

## Rhythmic Differences between Spontaneous and Read Speech of English

Kim, Sulki<sup>1)</sup> · Jang, Tae-Yeoub<sup>2)</sup>

### ABSTRACT

This study investigates whether rhythm metrics can be used to capture the rhythmic differences between spontaneous and read English speech. Transcription of spontaneous speech tokens extracted from a corpus is read by three English native speakers to generate the corresponding read speech tokens. Two data sets are compared in terms of seven rhythm measures that are suggested by previous studies. Results show that there is a significant difference in the values of vowel-based metrics (VarcoV and nPVI-V) between spontaneous and read speech. This manifests a greater variability in vocalic intervals in spontaneous speech than in read speech. The current study is especially meaningful as it demonstrates a way in which speech styles can be differentiated and parameterized in numerical terms.

**Keywords:** spontaneous speech, read speech, rhythm metrics, English rhythm, speech style

### 1. Introduction

Spontaneous speech and read speech are presumably different in speaking style from each other as they may have different prosodic characteristics including intonation, speaking rate, and rhythm. It is spontaneous speech that we are exposed to most of the time in our daily life rather than read speech. Nevertheless, most phonetics studies have focused on analyzing highly controlled read speech, and there remain various issues which call for further research with regard to spontaneous speech.

Some studies have investigated spontaneous speech on pauses (Rochester, 1973), temporal patterns (Henderson *et al.* 1966), F0 contour (Lieberman *et al.*, 1985), consonant reduction, and energy difference (van Son & Pols, 1996). Along with these studies,

Gibbon and Gut (2001) analyze rhythmic differences between British English and Nigerian English using semi-spontaneous speech and read speech. As a methodology, they employ a rhythm metric Rhythm Ratio, which is a calculation formula devised to capture durational variability. Even though their study investigates the rhythmic difference between the two varieties of English, the results imply the possibility of using computed metrics as a device to capture rhythmic characteristics of different speech types within a language.

In this study, we attempt to investigate whether rhythm metrics suggested by relatively recent research (Ramus *et al.*, 1999; Ramus, 2002; Grabe & Low, 2002; Dellwo & Wagner 2003; Dellwo 2006) can be used to capture rhythmic differences of spontaneous and read speech of English. We apply seven rhythm metrics %V,  $\Delta V$ ,  $\Delta C$ , VarcoV, VarcoC, nPVI-V and rPVI-C, which will be described in more detail in the following section, to spontaneous speech data extracted from the Buckeye Speech Corpus and the read speech utterances produced by three native speakers of English.

### 2. Definition of Rhythm Metrics

Recent studies such as Ramus *et al.* (1999), Ramus (2002),

1) Hankuk University of Foreign Studies, sulkikim@gmail.com

2) Hankuk University of Foreign Studies, tae@hufs.ac.kr,  
corresponding author

This study is extension of the preliminary version which was presented in *The 2009 Spring Conference of Korean Society of Speech Sciences* at Chonbuk National University on May 8.

Received: July 26, 2009

Revision received: August 20, 2009

Accepted: September 10, 2009

Grabe & Low (2002), Dellwo & Wagner (2003), Dellwo (2006) have devised rhythm metrics which are based on the acoustic measurements of vocalic and intervocalic variability. By applying the rhythm metrics, they attempt to capture rhythmical differences of various languages. Ramus *et al.* (1999) attempt to prove that vocalic and intervocalic intervals can account for the traditional rhythm dichotomy and investigate the possibility of other types of rhythm. They propose three rhythm metrics:

- (1) %V (the average proportion of vocalic intervals)
- (2)  $\Delta C$  (the average standard deviations of consonantal intervals)
- (3)  $\Delta V$  (the average standard deviations of vocalic intervals)

With these variables, they investigate eight languages (English, Dutch, Polish, French, Spanish, Italian, Catalan, and Japanese) with four speakers per language and five sentences per speaker, constituting a set of 160 utterances. They conclude that (%V,  $\Delta C$ ) projection best fits to the standard rhythm classes rather than (%V,  $\Delta V$ ) or ( $\Delta V$ ,  $\Delta C$ ) projection. For both %V and  $\Delta C$ , a significant effect of rhythm class is found ( $p < 0.001$ ) but no significant class effect is found with  $\Delta V$ . Based on the phonological accounts of speech rhythm (Dauer, 1983), they discuss that  $\Delta C$  and %V appear to be directly related to syllable structure. A language which permits more syllable types tends to have a greater variability in consonants, which would lead to higher value in  $\Delta C$ . Furthermore, this explains the difference of %V since the value would be lowered in a language where a greater consonant variability is permissible. Also, in a language such as French or Spanish which has only a simple CV structure (Dauer, 1983), it would result in higher value in %V.

Dellwo (2006) suggests a rate-normalized metric, VarcoC. Based on the observation by Dellwo and Wagner (2003) that  $\Delta C$  is susceptible to speech rate in the inverse direction, he calls for normalization for  $\Delta C$  through dividing the value by the mean consonantal duration.

$$(4) \text{VarcoC} = \frac{\Delta C \times 100}{\text{mean}C}$$

With VarcoC, Dellwo (2006) investigates three languages (English, German and French) with five versions of different speed: normal, slow, even slower, faster and the maximum speed. He argues that VarcoC is more effective than  $\Delta C$  in that the pattern is more clearly clustered since there is no overlap in the

values of VarcoC of English and German (stress-timed) and French (syllable-timed).

VarcoV is proposed in Ferragne and Pellegrino (2004) to monitor relative  $\Delta V$  variation across speech rates.

$$(5) \text{VarcoV} = \frac{\Delta V \times 100}{\text{mean}V}$$

They investigate British English dialects with the speech materials from the Accents of the British Isles corpus on durational aspects. The results suggest that VarcoV is a more appropriate metric, especially when the compared materials show a great variability in speech rates.

Low *et al.* (2000) and Grabe and Low (2002) suggest two other rhythm metrics, the raw Pairwise Variability Index (rPVI) and the normalized Pairwise Variability Index (nPVI). The equation for rPVI is:

$$(6) \text{rPVI} - C = \left( \sum_{k=1}^{m-1} |d_k - d_{k+1}| \right) / (m-1)$$

where  $m$  is the number of items in an utterance and  $d$  is the duration of the  $k^{\text{th}}$  item.

The equation for nPVI is:

$$(7) \text{nPVI} - V = 100 \times \left( \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| \right) / (m-1)$$

where  $m$  is the number of items in an utterance and  $d$  is the duration of the  $k^{\text{th}}$  item.

Grabe & Low (2002) further explain that normalization for vocalic intervals is due to the significant correlation between interval duration and speaking rate, which is in line with Dellwo (2003), Dellwo & Wagner (2006) and Ferragne & Pellegrino (2004). They examined the effect of speaking rate on the rPVI across speakers and across languages. In British English, it is shown that vocalic and intervocalic rPVI values increased significantly as the average interval duration increased across speakers. They conclude that these results make it reasonable to normalize vocalic intervals since the majority of vocalic intervals consists of a single vowel that is stretched or compressed when speech rate changes. However, they do not apply normalization for intervocalic intervals since they reflect the cross-language differences in syllable structure. In other words, the very difference of consonant structure between languages would be masked if the normalization is applied. Therefore, they normalize

vocalic intervals (nPVI-V) for speech rate but do not apply normalization on intervocalic intervals, using raw Pairwise Variability Index (rPVI-C).

Applying nPVI-V outlined above, Low *et al.* (2000) investigate differences between British English and Singapore English with regard to vowel duration. They successfully show that there is a lesser variability between a full vowel set and a reduced vowel set in Singapore English, which is considered to be a syllable-timed language and thus does not reduce vowels. Grabe & Low (2002) investigate eighteen languages (British English, German, Dutch and Thai as stress-timed; Tamil, Spanish, French and Singapore English as syllable-timed; Japanese as mora-timed; Polish and Catalan as mixed; and Estonian, Greek, Luxembourg, Malay, Mandarin, Rumanian and Welsh as unclassified languages according to the traditional categorization) with one subject for each language. They apply nPVI-V and rPVI-C to test if these metrics can capture the characteristics of languages that are categorized as different rhythm classes. They claim that the results show that nPVI-V provides discrimination between Dutch, German and British English, and French and Spanish, which is in line with the rhythmic categorization by Pike (1945) and Abercrombie (1967). nPVI-V scores for Catalan and Polish are located between the two groups. However, it is found that rPVI-C scores do not show a clear distinction for these languages.

The metrics outlined above have also been utilized with a various combinations in a number of comparative studies of rhythm between first and second language, or between different varieties of a language.

White & Mattys (2007) apply %V,  $\Delta V$ ,  $\Delta C$ , VarcoV, VarcoC, nPVI-V, and rPVI-C to compare Spanish, French, English and Dutch as first and second languages. They find the VarcoV as the most effective metric in quantifying the influence of first language on second language and %V to be the next effective metrics.

Jang (2008) applies above seven metrics and suggests three other metrics %FW (proportion of function words within the sentence), speech rate and Num-sil (number of silence/pause intervals within the sentence), to investigate whether various rhythm metrics can be used to capture the non-native rhythm structure of the utterances spoken by Korean learners of English. Speech tokens by 27 Korean learners are stratified to four levels of proficiency and compared with the production of 9 native speakers in terms of rhythm scores. He finds that five metrics %V, VarcoV, nPVI-V, rate and Num-sil are effective for distinguishing at least one proficiency level from others.

Gibbon and Gut (2001) apply the rhythm metrics to semi-spontaneous speech of British English and Nigerian English. Their study is especially meaningful in that speech materials are extended beyond highly controlled read speech. A corpus of one British English speaker reading and retelling a story, and a Nigerian English corpus of four speakers reading and three speakers retelling a story is used. They apply the Rhythmic Ratio (RR), a slightly modified version of nPVI, where  $RR = nPVI/2$ , to syllables and vowels. In their study, it is shown that the British English speaker shows higher values of RR in vowels in semi-spontaneous speech than in read speech, whereas Nigerian speakers show the results of opposite tendency. They conclude that the British speaker produces greater variation of vowels in semi-spontaneous speech than in read speech. Even though their purpose of study is to find rhythmic differences between the two varieties of English, the results indicate the possibility of using a rhythm metric as an apparatus to capture rhythmic characteristics of different speech types within a language. However, their study is restricted in that only one British English speaker has participated in the experiment. Also, segmental composition of spontaneous and read speech has not been secured, which could possibly affect the result values of RR.

### 3. Experiment

As previously stated, most of the previous studies that employ rhythm metrics have focused on categorizing languages based on the rhythm class hypothesis. In the current study, however, we attempt to investigate rhythmic differences between spontaneous and read speech of English by applying seven kinds of rhythm metrics.

Spontaneous speech tokens are extracted from the Buckeye Speech Corpus (henceforth BSC) (Pitt *et al.*, 2007), and their transcription is used to produce corresponding read speech materials obtained by having three American speakers read the transcribed prompt. Speech tokens of spontaneous speech are limited to continuous and fluent streams of speech. Even though filled or silent pauses and frequent hesitations during utterances are considered to be the important characteristics of spontaneous speech (Clark & Wasow, 1998; Rochester, 1973; Swerts, 1998), it becomes problematic when trying to conclude whether the rhythmic difference has to be attributed to the presence of such disfluencies, or to the fundamental prosodic difference between spontaneous and read speech once they are included in the tokens. In addition, it would be difficult to include the

disfluencies in the script and make subjects read them. The role of disfluencies on speech rhythm is beyond the scope of the current study, and we concentrate on investigating fluently and properly uttered, continuous spontaneous speech along with its corresponding read speech. Then, seven rhythm metrics %V,  $\Delta V$ ,  $\Delta C$ , VarcoV, VarcoC, nPVI-V and rPVI-C are applied to obtained speech tokens.

### 3.1 Spontaneous Speech Tokens

According to Kiesling *et al.* (2006), 40 talkers in the BSC are from Columbus, Ohio, and all are natives of Central Ohio. The sample is stratified for age (under 30 and over 40) and sex, and the sampling frame is limited to middle-class Caucasians. Before recording, participants are told that the purpose of the interview is to find out how people express about everyday opinions in conversation, and the true purpose, which is to build a speech corpus has been later notified.

Among the 40 speakers from the BSC, the first five male and five female speakers, respectively, are selected in the order of appearance to avoid a bias towards the subjects. These 10 speakers are referred to as 'BSC speakers' for brevity. Three females and three males are classified as old, and two females and two males are classified as young according to the speakers' information from the BSC.

A total of 120 utterances (12 tokens per speaker) are selected as tokens for spontaneous speech based on the following criteria. A set of tokens is composed of 12 tokens by each BSC speaker. First, a token consists of one continuous utterance of 8 to 25 prosodic words and is surrounded by silence or other labels indicating utterance boundaries. Second, a token does not contain any utterance-internal pauses labeled as <SIL> or other non-speech vocalizations labeled as <VOCNOISE>. Also, utterances containing filled pauses such as *uh* and *um* are not selected. However, when these filled pauses or discourse markers *and*, *but*, *you know*, *well*, *so* and *because* (Schiffirin 1987) are present at the beginning of the utterance, they are chosen as tokens but the duration of those words are not calculated in the rhythm metrics. Without this kind of selection criteria, tokenization would be extremely difficult due to the characteristics of spontaneous speech where discourse markers are ubiquitously located.

### 3.2 Read Speech Tokens

Two male and one female subjects, who are paid, have participated in the experiment. They are referred to as 'RS

subjects' (subjects of read speech) in this study. They are currently English professors of a university located in Seoul. The abbreviations of the subjects' names are JS, KL, ML and their ages are 25, 26, and 31, respectively. JS is from New Jersey, KL from Los Angeles and ML from Texas. In this study, it is assumed that the regional background, age and gender of subjects are not critical variables affecting speech rhythm. The purpose of the experiment is not disclosed to the subjects before the experiment. None of them have majored in linguistics or phonetics at their undergraduate or graduate education.

A script of the 120 utterance tokens chosen previously from the BSC has been created on the basis of the word level transcription provided by the BSC. Any contractions made by the BSC speakers such as *I will* to *I'll*, or *have not* to *haven't* are all reflected in the script in order to minimize discrepancies between the segmental composition of spontaneous and read speech.

Other than apostrophes as stated above, no punctuation marks are used even though some of the utterances might be seen as containing more than one sentence. This is in line with the findings of Guaitella (1999) that subjects who are asked to place periods at the end of the sentences in the spontaneous speech never agree with one another. They are able to recognize sentences in read speech but not in spontaneous speech. Therefore, it seems unsuitable to mark arbitrary commas or periods in the script. In addition, since the spontaneous speech tokens of the BSC speakers do not contain pauses as stated previously, the inclusion of such punctuation marks would bring about unnecessary pauses or lengthening to read speech.

The recordings are made in a sound-proof booth. An AKG digital microphone is used and speech data are saved to 16-bit mono sound at a sampling rate of 44,100 Hz, using the TASCAM US-122 audio interface.

The subjects are asked to read the script in their normal speed and regard each token as one utterance. They are allowed to read the script again if they make mistakes. After reading 48 tokens, a short break lasting two to three minutes is given, and then the subjects finish reading the rest of the script containing 72 tokens. The experiment lasts approximately 20 minutes for each of the subjects.

### 3.3 Calculation of Rhythm Metrics

Even though the subjects are allowed to read the script again when they make mistakes during the recording, errors have been found during the measurement procedure such as the omission or repetition of words, or inserting words which are not present on

the script. Since four mistakes are found for a set of BSC speaker 02, which is the maximum number of errors in a set, four tokens from each set have to be excluded to make the number of tokens identical for each set. For example, if one of the RS subjects makes one mistake but the other two do not make any mistakes in one set, then the three other tokens which have the longest utterances are excluded. Consequently, a total of 240 (8 tokens per one set of 10 BSC speaker read by 3 RS subjects) tokens are analyzed. The recorded materials are segmented and labeled manually using the speech analysis tool Praat version 5.0.35. The criteria for segmentation are based on the Buckeye Corpus Manual in Kiesling *et al.* (2006) to maintain consistency.

Based on the label files from the BSC (spontaneous speech) and from the segmentation process (read speech), the values for  $\Delta V$ ,  $\Delta C$ , %V, VarcoV, VarcoC, nPVI-V and rPVI-C are calculated automatically through a rhythm metrics calculation program developed in Jang (2008). The program has been slightly modified to exclude the utterance initial stop consonant and the rime of utterance final syllables. This modification, as is also attempted in Jeon (2008), is an effort to evade arbitrary determination of silence duration of utterance initial stop consonants and to eliminate effects of the utterance final lengthening. Along with this modified version, however, the original program is also used to calculate the rhythm metrics which include the durations of the utterance initial stop consonant and the rime of utterance final syllables, check the differences in results. Following Ramus *et al.* (1999), the duration of pre-vocalic glides are counted as consonants, whereas post-vocalic glides are counted as vowels.

#### 4. Results and Discussion

Table 1 shows values of the rhythm metrics calculated by averaging individual values over 80 tokens (8 utterances x 10 BSC speakers) for spontaneous speech, and 240 tokens (80 iterated utterances of BSC speakers x 3 subjects) for read speech. The average articulation duration and speech rate is 2644.28 ms, 5.64 syl/sec for spontaneous speech and 3263.20 ms, 4.64 syl/sec for read speech, which indicates that spontaneous speech is generally faster than read speech.

Table 1. Overall rhythm metrics and t-test results: Utterance initial stop consonant and the rime of utterance final syllables are excluded (SS=spontaneous speech, RS=read speech, \* $p < 0.05$  significant, \*\* $p < 0.01$  highly significant by one-tailed *t*-test)

Metric	SS(N=80)	RS(N=240)	p
%V	45.33	49.37	0.0232
$\Delta V$	52.25	47.33	0.0938
$\Delta C$	57.42	57.76	0.1101
VarcoV	55.64	45.74	0.0030**
VarcoC	54.10	58.55	0.3723
nPVI-V	60.73	49.81	0.0052**
rPVI-C	68.87	62.51	0.0360

According to the *t*-test results, two metrics VarcoV and nPVI-V are shown to be effective for distinguishing spontaneous speech and read speech, which suggests that there is a greater variability in vocalic intervals in spontaneous speech than in read speech.

As shown in Table 2, a *t*-test carried out against each pair of the RS subjects (*p*-values are all above 0.05) suggests that there is no significant difference among the RS subjects. Therefore, values of VarcoV and nPVI-V can be safely considered to reflect the differences of spontaneous speech and read speech. For other metrics %V,  $\Delta V$ ,  $\Delta C$ , VarcoC and rPVI-C, no significant differences are found between the two types of speech.

Table 2. *p*-values on nPVI-V and VarcoV for each pair of the RS subjects: Utterance initial stop consonant and the rime of utterance final syllables are excluded.

Metric	JS-KL	KL-ML	ML-JS
nPVI-V	0.2463	0.3396	0.1505
VarcoV	0.1178	0.2312	0.3995

Meanwhile, Table 3 shows the overall rhythm metrics and *t*-test results when the utterance initial stop consonant and the rime of utterance final syllables are included in the calculation. It is shown that the five metrics nPVI-V (highly significant),  $\Delta V$ ,  $\Delta C$ , VarcoV and rPVI-C (significant) are effective for distinguishing spontaneous speech and read speech. Compared with the rhythm metrics shown in Table 1, where the utterance initial stop consonant and the rime of utterance final syllables are not counted, the rhythm metrics of the additional  $\Delta V$ ,  $\Delta C$  and rPVI-C are shown to be significant. However, since  $\Delta V$ ,  $\Delta C$  and rPVI-C are not rate-normalized metrics, it is likely that those differences are due to the inclusion of the arbitrary durations from the utterance initial stops or the lengthened durations of the utterance final syllables. Rate-normalized nPVI-V and VarcoV, on

the other hand, still seem to be the promising metrics for distinguishing the two types of speech. To ensure that the results regarding VarcoV and nPVI-V are not originated from individual variance of speaking style or other unexpected factors, a *t*-test is again carried out against each pair of the RS subjects on both metrics.

Table 3. Overall rhythm metrics and t-test results: Utterance initial stop consonant and the rime of utterance final syllables are included. (SS=spontaneous speech, RS=read speech, \**p*<0.05 significant, \*\**p*<0.01 highly significant by one-tailed *t*-test)

Metric	SS(N=80)	RS(N=240)	p
%V	46.37	47.69	0.1195
ΔV	51.35	57.81	0.0367*
ΔC	58.51	67.04	0.0205*
VarcoV	55.08	51.39	0.0297*
VarcoC	55.35	57.30	0.1842
nPVI-V	58.22	52.88	0.0042**
rPVI-C	65.78	74.63	0.0199*

Table 4 shows that all the *p*-values are above 0.05 indicating that there is no significant difference among the RS subjects both in nPVI-V and VarcoV.

Table 4. *p*-values on nPVI-V and VarcoV for each pair of the RS subjects: Utterance initial stop consonant and the rime of utterance final syllables are included.

Metric	JS-KL	KL-ML	ML-JS
nPVI-V	0.1541	0.1140	0.0730
VarcoV	0.2172	0.3497	0.1246

The above results are in line with findings by Gibbon & Gut (2001) that the British English speaker shows higher Rhythmic Ratio in spontaneous speech and produces greater variation between vowel durations in semi-spontaneous speech than in read speech. It seems that greater vocalic variability in spontaneous speech stems from the role of prosodic stress in behalf of a speaker's effort to deliver information to an interlocutor as effectively as possible, which is also congruent with the argument of Barry & Andreeva (2001) that the need for and sensitivity to information structure in spontaneous speech allows more variability in rhythmic structure and in accenting and de-accenting. It is notable that the result of VarcoC is meaningful in that it implies there is no significant difference in intervocalic variability between the two types of speech whereas speakers of spontaneous speech do vary vocalic intervals significantly. For

%V, the result is not surprising in that this experiment is a not cross-linguistic investigation on languages that have different consonant cluster phonotactics. Still, the result suggests the possibility that the lengthened durations of stressed vowels in spontaneous speech have been shortened by highly compressed unstressed vowels, which results in similar values of %V for both spontaneous and read speech. In other words, it is possible that speakers of spontaneous speech have uttered unstressed vowels more quickly than in read speech even though it calls for more detailed investigation.

## 5. Conclusion

In this study, the possibility of using rhythm metrics in capturing the rhythmic differences between spontaneous and read speech of English has been investigated. The results show that VarcoV and nPVI-V are the effective metrics for capturing the different characteristics of spontaneous speech and read speech. Greater variability in vocalic intervals seems to represent the characteristics of spontaneous speech, in which a speaker has more freedom to convey information effectively to an interlocutor by giving greater contrast in vowel durations. The current analysis is especially meaningful as it brings up the possibility that speech styles can be modeled in terms of numerical values.

Further research is needed to investigate more detailed characteristics of spontaneous speech rhythm employing other rhythm metrics related to function word characteristics and utterance internal non-speech intervals. In particular, such metrics based on internal pauses and/or hesitation markers are expected to play an important role in characterizing spontaneous speech, which is quite likely to contain various types of disfluency. Finally, a more elaborate design for an experiment would be desirable such as applying rhythm metrics on spontaneous and read speech produced by identical speakers.

## References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Barry, W. & Andreeva, B. (2001). "Cross-language Similarities and Differences in Spontaneous Speech Patterns", *Journal of the International Phonetic Association*, Vol. 3, No. 11, pp. 51-66.
- Clark, H. H. & Wasow, T. (1998). "Repeating Words in Spontaneous Speech". *Cognitive Psychology*, Vol. 37, pp. 201-242.
- Dauer, R. M. (1983). "Stress-timing and Syllable-timing Reanalyzed", *Journal of Phonetics*, Vol. 11, pp. 51-62.

- Dellwo, V. (2006). "Rhythm and Speech Rate: A Variation Coefficient for DeltaC". In P. Karnowski and I. Szigeti (eds.). *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium*, pp. 231-241. Frankfurt: PeterLang.
- Dellwo, V. & Wagner, P. (2003). "Relations Between Language Rhythm and Speech Rate", In *Proceedings of the 15th International Congress of Phonetics Sciences*, pp. 471-474. Barcelona.
- Ferragne, E. & Pellegrino, F. (2004). "A Comparative Account of the Suprasegmental and Rhythmic Features of British English Dialects", In *Proceedings of Modelisations pour l'Identification des Langues*, pp. 121-126. Paris.
- Gibbon, D. & Gut, U. (2001). "Measuring Speech Rhythm", In *Proceedings of Eurospeech*, pp. 95-98. Aalborg.
- Grabe, E. & Low, E. L. (2002). "Durational Variability in Speech and the Rhythm Class Hypothesis", In N. Warner and C. Gussenhoven (eds.). *Papers in Laboratory Phonology* Vol. 7, pp. 515-546. Cambridge: Cambridge University Press.
- Guaïtella, I. (1999). "Rhythm in Speech: What Rhythmic Organizations Reveal about Cognitive Processes in Spontaneous Speech Production Versus Reading Aloud", *Journal of Pragmatics*, Vol. 31, pp. 509-523.
- Henderson, A., Goldman-Eisler, F., and Skarbek, A. (1966). "Sequential Temporal Patterns in Spontaneous Speech", *Language and Speech*, Vol. 9, pp. 207-216.
- Jang, T. (2008). "Speech Rhythm Metrics for Automatic Scoring of English Speech by Korean EFL Learners", *Malsori (Speech Sounds)*, Vol. 66, pp. 41-59.
- Jeon, H. S. (2008). "Rhythm Metrics and Second Language Speech: Korean Learners of English", Abstract. Workshop on Empirical Approaches on Speech Rhythm. UCL, London, 28 Mar 2008.
- Kiesling, S., Dille, L. & Raymond, W. D. (2006). The Variation in Conversation (Vic) Project: Creation of the Buckeye Corpus of Conversational Speech. [www.buckeyecorpus.osu.edu].
- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R. and Miller, M. (1985). "Measures of the Sentence Intonation of Read and Spontaneous Speech in American English", *Journal of the Acoustic Society of America*, Vol. 77, pp. 649-657.
- Low, E. L., Grabe, E. & Nolan, F. (2000). "Quantitative Characterizations of Speech Rhythm: Syllable-timing in Singapore English", *Language and Speech*, Vol. 43, pp. 377-401.
- Pike, K. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Pitt, M. A., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).
- Ramus, F. (2002). "Acoustic Correlates of Linguistic Rhythm: Perspectives", In *Proceedings of Speech Prosody*, pp. 115-120. Aix-en-Provence.
- Ramus, F., Nespore, M., & Mehler, J. (1999). "Correlates of Linguistic Rhythm in the Speech Signal", *Cognition*, Vol. 73, pp. 265-292.
- Rochester, S. (1973). "The Significance of Pauses in Spontaneous Speech", *Journal of Psycholinguistic Research*, Vol. 2, No. 1, pp. 51-81.
- Schiffirin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.
- Swerts, M. (1998). "Filled Pauses as Markers of Discourse Structure", *Journal of Pragmatics*, Vol 30, pp. 485-496.
- van Son, R. & Pols, L. (1996). "An Acoustic Profile of Consonant Reduction", In *Proceedings of 4th International Conference on Spoken Language Processing*, pp. 1529-1532. Philadelphia.
- White, L. & Mattys, S. L. (2007). "Calibrating Rhythm: First Language and Second Language Studies", *Journal of Phonetics*, Vol. 35, pp. 501-522.

- **Kim, Sulki**

Department of English Linguistics  
 Hanyang University of Foreign Studies  
 270 Imun-dong, Dongdaemun-gu, Seoul 130-791  
 Tel: 010-9642-7428  
 Email: sulkikim@gmail.com

- **Jang, Tae-Yeoub**, corresponding author

Department of English Linguistics  
 Hanyang University of Foreign Studies  
 270 Imun-dong, Dongdaemun-gu, Seoul 130-791  
 Tel: 02-2173-3119 Fax: 02-963-8066  
 Email: tae@hufs.ac.kr