Improving Japanese English pronunciation with speech recognition and feedback system

Kanta Igarashi[1,*] and *Ian* Wilson[1,**]

[1]CLR Phonetics Lab, University of Aizu, Tsuruga, Ikki-machi, Aizuwakamatsu, Fukushima-ken, Japan

Abstract. For Japanese people, communicating with English speakers from abroad has become more common because of internationalization, and there are many people who want to improve their English-speaking skills. However, there are few environments where we can speak English outside of the classroom, so Japanese students rarely have a chance to study English pronunciation. Even if students do have a chance to take an English pronunciation class, teachers do not have enough time to individually teach each student pronunciation in a big class. Because of that, computers and smartphones may be one good type of tool to solve this problem. In this research, we develop a web-based application to help Japanese learners with their English pronunciation.

1 Introduction

According to Dizon [1], intelligent assistants such as Siri or Alexa, which use speech recognition, could be helpful tools for language learners, especially Japanese second-language (L2) English learners, who usually do not have many opportunities to use English outside of the classroom. Daniels and Iwago [2] have compared Siri to Google Speech Recognition (GSR) and found that GSR is more accurate for transcribing Japanese undergraduates' L2 English speech. However, although intelligent assistants can recognize speech, they do not advise users about their pronunciation. So, the current research is focused on creating a web-based feedback system that would be helpful for language learners. The feedback system advises users from the point of view of articulation and correctness based on typical errors in real test results that we collected. In this research, the web application with feedback system to evaluate users' English pronunciation uses Mozilla Developer Network (MDN)'s Web Speech API [3]. In order to make the range of possible feedback more manageable, the words and phrases used for this web application are limited. They are chosen from the "Wolf Story" [4], a phonetic passage adapted from an Aesop fable, because of the wide range of English phonemes and different combinations of phonemes. It is easy for teachers and students to access the system, because the web application is hosted on the University of Aizu's CLR Phonetics Lab website at http://clrlab1.u-aizu.ac.jp/acoustics.html.

2 Speech Recognition Interface

There are many speech recognition applications, so it is possible to choose an interface according to one's individual requirements. For example, Google Speech Recognition (GSR), IBM Cloud, Siri, Alexa, and Speech Recognition API (MDN) are all popular. As previously stated, GSR is better for L2 English learners than Siri and Alexa. So, we compared GSR, IBM, and MDN's API. Table 1 shows a list of these APIs' advantages and disadvantages.

Table 1. Comparison of three APIs for speech recognition (GSR = Google Speech Recognition; IBM = IBM Cloud; MDN = Mozilla Developer Network's Web Speech API)

	Advantages	Disadvantages
GSR	•High accuracy	• Need account, many source files, must pay according to use
IBM		• Need account, must pay according to use
MDN	• Need only Javascript • Free	•Accuracy is lower •Only works with Chrome

3 Method

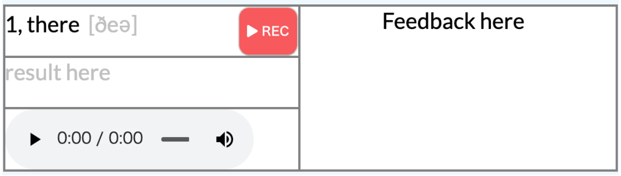
The web application created in this research consists of HTML, CSS and JavaScript. First, the HTML and CSS functions are "recognition start button", "sample mp3 files control", etc. Javascript controls the speech recognition and feedback system. The Web Speech API is an interface that Mozilla Developer Network (MDN) provides and this API can recognize any language. In this web application, we selected American English as the language setting because much of the English education in Japanese primary and secondary schools seems to focus on American English. In addition, the application needs users to install and use Google Chrome because MDN's Web Speech API works on Google Chrome only (this API does not work on iOS).

[*]e-mail: s1240221@u-aizu.ac.jp

[**]e-mail: wilson@u-aizu.ac.jp

4 Web Application Details

This web application uses Web Speech API [3]. A part of the web application is shown in Figure 1. The application includes a feedback system and a lecture page. The feedback system must give advice to users, so we created an advice dataset corresponding to words from test participants that were commonly misunderstood by the speech recognition system. In case the recognized word is not in the dataset, the feedback given is very general ("please try again and speak more clearly").



1, there [ðeə]

result here

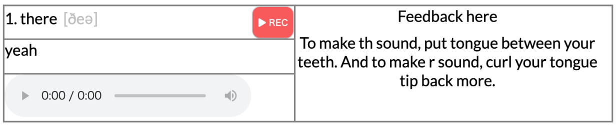
Feedback here

Figure 1. Part of the "word practice" page of the resultant web application

The lecture section of the application contains three pages. First, an explanation of what the International Phonetic Alphabet (IPA) is. Second and third, lessons teaching how to pronounce vowels and consonants, respectively. For each type of sound, static MRI images of articulators producing the sounds can be seen. Although this study's purpose is a complete web application with feedback system, we also made a lecture page to help with the study of English pronunciation.This application is mainly for Japanese L2 English learners, so the lecture page is written in Japanese. According to Ono [5], the way of effective pronunciation teaching is that the student repeats the teacher's English pronunciation and the teacher explains, in words and illustrations, how to pronounce.

4.1 Feedback System

The web application can return advice for a word from a list of common errors by Japanese speakers. Figures 2, 3, and 4 show examples of feedback. In case the recognized word is not in the list, the application returns common advice. For example, if the word is missing "th" sound, the application returns appropriate advice such as "Put your tongue tip between your teeth to make "th" sound."



1. there [ðeə]

yeah

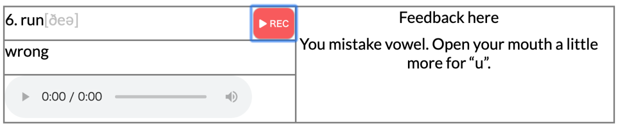
Feedback here

To make th sound, put tongue between your teeth. And to make r sound, curl your tongue tip back more.

Figure 2. Example showing feedback given to the user for "th" sound

4.1.1 Common Pronunciation Errors

There are many common errors that Japanese speakers make when speaking English as a second language. Table 2 and Figure 5 show the consonants and vowels, respectively, that exist in Japanese and English. In Table 2, consonants only used in English are shown in blue and consonants only used in Japanese are shown in red. Also, consonants used in both English and Japanese are shown in black. Common pronunciation errors by Japanese speakers of English usually occur in vowels and consonants that are not found in Japanese (i.e., the ones shown in blue). In Figure 5, the vowels underlined in pink are those found in North American English but not in Japanese.



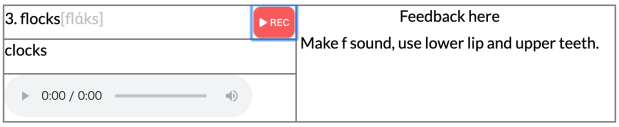
6. run[ðeə]

wrong

Feedback here

You mistake vowel. Open your mouth a little more for "u".

Figure 3. Example showing feedback given to the user for a vowel sound



3. flocks[flɑ́ks]

clocks

Feedback here

Make f sound, use lower lip and upper teeth.

Figure 4. Example showing feedback given to the user for "f" sound

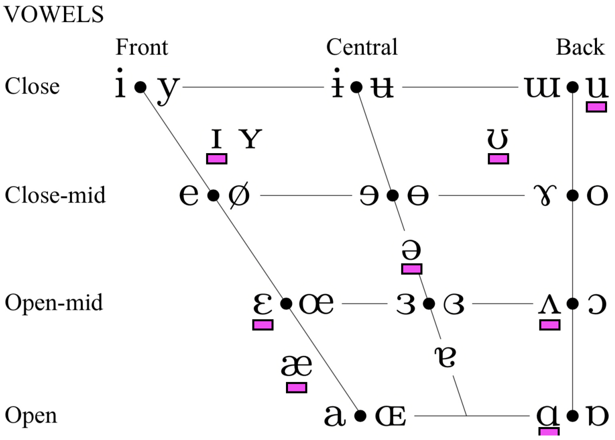


Figure 5. International Phonetic Alphabet chart of vowels showing (underlined in pink) English vowels not found in Japanese

4.1.2 Data Collection

In order to build a feedback system, we collected real examples of the Web Speech API's transcriptions of Japanese speakers' English utterances. In order to collect test samples, we created a website that recognizes users pronunciation and sends data by email to us. Participants were mainly students of the University of Aizu. Anyone could access the website easily because it is hosted on the University of Aizu's CLR Phonetics Lab website at http://clrlab1.u-aizu.ac.jp/acoustics.html. The data collected can be seen in Table 3. The target word is shown

Table 2. International Phonetic Alphabet chart of consonants showing differences between English and Japanese. Red phonemes are found only in Japanese, blue ones are found only in English, and black ones are found in both languages

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	
Nasal	m			n			ŋ	
Trill								
Tap or Flap				ɾ				
Fricative	ɸ	f v	θ ð	s z	ʃ ʒ			h
Lateral fricative								
Approximant	w			ɹ		j		
Lateral approximant				l				
Affricate				ts dz	tʃ dʒ			

in the left column, and for each participant (A–M) the resultant transcription by MDN's Web Speech API is shown. The number of mismatches ("errors") between target word and transcription can be seen in the bottom row. Note that even for the native speaker (participant L), the system indicated an error for two words: *threaten* and *fool*.

4.2 Lecture Page

In the lecture page, learners are taught how to pronounce English vowels and consonants with MRI pictures. Figure 6 shows the top page of the web application. This webpage has four tabs. In "About IPA", "Vowels" and "Consonants" pages, it explains how to read the IPA chart and how to make vowel and consonant sounds. In the "Word Practice" page, users can practice pronunciation and get feedback. I referred to [6] in making the "Vowels" and " Consonants" pages.

5 Source Code Description

5.1 HTML Source Code

The following is HTML source code. This code is for showing what is found in Figure 1.

```
<div class="box_2">
 <div class="practice-place">
  <div class="example">
   <div class="practice-right">1, there
    <span class="ipaDiv">[]</span>
   </div>
   <button  class="rec-btn"id="p1">
   <i class="fas fa-play"> REC</i></button>
  </div>
  <div class="resultDiv" id="practice-result1">
   result here</div>
  <div class="btn-box">
  <div class="audio-btn">
  <audio src="there.mp3" controls>
  </div>
 </div>
</div>
```

The above code is for the left side of Figure 1. This display example word and REC button in example class, recognized word place in resultDiv class and example sound in audio.

```
 <div class="feedback-place">
  <div class="fb-title">Feedback here</div>
  <div id="fb1"class="feedback-here"></div>
 </div>
</div>
```

The above section of code is for the right side of Figure 1.

5.2 Javascript Source Code

The following is part of the Javascript source code.

```
var misThere =[
  ["yeah","To make th sound, ..."],
  ["sarah","You are confusing ..."],
  ["date","Put your tongue ..."],
  ["dale","You confuse d sound ...."],
  ["zelle","You confuse z ...."]
];
```

This is part of arrays for making advice list.

```
 startBtn[0].onclick = () => {
  recognition[0].start();
}
recognition[0].onresult = (event) => {
  clickFunc(0,event);
}
```

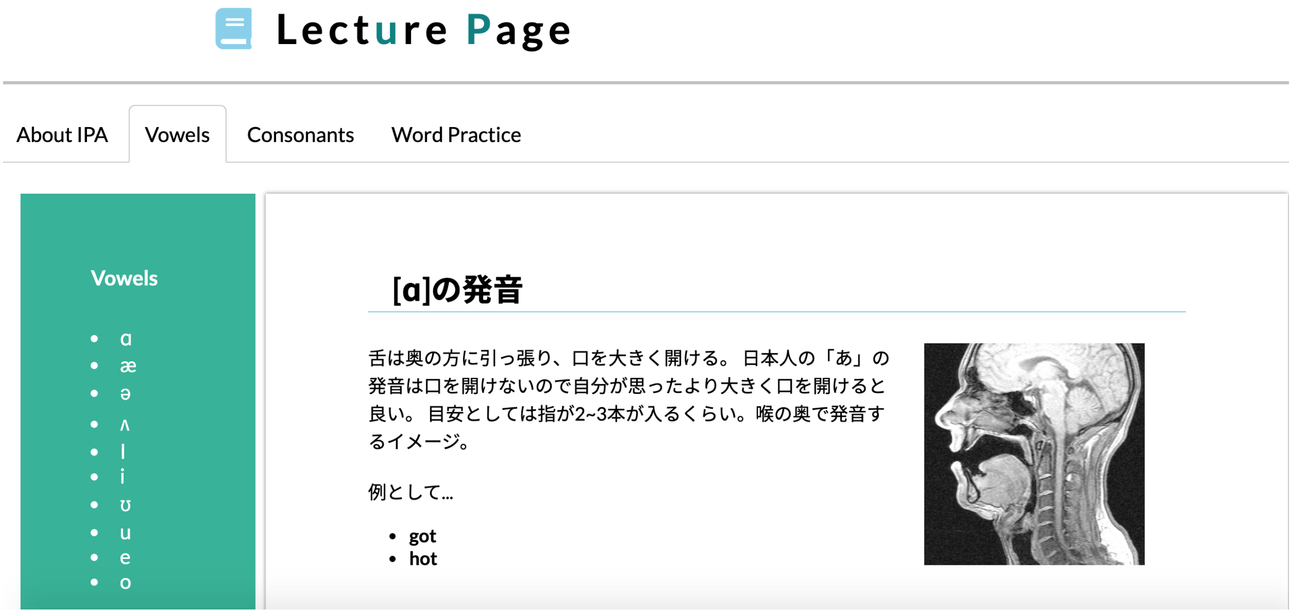
When the REC button is clicked, the speech recognition function is called and starts. Also, each button has a number to identify it.

```
var clickFunc = function(clickNum,event){
  finalTranscript='';
  let interimTranscript = '';
  for (let i = event.resultIndex;
  i < event.results.length; i++) {
    let transcript = event.results[i][0].
                             transcript;
    if (event.results[i].isFinal) {
      finalTranscript += transcript;
    } else {
      interimTranscript = transcript;
    }
  }
  resultDiv[clickNum].style.color = '#000000';
  resultDiv[clickNum].innerHTML = finalTranscript
  + interimTranscript;
  feedbackDiv[clickNum].innerHTML =
                        checkFunc(clickNum);
}
```

The clickFunc code above is for displaying the results of recognition and feedback.

Table 3. Results of MDN Web Speech API speech recognition for 13 participants including one native speaker

	Participants (A–M)												
sample	A	B	C	D	E	F	G	H	I	J	K	L (native)	M
there	there	yeah	yeah	there	sarah	date	there	there	dale	zelle	sarah	there	there
poor	cooler	frac		poor	poor poor	poor	pearl	cool	cool	hua	torah	poor	poor
flocks	flat	frock	trunks	crocs	lyrics	clocks	trucks	rocks	crooks	clocks	frocks	phlox	phlox
dark	dark	DAC	dak	dark	doc	doc	dirk	dark	dog	dogs	dark	dark	dogs
plan	brown	put on		prawn	rap	brown	plum	put on	brown		prom	plan	slime
run	noun	land		run	long	ren	long	la	wrong	wrong	wrong	run	wrong
village	ferrets	Dodge	petr cech	greenwich	spirits	we teach	pH	pH	spinach	philles	village	village	village
little	little	retina	little	little	critter	I don’t	little	later	little	little	little	little	beetle
fun	song	fan	front	fun	fang	phone	farm	farm	farm		phone	fun	bomb
third	todd	turn	sad	third	salt	surd	third	sade	sad	surag	shirt	third	todd
wolf	hello	all right			lowe’s	oops	bro	both	wolf	waltz	earth	wolf	wall
threaten	threatened	Toyota	threatened	threatened	samaritan	threatened	creighton	slaton	sheraton		veteran	threatened	310
sheep	sheep	sweet	seep	sheep	seat	see	sheep	sheep	seep		sheep	sheep	sheep
louder	louder	ladder	ladder	louder	browser	rhoda	louder	yoda	louder	louder	louder	louder	louder
exactly	exactly	exactly	exactly	exactly	echoes act 3	duck tree	exactly	exactly	you can suck 3	exactly	exactly	exactly	exactly
who	who	who	2	who	who	TRUE	who	food	who	crew	who	who	who
bother	bother	brother	bozeman	brother	weather	butter	bother	buzzer	father	squirrel sound	bouncer	bother	bother
fool	who	woo	food	full	food	coup	whole food	food	who	who	tune	cool	pool
cousins	cousins	cousin	cousins	cousins	cousins	consonance	cousin	cozy	cousins		cousin	cousins	cousin
fist	quit	quest	tryst	fist	fist	fixed	first	feast	fist	wrist	cost	fist	best
# OF ERRORS	11	18	13	5	16	19	12	16	14	12	13	2	11



[ɑ]の発音

舌は奥の方に引っ張り、口を大きく開ける。 日本人の「あ」の発音は口を開けないので自分が思ったより大きく口を開けると良い。 目安としては指が2~3本が入るくらい。喉の奥で発音するイメージ。

例として…

- **got**
- **hot**

Figure 6. The Vowels page of the web application

```
var checkFunc = function(clickNum){
  var result=resultDiv[clickNum].innerHTML;
  var advice="";
  var listFlag = 0;
  result = result.toLowerCase();
  if(result.indexOf(sample[clickNum]) != -1)
  return 'correct!';
```

The checkFunc code above makes and returns advice according to recognized word. If the recognized word is not collected, go to branch of switch.

```
  switch (clickNum) {

    case 0:
    for(let i = 0;i<misThere.length;i++)
    {
     if(result.indexOf(misThere[i][0]) != -1)
     {
        advice = misThere[i][1];
        listFlag = 1;
      }
    }
```

If the recognized word is not in the advice list, return common advice against the word.

```
    if(listFlag == 0){
      if(result.indexOf("th") == -1){
        advice+=commonAdvice["th"];
      }
      if(result.indexOf("r") == -1){
        advice+=commonAdvice["r"];
      }
    }
```

```
    break;

    case 1:
    .......
```

6 Future Work

We believe that this web application should be developed further, because the feedback system in the web application is based on too little result data. In the future, we must collect more recognized word data and improve the feedback system.

References

[1] G. Dizon, Computer Assisted Language Learning **29**, 1249 (2016)

[2] P. Daniels, K. Iwago, JALTCALL Journal **13**, 229 (2017)

[3] Mozilla Developer Network, *Web Speech API*, `https://developer.mozilla.org/ja/docs/Web/API/Web_Speech_API`

[4] D. Deterding, Journal of the International Phonetic Association **36**, 187 (2006)

[5] K. Ono, J. Fac. Edu. Saga Univ **17**, 57 (2012)

[6] A. Nogita, Japan Association of Foreign Language Education bulletin **21**, 1 (2018)