

## Characteristics of Contrast between the Stressed and the Unstressed in Rhythm Units Observed in Duration Structure in English Speech by Japanese Learners

Shizuka Nakamura\*

Waseda University / Japan Society for the Promotion of Science

**Nakamura, S. (2011). Characteristics of contrast between the stressed and the unstressed in rhythm units observed in duration structure in English speech by Japanese learners. *Journal of Pan-Pacific Association of Applied Linguistics*, 15(1), 177-189.**

English rhythm is related to contrast between the stressed and the unstressed in duration structure. In native English speech, in general, an intra-speaker average duration of stressed syllable is longer than that of syllable as a whole. On the contrary, that of unstressed syllable is shorter than that of syllable as a whole. In the previous paper by the present author, it was reported that stressed syllable durations of speech of learners tend not to be lengthened enough, and an unstressed ones tend not to be shortened enough as compared with that of native speakers (Nakamura, 2010). For these reason, learner speech does not have such a high ratio of intra-speaker average durations of stressed to unstressed syllables as native speech. Since this lower ratio affected subjective evaluation, a correlation coefficient between the ratios and subjective evaluation scores given by English language teachers was observed to be 0.48. In this paper, an indicator, which demonstrates more adequately a duration contrast between the stressed and the unstressed, was investigated, for the purpose of increasing the correlation with subjective evaluation. A rhythm unit was defined here as a stressed syllable connected to the preceding and succeeding unstressed syllables. Then, a value based on a ratio of stressed to unstressed syllable durations in the rhythm unit was treated as an indicator to represent learner characteristics. As a result, a correlation coefficient of the indicators with subjective evaluation scores was increased to 0.65. A substantial part of mechanism in subjective evaluation of rhythm in English speech was revealed and became possible to simulate reasonably by objective evaluation.

**Key Words:** rhythm in English speech, duration characteristics, the stressed and the unstressed, subjective evaluation, Japanese learners.

---

\* This study was supported in part by the Grant-in-Aid for Research Fellow No. 21 · 06162, Japan Society for the Promotion of Science. The research project “Objective evaluation of English speech uttered by learners based on mathematical model reflecting prosodic control and perceptual characteristics” including this work is supervised by Prof. Yoshinori Sagisaka (Waseda University). The author also wishes to thank Prof. Michiko Nakano (Waseda University) for her valuable advices on editing this article.

# Shizuka Nakamura

## 1 Introduction

As internationalization has advanced, the demand to acquire the ability to speak English has increased. When we consider ways of evaluating English speech of learners, it is desirable to develop the strategy of using subjective evaluation by English language teachers to a more precise and reliable stage. To this end, it is necessary to analyze multi-dimensionally the strategy of subjective evaluation by teachers and to find how they utilize the acoustical characteristics of learner speech in their hearing capacity. The analyzed strategy of subjective evaluation can be replaced by a more effective objective evaluation system by using a computer (Yamashita, 2005; Ito, 2006; Nakano, 2008). The present author has studied the strategy of this kind of evaluation (Nakamura, 2007 & 2009). In this process, intrinsic and significant knowledge about the relationship between the acoustical features of learner speech and subjective evaluation was obtained. These results are reported in this paper.

Stress characterizes rhythm in English speech. The physical quantities of acoustical features that relate to stress are duration, fundamental frequency, and intensity (Lehiste, 1970). They correspond to the psychological quantities of phone length, pitch, and loudness, respectively. Among the acoustical features related to the subjective evaluation of rhythm in English, durations, which are the basis of the duration structure, are focused on in this paper for the following reasons: 1) Duration can be thought to include most of the information of fundamental frequency and intensity, and 2) The information of duration, which is based at the start and end points of each phoneme unit, can be measured with relatively high reliability.

Learners aim to control rhythm in English as native speakers do. Therefore, characteristics of learner speech are analyzed by comparison with those of native speakers for the same texts. The obtained characteristics are indicators of proficiency levels of rhythm in English and can be used for simulating subjective evaluation. In this paper, this indicator to simulate subjective evaluation with higher precision is investigated.

## 2 Analysis Data

In this chapter, speech data and subjective evaluation scores used for analyses are presented.

### 2.1 Speech data

Speech data used for analyses were selected from the “English speech database read by Japanese students (hereafter “ERJ database”) (Minematsu, 2003)” which includes texts satisfying all requirements mentioned in the

## The Stressed and the Unstressed in Rhythm Units

following part of this section. This database consists of the English speech, which was uttered by learners of a wide range of English proficiency levels and recorded in a standardized recording environment.

Texts were selected as shown in Table 1. They satisfy all requirements of texts for evaluation of rhythm in English as set in the following part of this section. These requirements are explained by quoting examples from this table.

### 2.1.1 Texts

The target of evaluation in this study is a proficiency level of not a phonological aspect of a specific word but a prosodic one in a whole sentence, especially rhythm in English. For this reason, it is desirable that there is no difference between learners in their knowledge of the words included in the texts. Therefore, texts mainly consist of the simple words required in English classes in Japanese junior high schools and contain no proper noun are selected.

Stress characterizes rhythm in English speech. The reason that a stressed syllable is recognized as a syllable with a stress is the following: it is heard to stand out more prominently than its immediate unstressed syllables by longer duration, greater intensity, and higher pitch (Roach, 2009). Repetition of these stressed syllables alternating with unstressed syllables makes hearers perceive rhythm (Lehiste, 1970). In this study, focusing on the property of salient stressed syllables in the repetition, the number of stressed syllables included in a text is treated as an element of a requirement of texts.

In the case of a text including two stressed syllables, one interval

Table 1. Texts Used for Analyses with Symbols Indicating the Location and Degree of Stress.

Text
A I'm a • mused by the man and his ver • y fun • ny jokes. - - @ - - @ - - - - - @
B Why won't you wait un • til Fri • day when he's back? - - - @ - - @ - - - @
C I was ter • ri • bly an • noyed with the man for beat • ing the dog. - - - - - @ - - @ - - - @
D The boys have sold some of the flow • ers. - @ - @ - - @ -
E Thank you ver • y much for eve • ry • thing that you did for us. - - - @ - @ - - - @ -

Note. The symbol “.” stands for a syllable boundary. Stressed (@) and unstressed (-) syllables are based on the definitions described in 2.1.1. between these stressed syllables is formed and its absolute duration is

## Shizuka Nakamura

perceived. On the other hand, in the case of a text including at least three stressed syllables, at least two intervals between adjacent these stressed syllables are formed and its periodic repetition is also perceived. For these reasons, a text including at least three stressed syllables is set as a requirement of texts for evaluation of rhythm in English.

The location and degree of stress can be changed in some cases according to a general English rule to avoid having stresses too close, but to maintain regular intervals (Ladefoged, 1975). These phenomena are not necessarily done by all native speakers in common (Roach, 2009). Furthermore, the degree of stress is not paid attention to even by native hearers in an ordinary situation of hearing a speech (Jones, 1960).

Considering these facts, just a syllable with a primary stress, which stands out prominently in contrast to an unstressed syllable, is treated as a stressed syllable in this study. This can be useful for weakening an effect of a difference in the way evaluators recognize stress and having a clearer result. Hereafter, just a syllable with a primary stress is called a stressed syllable, and the other syllable is called an unstressed syllable. The expressions of stressed and unstressed syllables in tables and figures in this paper are also based on the definition.

Considering a limitation of accurate evaluation by human evaluators, a simple sentence or complex one which consists of two pairs of a subject and a predicate is selected as texts. It is desirable that the number of these words is up to about 7 in a simple sentence and 14 in a complex one in view of the structure of a normal sentence. The number of words in each text used in this study introduced in the next section is from 8 to 11 and within this limit. The number of syllables is from 9 to 15.

Texts were selected to meet the all above requirements. As shown in the second line of each text in Table 1, every selected text includes three stressed syllables indicated by the symbol “@.”

### 2.1.2 Speakers and the number of samples

One hundred and six samples were selected for speech samples of learners. Speakers were 106 university students whose native language was Japanese. The number of samples per text was approximately 21. In the process of constructing the ERJ database, speech samples uttered by native speakers were not presented as references during practices and recordings. Additionally, learners were given prosodic symbols indicating location and degree of stress in the texts and required to practice speaking them prior to the recordings. The location and degree of stress shown in Table 1 are based on the one given to the speakers.

The ERJ database also includes speech samples of native speakers for the same texts as those of learners. Fifty eight samples uttered by 20 native speakers, which were corresponding to those by learners mentioned above, were selected. The number of samples per text was approximately 11.

# The Stressed and the Unstressed in Rhythm Units

## 2.2 Subjective evaluation score

English language teachers were asked to give subjective evaluation scores to every selected speech samples of learners. The evaluators were five English language teachers who had knowledge of English phonetics and careers in teaching English to Japanese learners. Evaluators did not include the native speakers who uttered the selected speech described in the last section.

An evaluation measure of a 7-point scale (-3: Awful - +3: Excellent) representing the proficiency level of rhythm in English was used in subjective evaluation. Subjective evaluation scores were given to each sentence. Evaluators were allowed to listen to each speech sample multiple times.

One subjective evaluation score was given to each speech sample of whole sentence by one evaluator as mentioned above. As a result, five scores in total were given to each sample. Based on these row subjective evaluation scores, a representative subjective evaluation score was calculated for each speech sample. The method of calculating a representative score followed the previous paper by the present author (Nakamura, 2010). A representative subjective evaluation score calculated in this way is called just a subjective evaluation score hereafter.

## 3. Duration Characteristics of Learner Speech

In this chapter, characteristics of stressed and unstressed syllables of learners obtained by the present author are summarized. A syllable with a secondary stress was treated as not an unstressed syllable but a stressed syllable in the previous study.

### 3.1 Stressed and unstressed syllables

Learners tend to speak English slower than native speakers, that is, sentence durations of learner speech is longer than those of native speech because of their inexperience of English speech. For this reason, it is natural that an inter-speaker average syllable duration of learners is longer than that of native speakers. In the previous paper by the present author, it was reported that an intra-speaker average duration of stressed syllable uttered by learners tends not to be lengthened enough, and that of unstressed syllable tends not to be shortened enough as compared with that of native speakers (Nakamura, 2010).

As a result of correlation analyses of these durations of learners and subjective evaluation scores, correlation coefficients of -0.13 and -0.43 were obtained for stressed and unstressed syllables, respectively. It was revealed that durations of unstressed syllables rather than those of stressed syllables have a stronger correlation with subjective evaluation.

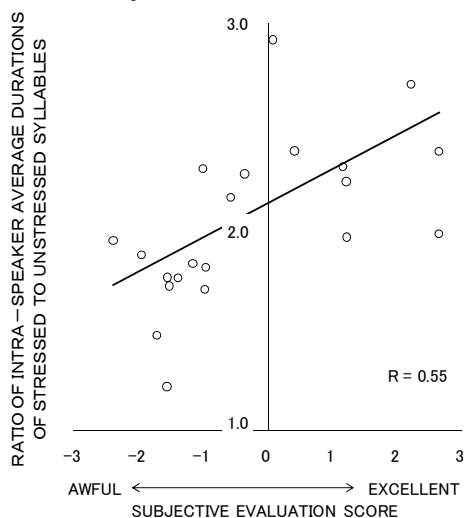
### 3.2 Contrast between stressed and unstressed syllables

Learner speech does not have such a high ratio of intra-speaker average durations of stressed to unstressed syllables as native speech. The relationship between this ratio of learners and subjective evaluation for the example of Text A is shown in Figure 1. Subjective evaluation scores are shown on the horizontal axis, and the ratios on the vertical axis. A correlation coefficient between ratios and subjective evaluation scores showing 0.48 was obtained for all five texts. Though this ratio affects subjective evaluation, the correlation was not strong enough. An effect of a weak correlation of stressed syllable durations with subjective evaluation scores mentioned in the last section is expected to be the reason that this correlation is not improved so much compared to a correlation of unstressed syllable durations with subjective evaluation scores also mentioned in the last chapter.

## 4 Ratio of Stressed to Unstressed Syllable Durations

In the previous study (Nakamura, 2010), relationship between stressed and unstressed syllable durations of learners and subjective evaluation was revealed. Both of them showed weaker correlation coefficients than -0.5. In the last chapter, relationship between ratios of intra-speaker average durations of stressed to unstressed syllables of learners and subjective evaluation scores was analyzed. However, a weaker correlation coefficient than 0.5 was also obtained.

Figure 1. Relationship between ratios of intra-speaker average durations of stressed to unstressed syllables and subjective evaluation scores for the example of Text A.



## The Stressed and the Unstressed in Rhythm Units

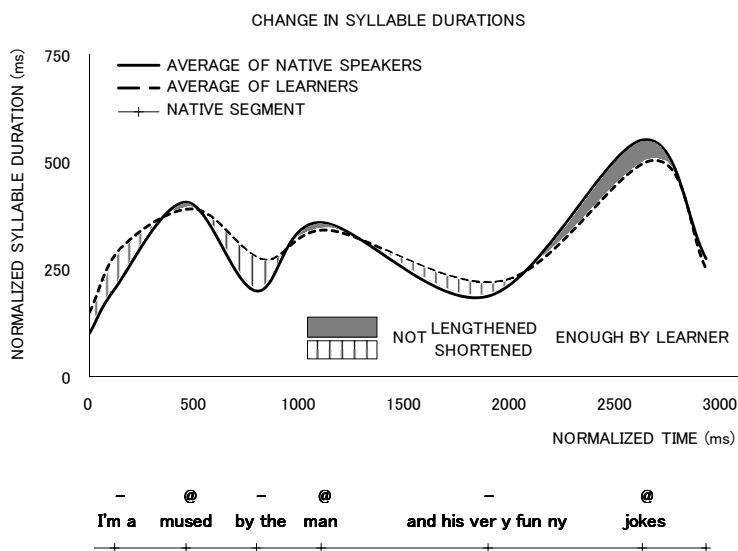
In this chapter, an indicator, which adequately shows a duration contrast between the stressed and the unstressed, was analyzed on the basis of a correlation with subjective evaluation to obtain a stronger correlation. The target correlation coefficient was stronger than 0.5.

### 4.1 Repetition of a set formed by stressed and unstressed syllables

As mentioned in 2.1.1, the repetition of the set of a stressed syllable and the preceding and succeeding unstressed syllables is greatly concerned with rhythm in English. Repetition of this set can make rhythm in English be perceived.

Figure 2 shows a change in each inter-speaker average syllable duration of native speakers and learners alongside time. To express contrast between the stressed to the unstressed methodically, in the case of a series of stressed syllables (hereafter “a stressed part”) or that of unstressed syllables (hereafter “an unstressed part”), these durations were normalized by the number of syllables consisting of the corresponding stressed or unstressed parts. Syllable durations shown on the vertical axis were plotted alongside time on the horizontal axis. Sentence durations of each speaker were normalized by an inter-speaker average of native speakers in order to analyze them after

Figure 2. Comparison of changes in inter-speaker average syllable durations of native speakers (solid line) and learners (dotted line) plotted alongside time for Text A “I’m amused by the man and his very funny jokes.”



Note. The sentence durations are normalized by the average of native speakers. Stressed and unstressed syllables are shown with the symbols “@” and “-” respectively.

## Shizuka Nakamura

eliminating a difference of sentence durations by speaker. It is clear that a set of a long stressed syllable and a short unstressed syllable is repeated in this way.

### 4.2 Learner characteristics

Learner characteristics of this repetition are different from those of native speakers. First, the characteristics of stressed syllable durations are focused on. As shown in Figure 2, three of stressed syllables, which are “mused,” “man” and “jokes,” exist in the example of Text A. They are relatively long and correspond to three of the peaks. The difference indicated by the gray area shows that stressed syllable durations of learner speech tend not to be lengthened enough as compared with those of native speech.

Next, the characteristics of unstressed syllable durations are focused on. To express the difference with stressed syllable durations clearly, unstressed syllable durations are shown at the point of each group of them. These values are averages. For example, at the point of “by the,” the average of “by” and “the” is shown there. As shown in Figure 2, three of groups of unstressed syllables, which are “I’m,” “by the” and “and his ver y fun ny,” exist in the example of Text A. They are relatively short and correspond to three of the troughs. The difference indicated by the area with vertical stripes shows that unstressed syllable durations of learner speech tend not to be shortened as native speech. In the following sections, an indicator to show more adequately these characteristics is investigated.

### 4.3 Calculating ratio in rhythm unit

Learner characteristics revealed in the last chapter can be concealed by calculating a ratio after averaging durations of stressed and unstressed syllables separately. To make the most of these characteristics, it can be helpful to set a rhythm unit and use a ratio of stressed to unstressed parts in each rhythm unit. However, the problem is that where the start and end points of each rhythm unit should be fixed to calculate ratio of stressed to unstressed parts in the repetition of the set formed by stressed and unstressed parts.

A stressed part is saliently perceived. One of the reasons is that an intensity of a stressed part is bigger than that of an unstressed part. Considering the fact, in this paper, a stressed part was assumed to be the center of each rhythm unit. Rhythm unit was formed by a stressed part connecting to the preceding and succeeding unstressed parts. At this point, each unstressed part was divided into two, and the former one connects to the preceding stressed part and the latter one connects to the succeeding stressed part. A rhythm unit was defined as follows:

$$\text{Rhythm unit} = \text{a half of preceding unstressed part} + \text{a stressed part} \\ + \text{a half of succeeding unstressed part}$$

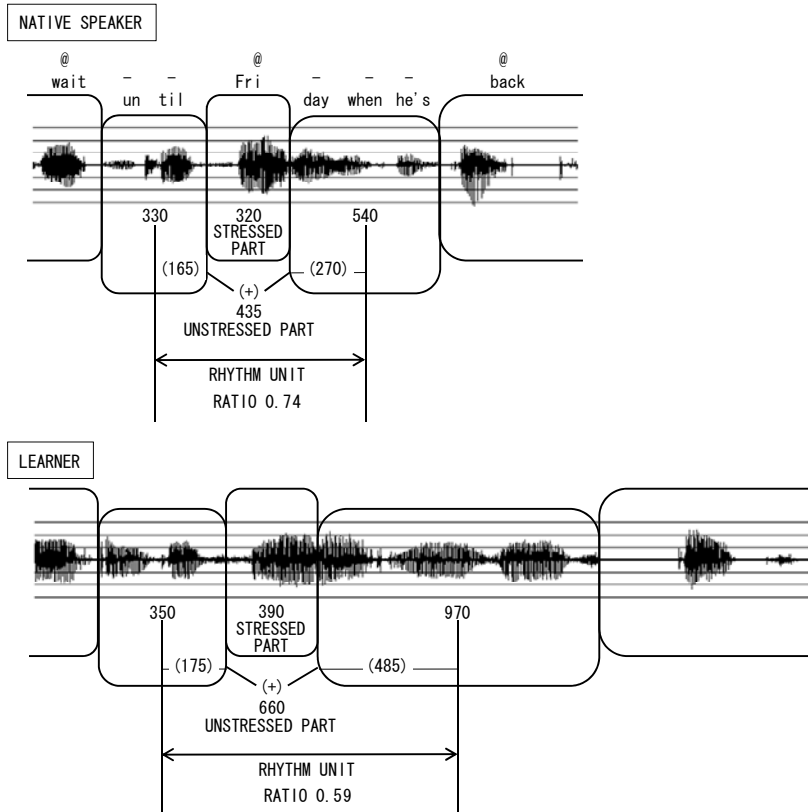


## The Stressed and the Unstressed in Rhythm Units

An intra-speaker average ratio of stressed to unstressed parts in this unstressed + stressed + unstressed (USU) rhythm unit shown in Figure 3 was defined as a basis of an indicator. In the example of this figure, the ratio of a native speaker and a learner are 0.74 and 0.59, respectively.

Figure 3. Speech waveform. Comparison of ratios of stressed to unstressed parts in a type USU rhythm unit of a native speaker (top figure) and a learner (bottom figure)

/ Why won't you wait un · til Fri · day when he's back? /



Note. The waveform is drawn alongside time for “wait until Friday when his back” which is the end portion of Text B “Why won’t you wait until Friday when he’s back?” Prosodic symbols are as follows: @: stressed syllable, -: unstressed syllable, and ·: syllable boundary.

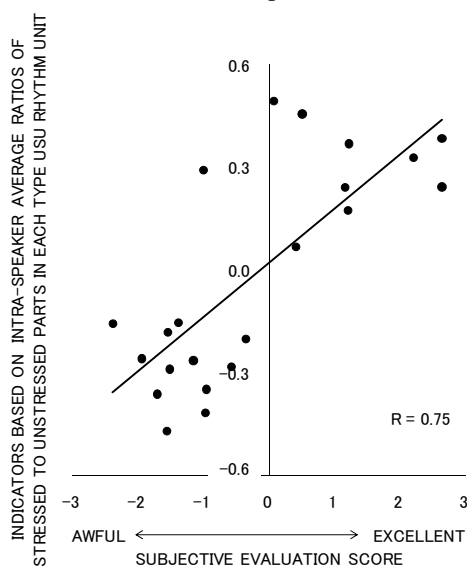
#### 4.4 Relationship between learner characteristics and subjective evaluation

To examine relationship between the basis of an indicator which shows learner characteristics mentioned in the last section and subjective evaluation, the correlation of them was analyzed.

The following arrangements were given to the basis of an indicator to show learner characteristics more adequately. First, to reflect the difference with native speech and the relationship with the other learner speech, the bases of the indicators of each learner were compared to the average of learners after normalizing by the average of native speakers. Next, to obtain the best correlation with subjective evaluation scores, the obtained difference keeping its negative or positive polarity was weighted by a power. An experimental result of the best power was 0.5. Calculated values in this way were defined as indicators to show learner characteristics.

A relationship between the indicators and subjective evaluation scores are shown in Figure 4 for the example of Text A. A correlation coefficient for all five texts was 0.65 as shown in Table 2. In figure 4, subjective evaluation scores are shown on the horizontal axis, and the indicators on the vertical axis. This result was 0.17 stronger than that of the previous study by the present author using the ratio of intra-speaker average durations of stressed to unstressed syllables mentioned in 3.2. There was a significant difference of the results at the 0.01 level

Figure 4. Relationship between the indicators and subjective evaluation scores for the example of Text A.



Note. The indicators are the values based on intra-speaker average ratios of stressed to unstressed parts calculated in each type USU rhythm unit.

## The Stressed and the Unstressed in Rhythm Units

Table 2. Correlation Coefficients between Three Kinds of the Indicators Showing Learner Characteristics and Subjective Evaluation Scores.

Indicator	Classification of a syllable with a secondary stress	Correlation coefficient
Ratios of intra-speaker average durations of stressed to unstressed syllables	Stressed syllable	0.48
Values based on intra-speaker average ratios of stressed to unstressed parts in type USU rhythm unit	Unstressed syllable	0.65

of significance.

Furthermore, the effect of treating a syllable with a secondary stress as not a stressed but an unstressed syllable was confirmed. By treating a syllable with a secondary stress as an unstressed syllable, a correlation coefficient of 0.65 was obtained shown in Table 2. This result was 0.08 stronger than the result by treating a syllable with a secondary stress as not an unstressed but a stressed syllable. There was a significant difference of the results at the 0.01 level of significance.

### 5 Conclusions

An indicator, which demonstrates more adequately learner characteristics of a duration contrast between the stressed and the unstressed, was investigated on the basis of a correlation with subjective evaluation. A type USU rhythm unit was defined as a stressed part connecting to a half of the preceding and succeeding unstressed parts. An indicator was defined as value based on an intra-speaker average ratio of stressed to unstressed parts in this rhythm unit. In addition, a syllable with a secondary stress was treated as not a stressed syllable but an unstressed syllable.

As a result, a correlation coefficient showing 0.65 of the indicators with subjective evaluation scores was obtained. This correlation coefficient was significantly improved compared to the correlation coefficient when treating a ratio of intra-speaker average durations of stressed to unstressed syllables as an indicator. Obtained results can be used for standardizing an evaluation measure and applying to CALL (Computer-assisted Language Learning) system of evaluating rhythm in English.

## Shizuka Nakamura

### References

- Ito, A., Nakagawa, T., Ogasawara, H., Suzuki, M., & Makino, S. (2006). Automatic detection of English mispronunciation using speaker adaptation and automatic assessment of English intonation and rhythm. *Educational Technology Research*, 29, 13-23.
- Jones, D. (1960). *An outline of english phonetics*. W. Heffer & Sons LTD.: Cambridge and Maruzen Company LTD.: Tokyo.
- Ladefoged, P. (1993). *A Course in phonetics*. Fort Worth, Texas: Harcourt Brace Jovanovich.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge: MIT Press.
- Minematsu, N., Nishina, K., & Nakagawa, S. (2003). Read speech database for foreign language learning. *Journal of Acoustical Society of Japan*, 59, 345-350.
- Nakamura, S., Tsubaki, H., Kondo, Y., Nakano, M., & Sagisaka, Y. (2007). Tempo-normalized measurement and test set dependency in objective evaluation of English learners' timing characteristics. *Proceedings of the 16th International Congress of Phonetic Sciences*. 1733-1736.
- Nakamura, S., Matsuda, S., Kato, H., Tsuzaki, M., & Sagisaka, Y. (2009). Objective evaluation of English learners' timing control based on a measure reflecting perceptual characteristics. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. 4834-4840.
- Nakamura, S. (2010). Analysis of relationship between duration characteristics and subjective evaluation of English speech by Japanese learners with regard to contrast of the stressed to the unstressed. *Journal of Pan-Pacific Association of Applied Linguistics*, 14(1), 1-14.
- Nakano, M., Kondo, Y., & Tsutsui, E. (2008). Fundamental research on automatic speech evaluation. *Proceedings of the 9th APRU Distance Learning and the Internet Conference—New Directions for Inter-institutional Collaboration: Assessment & Evaluation in Cyber Learning*. 207-212.
- Roach, P. (2009). *English phonetics and phonology*. Cambridge University Press: Cambridge.
- Yamashita, Y., Kato, K., & Nozawa, K. (2005). Automatic scoring for prosodic proficiency of English sentences spoken by Japanese based on utterance comparison. *IEICE Transactions on Information and Systems*, E88-D, 496-501.

## **The Stressed and the Unstressed in Rhythm Units**

Shizuka Nakamura  
Global Information and Telecommunication Studies,  
Waseda University  
1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan  
Tel: +81 3 5286 3841  
E-mail: shizuka@akane.waseda.jp

Received: February 25, 2011  
Revised: May 27, 2011  
Accepted: June 15, 2011