occur at fairly regular intervals. Unstressed syllables and function words are sufficiently reduced" (Anderson-Hsieh, Johnson, & Koehler, 1992, p.541).
3. Intonation
   "Intonation contours are appropriate and pitch range is sufficiently wide" (Anderson-Hsieh, Johnson, & Koehler, 1992, p.541).

4.3 Data Analysis

The mean value of the rater's evaluations was computed for each participant, after which the speech with the highest evaluation (SH) and that with the lowest evaluation (SL) were selected and analyzed with a speech analyzer *praat*[1], in order to compare them with the speech patterns of the native speaker (NS). The analysis

included pitch contours, waveforms, and phonological phrases. A phonological phrase typically consists of 4 to 7 syllables and has one melodic contour (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004). Because a phonological phrase is closely related to intonation, our analysis included this perspective. In order to examine whether there were any significant differences in duration of prominence, a one-way ANOVA was also performed.

5. Results

Speech samples of the two subjects at either end of the scale were then selected for comparison: the mean of SL was 1.67, while that of SH was 3.36. Table 1 presents the mean scores of SH and SL from perspectives of intonation, rhythm, and segmental. SL's mean scores of intonation and rhythm were 1.5 and that of segmentals was 2. On the other hand, the ratings of SH were 3.5 for intonation, 3.25 for rhythm, and 3.33 for segmentals.

Table 1

The Mean Scores of SL and SH

Subject	Intonation	Rhythm	Segmentals	Total
SL	1.5	1.5	2	1.67
SH	3.5	3.25	3.33	3.36

Note. Minimum value = 1; Maximum value = 5

The results of the analysis using the speech analyzer show the differences between NS's accenting of syllables and that of SH and SL (see Figure 1 to 3). NS and SH produced the stressed syllables by raising pitch, whereas SL pronounced accented syllables by lowering pitch. For example, when NS and SH pronounced "response", they raised pitch in the syllable of "ponse". On the other hand, SL pronounced the syllable "res" flatly and reduced pitch abruptly on the syllable "ponse". This trend was also reflected in the production of "complaints" and "residents". The direction of pitch differed for the accented syllables: Pitch increased in NS and SH, whereas it decreased in SL.

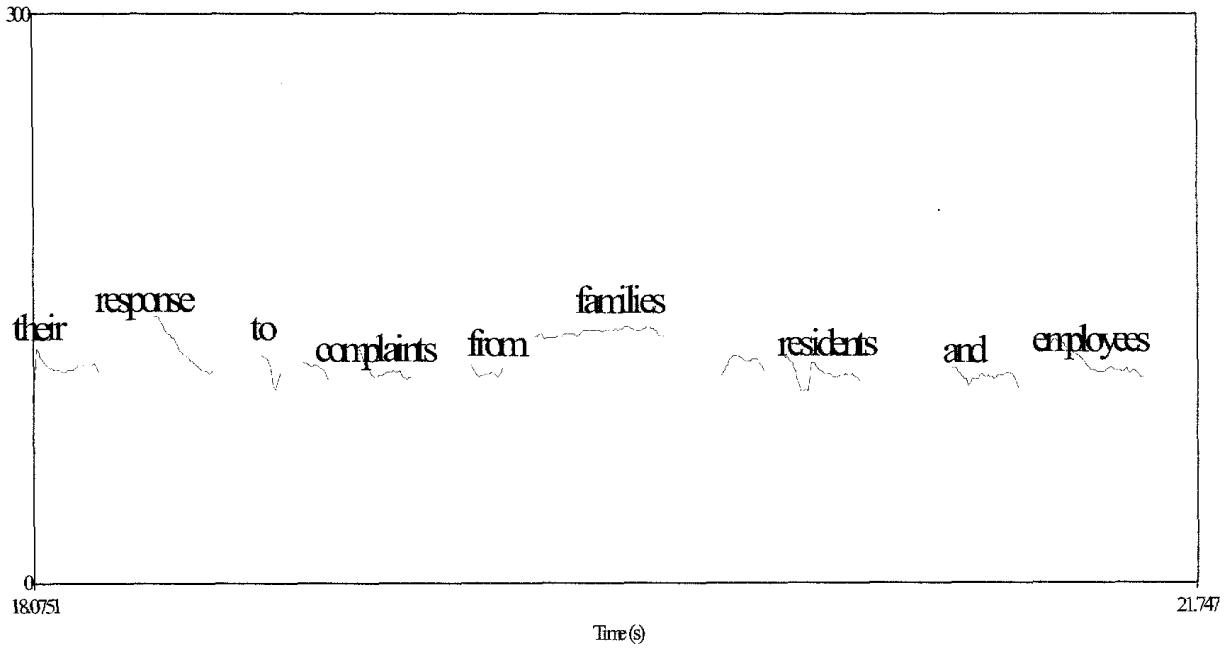

Figure 1. Pitch contours of NS

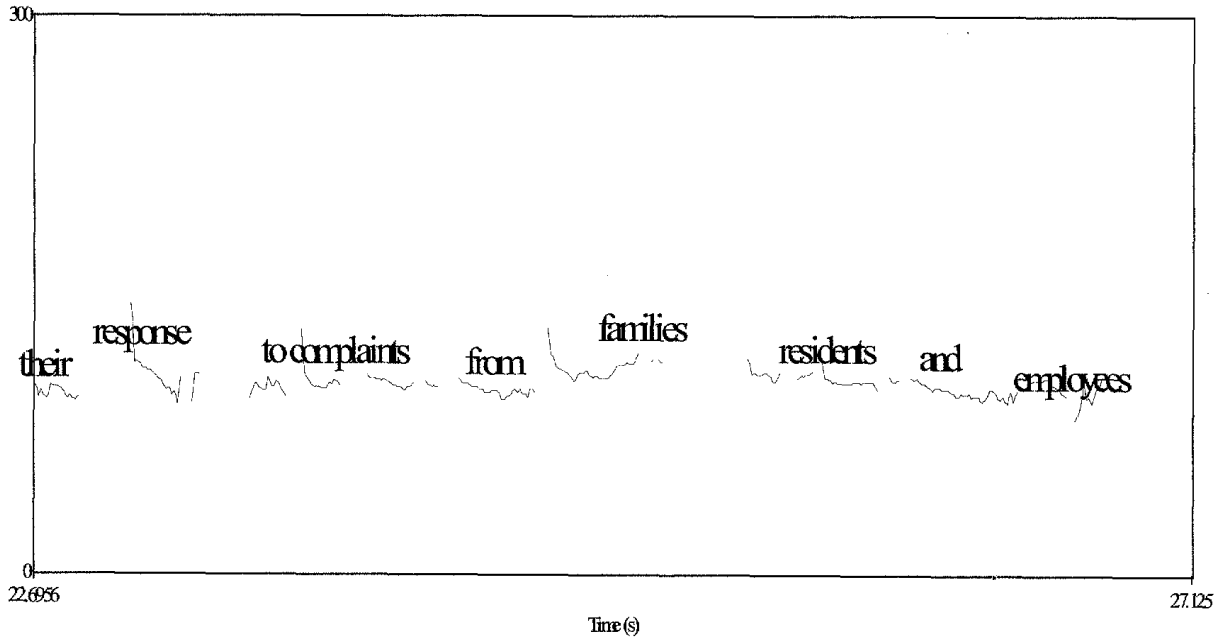

Figure 2. Pitch contours of SH

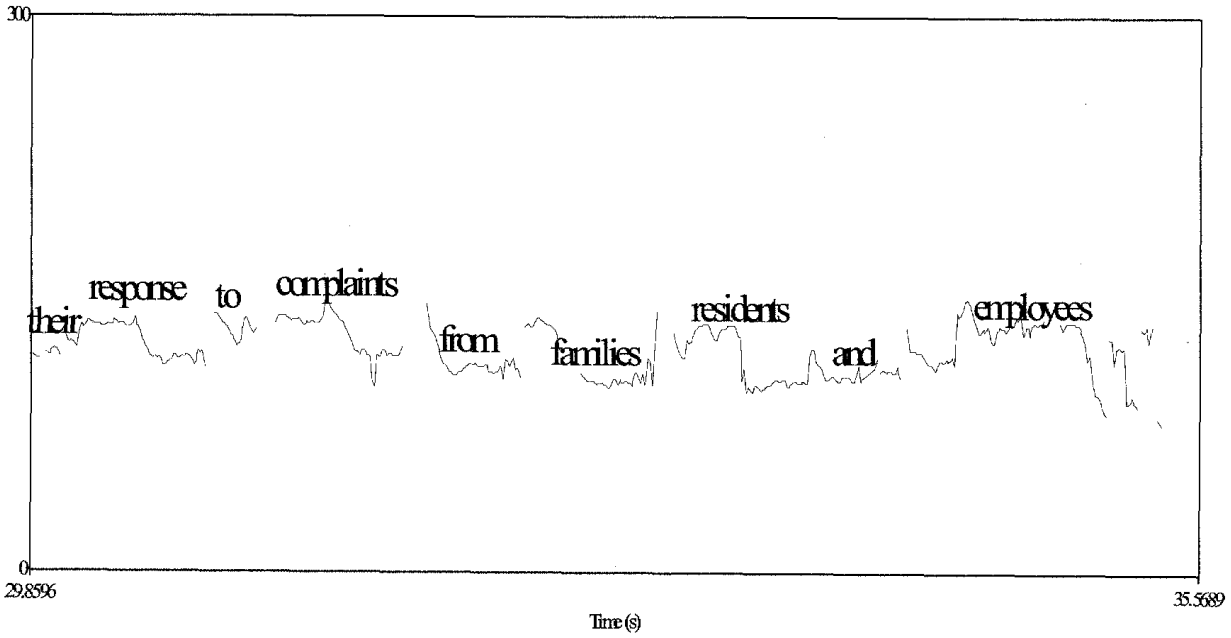

Figure 3. Pitch contours of SL

Next, the speakers differed in the number of phonological phrases that they produced. Table 2 shows the number of phonological phrases produced by the three speakers. SL produced more phonological phrases than NS and SH. This means that SL tended to articulate each word distinctively and produce phonological boundaries between every word. On the other hand, NS and SH organized words and produced units of phonological phrases.

Table 2

The Number of Phonological Phrases

Speaker	Sentence 1	Sentence 2
NS	6	9
SH	6	13
SL	12	22

Furthermore, duration of prominence in pitch is longer in SL (M = 357.3ms, SD = 244) than in NS (M = 108ms, SD = 199.9) and SH (M = 209.6ms, SD = 92.3). As a result of a one-way ANOVA, there were significant differences in duration of prominence between the groups (F (2, 67) = 8.31, $p < .01$). In particular, SL differed significantly from NS ($p < .01$) and SH ($p < .05$). There was no significant difference between NS and SH ($p = .34$). In summary, SL produced more phonological boundaries and longer duration of the pitch than NS and SH. This is because NS and SH raised pitch only on the stressed syllables,

whereas SL retained high pitch in general and tended to lower pitch on the word accents. The results suggest that pitch of stressed syllables may affect native speakers' evaluation of the pronunciation of Japanese learners of English.

In order to obtain more information about the pronunciation of SL and SH, the researcher interviewed the three native English raters. Each rater had similar impressions about the two Japanese subjects' speech. Although they completely understood the content of SH, none of them understood what SL said. All three raters reported that although SL's intonation showed some variation in pitch, it was flat and lacked a sense of rhythm. Yet, SL had a pitch range of 67.75 Hz, while NS had a pitch range of 40.24 Hz and SH had a range of 54.24 Hz. Thus, despite the fact that SL had the greatest range in pitch, the native speakers nonetheless judged SL's intonation to be "flat." We can consequently suppose that this must be because SL accentuated speech in a way that differed from the habits of native speakers. In short, SL pronounced stressed syllables by lowering pitch, while NS produced stressed syllables by raising pitch.

Figure 4 to 6 present respectively the waveforms of Sentence 1 by NS, SH, and SL. In NS's waveform (see Figure 4), each word has one or two areas where the amplitude is large. The same tendency can be seen in the waveform of SH (see Figure 5). On the other hand, in the waveform of SL (see Figure 6), more than two areas of some words display large amplitude. This is because the subject with low evaluations inserted vowels after

consonants and pronounced each syllable with equally strong intensity. That is, SL employed a Japanese syllable structure when pronouncing English.

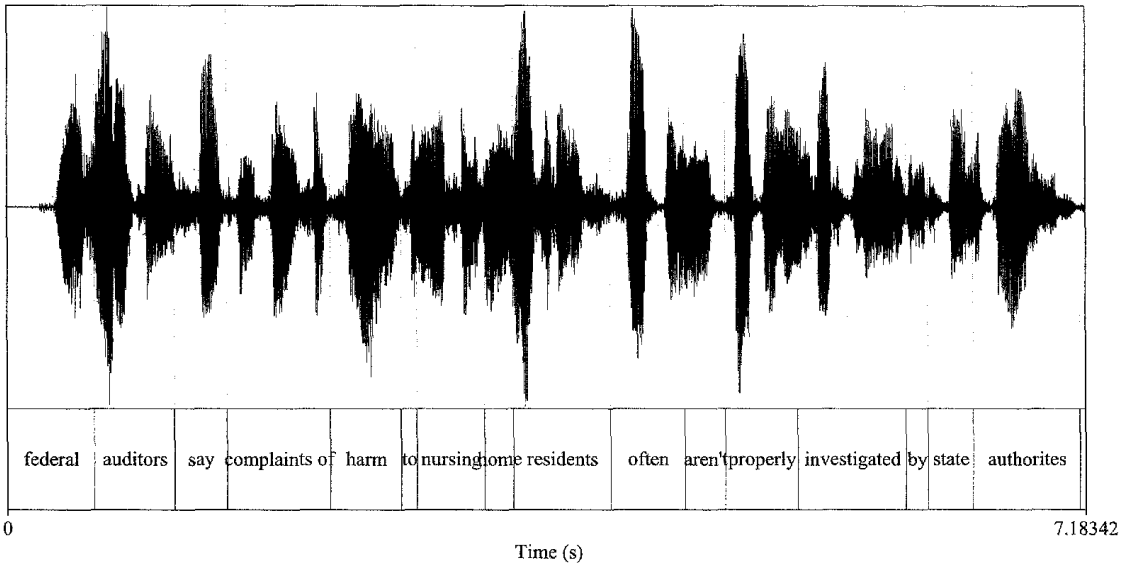

Figure 4. The waveform of NS.

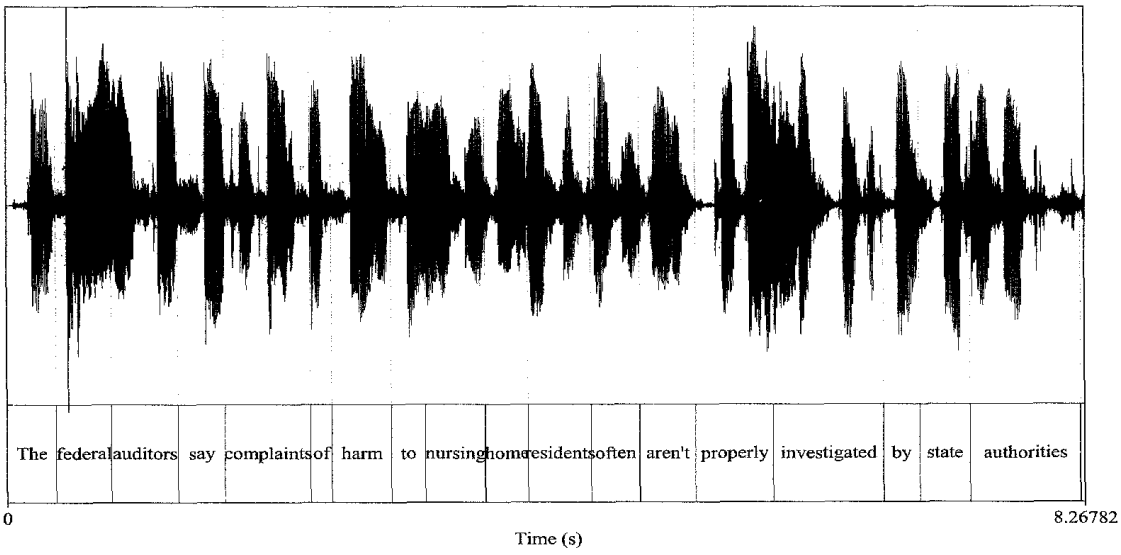

Figure 5. The waveform of SH.

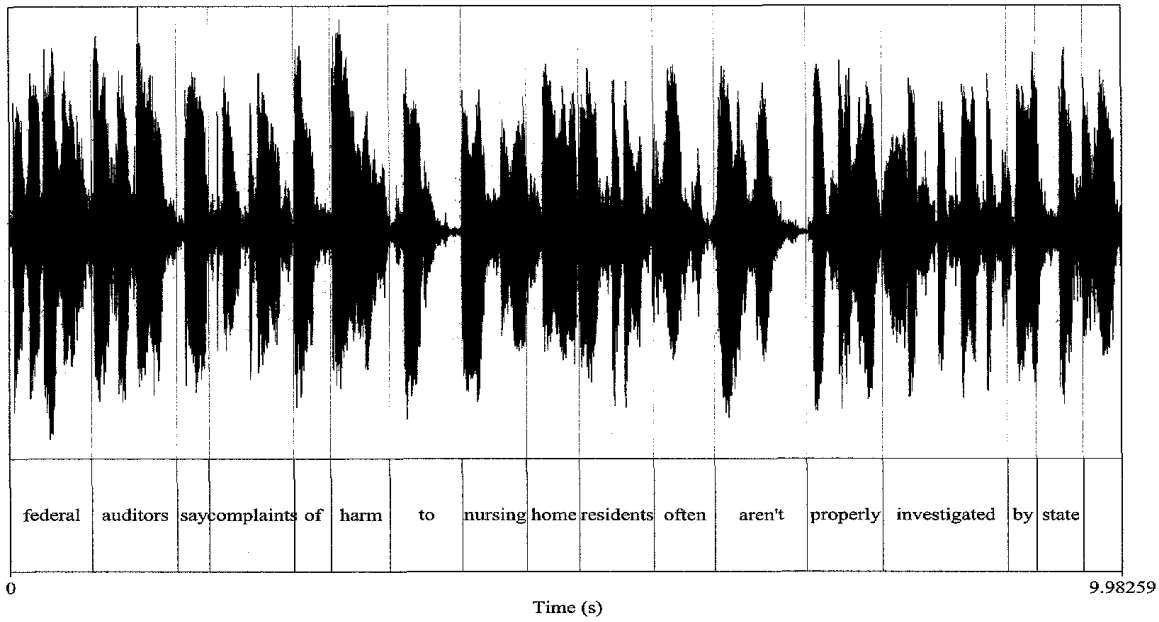

Figure 6. The waveform of SL.

6. Discussions

What this study indicates is that SL fails to pronounce English stresses and syllables, but, instead, employs Japanese pitch accent for stress and inserts vowels after consonants for English consonant clusters. As a result, his speech is lengthened to accommodate the number of equally accented syllables. A difference may also be noted between SH and SL evaluations as to the production of stress. On stressed syllables, SH raises the pitch, and his stressed and unstressed syllables are clearly contrastive; He pronounces consonant clusters in the proper way, while producing distinctive sentence stresses. This makes the speech of SH close to that of the native English speaker. These results also support the predictions: when

Japanese learners begin to speak English, they use pitch accents to denote stresses, and that this makes it hard for native speakers to comprehend their speech. In addition, when Japanese learners articulate each word distinctively, as they do in their own language, they tend also to produce phonological boundaries between every word, which can cause a breakdown in communication. Although learners are encouraged to pronounce words clearly, a focus on the segmental level might actually hinder, rather than help, native speakers' comprehension.

Next, this study has raised the following research question: What is the threshold level of suprasegmentals in L2 pronunciation? The results show that SL decreased pitch on stressed syllables, inserted vowels after consonants, and broke up an utterance into discrete phonological phrases. Thus, as an answer of the research question, learners should highlight stressed syllables, as well as produce phonological phrases. The results of this study also support the previous studies. Japanese speakers recognize an accent in the syllables with falling pitch (Masaki et al., 2000) because, in Japanese, pitch falls rapidly between a syllable with an accent and the syllable that follows it (Kubozono, 1997). Because pitch falls characterize the speech of SL, we can assume that the speaker produced the stressed syllables as he does in Japanese, which may be one of the reasons why native speakers could not understand his speech. We may conclude that SL was reproducing the pronunciation patterns of his first language, and since the native English speakers were not able to perceive SL's pitch falls

in the stressed syllables, they might well have felt, wrongly, that the speech lacked rhythm.

In summary, the poor suprasegmental management of SL's speech exhibits interference by the speaker's first language. When Japanese learners of English produce utterances, a failure to manage the stressed syllables, consonant clusters, and phonological phrases of the target language can affect the rating by native speakers of the meaning of the utterance. SL's way of producing these factors did not accord with the way native speakers use or perceive them.

7. Conclusion

This study has attempted to examine how suprasegmental features affect an L2 learner's pronunciation and cause a breakdown in communication with native English speakers. It has confirmed that incomprehensible pronunciation can be caused by negative transfer from the first language, such as the lowering of pitch on stressed syllables, inserting vowels after consonants, and the breaking up of an utterance into discrete phonological phrases.

This study has analyzed only one lowly evaluated speaker, and this is obviously a serious limitation since different speakers are likely to deploy suprasegmental features in different ways: if we hope to confirm our results, we shall therefore need to investigate a much greater number of participants. Another limitation has been that since the participants read the English text aloud, other linguistic factors, such as knowledge of

orthography and vocabulary, might have affected the participants' performances and hence their evaluations. We shall need to look later at spontaneous utterances as well. I hope nevertheless that this study will shed light on at least one aspect of the pronunciation of L2 speakers. In order to isolate the pedagogical implications, we shall require further study of what makes for the intelligible pronunciation of L2 speakers.

Notes.

1. available from www.praat.org

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgment of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42(4),* 529-555.

Celce-Murcia, M., Brinton, M. D., & Goodwin, M. J. (2003). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and language, 51*, 523-547.

Kubozono, H. (1997). *Gokeisei to Oninkozo* [Word formation and phonological structure]. Tokyo: Kuroshio-shuppan.

Kubozono, H. (1998). *Onseigaku oninron* [Phonetics and

phonology]. Tokyo: Kuroshio-shuppan.
Kubozono, H. (2001). *Oninkozo to akusento* [Phonological structure and accent]. Tokyo: Kenkyusha-shuppan.
Ladefoged, P. (2001), *A Course In Phonetics/ Fourth Edition.* Boston: Thomas Learning.
Masaki, A., Takazawa, M., & Arai, T. (2000). *Goakusento no chikaku niokeru bogo no eikyo nitsuite—nihongo, doitsugo, supeingowasha no hikaku*—[Effects of first language on perception of word accent—comparison between Japanese, German, and Spanish speakers]. *Proceedings of the Phonetic Society of Japan, Japan,* 165-170.
McCrostie, J. (2004). Review of the booklet. Teaching pronunciation: Why, what, when, and how. *JALT Journal, 26(2),* 271-272.
Munro, J. M. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17,* 17-34.
Riney, J. T., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly, 34(4),* 711-737.

Appendix:

The sentences below is the text to read aloud for the participants.

Sentence 1

Federal auditors say complaints of harm to nursing home residents often aren't properly investigated by state authorities.

Sentence 2

The General Accounting Office report says that although the federal government provides most of the money to states for nursing home inspections, it has done little to monitor their response to complaints from families, residents and employees.

The sentences below show the enunciation patterns of the three speakers. The slashes indicate the boundary of phonological phrases, while bold letters show the segments with prominence in pitch, followed by their duration in the parentheses.

Sentence 1

The native English speaker:

Federal (59ms) auditors say / complaints of harm (89ms) to / nursing home residents (70ms) / often (35ms) aren't / properly (44ms) investigated by / state authorities (135ms).

SH:

Federal (180ms) auditors say / complaints of / harm (50ms) to nursing / home (60ms) residents / often aren't / properly (140ms) investigated (60ms) / by state (86ms) authorities.

SL:

Federal (290ms) / auditors (170ms) / say (260ms) / complaints (610ms) of / harm (640ms) to / nursing home (1000ms) / residents (270ms) / often (390ms) / aren't (740ms) / properly (220ms) / investigated (40ms) by / state authorities (40ms).

Sentence 2

The native English speaker:

The General (63ms) Accounting Office report / says that although (110ms) the federal government / provides (193ms) most of the money to states / for nursing home inspections (78ms), / it has done little (69ms) to monitor / their response (37ms) to complaints / from families (401ms), residents (150ms) / and employees (87ms).

SH:

The General (162ms) Accounting Office / report (434ms) says / that although (148ms) the federal government / provides (748ms) / most (52ms) of / the money (85ms) to / states (343ms) / for nursing home (196ms) inspections, / it has (110ms) done / little (61ms) to monitor / their response (79ms) / to complaints from families (416ms), / residents (573ms) and employees.

SL:

The General (217ms) / Accounting (775ms) Office / report (126ms) says that / although (155ms) the / federal (882ms) / government (253ms) / provides (204ms) / most (200ms) of / the money (168ms) to / states (50ms) for / nursing (360ms) / home (283ms) / inspections (549ms),/ it (301ms) / has done (457ms) / little (149ms) to / monitor (400ms) / their response (267ms) / to complaints (553ms) / from families (281ms), / residents (239ms) / and employees (608ms).