THE CLASSIFICATION OF ENGLISH RHYTHM ACQUISITION LEVEL OF CHINESE LEARNERS BASED ON SVM

Pengfei Shao
Institute of Linguistics
Chinese Academy of Social Sciences,
Beijing, China
feipengshao@163.com

Yuan Jia
Institute of Linguistics
Chinese Academy of Social Sciences,
Beijing, China
summeryuan_2003@126.com

Aijun Li
Institute of Linguistics
Chinese Academy of Social Sciences,
Beijing, China
liaj@cass.org.cn

***Abstract*—the present paper examines whether the rhythm measures has been widely used in recent years can reflect the rhythm acquisition ability of second language learners or not. English learners from Shandong province and Jiangsu province of China are chosen here. Aiming at this problem, we put forward to look upon several rhythm parameter as eigenvector, using the SVM (Support Vector Machine) method to model the rhythm acquisition level and classify to different group automatically. First, several rhythm measures are chosen to compare to the former results about English. No significant difference was observed across the former study by %V and Varco_C, whereas the case was not by nPVI_V and rPVI_C. Then we built SVM training model and machine classification showed that rhythm measures could separate the groups at the maxima accuracy of 61%.**

Keywords—rhythm measures; English acquisition; Support vector machines; duration variation

I. INTRODUCTION

Traditional theory considers language into three major categories as stress-timed, syllable-timed and mora-timed from rhythm. English, of course, is described as stress-timed language, while Chinese is thought to be a typical syllable-timed language. Numerous quantitative indices have been developed in attempts to capture the variation in duration that proves the experimental findings [1]. Nevertheless, many other kinds of languages are not so clearly to be classified as one of these three categories. Some of them possess both the features of different kinds. We follow Barry and Russo [2] in calling these indices of durational variation 'rhythm measures' (RMs). Among these RMs, Ramus [3] used the proportion of vocalic material (%V), and the standard deviation of consonantal intervals (ΔC)). Dellwo [4] argues that if ΔC was determined by speech rate it would describe speech rate rather than rhythm. For this reason a variation coefficient (Varco_C) will be calculated in order to monitor relative ΔC variation across speech rates. The Pairwise Variability Index (PVI) is a quantitative measure of acoustic correlates of speech rhythm which calculates the patterning of successive vocalic and intervocalic (or consonantal) intervals showing how one linguistic unit differs from its neighbors. It was first applied by Low [5] in her study of Singapore English rhythm. Results showed that Singapore English had a lower average PVI over utterances than British English. This fits in with the impressionistic observation that Singapore English is more 'syllable-timed' than British English. Detailed descriptions of these RMs mentioned above are listed in Table Ⅰ.

Much work has tried to determine which particular RMs best distinguish languages of different rhythms [6] [7] [8] [9]. Few of them focus on the acquisition level reflected by the classification of some extent rhythm measures. The learners where influenced inevitable by the native language when learning a foreign language. Considering the rhythm, the negative effect was obvious when it comes to the two distinguished rhythms. Loukina [1] compared languages on more than two measures at a time; he got the conclusion that the differences between the five languages can not be captured by only one RM.

Aiming at evaluating the rhythm of English spoken by Chinese learners, we employed several features and parameters that are widely used and accepted in describing rhythm or distinguishing languages of different rhythm classes. The basic idea of this present paper is to classify the category using a SVM(Support Vector Machine) methods.

TABLE I. RMS USED IN THIS STUDY

RMs	Description
%V	Percentage of vocalic intervals
ΔC	Standard deviation of consonantal intervals
Varco_C	ΔC/mean intervocalic duration
nPVI_C	Normalized PVI of consonantal intervals
nPVI_V	Normalized PVI of vocalic intervals
rPVI_C	Raw PVI of consonantal intervals
rPVI_V	Raw PVI of vocalic intervals

II. METHOD

Phonetic analysis is adopted to approach the research goal. To quantify variation in RMs between different groups of level, we applied SVM to classify. Support Vector Machines (SVM) was proposed by Vapnik based on the theory of structural risk minimization from the statistical learning theory [10]. SVM are classifier separating two classes which yields a good generalization performance [11].

A. Speech Data

The present study contained 5 sentences recorded from 30 speakers distributed across Shandong Province (SD, N=15 speakers) and Jiangsu Province (JS, N=15 speakers) from China. Each speaker read the same 5 sentences in English (see below). All the 5 sentences included declarative sentence (N=3), questions (N=1) and Exclamation sentence (N=1) extracted from the AESOP corpus organized by CASS. Each of these sentences contained about 10 syllables. The selected sentences are listed in Table Ⅱ.

TABLE II. FIVE SENTENCES USED IN THIS STUDY

order	sentences
1	It would be better that he give up smoking.
2	Did the ship departed from Germany in the morning?
3	It is dangerous to ride fast on a busy road.
4	What great fun it is to go skating outdoors!
5	The fact that he was killed was a serious matter.

Speakers were 20-26 years old, all born and had grown up in their hometown. At the time of the recording, all speakers were living in the colleges of each district.

B. Speech Segmentation

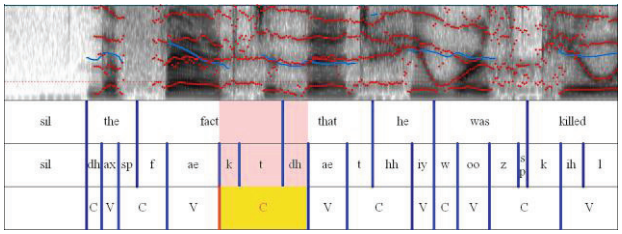

Fig. 1. "*the fact that he was killed*", CV was labeled as the third tier with Praat

The consonant and vowel boundaries discussed here were referred to vocalic and inter-vocalic intervals. We use 'C' indicates the inter-vocalic intervals and 'V' indicates the vocalic intervals. The CV boundary was segmented automatically and then modified by several masters majoring in phonetics. The target tier was labeled as C and V to illustrate, as shown in Fig. 1. The standard criteria were adopted to carry out the decision of the boundary. Additional criteria which facilitated the location of the boundary in certain contexts include:

- Glides followed the vowels or between the vowels don't changes in the shape of successive pitch periods or formants were labeled as V. otherwise, for examples glides initial were labeled as C.
- Nasal voice were labeled as C
- Short pauses in the speech as the inter words as C
- Other pauses such as the hesitations by speakers were excluded from the research

C. Data Extraction

Several rhythm metrics included in Table 1 were all associated with duration variation. So we simply extracted the duration of each C and V labeled by the Part 2.3 and calculated for each of the five sentences spoken by each of the 31 speakers by Praat script automatically. Each sentence got 7 results separately across the RMs in Table 1. Then we totally got 30*5*7=1050 samples.

D. Perceptual experiment

At the same time, we employed 10 students of English majors who are fluent in English and all got the TEM-8 of China to scale from 1 to 5 for all the speakers. One point to be underlined here is that all the graders are told to focus on rhythm only and ignore all the other factors of a speech. We then group all the speakers into two, by their average scores, which are higher group (Group-H) and lower group (Group-L).

An effective method to evaluate the performance of the automatic classification model and the relationship between human raters is consistency coefficient calculation. The correlations between all the graders are all significant at the 0.01 level (2-tailed).

E. SVM Training

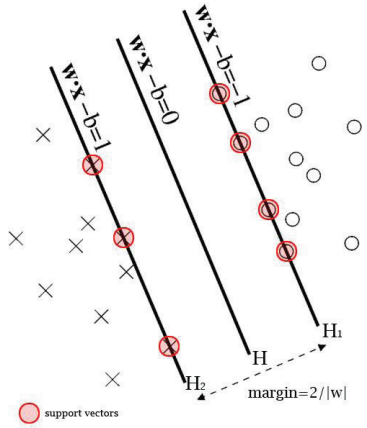

Fig. 2. Illustration of SVM, Line H separates them with the maximum margin. Samples on the margin(on the Line H1, H2) are called the support vectors.

SVM is originally introduced under the condition of linear and divisible and two kinds and is developed to an effectual means to solve the problems of nonlinear and many kinds of pattern recognition. The basic idea is available shown in Fig. 2 of the two dimensional cases. In Fig. 2, the solid and hollow points

represent two types of samples, H for the correct line separate two kinds of sample classification, H1 and H2, respectively, for all kinds of the nearest classification line samples and classification line parallel to the line, the distance between them is called classification interval. The line is called the optimal classification requirements classification line not only can separate two kinds of right (training error rate is 0), and the largest classification interval.

Then we use the 1050 samples to train SVM models to classify the group of the level of English learners. Training process is separated into the following three kinds:

- First, using a combination of several different parameters, corpora in SD area as the training set, corpora in JS area as the testing set
- Second: the same parameter to what JS corpus had as the training set, in SD region for the testing set
- Third: to find out the best parameters combination, part respectively selected as the training set in each place and others as a testing set

III. Result and Discussion

This section compares the RMs between the present study and the former result of English. The aim is to investigate the difference corpus influence on the RMs. Then on the second section, the classification result was given.

A. Comparison with the former results of English rhythm

We firstly compared the results of several parameters combinations to the former analysis. Dellwo [4] computed three languages' rhythm measures on Varco_C and %V, as is shown in Fig. 3. Result of this study is shown in Fig. 4. In this case, Group-L and Group-H are decided by the score perceived through Part 2.4. Group-H stands for the group that their score is relatively higher than Group-L. There is nearly no difference between the result of Dellwo [4] and result from Chinese learners. Group-H has the higher Varco_C than Group-L and the result of English from Dellwo [4].

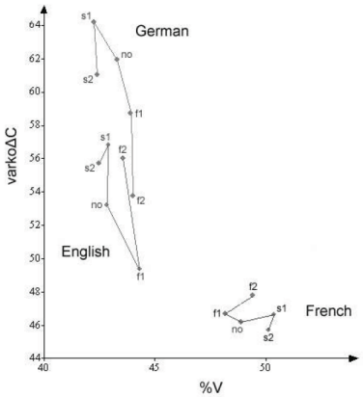

Fig. 3. Results for %V and Varco_C under different intended speech rate conditions(s2,s1,no,f1,f2)for the languages German, French, and English.[4]

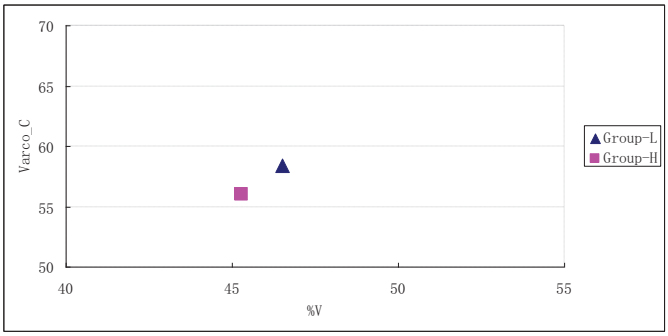

Fig. 4. Results for %V and Varco_C of English spoken by Chinese learners

Fig. 5 shows the data on languages that have often been cited as prototypical examples of stress-, syllable- and mora-timing: British English, Dutch and German (stress-timed), French and Spanish (syllable-timed) and Japanese (mora-timed). Vocalic *nPVI* values are plotted on the vertical axis against intervocalic *rPVI* values on the horizontal axis. Fig. 6 indicates that the PVI measures of English by Chinese learners are quite different from that of English showed by Grabe[9]. If we compare these measures with Fig. 5, we can see English by Chinese learners are closer to the Polish or Catalan. These two languages are described as mixed or unclassified languages.

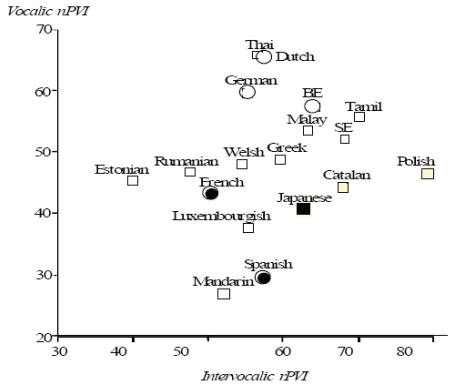

Fig. 5. PVI profiles for data from eighteen languages. Prototypical ○=stress-timed, ●=syllable-timed, ■=mora-timed, □=mixed or unclassified[9]

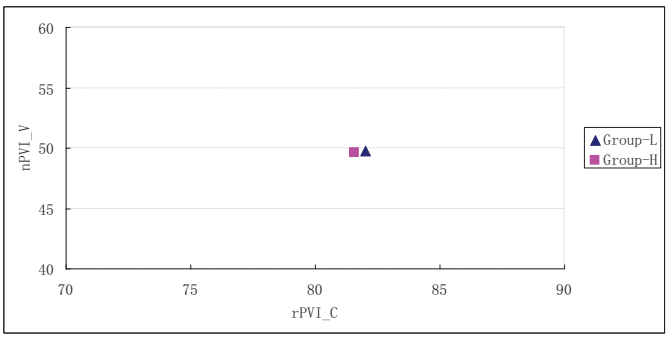

Fig. 6. PVI profiles for data from Chinese learners

B. Results of SVM Classification

From the results, the % V and Varco_C combination has the highest classification accuracy which is 61%. This means that the two parameters are suitable for the classification of Chinese

students' rhythm acquisition level of English. The measure % V refers to the period of the proportion of a vowel sound, for Chinese students, a very prominent problem is due to is a typical Chinese syllable timed language, is more noticeable in syllable pronunciation, and English is stress language, so in the process of the acquisition, obviously in terms of the way of Chinese will spoke each syllable clearly, and a main load-bearing part of each syllable is vowel, which makes every vowel proportion relatively longer. Likewise, consonant changes, sound reduction phenomenon is more saving in English, Chinese consonants change rate is very easy to reflect the degree of English acquisition. So the larger the difference between these two parameters is also is very easy to understand. Fig. 7 shows the best classification result of SVM training diagram.

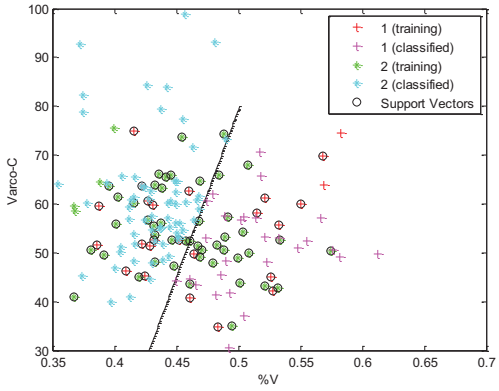

Fig. 7. The SVM training result on Varco_C vs. %V

TABLE III. CLASSIFICATION ACCURACY OF THE PRESENT EXPERIMENT

Exp. Number	vectors	accuracy
Proc. 1	nPVI_C 和 rPVI_V[a]	54%
Proc. 2	nPVI_V 和 rPVI_C[a]	41%
Proc. 3	nPVI_V 和 nPVI_C[a]	48%
Proc. 4	rPVI_V 和 rPVI_C[a]	52%
Proc. 5	%V 和 Varco_C[a]	61%
Proc. 6	%V 和 Varco_C[b]	59%
Proc. 7	%V 和 Varco_C[c]	61%

[a.] means data in SD area as the training set, data in JS area as the testing set

[b.] means data in JS area as the training set, data in SD area as the testing set

[c.] Means 50 samples of JS and 40 samples of SD as the training set, others as the testing set.

This research through to the rhythm of several common parameter classification performance was tested, and experiments show that the combination of rhythm can very good classifying data set, but the classification rate is not very satisfactory, the possible reasons about:

First may be the parameters will be influenced by the speed and the other researchers also note that. While this article requires pronunciation when recording natural read sentences as far as possible, but those who the nuances of the language fast is inevitable, the performance of these for the entire model caused much impact is still unknown.

Because of using the SVM only in 2 dimensional space, that is to say only two parameters for a single use, if you can get a better classification model in the multidimensional space, add more some model test, some believe the result will be better.

IV. CONCLUSION

No significant difference was observed across the former study by %V and Varco_C, whereas the case was not by nPVI_V and rPVI_C. On the one hand, the syllable number of the chosen sentences is about 10 which are not so clear to produce the duration variation. The learners of English can relatively well master English rhythm. The influence of the native language is not as obvious as the long utterances. On the other hand, there are still some controversy, we can't use these parameters absolute rhythm the distribution patterns of the existing problems.

All the samples as the training sample or test sample respectively using the SVM classification model is established. Results show that the combination of the rhythm measures rPVI and nPVI which generated four two-dimensional parameter model classification accuracy was 54%, 41%, 48% and 52%, while using % V and delta_C correct classification rate of 61%. We think this is caused by the influence by the native language of Chinese learners. As Chinese is a syllable-timed language, which is quite different from the stress-timed language English. The feature of syllable-timing is to pronounce every syllable very clearly. The measures % V and delta_C are better reflect this feature. On the contrary, PVI parameters focuses on the adjacent units, for Chinese learners, they tend to speak every syllable clearly, this change is not obvious. So the latter parameter combination of relatively good classification results has been achieved.

The next step, if one can consider some more parameters using multidimensional vector classification method may be able to better improve the rhythm of this classification method.

ACKNOWLEDGMENT

Many thanks are given to Weijing Zhou and Honghua Zhai for their help in data collection. This research is financially supported by Innovation Program of Chinese Academy of Social Sciences (AESOP—Corpus and Phonetic Study of English Learners from Chinese Dialect Regions), Chinese Social Sciences Foundation for Youth 'Phonetic Characteristics and Phonological Expression of Chinese Discourse' (No. 10CYY036), CASS Youth Foundation 'Level of Accents in Standard Chinese'.

[1] A. Loukina, G. Kochanski, B. Rosner, and E. Keane, "Rhythm measures and dimensions of durational variation in speech", J. Acoust. Soc. Amer., vol. 129, no. 5, pp.3258 -3270, 2011.

[2] J.W. Barry and M. Russo, "Measuring rhythm: Is it separable from speech rate", Actes des interfaces prosodiques, pp. 15-20, 2003.

[3] F. Ramus, M. Nespor and J. Mehler, "Correlates of linguistic rhythm in the speech signal", Cognition, 73, pp. 265-292, 1999.

[4] V. Dellwo, "Rhythm and speech rate: A variation coefficient for DC", in Language and Language Processing: Proceedings of the 38th

Linguistic Colloquium, Piliscsaba 2003 (Peter Lang, Frankfurt), pp. 231–241, 2003.

[5] E.L. Low, “Prosodic Prominence in Singapore English”, University of Cambridge, 1998

[6] L. White and S. L. Mattys, “Calibrating rhythm: First language and second language studies”, J. Phonetics 35, pp. 501–522, 2007.

[7] F. Nolan and E.L. Asu, “The pairwise variability index and coexisting”, Phonetica, 66, pp. 64-77, 2009.

[8] A.L. Eva and F. Nolan, “Estonian and English rhythm: a two-dimensional quantificatiion based on syllables and feet”, Speech Prosody 2006, City, pp. 25-27, 2006.

[9] E. Grabe and E.L. Low, “Durational variability in speech and the rhythm class hypothesis”, Laboratory Phonology, Jul, pp. 515-546, 2002.

[10] V. N. Vapnik, “The Nature of Statistical Learning Theory”, Springer Verlag, New York, 1995.

[11] A. K. Jain, R. P. W. Duin and J. Mao, “Statisitical Pattern Recognition: A Review”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, January, pp. 4 – 37, 2000.

[This paper was published in O-COCOSDA,2013]