

OPTIMIZING COMPUTER-ASSISTED PRONUNCIATION INSTRUCTION BY SELECTING RELEVANT TRAINING TOPICS

Antoine RAUX and Tatsuya KAWAHARA

School of Informatics, Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan
{raux, kawahara}@kuis.kyoto-u.ac.jp

ABSTRACT

We introduce a novel method to diagnose pronunciation errors that are most critical to the intelligibility of L2 learners. A preliminary study showed that error rates computed by a speech recognition-based system can be used to characterize intelligibility. We deduce a probabilistic algorithm to derive intelligibility from error rates. We also define an error priority function that indicates which errors are most critical to intelligibility. Experimental results proved the validity of the approach.

1. INTRODUCTION

Recent Computer-Assisted Pronunciation Learning (CAPL) research, stirred up by the improvement of computer hardware and Automatic Speech Recognition (ASR), has focused on two areas: evaluation and instruction. While studies on the correlation between acoustic features of speech and human judgements of intelligibility opened the way to automatic evaluation of intelligibility ([4],[2]), innovative approaches to instruction were developed, using ASR to detect segmental or prosodic errors ([5],[3]). However, little has been done to relate instruction to evaluation and provide learners with feedback on which errors are most critical to their intelligibility. This can result in sub-optimal learning as students spend time on aspects of pronunciation that do not noticeably affect intelligibility. In this paper, we propose a new method to assess intelligibility from pronunciation error rates and spot the errors that are most critical to each learner's intelligibility.

2. RELATIONSHIP BETWEEN ERROR RATES AND INTELLIGIBILITY

2.1. Experimental Protocol

We conducted an experimental study of the relationship between 10 selected pronunciation errors common among Japanese speakers of English and human ratings of intelligibility. The list of errors is given in Table 1. Subjects were

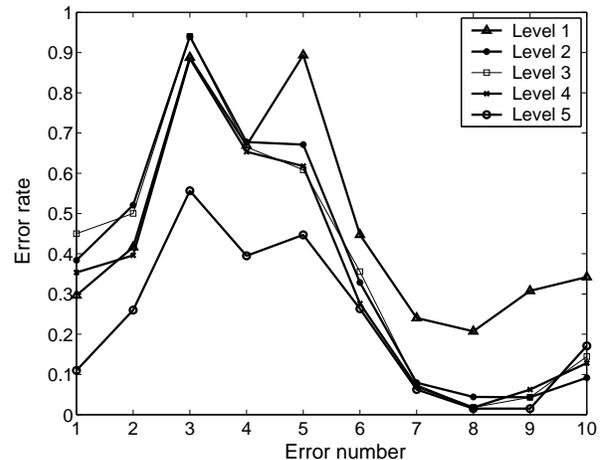


Fig. 1. Error rates averaged for each intelligibility level.

16 Japanese students and university personels. We recorded their reading of a passage designed for pronunciation evaluation[6]. The recordings were sent to a qualified linguist who rated each subject's intelligibility from 1 (hardly intelligible) to 5 (perfectly intelligible). Pronunciation errors in the recordings were then detected using ASR, and each subject's error rates were computed. We computed the average error rates of subjects of each intelligibility level. Figure 1 shows the error rates of the 5 levels.

2.2. Results

Figure 1 shows that the way error rates vary across levels depends on the error. Three types of errors can be distinguished.

2.2.1. Phonemic substitutions and deletion

For these errors (number 1 to 4), which concern phonemic substitutions and deletion, only level-5 students have a low

Table 1. Errors detected by the system

Number	Description	Example word	Erroneous pronunciation
1	Word-initial w/y deletion	<u>w</u> ould	u! d
2	SH/CH substitution	<u>ch</u> oose	∫ u! z
3	ER/A substitution	<u>er</u> paper	p e! p a:
4	R/L substitution	<u>r</u> oad	l o ∅ d
5	Vowel non-reduction	stud <u>e</u> nt	s t j u! d e n t
6	V/B substitution	prob <u>l</u> em	p r a v l e m
7	Word-final vowel insertion	le <u>t</u>	l e t ∅:
8	CCV-cluster vowel insertion	stud <u>y</u>	s u t ∆ d i
9	VCC-cluster vowel insertion	act <u>i</u> ve	a k u t i v
10	H/F substitution	<u>h</u> ire	h a! e ^r

error rate¹. The other students have an error rate of about 10 to 30% greater depending on the error. The error rate among students of level 1 to 4 are equivalent.

2.2.2. Vowel non-reduction

This error (number 5) divides the students in 3 groups: level-1 students have a 90% error rate, level-2 to 4 students have a 60-70% error rate and level-5 students have a 40% error rate.

2.2.3. Vowel insertion, H/F, V/B substitutions

For these errors (number 6 to 10), syllable structure and two consonant contrasts (/v-b/ and /h-f/) are considered. Level-1 students have comparatively large error rates compared to all other students. Level 2 to 5 students have a similar error rates.

2.3. Linguistic Interpretation

These observations show that mastering such aspects of pronunciation as consonant clusters (affected by vowel insertion) or vowel reduction is necessary to reach even average levels of intelligibility. On the contrary, phonemic accuracy, except for the two pairs H/F and V/B, does not prevent speakers to be largely intelligible (level 4). These results are consistent with the position of most recent linguists regarding the teaching of pronunciation[1]. Errors such as vowel insertion and non-reduction, which are related to prosodic features such as syllable structure and stress are considered to be more crucial to intelligibility than purely segmental errors.

¹For all errors except error 6 ($p < 0.2$), all differences between average error rates are significant at the 0.05 level.

3. INTELLIGIBILITY ASSESSMENT

3.1. Probabilistic Models

Based on the findings preliminary study, we propose a probabilistic approach to intelligibility assessment. Given the observed error rates O , our goal is to obtain the probability that the learner's intelligibility level is i , $\forall i \in \{1..5\}$. These probabilities (noted $P(i|O)$) can be computed using Bayes formula:

$$P(i|O) \propto P(i)P(O|i) \quad (1)$$

where $P(i)$ is the ratio of level- i students in the considered population and $P(O|i)$ the probability distributions of the error rates for level- i speakers. Under the assumption that all error rates are statistically independent *given the intelligibility level*, the overall probability distribution is given by $P(O|i) = \prod_j P(r_j|i)$. We model each $P(r_j|i)$ by a Beta distribution, defined on $[0, 1]$ by:

$$\beta_{(a,b)}(x) = B(a,b)x^{(a-1)}(1-x)^{(b-1)} \quad (2)$$

where a and b are parameters and $B(a,b)$ is a normalizing constant. Parameters are computed using data rated for intelligibility by a human judge. Combining equations 1 and 2 leads to the following formula for the probability of level i :

$$P(i|O) = K \prod_j \beta_{(a,b)_{i,j}}(r_j) \quad (3)$$

where K is a normalizing constant. We define the intelligibility score as the expected value of the level:

$$I = \sum_i iP(i|S) \quad (4)$$

Thus, the score can take any value in the range $[1, 5]$.

3.2. Evaluation

3.2.1. Speech data and intelligibility ratings

We evaluated the algorithm on data containing both reading passages and prepared oral presentations by a total of 42 Japanese university students. Each subject was rated for intelligibility by 3 associate professors of English at Japanese universities. Inter-rater correlations were surprisingly low, even after normalizing the scores to compensate for individual raters' bias (0.6 in average). We suspect that this is a consequence of the differences in the type of data (reading or speech) and the lack of training of the raters to this particular task. Therefore, evaluation was conducted on each rater separately.

3.2.2. Closed evaluation

We trained the model (i.e. computed the error rate distributions for each level) on the whole set of speakers and estimated the level of each speaker. The correlations between the each rater and the corresponding machine score were respectively 0.73, 0.84 and 0.78, confirming the validity of the approach.

3.2.3. Take-one-out evaluation

In this case, we trained the model using 41 subjects and estimated the intelligibility of the remaining subject, and repeated this 42 times. Correlations were much lower, respectively 0.52, 0.59 and 0.47. We think that the high variability of the evaluation data is partly responsible for these results. The fact that some levels were given to very few subjects (< 4) tends to increase each speaker's weight in the prior distributions and degrade the performance of the system.

3.2.4. Improvements of the model

To compensate for these weaknesses of the data, we tried two approaches: using a probabilistic model of error rates and constraining the error rates' probability distributions. In the former case, instead of considering the error rate as a deterministic variable, we model it by a belief distribution on $[0, 1]$ whose mean is the measured error rate and whose variance depends on the number of observed patterns (the more observations, the smaller the variance). We used Beta density functions to model these as well. Calculus yields an exact formula for $P(i|O)$ as well, which replaces equation 3:

$$P(i|O) = KP(i) \prod_j \frac{B(a_j + a'_j, b_j + b'_j)}{B(a_j, b_j)} \quad (5)$$

where a'_j and b'_j are the parameters of the belief distribution.

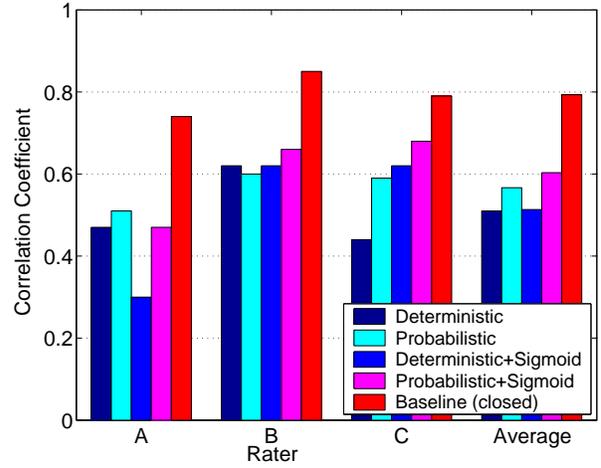


Fig. 2. Correlation coefficients on cross-validation

The second improvement is based on the assumption that all error rates must decrease as intelligibility increases. We apply this by constraining the rates of each error to be on a decreasing sigmoid function of the level, instead of computing the exact average from the training data.

The results of these modifications are summarized in Figure 2. Both modifications slightly improved the correlations, the best result being an average correlation of 0.6, obtained by combining the two.

4. DIAGNOSIS OF CRITICAL PRONUNCIATION ERRORS

To determine which errors should be studied by a given learner, we define the priority $\pi(j, i)$ of error j at level i as the difference between the learner's error rate and the average error rate of level- i students, that is:

$$\pi(j, i) = r_j - \langle r_j \rangle_{\text{level-}i \text{ students}} \quad (6)$$

The priority $\pi(j)$ of error j is defined as the expected value of each level's priority:

$$\pi(j) = \sum_i P(i|O) \pi(j, i) \quad (7)$$

Figure 3 gives an illustration of how error priority is affected by the overall intelligibility of the student. In this example, we consider a student whose error rate on error 9 (vowel insertion in VCC clusters) is 0.15. Circles represent students from the training sample and the dashed line connects the average value for each intelligibility level. The distance between the horizontal line and the dashed line represents the absolute value of the priority for this level. Priority is negative for level 1 (the student's rate is below the

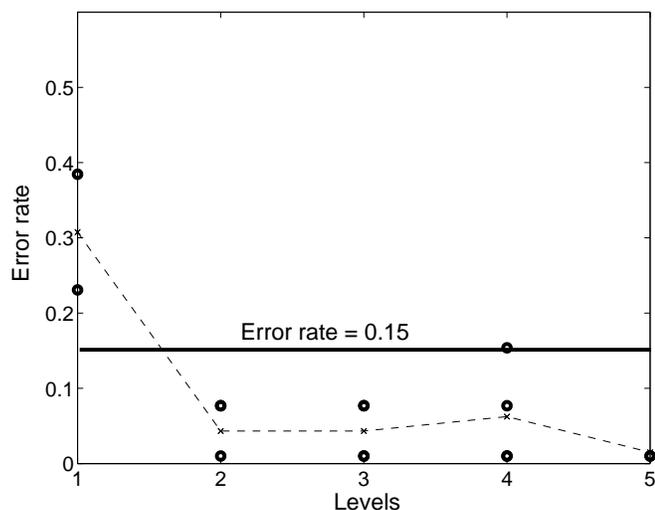


Fig. 3. Average error rates and priority for VCC vowel insertion (because some points are superimposed, less than 16 points appear).

average) and positive for all other levels. Thus, this error is likely to be proposed for study if the learner's intelligibility level is 2 or more, because speakers of these levels usually master this error. On the other hand, if the learner's level is 1, other errors are more likely to be proposed for study.

We computed error priorities for the 16 speakers of the preliminary experiment and found that they agreed with our own impression of the strengths and weaknesses of each speaker. To conduct a formal evaluation, we have developed a prototype CAPL system that uses the algorithms presented in this paper. Using this prototype, we hope to demonstrate the validity of this algorithm and its applicability to real world situations.

5. CONCLUSION

We proposed a probabilistic method to assess non-native speakers' intelligibility. Our approach yields an explicit model of the relationship between intelligibility and error rates. We showed that it can be used to provide meaningful feedback to the learners on their strengths and weaknesses. Although performance needs to be improved before real world applications can be developed, the promising results were obtained, which could find applications in CAPL systems, and more generally, Intelligent Tutoring Systems.

6. REFERENCES

[1] Celce-Murcia, M., Brinton, D. M. and Goodwin, J. M.: *Teaching Pronunciation: A Reference for Teachers of*

English to Speakers of Other Languages, CUP (1996).

- [2] Cucchiari, C., Strik, H. and Boves, L.: Automatic Evaluation of Dutch Pronunciation by using Speech Recognition Technology, *IEEE workshop ASRU*, pp. 622–629 (1997).
- [3] Eskenazi, M.: Using Computer in Foreign Language Pronunciation Training: What Advantages?, *CALICO Journal*, Vol. 16, No. 3 (1999).
- [4] Franco, H., Neumeyer, L., Kim, Y. and Ronen, O.: Automatic Pronunciation Scoring for Language Instruction, *ICASSP'97*, Vol. 2, pp. 1471–1474 (1997).
- [5] Jo, C.-H., Kawahara, T., Doshita, S. and Dantsuji, M.: Japanese Pronunciation Instruction System Using Speech Recognition Methods, *IEICE Trans.*, Vol. E83-D, No. 11, pp. 1960–1968 (2000).
- [6] Prator, C. and Robinett, B.: *Manual of American English Pronunciation*, HRW International Editions (1985).