Pronunciation Scaffolder: Annotation accuracy

Takumi Kondo, Jun Inoue and John Blake

University of Aizu, Japan
{s1240029 | s1240024 | jblake }@u-aizu.ac.jp

Abstract

This paper reports the annotation accuracy of the Pronunciation Scaffolder (ver. 2.0), an online tool to help learners of English read presentation scripts aloud. Scripts are automatically annotated using rule-based pattern matching. Colour, font sizes and symbols are used to provide intuitive visual prompts. Annotation accuracy was evaluated by systematically checking the accuracy of all the annotation functions on a political speech. Each instance of false positive, false negative and true positive results was tagged using predefined hashcodes. The hashcodes were counted. The results show that the Pronunciation Scaffolder performed extremely well for content words and reasonably well for most other functions. However, the word stress function resulted in the many false negative results and the linking function resulted in many false positive results. These results inform the development of version 3.0.

Keywords: automated annotation, visualization, pronunciation aid, reading aloud

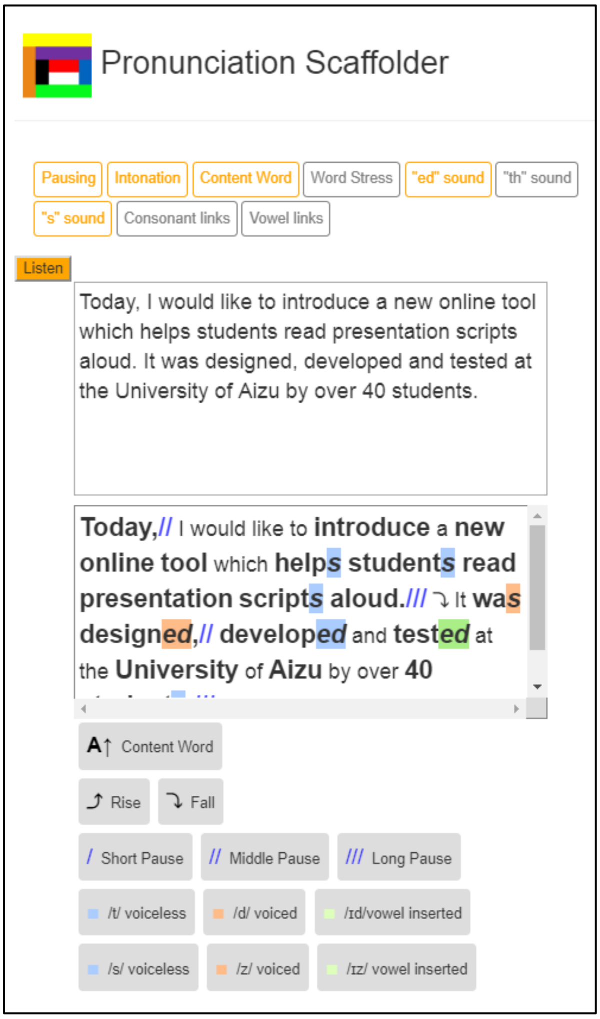
1. Introduction

Both undergraduate and graduate students in the University of Aizu are required to present and discuss their graduation thesis in English. This is a challenging task for many Japanese students. Although their language knowledge is extensive, few students are able to deliver spoken presentations with confidence. The majority of students rely on reading prepared scripts aloud.

Teachers and students frequently annotate presentation scripts by hand to enable the scripts to be read aloud more easily. This is particularly common when presenting in a second or additional language. Given that pronunciation rules follow discernable patterns, this annotation process could be automated. The main purpose of the Pronunciation Scaffolder (ver. 2.0) [1] is to help students read out presentation scripts more appropriately by visualizing pronunciation.

Users input a text and pronunciation features are visualized using colour, size and symbols. A screenshot of the interface and sample output of the Pronunciation Scaffolder is shown in Fig. 1. The visualizations are designed to be intuitive to reduce the memory burden when reading scripts aloud.

The Pronunciation Scaffolder has been deployed since December 2017 and has received very positive anecdotal feedback from both teachers and students in Japan and across Asia. To identify areas for improvement, an evaluation of the annotation accuracy was conducted.



Today, I would like to introduce a new online tool which helps students read presentation scripts aloud. It was designed, developed and tested at the University of Aizu by over 40 students.

Today,// I would like to introduce a new online tool which helps students read presentation scripts aloud./// ⤵ It was designed,// developed and tested at the University of Aizu by over 40

Figure 1: Screenshot of Pronunciation Scaffolder (mobile viewport)

10.21437/ISAPh.2018-15

1.1. Pronunciation scaffolder

The Pronunciation Scaffolder was designed primarily to be used by Japanese speakers of English who need to read presentation scripts aloud in order to fulfill their graduation requirements. This free web application automatically tags presentation scripts for various pronunciation features.

To the best knowledge of the authors, there is no other online tool that helps users read texts aloud by visualizing key pronunciation features such as the length of pauses, whether to use rising or falling intonation and which words to stress. A small-scale study of the efficacy of the first prototype of the Pronunciation Scaffolder (ver. 1.0) was conducted in 2017. The study reported positive results, particularly for lower-level learners [2]. Usability studies were conducted to improve the interface as reported in [3].

The Pronunciation Scaffolder provides users with the choice of which features they want visualized. There are three categories, six features and nine function buttons. Pronunciation features are divided into three categories: core, optional and advanced. The core features are recommended for lower-level learners and include pausing, intonation and content words. The optional features are word stress, sounds of the letter “s” and two graphemes, “-ed” and “th”. These features are designed to provide additional assistance to learners with specific difficulties. The advanced feature is linking, which divided into linking from consonant-ending words and vowel-ending words.

Table I shows a typical sequence of script reading practice activities used by teachers of English for presentations. When helping students read scripts aloud, teachers tend to focus on each pronunciation feature in isolation before combining them with other features. The first three features in this sequence form the core category.

Table 1: Pronunciation features visualized

#	Feature	Purpose of annotation
1	Pausing	show short, medium and long pauses
2	Intonation	show falling and rising intonation
3	Rhythm	stress content words more than function words
4	Word stress	locate syllable carrying primary stress
5	Sounds	disambiguate between letters realized by different sounds
6	Linking	shows elision, intrusion and linking

1.2. Annotation accuracy

Software development teams of computer science students developed regular expressions to match pronunciation features and annotate a discrete function. Teams conducted their own accuracy evaluations. However, no evaluation of overall accuracy of the final deployed version (ver 2.0) was conducted. This research project aims to fill that gap and identify areas for improvement. Annotation accuracy can be measured by comparing the results of automated annotation with manual annotation. The results that are of particular importance are the false positives and true positives because users pay more attention to these categories. The false negative results are also of concern, but are not the primary focus of this evaluation.

Table 2 shows the contingency table, or confusion matrix, for matched outcomes and language features. True positive results are the pronunciation features that are correctly identified and annotated. False positive results are those that are annotated but are not correctly matched. False negative results are those that contain the pronunciation feature, but are not matched.

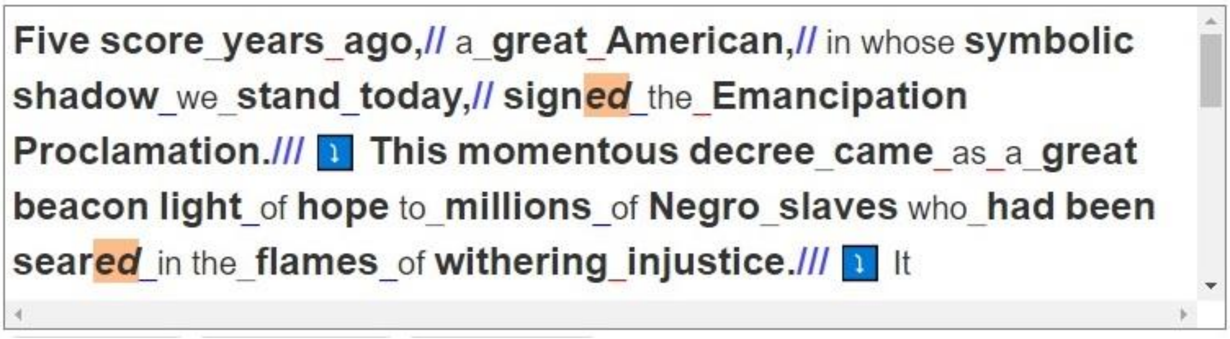
Table 2: Matrix for type of results

	True	False
Positive	True positive	False positive
Negative	True negative	False negative

Regular expressions match strings of characters that conform to the rule, but cannot distinguish between true and false positives. The problem is inherent in the rule. If the rule is coded to exactly reflect natural language usage, then there should be no false positives; however, this is a non-trivial task. The primary aim when creating regular expressions is to maximize true positive results while minimizing false positive results. Although false negative results are unwanted, they will most likely go unnoticed by many users.

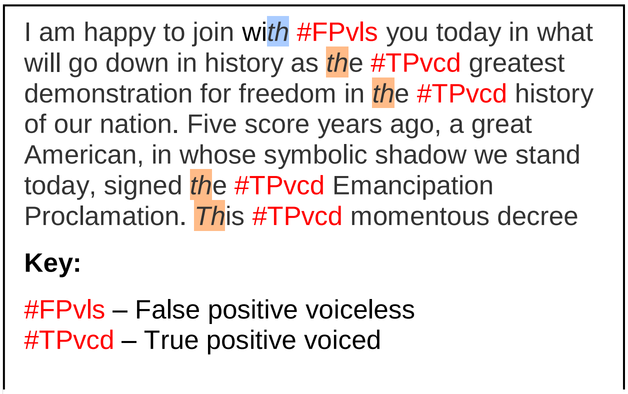
2. Method

Based on its worldwide fame, the full text of “I have a dream” speech delivered by Martin Luther King Jr. was selected as the vehicle for evaluation. The text (n=1651 words) was submitted into the Pronunciation Scaffolder, and the output of each function button was saved. Fig. 2 shows a screenshot of an extract of the output. The true and false positive results were manually tagged by an experienced teacher of English, and then checked by another annotator. Fig.



Five score_years_ago,// a_great_American,// in whose symbolic shadow_we_stand_today,// signed_the_Emancipation Proclamation./// ⤵ This momentous decree_came_as_a_great beacon light_of hope to_millions_of Negro_slaves who_had been seared_in the_flames_of withering_injustice./// ⤵ It

Figure 2: Screenshot of output for "I have a dream" speech



I am happy to join with #FPvls you today in what will go down in history as the #TPvcd greatest demonstration for freedom in the #TPvcd history of our nation. Five score years ago, a great American, in whose symbolic shadow we stand today, signed the #TPvcd Emancipation Proclamation. This #TPvcd momentous decree

Key:

#FPvls – False positive voiceless
#TPvcd – True positive voiced

Figure 3: Extract from "th" function output coded for TP, FP and FN results

3 shows an extract of the "th" function output coded for true positive, false positive and false negative results. The annotations were automatically tallied using a specially written python script. The underlying causes for false positive and false negative results were investigated by close analysis of the regular expressions in the program code.

3. Results

The results of the accuracy analysis are shown in Table 3. Eighty percent of all the false positive results occurred in the linking function, and just under ninety percent (87%) of false negatives occurred in the word stress function. True positives account for 95% of all the visualized annotations. For the core functions of pausing, intonation and content words, the accuracy rate is substantially higher with a true positives accounting for 99.5% of all the results.

3.1. Pausing

The current pausing function draws heavily on punctuation as a guide for reading aloud. The regular expressions need to be expanded to indicate pauses following exclamation marks and double dashes (--). The number of false negative results could be reduced by indicating short pauses after more multiword noun phrases and sentence-initial adverbial phrases.

Table 3: Total counts of false positive, false negative and true positive results

Feature	**Count**		
	False positives	*False negatives*	*True positives*
pausing	2	44	108
intonation (sentence-final)	0	0	84
content words	0	0	1651
word stress	0	334	67
"-ed" sounds (word final)	6	0	65
"th" sounds	25	0	190
"-s" sounds (word final)	8	5	171
linking	165	0	1485
Total	206	383	3821

3.2. Intonation

Intonation has both grammatical and attitudinal functions enabling speakers to add another layer of meaning to the words that are said [4]. However, the Pronunciation Scaffolder can only access endotextual linguistic features and so visualizes the grammatical-type of intonation. The current version, however, only indicates sentence-final intonation. This function could be enhanced by annotating rising intonation before conjunctions, such as: *or, and, but, so* and *because*.

3.3. Content words

Content words are shown in a larger font while function words are shown in a smaller font, making is easy for readers to add rhythm when reading aloud. The content word function, which harnesses nlp-compromise [5], a natural language processing JavaScript library, was the most accurate function overall with no false positive or negative results.

3.4. Word stress

The word stress function produced the highest number of false negative results. This is unsurprising given the complexity of demarcating syllables and identifying word stress [6]. The rule-based pattern-matching algorithm currently harnesses syllable counting from the word end to identify word stress, and does not take into account other factors, such as part of speech and word origin. This function will be replaced by one that uses a python script to access an online dictionary and display the primary stress using a raised red vertical line before the stressed syllable.

3.5. "-ed" sounds

This function was designed to help learners pronounce the *-ed* suffix used to form the past for regular verbs. The false positive results occurred in different uses of *-ed*. The past participle adjective *crooked* was annotated incorrectly as /t/ rather than /id/. The other false positives were caused by annotating *ed* occurring within words, such as *deeds* and *redemptive*.

3.6. "th" sounds

Of the 25 false positive results, 17 were caused by *with*, seven by *their* and one by *though*. In each case due to a coding error, a voiceless sound was indicated instead of the correct voiced sound.

3.7. "-s" sounds

Over 95% of the instances of letter *-s* were annotated correctly. Five out of the eight false positive instances were caused by the substring *-ous* which should end with a voiceless sound but was incorrectly annotated as voiced. All five false negatives were caused by the code not taking account of apostrophes used before s to show possession (e.g. father's).

3.8. Linking

The linking function resulted in the highest number of false positives. Figure 2 shows the annotation key for most of the pronunciation features. However, during the numerous updates to create an easy-to-discriminate colour scheme, the final version of the deployed scheme for linking was not updated in the key, shown in Fig. 4. This is, however, easily rectified. In addition, some regular expressions need further refinement to deal with silent letters and some overlooked cases.

3.9. Miscellaneous

There were some bugs in the code when the *th* and *-s* function buttons were used simultaneously or consecutively. The bugs occurred when a voiceless sound followed a voiced sound in the same word, such as in the word *this*.

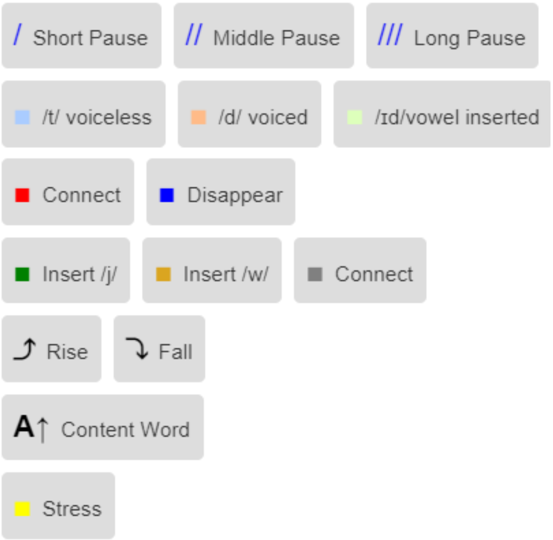


Figure 4: Annotation key for Pronunciation Scaffolder

4. Conclusions

Most of the false positive results can be reduced by simple alterations to the program code. These changes should result in a much more accurate tool that users can access online anytime. We also recommend that the next version of the Pronunciation Scaffolder (ver. 3.0) should incorporate a new word stress function. The intonation function could also be extended by adding in mid-sentence intonation, when intonation can be predicted by signpost words, such as conjunctions.

5. Acknowledgements

We would like to thank Keigo Ebihara for his leading role in the software development and creation of the current interface, and Hiroki Inoue for conducting a series of usability studies.

6. References

[1] Blake, J. 2017. Pronunciation Scaffolder (ver. 2.0) [online tool]. Available at: *www.jb11.org/pronunciationScaffolder.html*

[2] Blake. J. 2017, November 24-26. Automated presentation script annotation tool: Development and evaluation. Paper presented at the 15th International Conference of AsiaCALL. Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam.

[3] Inoue, H. 2018. *Verification and improvement of software to support reading English aloud.* Graduation thesis. University of Aizu.

[4] Tench, P. 1994. *The Intonation Systems of English.* London: Cassell.

[5] Spencer K. 2017. nlp-compromise [software]. Available at: *https://github.com/nlp-compromiose/compromise*

[6] Fudge, E. 1984. *English word-stress.* London: Routledge.