

# VOWEL-REDUCTION FEEDBACK SYSTEM FOR NON-NATIVE LEARNERS OF ENGLISH

*Jeesoo Bang, Kyusong Lee, Seonghan Ryu, and Gary Geunbae Lee*

Department of Computer Science and Engineering  
Pohang University of Science and Technology, South Korea

## ABSTRACT

In spoken English, vowels in non-stressed syllables are often reduced to a brief neutral vowel (e.g., ə or ɪ). Non-native speakers of English may not use this ‘vowel reduction’ correctly, so their utterances may sound unnatural. We propose an automatic system to provide feedback about vowel-reduction to non-native speakers of English. The system has three parts: it (1) predicts vowel reduction, (2) detects vowel reduction in speech, compares the prediction to the detected sound to generate a score then (3) uses this score to provide corrective feedback to the speaker. The system had good accuracy and provided positive learning results for the user. The proposed system can be used as a part of a computer-assisted language learning system.

**Index Terms**— CALL, CAPT, vowel-reduction training, vowel non-reduction, language assessment

## 1. INTRODUCTION

Computer-assisted language learning (CALL) allows learners to study a language at their convenience. CALL offers the user one-to-one training: it identifies learners’ errors, including those of grammar, vocabulary, and speaking [1], [2]. Learning to speak correctly is a difficult part of language acquisition. In the classroom, a language learner can get assistance and corrective feedback from tutors, but instruction in the classroom is restricted in time and space. CALL can allow learners of speaking to practice it at their convenience.

Computer-assisted pronunciation training (CAPT) is a type of CALL that trains the language learners’ pronunciation. Beyond pronunciation, prosody, i.e., sentence stress and intonation, is as important as pronunciation in speaking [3-5]. CAPT systems initially focused on pronunciation [6, 7], some now also correct speakers’ prosody [2, 8] and lexical stress [9].

In stress-timed languages such as English, vowels in non-stressed syllables are often reduced to a brief neutral vowel (e.g., ə or ɪ). This ‘vowel reduction’ is common in stress-timed languages as English [10]. Non-native learners of English, whose L1 is a syllable-timed language may have

difficulty reducing vowels, i.e., they may stress all vowels; as a result, their pronunciation is unnatural.

To acquire native-like English speaking skill, learners of English should learn how to pronounce the reduced vowels. CAPT systems that focus on the vowel reduction phenomenon have not been developed yet, but non-native speakers of English definitely need such a system.

We describe an automatic system to provide feedback about vowel-reduction. The system combines feedback about pronunciation and sentence stress. Experiments with human learners verify that the system helps them to recognize vowel reduction and to use it correctly.

## 2. VOWEL REDUCTION

Vowel reduction is a monothong or diphthong loses its own phonetic value and is pronounced short and reduced when it does not get stressed. Because English is a stress-timed language, unlike syllable-timed language, vowel reduction occurs a lot. For example, two-syllable word ‘lemon’ is not pronounced as /'lɛ'mɔwn/ but pronounced as /'lɛmən/; the unstressed second vowel ‘o’ is reduced to /ə/. English vowels are usually reduced to a mid-central vowel ‘schwa’ /ə/. This phenomenon is called as neutralization of vowel or centralization of vowel [11].

Previous studies have considered vowel reductions, especially acoustic approaches to them [11], [12]. However, no CALL system has been developed that can capture vowel reduction and provide feedback on it. To the best of our knowledge, no approach exists that applies vowel reduction feedback to language education.

We adopted sentence stress to predict and detect vowel reductions. By predicting and detecting sentence stress we can learn which words in a sentence receive a stress. Frequently-used words (e.g., ‘a’, ‘the’) usually have reduced vowels [13], but if we find these words in the pronunciation dictionary, there are phoneme sequences which have lexical stress in it. Therefore, we concluded that including sentence stress in the vowel reduction feedback system could improve its performance (here measured as accuracy, precision and F1 score), so we designed the system to consider sentence stress.

### 3. SYSTEM ARCHITECTURE

The proposed vowel-reduction feedback system (Fig. 1) is composed of three parts: a vowel reduction prediction model, a vowel reduction detection model, and a feedback model.

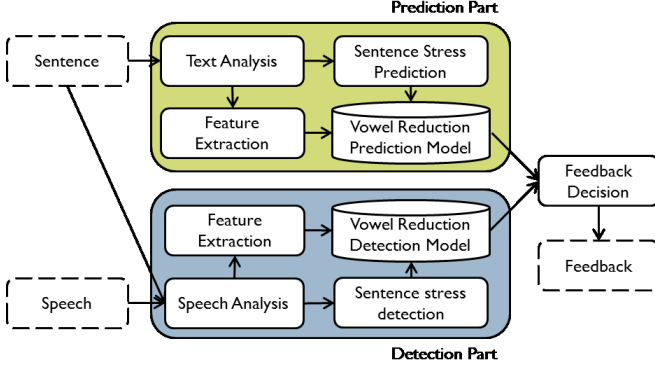


Fig. 1. Architecture of the vowel reduction feedback system

#### 3.1. Prediction model

The vowel-reduction prediction model analyzes a sentence and calculates the probability of vowel reduction for every vowel in it. First, a text analysis module parses each sentence and marks it with appropriate phoneme sequences and lexical stress from the CMU pronouncing dictionary<sup>1</sup> and part-of-speech (POS) tag; then a sentence stress prediction module [2] provides a sentence stress confidence score for every word in the sentence. The resulting text information (phoneme, lexical stress, POS tag) and sentence stress information are used as features to train the vowel-reduction prediction model.

We built the prediction model based on the Boston university radio news corpus (BURNc) [14]. We labeled the arpa phonemes 'AX' (= /ə/) and 'AXR' (= /ər/) in the BURNc as 'vowel reduction'. Every phoneme was labeled as either a reduced vowel, a full vowel, or a consonant. BURNc had the AX phoneme, but the CMU dictionary did not, so the extracted phoneme sequence did not have AX phoneme. For that reason, we needed to train the model. We adopted the conditional random fields (CRF) model, which has been widely used in natural language processing [15-17].

#### 3.2. Detection model

The vowel-reduction detection model analyzes the input utterance and calculates the probability of vowel reduction for every vowel in the utterance. Differently from the prediction model, the detection model analyzes speech (Fig. 1). A speech analysis module marks the input speech with its

actual phoneme sequence and time alignments. Because the speech analysis module includes the text analysis module, the text information from the prediction model is used as features for the detection model. Then a sentence-stress detection module outputs sentence stress confidence for every word in the sentence. These outputs of the speech analysis module (text information described in 3.1, actual phoneme, duration of each phoneme) and the sentence stress detection module are used as features for the vowel-reduction detection model.

We built the detection model based on the Korean learners' English accentuation corpus (KLEAC) [18], because non-native English learners have their own prosodic habits when uttering English sentences. KLEAC is composed of six hours of speech with 5,500 English sentences produced by 75 native Korean speakers; it includes orthographic transcription, rhythmic word marks and proficiency labels. KLEAC was annotated based on the CMU phoneme set, which does not have the 'AX' phoneme, so we labeled the human-labeled 'AH' phoneme as 'vowel reduction' for training the detection CRF model. The 'AH' phoneme is the most similar phoneme with 'AX' phoneme: both /ʌ/ and /ə/ are annotated with the same arpa 'AH', where only /ə/ is annotated with the 'AX' phoneme.

#### 3.3. Feedback model

We designed the feedback model to provide corrective vowel-reduction feedback to system users. This feedback is determined by comparing the predicted and detected vowel-reduction patterns, which are categorized as either correct or incorrect. When the pattern was incorrect, we provided feedback (an 'X').

To measure the confidence in the feedback, we designed an adjusted score by adopting the output probability of the CRF classifier for each vowel-reduction label. The adjusted score is the absolute difference between the probabilities of the predicted and detected vowel reductions. If the adjusted score exceeds a threshold, the system gives negative feedback. We set the threshold to 0.5 heuristically.

We also assessed the utterances on a scale of 1 to 100, using the feedback information:

$$\text{Utterance Score} = \frac{\text{Feedback count of "positive" type}}{\text{Entire feedback count}}. \quad (1)$$

The system provides negative feedback (if necessary) and the utterance score for every utterance. We also graded the utterance on a scale of A to D, and F. The grade was automatically generated using the score. If the score is higher than 90, the system generates 'A' grade; and for every 10 score the grade downs to F.

<sup>1</sup> Available at <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>, version 0.7a is used in this work

## 4. EXPERIMENT AND RESULT

The proposed system was measured using two criteria: accuracy and user satisfaction. These measures evaluate the proposed system in different ways and infer the usability and the appropriateness of the system as an effective CALL system.

### 4.1. Prediction and detection accuracy

**Table 1.** Accuracy, precision, recall and F1-scores (percentages) of the system's prediction and detection module

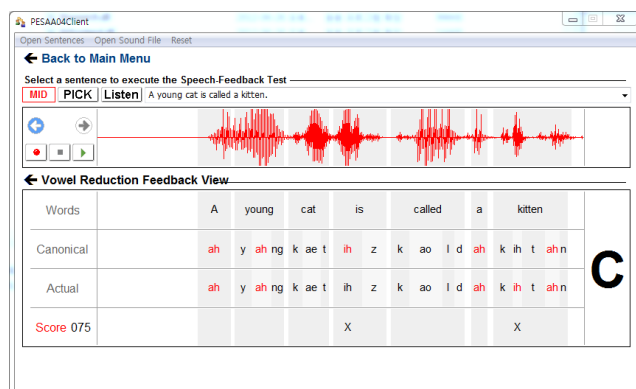
Models	Accuracy	Precision	Recall	F1-score
Prediction	90.6	84.1	82.7	83.4
Detection	95.9	87.9	89.5	88.7

We trained the prediction and detection models with the BURNC and KLEAC corpora, respectively. We measured the accuracies, precisions, recalls and F1-scores of our models using five-fold cross-validation. The numbers indicate how exactly the proposed models can predict and detect the vowel reduction in given sentences.

The precision, recall and F1-score values of our models indicate that they to predict and detect vowel reductions well (Table 1). The accuracies were high enough that the proposed feedback system can be adopted in a CALL system.

### 4.2. User satisfaction

We evaluated the user satisfaction in two ways: satisfaction with feedback and satisfaction with the user interface (UI).



**Fig. 2.** A screen capture of the system

Feedback is a very important part in a CALL system, because based on the feedback the students identify and correct their mistakes. The system shows the canonical phoneme sequence and the actual phoneme sequence. The system renders vowel-reduced phonemes in red to help the students visually identify the vowel reduction. The system generates the 'X' feedback for error phonemes, if the

**Table 2.** Learners' questionnaires for expected learning effectiveness and feedback satisfaction on a scale of 1 to 5.

Questions	Mean	s.d.
Pre-test		
Do you know about vowel reduction?	1.40	0.66
Post-test		
Does the system help you to understand about vowel reduction	3.30	1.00
Does the system point correctly to your problems in vowel reduction?	3.10	0.83
Does the system help you to improve English proficiency?	3.50	0.67
Are you interested in using this system to improve your English skills?	3.20	0.60
Are you satisfied with the score assessment?	3.00	0.63
Are you satisfied with the grade assessment?	3.70	1.10
Are you satisfied with the 'X' feedback?	3.80	0.87
Are you satisfied with the red-colored phoneme feedback?	3.70	0.78

predicted and detected vowel reductions are different, and the proposed system assesses the utterance with score and grade (Fig. 2).

To evaluate whether the system can be applied in real learning, we designed an experiment for 10 Korean university students to utilize the system. Each student used the system for one hour, then completed a questionnaire that evaluated expected learning effectiveness on a scale of 1 - 5, with positive adjectives anchoring the high end and negative adjectives anchoring the low end. With the questionnaire, the students also made free-form written comments on the system. The students' English skill was intermediate, and they were encouraged to try several times to get a sufficient grade (the grade 'A') for each sentence. They used a server-client system and uttered 70 sentences. The testers were allowed to use the system whenever and wherever they wanted.

Initially, the students were not much aware of vowel reduction (Table 2). After using the proposed system, they answered that they understood vowel reduction. The effectiveness of our system was deemed to be good; many of the students responded that our system helped them to improve their English proficiency.

The testers' subjective opinions about the usability of the UI were surveyed by using a questionnaire for user interaction satisfaction (QUIS) [19]. QUIS focuses on the user's perception of the usability of specific aspects of the UI. Each of the specific interface factors and optional sections has a main component question followed by related sub-component questions. Each item is rated on a scale from

**Table 3.** Learners' questionnaires for user interface satisfaction on a scale of 1 to 5 for the system. Each major item comprises 4-6 detailed questions.

Major items for satisfaction	Mean	s.d.
Overall system	3.38	0.15
Display content	3.38	0.19
Terminology	3.68	0.38
Easy use	3.83	0.47
Processing speed	3.30	0.42
Total	3.52	0.32

1 to 5, with positive adjectives anchoring the high end and negative adjectives anchoring the low end.

We used the short form of QUIS 5.0 to evaluate the usability of the system (Fig. 2). The sections of the original QUIS were retained, but some irrelevant items were omitted. The overall satisfaction score was 3.52/5, with a standard deviation (s.d.) of 0.32 (Table 3). This score can be considered to mean that the system is useful in real English learning and that it has an appropriate UI that helps the learners to understand their English assessment results easily.

## 5. DISCUSSION

The experiment shows that the automatic vowel-reduction feedback system can be used as computer-assisted system to facilitate learning of spoken English. The models' accuracies are high enough to pinpoint vowel-reduction errors. The overall system satisfaction of the students was positive.

Before the system utilization, the students did not know much about vowel reduction, but after using the system they did. We are satisfied that the students became familiar with vowel reductions. However, they were only slightly positive (3.10, 3.50) about the system. We consider that this relatively low evaluation is due to the low feedback scores, and conclude that vowel-reduction instruction alone does not constitute an attractive CALL system for language learners.

The students preferred the grade assessment over the utterance score assessment. The grade is calculated from the utterance score, so the two assessments are not much different. However, the grade assessment is more generous than the utterance score assessment, because grade had only five grades (A-D, F) and utterance score had a 1-100 scale. Therefore, when we upgrade the system, we should use the grade to assess utterances.

Our system's feedback seems to have worked effectively for the students (3.70, 3.80). The red phoneme helped the students to learn what vowel reduction is. The 'X' feedback also helped the students to see their vowel phoneme error quickly. Because our experiment participants' L1 was Korean, a syllable-timed language, the students just needed to reduce the "X"-positioned vowel, not the opposite much.

With the questionnaire, the students also made free-form written comments on the system. Some of the students said that if they considered only vowel reductions, the pronunciation gets unnatural. As the students wrote, the vowel reduction feedback system could be combined with other components of a CALL system, and could give a synergic effect.

## 6. CONCLUSIONS

The purpose of this paper was to build an automatic vowel-reduction feedback system for non-native speakers of English. The system had three parts: a prediction model, a detection model, and a feedback model. The prediction model and the detection model had high enough accuracies (90.6% and 95.9%). Ten Korean-speaking university students who were learners of English each utilized the system for an hour; they reported that the system helped them to learn about vowel reduction, and that they were willing to use the system to improve their English skills. The students were more satisfied with grade assessments than with score assessments, and were satisfied with the feedback styles: "X" for incorrect vowel reduction and red-colored phoneme for reduced vowels. The experiments showed that the vowel reduction system can be used effectively as a part of a CALL system.

## 7. ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0019523).

This work was supported by the IT R&D program of MSIP/KEIT. [10044508, Development of Non-Symbolic Approach-based Human-Like Self-Taught Learning Intelligence Technology]

## 8. REFERENCES

- [1] S. Lee, J. Lee, H. Noh, K. Lee, and G. G. Lee, "Grammatical error simulation for computer-assisted language learning," *Knowledge-Based Systems*, vol. 24, pp. 868-876, 2011.
- [2] J. Bang, S. Kang, and G. G. Lee, "An automatic feedback system for English speaking integrating pronunciation and prosody assessments", in *Proceedings of the 2013 SLaTE Workshop*, 2013.
- [3] A. Cutler, "Stress and accent in language production and understanding," *Intonation, accent and rhythm: studies in discourse phonology*, vol. 8, pp. 76-90, 1984.
- [4] J. Anderson-Hsieh, R. Johnson, and K. Koehler, "The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure," *Language learning*, vol. 42, pp. 529-555, 1992.

- [5] A. Cutler and S. Butterfield, "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *Journal of Memory and Language*, vol. 31, pp. 218-236, 1992.
- [6] Y. Tsubota, M. Dantsuji, and T. Kawahara, "Practical use of autonomous English pronunciation learning system for Japanese students," in *InSTIL/ICALL Symposium 2004*, 2004.
- [7] E. Fosler-Lussier, "Contextual word and syllable pronunciation models," in *Proceedings of the 1999 IEEE ASRU Workshop*, 1999.
- [8] K. You, H. Kim, H. Chang, J. Lee, and W. Sung, "A handheld english pronunciation evaluation device," in *Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers. International Conference on*, 2005, pp. 267-268.
- [9] A. Chandel, A. Parate, M. Madathingal, H. Pant, N. Rajput, S. Ikbali, O. Deshmukh, A. Verma, "Sensei: Spoken language assessment for call center agents," in *Proceedings of the 2007 IEEE ASRU Workshop*, 2007.
- [10] L. Shockey, *Sound patterns of spoken English*: Wiley. com, 2008.
- [11] C. P. Browman and L. Goldstein, "Targetless' schwa: An articulatory analysis," *Papers in laboratory phonology II: Gesture, segment, prosody*, pp. 26-56, 1992.
- [12] L. Burzio, "Phonology and phonetics of English stress and vowel reduction," *Language Sciences*, vol. 29, pp. 154-176, 2007.
- [13] D. Jurafsky, A. Bell, M. Gregory, and W. D. Raymond, "Probabilistic relations between words: Evidence from reduction in lexical production," *Typological studies in language*, vol. 45, pp. 229-254, 2001.
- [14] M. Ostendorf, P. Price, and S. Shattuck-Hufnagel, "The Boston University radio news corpus," *Linguistic Data Consortium*, 1995.
- [15] J. H. Jeon and Y. Liu, "Automatic prosodic events detection using syllable-based acoustic and syntactic features," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 4565-4568.
- [16] Y. Qian, Z. Wu, X. Ma, and F. Soong, "Automatic prosody prediction and detection with Conditional Random Field (CRF) models," in *Chinese Spoken Language Processing (ISCSLP), 2010 7th International Symposium on*, 2010, pp. 135-138.
- [17] V. Rangarajan Sridhar, S. Bangalore, and S. S. Narayanan, "Exploiting acoustic and syntactic features for automatic prosody labeling in a maximum entropy framework," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, pp. 797-811, 2008.
- [18] H. Lee, "Evaluation of Korean Learners' English Accentuation", Keynote Address, in *Proc. of the 16th National Conference of the English Phonetic Society of Japan and the Second International Congress of Phoneticians of English*, 2011.
- [19] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of an instrument measuring user satisfaction of the human-computer interface," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1988, pp. 213-218.