Journal of the Korean Society of Speech Sciences Vol.7 No.2 (2015.06.30) www.speechsicences.or.kr
ISSN 2005-8063, pp. 149~156 http://dx.doi.org/10.13064/KSSS.2015.7.2.149

Fluency Scoring of English Speaking Tests for Nonnative Speakers Using a Native English Phone Recognizer

Jang, Byeong-Yong[1)] · Kwon, Oh-Wook[2)]

ABSTRACT

We propose a new method for automatic fluency scoring of English speaking tests spoken by nonnative speakers in a free-talking style. The proposed method is different from the previous methods in that it does not require the transcribed texts for spoken utterances. At first, an input utterance is segmented into a phone sequence by using a phone recognizer trained by using native speech databases. For each utterance, a feature vector with 6 features is extracted by processing the segmentation results of the phone recognizer. Then, fluency score is computed by applying support vector regression (SVR) to the feature vector. The parameters of SVR are learned by using the rater scores for the utterances. In computer experiments with 3 tests taken by 48 Korean adults, we show that speech rate, phonation time ratio, and smoothed unfilled pause rate are best for fluency scoring. The correlation of between the rater score and the SVR score is shown to be 0.84, which is higher than the correlation of 0.78 among raters. Although the correlation is slightly lower than the correlation of 0.90 when the transcribed texts are given, it implies that the proposed method can be used as a preprocessing tool for fluency evaluation of speaking tests.

Keywords: speaking fluency, support vector regression

1. Introduction

To evaluate the fluency of English speaking, substantial linguistic knowledge and sufficient data are required [1]-[3]. Despite this difficulty, more speaking tests tend to be included in English tests, and the evaluation of speaking fluency is usually done manually by native English raters. Whereas manual evaluation conducted by expert raters has its strengths in accuracy and validity, it has high dependency on the rubrics and has a risk of becoming subjective. In addition, manual evaluation requires abundance of time and expense.

In the previous works on fluency, Fillmore defined the 4 elements of fluency: the ability to talk at length with minimal pauses, the ability to talk cohesively and logically, the ability to talk in a wide range of contexts or situations, and the ability to create talk [4]. Crystal defined the fluency as ‘smooth, rapid, effortless use of language’ [5]. Chamber established the definition of fluency in qualitative and quantitative aspects and proposed the evaluation guide for foreign language speaking tests. Chamber’s experiments showed that the important elements for fluency evaluation are the rate of speech, the frequency or position of pause, and hesitations, which are temporal and quantitative features [2]. Kormos investigated the effects of temporal and lexical features on fluency evaluation and asserted that important features are the speech rate, the phonation time ratio, the number of stressed words, and the accuracy [3]. In the Deshmukh et al.’s study, 8 prosodic and 8 lexical features were extracted for fluency evaluation, and good performance was generally achieved with the lexical features among which the

1) Chungbuk National University, byjang@cbnu.ac.kr
2) Chungbuk National University, owkwon@cbnu.ac.kr, corresponding author

This work was supported by the research grant of Chungbuk National University in 2014.

Received: May 21, 2015
Revised: June 14, 2015
Accepted: June 16, 2015

number of unique words and the number of closely-occurring unigrams are best [6]. They classified a speaker's response with respect to 3 questions into 4 levels by using the support vector machine (SVM) and achieved classification accuracy of 53.6% and correlation of 0.68 by using the regression analysis based on optimal linear combination. Similarly, Xi and others scored speaking proficiency on the assessment data of Test of English as a Foreign Language (TOEFL) Practice Online (the TPO data set) and the response data from a TOEFL Internet-based Test (iBT) field study (the iBT data set) [7], [8], where the scores were computed by the classification and regression trees (CART) and multiple regression (MR) for 29 features. They reported that the correlation between the MR-computed score and human score was 0.57 and 0.68 for the TPO data set and the iBT data set, respectively. The CART also showed a similar level of performance. They argued that the performance was significant compared to the correlation among raters was 0.74 and 0.94 for the corresponding data sets.

The majority of previous studies listed relevant features for fluency evaluation, but did not suggest automatic methods for extracting such features. Speech recognition technology is the key to the automatic fluency evaluation [9]. Because extracting features solely by speech recognition encounters difficulties, recent studies have attempted to utilize the transcribed text for feature extraction [8], [9]. Although these approaches allow the analysis of the contribution of phonetic features to fluency evaluation, they still have definite constraints in developing an automatic algorithm for fluency evaluation. In most of speaking tests, a testee is sometimes asked to read the sentences given in the text or to repeat along the sentences heard, but is asked more often to express one's thoughts freely. The diversity of the context in the testee's response has made it inevitable to undergo a manual procedure to make a transcribed text. Accordingly, the preparation of the transcribed text is an essential step in order to develop an automatic method for fluency evaluation.

In the recent study of Xi and others [7], [8], they used a speech recognizer to obtain word sequences and their duration information without transcribed texts. Then, they extracted silence-related features and word-based features. However, for nonnative speech, the speech recognizer showed word accuracy of 50%, and accordingly the extracted features were not sufficiently reliable for fluency evaluation. In the recent study of Wang and others [11], they automatically scored the scene question-answer in an English spoken test by using speech recognition technologies. They claimed that the recognition accuracy is the key to the automatic scoring system because the answers of scene question-answer test are not unique and unknown. In order to increase the recognition accuracy, they used the mix-based language model that was trained by combining reference answers and other English corpora involved with junior high school with correct grammar. In addition, they extracted fluency features from the results of keyword recognition. Wang extracted the phonetic features as well as the features related keywords. The features are extracted by the results of speech recognition using the 3 grammars: strict grammar, free grammar, and keyword grammar. By using the SVR for fluency scoring, they achieved the correlation of 0.72 between machine scores and raters.

Jang and Kwon recently proposed a method for fluency and pronunciation evaluation using an aligner in case when the transcribed text is given [10]. But, it could not evaluate the fluency of free-talking utterances if the transcribed text is not available.

In contrast, in this paper, we attempt to compute fluency scores based on phone recognition techniques when a transcribed text is not given. We adopt the conventional phonetics-based features with a few modifications and then apply support vector regression (SVR). The feature extraction module is modified so that the performance degradation due to not using the transcribed text can be minimized.

Generally, vowel segments of speech signals have stronger frequency characteristics than consonant segments. Since acoustic-phonetic differences among vowels are more prominent than consonants, vowels have higher segmentation accuracy than consonants. In addition, the energy of vowel/consonant is clearly higher than the energy of silence, vowel/consonant/silence is well segmented. Thus we can extract phonetics-based features from segmentation information of vowel/consonant/silence even though the speech recognition accuracy for non-native speech is not high enough.

Unlike the previous studies that extracted fluency features using the number of phones and the silence factor between words [7], [8], we use the number of vowels instead of the number of syllables, and use a sigmoid function to consider pause duration. In addition, we calculate the final fluency score by combining the fluency features using SVR.

The paper is organized as follows. Section 2 proposes the method for fluency scoring by using phone recognition and SVR. Section 3 reports the experimental results and discussion. Conclusions and future works are presented in Section 4.

2. Proposed Method

2.1 Overall Structure

<Figure 1> shows the overall structure of the proposed method. The proposed method is composed of 3 parts: Extraction of phone sequence and duration, feature extraction for fluency scoring, and fluency scoring using SVR. To extract phone sequence and duration, we use a phone recognizer with phone bigram and phone model. Details about the algorithms to extract the phone sequence and duration will be elaborated in Section 2.2.2. To train the acoustic model, we used the native corpora. We extracted the phone segmentation information (phone sequence and phone duration) from the outputs of recognizer. Features are computed by analyzing the phone sequence and duration for fluency scoring. In the last part, we train the SVR model, and compute the fluency score with the SVR model.

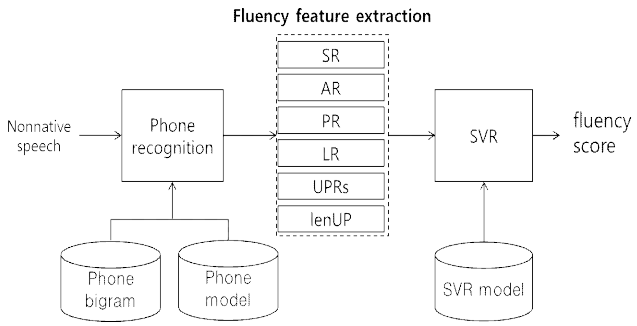

Figure 1. The proposed method for speaking fluency scoring

2.2 Phone Recognition

2.2.1 Phone Model and Phone Bigram

For feature extraction, we utilize the phone sequence and phone duration of speech samples obtained by using a phone recognizer. Although the phone recognizer gives worse performance than a phone aligner in obtaining the phone sequence and phone duration, the phone recognizer has the advantage that it does not require the transcribed text and thus can save the cost and time for fluency scoring. The phone recognizer operates with an acoustic model based on hidden Markov model (HMM). We train the acoustic model by using the HTK_Recipe toolkit [12] and 2 American English speech databases: WSJ0 [13] and TIMIT [14].

A feature vector consists of 39 mel frequency cepstral coefficients (MFCCs). All triphone models have the left-to-right topology with 3 states as shown in <Figure 2>. In detail, <Figure 2> (a) is the HMM topology of all triphones except silence and short pause, and (b) and (c) are the topology of silence and short pause, respectively. We use 8 Gaussian mixtures for all triphones except silence and 16 Gaussian mixtures for silence. The CMU dictionary [15] is used for pronunciation dictionary where a word has 2 pronunciation entries with silence or with short pause at the end of the phone sequence. As a result, the pronunciation dictionary has about 2.6 million entries. The base phone set in this work consists of 15 vowels ('ah', 'ey', 'iy', 'ay', 'ih', 'aa', 'ae', 'er', 'aw', 'uw', 'ao', 'eh', 'ow', 'oy', 'uh') and 24 consonants ('jh', 'zh', 'sh','hh', 'd', 'y', 'r', 'k', 's', 'ng', 'g', 'w', 'l', 'n', 'm', 't', 'dh' ,'z' ,'th , 'b', 'f', 'v', 'p', 'ch'), silence ('sil'), and short pause ('sp'). In this paper, the phonetic symbols are represented according to ARPABET [16]. The phone sequence was expanded to a triphone sequence to train the triphone-based acoustic model. The phone recognizer used a phone bigram trained by phone sequence of WSJ corpus.

2.2.2 Phone Recognizer

We trained the acoustic model by using the 2 databases: WSJ0 and TIMIT. Because the databases were recorded by native speakers, the phone recognition accuracy degrades when recognizing nonnative English speech. However, it is quite enough to have segment information on vowels, consonants, and silence instead of the exact phone sequence information. In our method, all features can be computed by using the segmentation information mostly of vowels. We used a phone recognizer to obtain phone segmentation information without a time-consuming and expensive manual labeling process. The recognition accuracy with respect to all phones is inadequate. However, because

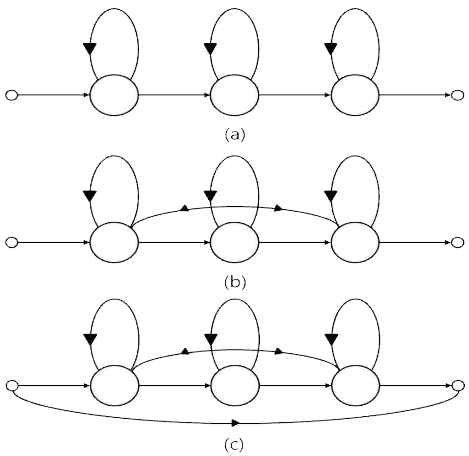

Figure 2. HMM topology: (a) All phones except silence and short pause, (b) silence, and (c) short pause

vowels have long duration and distinct acoustic features compared to consonants, they are generally recognized with high accuracy. Whereas the phone recognizer showed phone accuracy of 40%, it achieved the accuracy of 70% in recognizing vowel/consonant/silence.

From this reasoning, we assert that the phonetics-based features can be computed reliably for fluency scoring. This assertion will be justified in the experiments. The base phone set in the phone recognizer included 41 phones (15 vowels, 24 constants, 1 silence, and 1 short pause).

2.3 Feature Extraction

For fluency scoring, the syllable-based features are often used. In this work, we regard the vowel units of pronunciation sequence as the nucleus of a syllable for convenience and simplicity. Although the boundary of vowels is not aligned to the boundary of syllables in a strict sense, the vowel units can still serve as an acceptable framework to extract fluency features. For example, the word 'student' (/s t uw d ah n t/) has 2 vowels, which is the same as the number of syllables. Table 1 is the list of phonetics-based features which were known to produce good performance in fluency evaluation [1], [3]. Because the syllabication of English speech signals requires a complex process, we used vowel units instead of syllable units to extract the features of 'SR', 'AR', and 'LR'. We also extracted the modified 'SUPR' feature by applying a sigmoid function.

Table 1. List of extracted features

No.	Acronym	Full name	Explanation
1	SR	Speech rate	The total number of vowels in a speech per second
2	AR	Articulation rate	The total number of vowels in a speech per second without silence duration
3	PR	Phonation time ratio	A percentage proportion of the time to speech
4	LR	Mean length of runs	The amount of continuous speech of a speaker
5	SUPR	Smoothed unfilled pause rate	The number of unfilled pauses per second with sigmoid function
6	lenUP	Mean length of unfilled pauses	The mean length of unfilled pauses

The rate of speech is strongly associated with fluency. In the case of news anchors or fluent speakers, they look fluent by giving their speech in a steady rate without hesitation. Conventionally the rate of speech was computed as the number of syllables in a spoken speech, but we compute the rate of speech as the number of vowels in our study. The rate of speech is represented as 'speech rate (SR)' or 'articulation rate (AR)' according as the silence of spoken speech is considered or not [3]. SR is calculated as the total number of vowels produced in a spoken speech divided by the amount of total time in seconds. AR is similar to SR, but AR uses the amount of total time excluding pause time. In general, pauses can be classified into filled pause and unfilled pause. In our study, the filled pause refers to hesitation or repetition such as 'uh' or 'um', whereas the unfilled pause refers to silence without any sound.

In speaking tests, we generally perceive a speech as fluent when the speech was not cut and smoothly continued. The feature related with this characteristic is 'phonation time ratio (PR)', the percentage of time spent in speaking. PR is calculated by the total time without unfilled pauses divided by the total time [3]. 'Mean length of runs (LR)' indicates the amount of continuous speech of a speaker. This is calculated as an average number of syllables produced in utterances between pauses of 0.25 seconds and above [3]. This 0.25 second is the cut-off point; if the cut-off point is too low, an apparent pause may be confused with the stop phase of gemmated plosives or other normal phenomena, if it is too high, significant amounts of pause time may be omitted [17].

Unfilled pause is an important factor in itself as well as LR for fluency evaluation. We extract the 'Smoothed unfilled pause rate (SUPR)' and the 'mean length of unfilled pauses (lenUP)'. The SUPR is calculated as the total number of pauses divided by the total amount of time spent in speaking expressed in seconds and is multiplied by 60 [3]. The pauses here do not consider the ones between sentences but the ones inside the sentences. To avoid the problem due to hard clipping on pause duration and obtain the SUPR in a continuous value, a sigmoid function was applied to the duration of silent segments [10]. The lenUP is related to the length of unfilled pauses, and calculated as the total duration of unfilled pauses divided by the number of unfilled pauses [3].

2.4 SVR

As in the previous paper [10], we selected SVR as a predictor for fluency score because SVR is known to achieve good

generalization performance by virtue of a nonlinear prediction function [18]-[21]. As shown in <Figure 3>, the basic goal of SVR is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets [21].

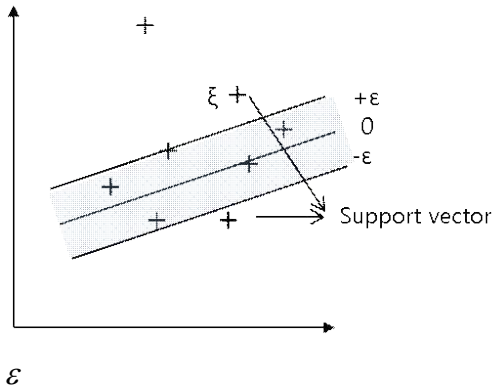

Figure 3. Function on ε deviation

3. Experimental Results and Discussion

3.1 Speaking Tests and Speech Database

The speech database consisted of 3 sets of English speaking test sets taken by 22 male and 26 female students in Korean universities. Each test set included 4 tasks and the total recording time was about 8h 50m. The type and time limit for each test set is shown in <Table 2>. Detailed information on the speech database is described in [10]. However, Task 1 has been excluded in this work because the utterances mostly have a length of 9 s and accordingly do not have discrimination in phone sequence and phone duration.

Table 2. Type and time limit for each task in a test set

Task	Type	Time (s)
1	Read aloud a given sentence	9
2	Speak about the given topic for myself	60
3	Make up a story for the given pictures	50
4	Make up a story for the given situation	50
5	Describe the given graph or chart	50

<Figure 4> shows a testee's transcribed answer to a question in Task 4 of 'Your sister is a shy person. She feels nervous whenever she talks in front of many people. What would you like to say to her?' In transcribed text, +<R>*word*+ denotes repetition of the *word* and +<H>*word*+ denotes hesitation of the *word*.

```
I would suggest her to buy a bear doll and then practice to him,
but since she could have a +<R>major+ major ah public speaking
project, it is better for her to try out a minor things before
she tries a major things so that it won't effect her life.
And I would suggest her to grab up eraser +<H>when+ whenever
she talks in front of everybody, so that she won't fidget.
But also what's the best thing is to practice.
```

Figure 4. Example of a transcribed answer to a question in Task 4

3.2 Analysis of Rater Scores

The speech database for speaking test was scored by 3 raters: an English lecturer in university (rater id 1) and 2 bilingual Korean adults (rater id 2 and 3). All tasks were rated according to the rubrics in the 5-point scale. The 10 rubrics used in this work were categorized into holistic, pronunciation, fluency, and language usage, as shown in <Table 3>. Before evaluation, the raters listened to 2 standard speech samples for each speaking proficiency level, and then performed independent scoring. In our experiment, we use the sum of 3 scores in the fluency category as the rater score.

Table 3. Rubrics used by raters

Category	Rubrics
Holistic	What is the holistic level of speaking proficiency?
Pronunciation	Is pronunciation clear and intelligible? Are accent, stress, and intonation natural?
Fluency	Is the rate of speech appropriate? Is there any pause or non-speech in utterance? Is the flow of sentence continuous without repetition?
Language usage	Are there no grammatical errors? Is vocabulary and expression appropriate? Is only English used? Are the sentences complete?

We used IBM SPSS Statistics to measure the correlation of fluency scores among 3 raters. The statistical analysis method was cross correlation with the Pearson coefficient. As shown in <Table 4>, the correlation among raters was 0.78 on average, which is comparable to the previous studies [3], [6], [9]. We note that the correlation coefficients in <Table 4> are a little

different from the previous paper [10] because the scores were computed only for the fluency category.

Table 4. Correlation among rater scores

Rater id	1	2	3
1	-	0.76	0.81
2	-	-	0.78
3	-	-	-

3.3 Analysis of features

The correlation among the fluency features is shown in <Table 5>. 'SR' is highly correlated with other features but 'AR' is low correlated with other features except 'SR'. This is because pause information is not used in computing 'AR'. We note that 'PR' and 'SUPR' are highly correlated with other features. In order to measure the contribution of each feature to fluency scoring, we performed the statistical analysis between rater score and feature. The results indicate that the phonetics-based features of SR, PR, and SUPR are best relevant to rater score. The correlation between feature and rater score is shown in <Table 6>. We extracted 'UPR' (unfilled pause rate without sigmoid function) feature to investigate the effect of sigmoid function used for 'SUPR'. From the fact that the correlation coefficient of SUPR was highly than the correlation coefficient of UPR, we can know that the features extracted by using a phone recognizer are closely related to the rater score.

Table 5. Correlation among the fluency features

Feat.	SR	AR	PR	LR	SUPR	lenUP
SR	1.00	0.78	0.88	0.78	-0.84	-0.67
AR	-	1.00	0.41	0.50	-0.38	-0.28
PR	-	-	1.00	0.76	-0.95	-0.79
LR	-	-	-	1.00	-0.62	-0.38
SUPR	-	-	-	-	1.00	0.88
lenUP	-	-	-	-	-	1.00

Table 6. Correlation between feature and rater score

Feature	SR	AR	PR	LR	SUPR	UPR	lenUP
score	0.76	0.52	0.74	0.68	-0.72	-0.68	-0.49

3.4 Regression results

In our experiments, the effect of the speech recognizer on fluency scoring is compared with a phone aligner. For this purpose, we trained the acoustic model for the phone aligner with transcribed text [10]. When a transcribed text is available, the phone aligner produces the best results for phone segmentation. We performed our experiments in 2 modes: Using the phone recognizer ('Recog') and the phone aligner ('Align').

The 48 testees were divided into 8 groups each of which has 6 speakers. The 7 groups were used for training the SVR model, the remaining group was used for evaluating performance. Cross-validation was performed so that 8 experiments were repeated with the testing group for each trial. To train the SVR model, we used the mean rater score as the desired target value. For parameters of SVR, we used the linear kernel function, set epsilon to 0.1, and set the cost constant to 1. In order to evaluate the performance of SVR, we measured the correlation between the SVR (computer) score and the rater score for each test and for overall test.

<Table 7> shows the correlation between the SVR score and the rater score depending on task kinds. The result indicates that the task 4 of 'make up a story for the given situation' is the best for fluency evaluation.

Table 7. Correlation between the SVR score and the rater score depending on task kinds

Task	2	3	4	5
score	0.73	0.75	0.76	0.73

The final scores of raters and SVR were computed in the same manner so that the final fluency score of a testee is computed by averaging the scores of 4 tasks in a test. <Table 8> shows the correlation between the final SVR score and the final rater score depending on whole test sets. The correlation in the 'Recog' mode was 0.84, which is a reliable result considering that the average correlation among rater scores was 0.78. On the other hand, the correlation in the 'Recog' mode was slightly less than the correlation of 0.90 in the 'Align' mode by 0.06. This fact tells that the phone recognizer achieves comparable performance with the phone aligner.

Table 8. Correlation between the SVR score and the rater score depending on whole test sets

Test set	1	2	3	Average
'Recog' mode	0.84	0.85	0.81	0.84
'Align' mode	0.89	0.92	0.90	0.90

<Figure 5> shows the scatter plots of the SVR score and the mean rater score in the 'Recog' and 'Align' modes, respectively.

Although both plots show similar tendency, the SVR score is slightly more correlated with the mean rater score in the 'Align' mode rather than in the 'Recog' mode as shown in <Table 8>.

5. Conclusion

We proposed a new method for speaking fluency scoring by using phonetics-based features. The proposed method is different from the previous research works in that it does not require any transcribed texts. By using a phone recognizer, we extracted phonetics-based features based on syllabic structures of input utterances. The feature extraction module was also modified in order to reduce the performance degradation due to the usage of a phone recognizer instead of a phone aligner. Experimental results showed that SR, PR, and SUPR are best relevant to rater scores. The final fluency score showed a high correlation value of 0.84 with the rater scores, which is similar to the correlation among raters.

Our work has original contributions on the following points. First, we reduce the loss of information between phones or silence by using a phone recognizer. Second, we extract robust phonetics-based features by utilizing segmentation information of a phone recognizer trained with native databases. Finally, we applied a sigmoid function for computing the number of unfilled pauses.

From experimental results, we show that the proposed features produce high correlation to the score of raters despite the phone recognition accuracy is less than 40%, and then yield high correlation between fluency scores of raters and SVR scores by effectively combining extracted features. It is also shown that the performance of the proposed method based on a phone recognizer is comparable to the performance obtained by using a phone aligner assuming that the transcription of speech signals is given.

For further study, we plan to refine phonetics-based features for accuracy improvement, and find additional acoustic or linguistic features.

Acknowledgements

This work was supported by the research grant of Chungbuk National University in 2014. The authors would like to thank SK Telecom for allowing us to use the speech database of English speaking tests.

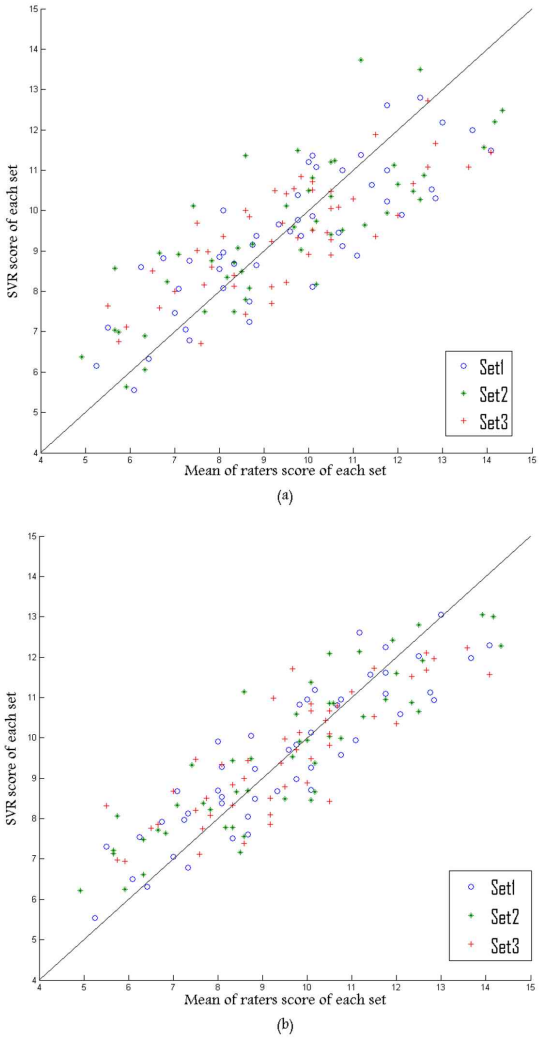

Figure 5. Scatter plots of the SVR score and the mean rater score (a) in the 'Recog' mode and (b) in the 'Align' mode.

References

[1] Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, Vol. 14, No. 4, 423-441.

[2] Chambers, F. (1997). What do we mean by fluency?. *System*, Vol. 25, No. 4, 535-544.

[3] Kormos, J. & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, Vol. 32, No. 2, 145-164.

[4] Fillmore, C. J. (1979). *Individual differences in language ability and language behavior*. Academic Press, 85-101.

[5] Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge.

[6] Deshmukh, O. D., Kandhway, K., Verma, A. & Audhkhasi, K.

(2009). Automatic evaluation of spoken English fluency. In Proc. ICASSP 2009, 4829-4832.

[7] Zechner, K., Higgns, D., Xi, X. & Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, Vol. 51, No. 10, 883-895.

[8] Xi, X., Higgins, D., Zechner, K. & Williamson, D. (2012). A comparison of two scoring methods for an automated speech scoring system. *Language Testing*, 1-24.

[9] Neumeyer, L., Franco, H., Digalakis, V. & Weintraub, M. (2000). Automatic scoring of pronunciation quality. *Speech Communication*, Vol. 30, No. 2-3, 83-93.

[10] Jang, B. Y. & Kwon, O. W. (2014). Computer-based fluency evaluation of English speaking tests for Koreans. *Journal of the Korean Society of Speech Sciences*, Vol. 6, No. 2, 9-20.
(장병용 & 권오욱 (2014). 한국인을 위한 영어 말하기 시험의 컴퓨터 기반 유창성 평가. 말소리와 음성과학 제6권 제2호, 9-20.

[11] Wang, L., Liu, Y., Pan, F., Dong, B. & Yan, Y. (2014). Automatic scoring of scene question-answer in English spoken test. In Proc. Information Science, Electronics and Electrical Engineering (ISEEE), 712-715.

[12] Vertanen, K. (1994). HTK Wall Street Journal Training Recipe, www.keithv.com.

[13] Paul, D. B. & Baker, J. M. (1992). The design for the Wall Street Journal-based CSR corpus. In Proc. Workshop on Speech and Natural Language, 357-362.

[14] Garofolo, J. S. (1988). Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. National Institute of Standards and Technology (NIST), Gaithersburgh, MD, 107.

[15] Lenzo, K. (2007). The CMU pronouncing dictionary, www.speech.cs.cmu.edu/cgibin/cmudict.

[16] Rabiner, L. & Juang, B. (1993). *Fundamentals of Speech Recognition,* Prentice Hall.

[17] Towell, R., Hawkins, R. & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, Vol. 17, No. 1, 84-119.

[18] Müller, K. R., Smola, A. J., Rätsch, G.., Schölkopf, B., Kohlmorgen J. & Vapnik, V. (1997). Predicting time series with support vector machines. In Proc. Artificial Neural Networks－ICANN'97, 999-1004.

[19] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 155-161.

[20] Haykin, S. (1999). *Neural Networks*, Prentice Hall.

[21] Smola, A. J. & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, Vol. 14, No. 3, 199-222.

[22] Cucchiarini, C., Helmer, S. & Lou, B. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, Vol. 172, No. 2, 989-999.

• **Jang, Byeong-Yong**
Dept. of Control and Robot Engineering, Chungbuk National University
Chungdae-ro 1, Seowon-Gu, Cheongju, Chungbuk 362-763, Korea
E-mail: byjang@cbnu.ac.kr
Areas of interest: speech recognition, pattern recognition, and automatic scoring of speaking tests

• **Kwon, Oh-Wook**
School of Electronics Engineering, Chungbuk National University
Chungdae-ro 1, Seowon-Gu, Cheongju, Chungbuk 362-763, Korea
Tel: 043-261-3374
E-mail: owkwon@cbnu.ac.kr
Areas of interest: speech recognition, speech and audio signal processing, and pattern recognition