Sounds of Englishes: English pronunciation database [*]

○ Noriko Nakanishi (Kobe Gakuin University)

1 Introduction

This paper is a preliminary report of an ongoing project that seeks to build a database for elucidating the trends of various accents of English across the world (World Englishes, WE). As of July 2020, the database contains over 70,000 audio files read by approximately 600 participants from 48 countries, and they are accessible and downloadable in mp3 format for pedagogical and research purposes [1]. The files are tagged with speakers' home region, generation, and gender.

Regarding speakers' home region, most of the English educational materials focus on the accents of native speakers (NSs) [2]. However, as suggested by Kachru's concentric circles [3], non-English native speakers (NNSs) outnumber NSs when WE communication takes place. The database includes pronunciations of NNSs as well as NSs that meet communication needs in this era of globalization.

The pronunciation trends of different generations among British English (BrE) and General American English (AmE) speakers are considered in the Longman Pronunciation Dictionary (LPD3) [4] (Fig. 1). However, the data shown in LPD3, which is the most recent version, is based on the results of the Pronunciation Preference Poll conducted two decades ago [5,6]. Since the generation considered to be *younger* in the 1990s is *older* now, the current project aims to update the data. Information regarding speakers' birth year (generation) can be utilized for tracing the language change.

Speakers' gender has also been provided as a possible indicator of language change, since females are often "quick to adopt new trends in speech and lead the linguistic change [5]." It is also reported that, all other factors being equal, females tend to speak with a more prestigious accent than their male counterparts [7].
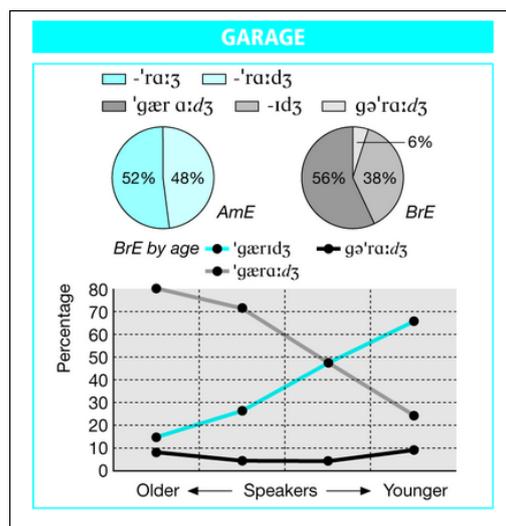

Fig. 1 A sample of the preference poll for the word *garage* shown in LPD3.

2 Selection of Words and Sentences

Following pilot studies in 2018 and 2019, a total of 359 words and sentences were selected primarily from three sources.

2.1 LPD3 and Wells (1998)

In addition to the 263 words with "uncertain or disputed pronunciation [4]" introduced in LPD3, nine words were taken from Wells' Pronunciation Preference Poll [8] conducted in 1998. Table 1 summarizes the number of pronunciation features in question for BrE and AmE accents. The total number of focused features exceeds the target words, since in some cases, a single word has more than one debatable feature. For example, for the word *garage* (Fig. 1), participants' pronunciation preference is classified based on the place of primary stress and also the articulation of the consonant at the end of the word.

Table 1 Number of features focused in LPD3

	BrE	AmE
Stress	52	41
Consonants	62	49
Vowels	101	56

2.2 Lexical sets for vowels and consonants

In dealing with the different variants of WE, it is necessary to take into account the possible differences in the quality of sounds that are considered to fall under the same phoneme. For example, the sound qualities of the diphthong in *face*, or the consonant at the beginning of *wet*, may vary in different parts of the world. Thus, the lexical sets for 23 vowels [7] and 36 consonants [9] were added to the vocabulary to be examined.

2.3 Sentences

Besides vowels and consonants, connected speech and suprasegmental features are also essential elements of speech. In order to observe the phenomena of linking, elision, assimilation, and weak forms of the words, 28 sentences were selected from an introductory English phonetics textbook [10]. Furthermore, for introducing variety to the intonation patterns, descriptions were prepared to indicate the intention of each sentence and the emotions of the speaker.

3 Data Gathering

During the pilot study in 2018 and 2019, data were gathered individually using PowerPoint slides and the Praat recording function. However, due to the ongoing Covid-19 pandemic, the sound data gathering process was moved online, and participants were contacted through email correspondence and social media.

3.1 Survey form

A pronunciation survey form was prepared online (Fig. 2) for participants across the globe to access and record their pronunciations. MediaRecorder API was used for the recording and saved in mp3 format with 128 kbps, mono.

After being informed of the instructions on suitable devices, PC settings, and ways of recording, the participants are asked to enter their language background, gender, and their year of birth. For the language background, a list of United Nations' classification [11] was prepared, from which the participants can select the region that has most strongly influenced their English accent.

For gender and birth year, a choice of "I prefer not to answer" is given as an option. All other personal information regarding participants' identity was kept confidential.
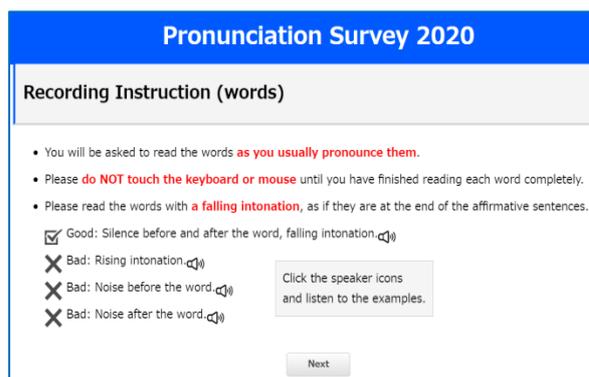

Fig. 2 An excerpt from the online survey form

The survey is divided into three sections, each of which consists of 105 to 121 words and 9 or 10 sentences. On average, it takes 10-15 minutes to complete each section. To counterbalance the order effect (participants' articulation gets better/worse as they get used to/tired of the recording), the words and sentences are shown in a random order each time they open a new section.

3.2 Participants

Currently, calls for participants are made online, mainly through social media networks such as Facebook and Twitter, and online forums hosted by LibriVox [12] and personal communication with the users of Forvo [13].

A limitation of inviting participants to the remote survey is that the sound quality cannot be controlled. An automatic play-back function is added after the recording of each word, which encourages the participants to re-record when they notice too much noise during the first recording.

An advantage of the online survey is that it attracts a wide range of participants. The online research has made it possible to recruit participants of different age groups from various countries around the world, while the data from the face-to-face survey can be acquired only at the location where the researchers are present.

4 Analyses

While the Pronunciation Preference Poll [5, 6] asked directly about the participants' preferred pronunciation in written form, the current survey asks for recordings of their pronunciation. To compare the results of the current survey with the previous data, each audio file needs to be labeled corresponding to the contrasts reported in the previous research. At this juncture, acoustic analysis is set aside, and labeling is being done aurally for the following reasons:

・ It is uncertain whether the sound quality of individual recordings withstands acoustic analysis.

・ It is impossible to judge various accents of WE together on a unified criterion.

4.1 Analysis form

The 359 words and sentences used in the survey were divided into four categories, in accordance with the features that are likely to vary among speakers; i.e., word stress, vowel, consonant, and intonation, and further divided into subcategories as follows. The numbers in parentheses indicate the number of words or sentences to be analyzed in each subcategory.

・ Word stress: location of primary stress in 2-syllable (30), 3-syllable (25), 4-syllable (13), 5-syllable (7), and 6-syllable (1) words

・ Vowel: tense - lax (27), open - mid - close (63), front - back (23), diphthong (74), r-color (23)

・ Consonant: voiceless - voiced (36), /j/ - fricative - affricate (37), epenthesis / elision (26), voicing / glottal (16)

・ Intonation: fall / rise / fall-rise / rise-fall / level (28)

4.2 Labeling the pronunciation features

The ongoing analysis is being carried out by 27 undergraduate students supervised by the author. The students were initially recruited based on their understandings of the research objectives and pronunciation score of the Versant English speaking test.

After receiving one year (30 sessions) of phonetics training, the students were given access to the analysis form (Fig. 3). The filename "POL_f_1998_190704_B7130_13_address (n)" indicates that the noun form of *address* was gathered by a researcher #B7130 on July 04, 2019, pronounced by a Polish female born in 1998. The length of each sound is 1 to 4 seconds. The students listen to each word and indicate how it is pronounced by ticking one of the given options. In the case of *address* (n) in Fig. 3, the option is ADDRess /ˈædr es/ or adDRESS /əˈdres/. The four boxes aligned vertically at the right are for reporting problem(s) with each audio file, i.e., suggestion(s) for adjusting noise, volume, length before and after the sound. The audio files with rising intonations (the fourth box) will be given a tag and kept as they are. All the files are aligned in a random order every time the students access the page. At the beginning of each labeling process for new sets of words (word stress, vowel, and consonant), a 90-minute orientation session is held to confirm the labeling criteria. Labeling on intonation patterns will be considered independently later this year.

Fig. 3 A sample of the analysis form

As of July 2020, a total of 440,006 tentative labels are given to 47,915 audio files (73 for stress, 108 for consonant, and 156 for vowel). Each of these tentative labels is counted as one vote, and when a word recorded by one participant receives ten or more votes, the labeling process goes to the final stage. The criteria for the final labeling are as follows:

・ When more than 80% of the votes go to the same option, the file will be labeled with that feature.

- When none of the options for a word receives more than 80% endorsement, the file will be returned to the analysis form.
- When the returned word receives 20 or more votes, and still none of the options receives more than 80% endorsement, the file will be given a label to indicate that the feature is uncertain.

5 Database

As shown in Fig. 4, the users of the database can access and download the pronunciations of the words read by speakers of different genders and generations from various regions in the world. Currently, a follow-up survey is being conducted for British and North American English speakers, for adding the tags that indicate the regional classification within the country. The audio files for the 28 sentences will also be uploaded later, after adding the tags to indicate the instructed emotion at the time of recording.
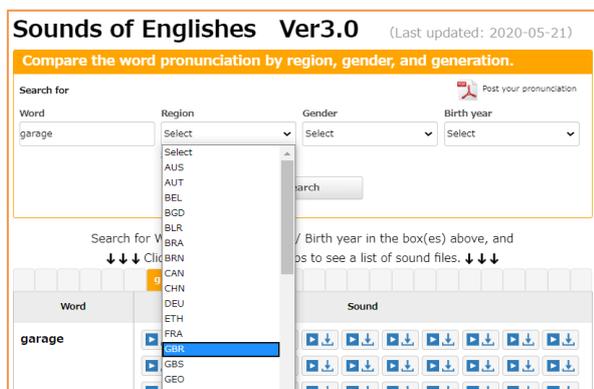

Fig. 4 Database "Sounds of Englishes"

The database can be used for pedagogical and research purposes. It can be especially useful as an educational tool for English learners in monolingual classrooms. It will prepare them to deal with various accents before facing the WE in the future, such as in study-abroad programs or at their workplace. Familiarity with, and acceptance of, linguistic diversity will give them confidence to speak with their own accent in WE communication. Simultaneously, it will prevent discriminatory attitudes toward NNSs English accent. For research purposes, the database has the potential to cover a wide range of English speakers around the world who pronounce the same words and sentences. However, it is imperative to recognize that the sound quality may vary depending on the recording environment. As the number of audio files continues to increase, statistical analysis will be conducted to compare the pronunciation trends by speakers' home region, gender, and age when adequate data are gathered.

References

[1] Nakanishi, Sounds of Englishes (Ver3.0), https://noriko-nakanishi.com/sounds/.

[2] Jenkins, Which pronunciation norms and models for English as an International Language? *ELT Journal, 52(2)*, 119-126, 1998.

[3] Kachru *ed*, *The Other Tongue: English Across Cultures*, UI Press, 1992.

[4] Wells, *Longman Pronunciation Dictionary. 3rd ed*. Longman, 2008.

[5] Shitara, A survey of American pronunciation preferences. *SHL, 7*, 201-232, 1993.

[6] Wells, Pronunciation preferences in British English: A new survey, *Proc. ICPS*, 1999.

[7] Wells, *Accents of English 1: An Introduction*, Cambridge, 1982.

[8] Wells, LPD pronunciation preference poll, https://www.phon.ucl.ac.uk/home/wells/poll 98.htm.

[9] Jones, *Cambridge English Pronouncing Dictionary*, Cambridge, 2011.

[10] Imai, *et al.*, *Sounds Make Perfect*, Eihosha, 2010.

[11] UNSD,http://unstats.un.org/unsd/methods/m 49/m49.htm.

[12] Librivox, https://forum.librivox.org/.

[13] Forvo, https://forvo.com/languages/.