

Model-based duration analysis on English natives and Thai learners

Chatchawarn Hansakunbuntheung¹, Hiroaki Kato² and Yoshinori Sagisaka¹

¹GITI/Language and Speech Science Research Labs, Waseda University,

²NICT/ATR Cognitive Information Science Labs, Japan

<https://doi.org/10.36505/ExLing-2008/02/0026/000085>

Abstract

This paper presents a model-based statistical duration analysis on the comparison between English natives and Thai-native English learners. The analyses were carried out to characterize non-native's duration control on (1) control differences between natives and learners and on (2) the relationship between the duration characteristics and learners' background reflecting English experiences. The analyses showed lower speech rate and strong influence of the first language characteristics in beginner's speech and gradual improvements according to their English experiences quantitatively.

Keywords: English duration, computational duration modeling, second language

Introduction

Timing control is one of the most fundamental and essential issues in spoken language education and in smooth speech communication. For English learners, differences in timing control characteristic should be clearly understood and self-evaluated to improve their own English-speaking skill. Thus, we need a scientific quantitative evaluation method of English timing and the knowledge of timing differences based on the model for the improvement of non-native English learner's timing control skill.

In this paper, we proposed a method of model-based statistical analysis on timing characteristics for Thai-native English learners. The analysis was carried out by comparing durational differences between learners and English natives. In the analysis, we quantified the differences in speech rate and segmental duration using a statistical model. Furthermore, the prediction errors between another statistical model and Thai learners were employed to associate with learners' English experiences quantitatively.

Model-based duration analysis

For learner's proficiency evaluation in timing control, we adopted an objective measure of an average difference between each segmental duration of a target learner and native's one instead of conventional subjective measures. To be free from subject dependent durational characteristics, a statistical duration model was trained using English speech data uttered by multiple native speakers. We measured the durational differences between

statistically predicted values, and, used the resultant prediction error as a measure to evaluate proficiency in English timing control.

Specifications on database

English speech database. We used three English speech databases. The first one was the ARCTIC database (Kominck 2003) read by English-speaking natives. The database was separated into two sets: set A of 593 sentences, and, set B of 539 sentences. The second database was a read English speech database of the fairy tale “The north wind and the sun” from CUCHLOE corpus that read by English natives, and, speakers from English-as-an-official countries. The third one from NECTEC was a speech database of the fairy tale that read by 45 Thai learners of English, and, one Indian English speaker.

The above databases were grouped into four sets for modeling and analysis. The first set consisted of speech data from ARCTIC set A uttered by four US speakers. This set was used as a training set for the reference English duration model. The second set consisted of speech data from ARCTIC set B by the same four speakers. We referred this set as a (speaker) close set, and, used to evaluate consistency of the model. To evaluate the validity with various English accents, we made an open set. The third set contained 3 non-US-accent speakers from CMU ARCTIC set B, 7 speakers from CUCHLOE, and, one Indian English speaker from NECTEC. The last set contains 45 Thai learners of English from NECTEC. This set was used as a test set to evaluate English duration characteristics of the Thai learners.

Learner’s information. For Thai learners, the learner’s information related to English experience were collected. It consists of age (years), English-educated period (years), educating period in English-as-an-official-language countries (years), and, average duration for English usage in a week (hours).

Factors for statistical modeling. For a statistical model, the following categories of control factors were employed; current and four context phones, stress, phone position and numbers of constituent phones in syllable, word and phrase, syllable position and numbers of syllables in word and phrase, narrow and board parts of speech.

A statistical model for English duration. Before modeling, we normalized phone duration for each speaker using z-score with mean and standard deviation to eliminate speech rate effect for inter-speaker comparison. For the modeling, we adopted a linear regression based on categorical factors (Hayashi 1950). In this model, each sub-category of the factor categories was encoded as “1” if the considering sub-category exists in current segment. Otherwise, it was encoded as “0”. By adopting least-square-error minimization criteria, the model coefficient representing the contributions of the control factors were calculated. Finally, root-mean-squared errors of

differences between measured and estimated duration were calculated for each speaker for comparison.

Experimental results

Prediction error analysis on segmental duration

An average of duration prediction errors was employed as an objective measure for the proficiency in English timing control. Figure 1 shows noticeable grouping of prediction errors by speaker profiles. For the close set consisting of the same speakers employed in model training, the errors showed the closest distant from the error of the training set. The close set also showed the lowest errors in all groups of speakers. Thus, the results showed consistence accurate prediction of the model on both training and close sets. For other native speakers, their prediction errors were much close to ones of speakers used in model training and smaller than most of Thai learners. The learners living in English-as-an-official-countries for more than 10 years showed salient decreasing of the distances from the reference model. Furthermore, the learners having no experience in English-as-an-official-language country showed bigger prediction errors with large variation of English skills.

English duration characteristics of the Thai learners

In Figure 2, “S”, “I”, “AI”, “M”, “BF”, “F” labels represent syllable position in monosyllabic word, initial, after-initial, mid, before-final, and, final position in a word, respectively. Big errors of the Thai learners were found at the ends of a word and a phrase. This is resulted from the difference in stress placement between Thai and English. In Thai, primary stress syllable is always located at the last syllable of words, while, stress placement in English need not to be the case.

Analysis based on speaker’s English learning experience

By observing from the model coefficient of the model between the prediction errors and the experience, Table 1. shows a noticeably corresponding trend between Educating period in English-as-an-official-language country and the errors. It shows gradual improvements according to their English experiences when the period more than five years. This result corresponds with the results in Figure 1.

Table 1. Model coefficients of the Educating period control factor.

Period (years)	0	0<Y<= 5	6<=Y<= 10	11<=Y<= 15
Model coefficient	0.506	0.570	0.0	-0.780

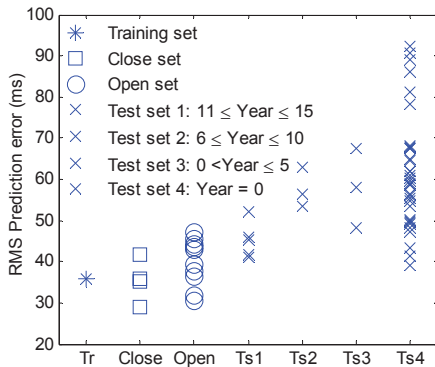


Figure 1. Comparison of prediction errors between English natives and Thai learners with different educating period in English-as-an-official-language countries.

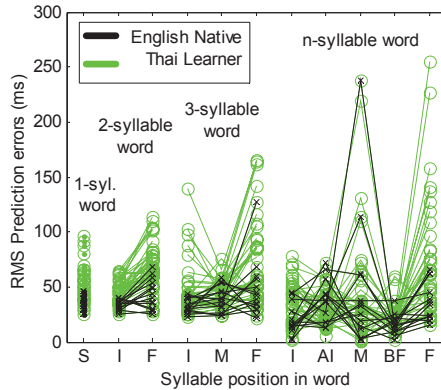


Figure 2. Prediction errors comparing timing control of syllable position in word with different word lengths between the native and the learners.

Conclusions

We proposed a method to analyze English duration characteristics of Thai learners of English for the proficiency evaluation of English timing control. Experimental analyses on duration prediction errors showed the effectiveness of the proposed objective evaluation using statistical duration characteristics based on the English duration model as a reference. The results suggested the use of global prediction errors, prediction errors from timing controls, and, the learners’ background as objective measures. These results could lead to further study on quantitative evaluation of English skill.

Acknowledgements

We would like to thank our colleagues at National Electronics and Computer Technology Center (NECTEC), for collecting the English speech data of Thai learners and native. We also would like to express our special thanks to Prof. Helen Meng (Chinese University of Hong Kong) for providing the CUCHLOE English speech corpus. This work was supported in part by Waseda University RISE research project of “Analysis and modeling of human mechanism in speech and language processing” and Grant-in-Aid for Scientific Research B, No. 20300069 of JSPS.

References

Kominek, J, and, Black, A. W., 2003, CMU ARCTIC database for speech synthesis.
 Hayashi, C. 1950. On the quantification of qualitative data from the mathematico-statistical point of view. *Annals of Institute of Statistical Mathematics*, vol. 2.