The American English Transcription Trainer User's Manual

Version 1.0

Roslyn Burns

January 31, 2024

Abstract

The American English Transcription Trainer by DiversiPHy is a free online tool designed to aid in the learning and teaching of American English transcription of the International Phonetic Alphabet. This tool contains an virtual keyboard with English-specific symbols from the International Phonetic Alphabet, audio of native speakers of American English, and an answer checking feature for audio-paired transcriptions.

Contents

1 Overview **2**

2 Examples of Specific Uses **3**
2.1 Dialect Hobbiest, EFL learner 3
2.2 Instructor Teaching EFL 4
2.3 Student Learning IPA 4
2.4 Instructor Teaching IPA 4

3 Word List Development **5**

4 Sound File Development **5**
4.1 Speakers 6
4.2 Methodology of Recording 6
4.3 Sound File Processing 7

5 Pronunciation Analysis **7**
5.1 Consonants 9
5.2 Vowels 11
5.3 Prosody 14

6 The Tool **14**
6.1 ABOUT button 14
6.2 KEYBOARD button 15
6.3 PLAY button 16

6.4 Answer Checking 16
6.4.1 Diacritics 17
6.4.2 Full Symbols 18
6.4.3 Syllable Options 20
6.5 Troubleshooting Answer Checking 20

References **21**

Appendix A: Word List **22**

List of Tables

1 Speaker List and Devices 6
2 Consonant Contrasts of American English 7

List of Figures

1 Vowels of American English 9

1 Overview

The American English Transcription Trainer is a free web-based product by DiversiPHy that aims to provide students and instructors with a resource for learning (or teaching) how to use the *International Phonetic Alphabet* (henceforth IPA) to transcribe speech. It is a sister product to the American English Sound Bar by DiversiPHy (Burns 2023).

The American English Transcription Trainer allows users to:

- Listen to audio of native speakers of different varieties of American English
- Type on a virtual keyboard using a variety of symbols used when teaching IPA transcription of American English
- Check your answers against the program's answers based on the speech features that you have selected

The intended use of this tool is didactic in nature. It should primarily be used to supplement one's studies of American English and IPA transcription. The transcription checking function is highly flexible and allows users to access both *broad transcription* (i.e., transcription of contrastive segments) and a variety of different *narrow transcriptions* (i.e., transcriptions with finer phonetic detail than the contrastive elements alone). This tool is not, however, intended to be the final say on which transcriptions your instructor will accept as valid transcriptions, so bear that in mind when using this tool.

One might wonder why this tool does not present a single "correct" standardized method of transcription in the same way that national standards of spelling often have one definitively "correct" way to spell something. In truth, no orthography designed to capture phonetic detail can meet

the same standard of objective correctness as a nationally standardized orthography. This is because all writing systems involve discrete entities (e.g., letters, or in the case of the IPA, characters which represent *phones* or sounds) but all sound systems involve processing continuous information (e.g., an acoustic signal) into discrete sound categories. While one may assume that matching discrete letters with discrete sound categories should be unambiguous, there is grey area because sound category boundaries themselves apply to inherently gradient entities meaning that the cut-offs for category boundaries are inherently gradient. These boundaries can and do vary from person to person even among speakers of the same language. As a result, there is always an element of subjectivity and discretion when assigning transcriptions even among professionals who have been trained in understanding these continuous vs. categorical relationships.[1]

The rest of the user manual will describe the uses of the tool, the methodology used to create the tool, and the features and functions of the tool.

2 Examples of Specific Uses

This tool may be used in part or in whole as transcription of speech may not be your final goal. This section provides information about how to use the American English Transcription Trainer tool for different goals.

- Dialect Hobbiest, EFL Learner
- Instructor Teaching EFL
- Student Learning IPA
- Instructor Teaching IPA

While it is fine to use this tool for pedagogical purposes, replication of this tool in part or in whole is prohibited.

2.1 Dialect Hobbiest, EFL learner

If you are just learning English, you may have noticed that people say things differently even if it is written the same. If you are a dialect hobbiest, you have likely already noticed that there are differences in speech and find it fun and fascinating! For both of these interest areas, this tool can help you get a sense of some of the variation in pronunciation found in different types of American English.

[1] If our own processing were not allowed to influence our transcriptions, then we may not even be writing phonetic detail, but rather just symbols that reflect the history of different standardization attempts. This is the core issue that exists with modern English orthography; different spelling conventions within the language reflect various historical political events which introduced new orthographic standards to the language without necessarily changing the older conventions. The result is that the English writing system in particular is less alphabetic (and more logographic) than alphabetic writing systems used in other related languages.

Use the page navigation and audio features described in §6.3 to listen to audio of native speakers saying these words.

2.2 Instructor Teaching EFL

If you are teaching EFL, you can use this tool to play consistent example words to students. An advantage of using this tool is the ability to provide your students with real examples of accents which you yourself may not speak.

When using this tool, be aware that the words were not recorded as isolated words in a list. They were collected in a carrier phrase and have the intonation of that phrase.

Before presenting examples in class, familiarize yourself with what you want to present. During class, use the page navigation and audio features described in §6.3 to listen to audio.

2.3 Student Learning IPA

If you are learning IPA there are different ways to use it depending on how it is integrated into your class.

The first way is to use the IPA Keyboard described in in §6.2 to write out character strings and copy and paste the strings into a document. If you are using LaTeX, you will need to use special packages to be able to type IPA directly into the document.

If you are using audio, familiarize yourself with page navigation and audio play described in §6.3.

If your instructor is using the tool in class, learn the settings which they are using for transcription before checking work on your own. Make note of any differences that they may have with the transcription. If you are not using this tool as a part of your class, it may still be helpful to hear different examples and practice, but remember that your instructor may have different preferences and intuitions than what is represent in this tool.

If you are using the answer checking functions described in §6.4, remember that the answers are specific to each speaker's pronunciation. A mistake sometimes made by students is providing IPA transcriptions of their own pronunciation even when the task is to transcribe someone else's speech.

2.4 Instructor Teaching IPA

If you are teaching IPA transcription, you can use this tool to provide students with examples to transcribe. This may be especially useful during a transcription test setting if you want access to consistent audio for students to transcribe.

You may use the tool with or without the answer-checking function. If you want to use the tool without the answer-checking function, familiarize yourself with the navigation and audio selection described in §6.3.

To use the tool with the answer checking function, familiarize yourself with the features that are encoded in §6.4 and how they are written in this tool (details are in §5). For **broad transcription** turn off all of

the buttons in the transcription options. The exceptions are preferential features like the CONSONANT TIE BAR button, DIPHTHONG TIE BAR button, DIPHTHONG OFF GLIDE button, etc.

For **narrow transcriptions**, all features are à la carte with the exception of the dependency-based features described in §6.4. If you want your students to use this tool outside of class, make sure to let them know which features they should have active.

3 Word List Development

This resource contains a list of 300 commonly used words in American English in alphabetical order. The word list was created from a list of the 5,000 most frequently used words in American English according to the Corpus of Contemporary American English (COCA).[2]

The word list from COCA was subset in R to capture only common nouns (i.e., no pronouns) and verbs (n = 3,204). Words in the noun/verb list were matched with different pronunciation variants in the Carnegie Mellon University's (CMU) Pronouncing Dictionary (n = 3,797) and sorted by the primary stressed vowel category (category n = 15).[3] 20 words from each of the 15 vowel categories were selected as representative members. Within each vowel category, insofar as it was possible, pre-nasal, pre-lateral, and pre-rhotic samples had a minimum of 3 members each. Other segments following the vowel were selected based on known environments of regional dialectal variation (see Labov et al. 2006).[4]

The word list sample tried to accommodate a variety of different considerations mentioned by instructors of introductory American English IPA based on the results of an online survey administered in Fall of 2023. These considerations included variation in syllable count, onset consonant behavior, and primary and secondary stress patterns (both lexical and dialectal) to name a few. The full word list may be found in Appendix A: Word List.

4 Sound File Development

This section explains the methodology used in developing the audio samples used in the tool including recruiting speakers, recording materials, and processing materials.

[2] Although the owners of COCA publish the list, it is listed on a different website than the main corpus.

[3] Pronunciation in the CMU dictionary is written in the ARPABET system; version of IPA for American English that does not require access to special unicode characters.

[4] This tool does not use the Wells (1982) vowel class system commonly used in British English dialectology (and sometimes American English dialectology). Relevant consonant and vowel behavior primarily reflects the Labovian vowel class system which was specifically designed to capture variation observed in North America (see Labov 1994, Labov et al. 2006).

4.1 Speakers

The speakers who are featured in this speech sample are all professional scholars of language, predominantly in the field of Linguistics. Although the scope is limited to native speakers of American English, this tool highlights that there is not a single way that American English should be spoken or that a professional should speak.

Speakers were recruited online by the snowball method (i.e. one participant refers another) and by quota sampling (i.e. trying to fill out different regions and known dialects).

4.2 Methodology of Recording

Speakers who agreed to participate were given a list of the 300 target words in a carrier phrase *'I say ___ again'* [a͡ɪ se͡ɪ ___ əgɪ̃n].[5] The order of the list was randomized in R. Full sentences (consisting of the carrier phrase plus target word) were 1 line each in 15 point font totaling 11 pages of text.

Speakers were asked to record themselves saying the randomized word list in order making 1 sound file for each page of the 11 page word list. They were instructed make note of their recording equipment and to make sure that their device sampled at a rate of 44.1 kHz. Speakers were told that they could record in either mono or stereo. The summary of speakers and their devices is shown in Table 1 below.

Speaker	**Region**	**Device**	**Microphone**	**File Type**
Burns	Seattle, WA	Zoom Hn4 Recorder	Device internal	wav
Khan	So Cal	iPhone 13 Mini	Device internal	m4a > wav[6]
Mayer	Vancouver, BC	Computer with Focusrite Scarlet 2i2 USB	XLR Shure WH20 headset	wav
Vais	Akron, OH	Mac Book Air	Blue Yeti	wav
Browne	Cambridge, MA	Mac Book Air Praat	Shure MV5	wav
Preseau	Detroit, MI	Zoom Hn4 Recorder	Device internal	wav

Table 1: Speaker List and Devices

Recording the randomized word list was estimated to take about 20 minutes. After recording, speakers returned the files online in wav format for processing and provided basic demographic information for the online tool. They were also asked to report any know variation in their speech, especially concerning known vowel chain shifts, and any preferred transcriptions for those features.

[5] Speakers of R-less varieties who reported either linking or intrusive rhotics were told that they could say the phrase with a short pause around the target word. This does not always result in the target word being isolated from the following vowel and the rhotic will sometimes still appear.

[6] The original files were recorded with VoiceMemos and converted using MediaHuman Audio Converter.

4.3 Sound File Processing

Files received from the speakers were annotated in Praat (Boersma & Weenink 2019) as TextGrids. After annotating, the sound files were checked for stereo vs mono setting and converted to mono if necessary. The TextGrids and wav files were forced aligned using the Montreal Forced Aligner (McAuliffe et al. 2017) with the English pronunciation model and the CMU pronouncing dictionary.

After forced alignment, the target word boundaries for the new TextGrids were manually checked in Praat. Boundaries were adjusted if forced alignment either failed to align to the word or included portions of the carrier phrase within the target. All target word boundaries were shifted to the nearest zero crossing in the air pressure signal. Individual sound files for each target word were extracted using a Praat script and then the amplitude of each sound file was adjusted to 65 dB using a Praat script.

In addition to processing the individual recordings for integration with the tool, sound files were processed for formant extraction. Formants were extracted using the Berkeley Phonetics Machine (Sprouse & Johnson 2016) and a pre-programmed script based on the Inverse Filter Control method (Ueda et al. 2007). The script extracted F1-F4, and F0 at 7 evenly spaced intervals throughout the duration of the vowel. These measurements were exported to a tab-delimited plain text file.

5 Pronunciation Analysis

This section describes the procedure for assessing the pronunciation of the different speakers. It is assumed that speakers maximally exhibit the core set of consonant contrasts in Table 2 below.

	Bilabial	Labio-Dental	Inter-Dental	Alveolar	Post-Alveolar	Palatal	Velar	Glottal
Plosives	p b			t d			k g	
Nasals	m			n			ŋ	
Fricatives		f v	θ ð	s z	ʃ ʒ			h
Affricates					t͡ʃ t͡ʒ			
Central Approximants	ʍ w			ɹ			ʍ w	
Lateral Approximants				l				

Table 2: Consonant Contrasts of American English

While it is assumed that Table 2 represent the maximal number of distinctions, it is also assumed that many speakers will have mergers (e.g., /ʍ/ > /w/).

It is assumed that speakers maximally exhibit the structures that are associated with the core set of vowels shown in Figure 1 below. All vowels

are assumed to be contrastive except for [ə] which is a generic realization of different unstressed vowels.

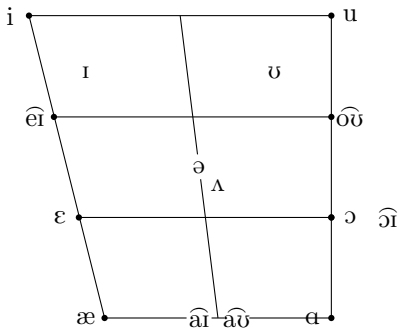

Figure 1: Vowels of American English

In addition to the core set of contrastive sounds, also called *phonemes*, this tool tracks non-contrastive sound, or *allophones*, based on responses to a survey administered in Fall of 2023 to instructors of introductory American English IPA transcription.

As a note: Although all analysis involved using Praat and in some cases R, none of the assigned labels were checked for statistical significance.

5.1 Consonants

Consonants were analyzed in Praat and examined aurally. This section describes the assignment of allophone transcription assuming that the reader has a basic understanding of the core allophones of American English.

Aspiration is written on voiceless plosives and affricates with lag following a release burst if the lag is stronger than what would naturally occur at a particular place of articulation. For example, dorsal segments like [k] naturally have longer lag than segments further forward in the oral cavity regardless of syllable stress. Use of the symbol [k^h] is for a segment which has an even longer lag than this naturally occurring lag in the unstressed syllable.

The glottal segment /h/ has two allophones, in the dataset. The allophone [ɦ] is audible and visually detected by a long-tailed voicing bar in the spectrogram during the production of the segment. The allophone [ç] is audible and detectable through a band of energy around 3000 – 6000 Hz. The trigger for the allophone [ç] in /h/ is the same trigger for the advanced allophone [k̟] in /k/, the high front vowel /i/ and the glide /j/.

In addition to being an allophone of /h/, [ç] is an allophone of /j/ when partial sonorant devoicing is triggered by aspiration. Similarly, the glide /w/ has a voiceless allophone [ʍ] which is triggered by aspiration. The symbols [ç̥] and [ʍ] are used instead of [j̥] and [w̥] respectively because these independent symbols represent the same sound as the symbol with the diacritic for physical reasons. As the airflow of a voiceless glide quickly shifts from laminar to turbulent, a fricative is produced.

Partial devoicing of other sonorants, including /ɹ/ and /l/, can be

triggered by aspiration or co-occurrence in an onset cluster with /s/ and /f/.[7] Although /t͡ʃ/ can also arguably also trigger partial devoicing when in an onset cluster with /ɹ/, this is not represented in the transcriptions unless aspiration would also otherwise trigger partial devoicing. This consideration is taken mostly because the partially-devoiced rhotic is audible during aspiration, but absent of aspiration, a partially devoiced rhotic is not audible as a separate segment independent of the normal lip rounding associated with [t͡ʃ].

The only case where partially-devoiced rhotics are transcribed next to an affricate without aspiration is in the case of partial-affrication. While some speakers have /tɹ/ sequences that are rendered with a full affricate [t͡ʃɹ] (e.g., Burns, Mayer, Vais, Preseau), other speakers tend to produce partial affricates (e.g., Khan, Browne). Although partial affrication is sometimes represented with a superscript symbol (e.g., [tʃ] and [dʒ]), this tool represents partial affricates through their effect on the following segment as [t͡ɹ̥ɹ]. This is even shown with partial affrication of /d/ as the partial affrication leaches the voicing of the sonorant.

In addition to sonorants, voiced obstruents may undergo partial devoicing, usually at the end of words. These are not represented with voiceless segments as they are still qualitatively distinct from the true voiceless counterpart (i.e., [z̥] ≠ [s]).

Rhotics have multiple allophones. Although it is possible to track whether a speaker uses the [ɹ] or [ɻ] allophone of /ɹ/ by analyzing F4, this analysis was not done. This is because the nature of the tool is to teach transcription based on audio samples and this distinction is known to not be audible despite the fact that it leaves acoustic artifacts in a recording. For this reason, both [ɹ] and [ɻ] are encoded as /ɹ/ and the options described in Full Symbols handle this feature. The tap allophone [ɾ], triggered by co-articulation with /θ/, is represented in the dataset.

Labialization is represented on /ɹ/ when it exhibits extreme dips in F3. Although both allophones [ɹ] and [ɻ] are characterized by low F3, the F3 is usually particularly low in initial onset position of stressed syllables and word initially.

Only one speaker, Browne, uses an r-0 (also called r-less) dialect. Despite speaking an r-0 variety, rhotics can be found in his speech due to linking with the vowel-initial word in the carrier phrase following the target and a couple of other isolated cases. When vocalization of the rhotic occurs, it is represented as [ə] unless vowel coalescence occurs in which case nothing is written where the rhotic used to be (e.g., dark).

Taps are represented for coronal stops /t, d, n/ and coronal stop clusters /nt, nd/ if the closure is 36 ms or shorter. Nasalized taps, [ɾ̃], sometimes have a slightly longer closure duration (e.g., around 45 ms). Mayer often exhibited a timing distinction between oral taps derived from

[7] Although the quality of the /s/ in onset clusters with either /m/ or /n/ is notably different than /s/ outside of this context, voiceless nasals are not transcribed in this tool because the pressure properties of sibilant fricatives are inherently in conflict with the pressure properties of nasal stops. Sibilant fricatives have high oral pressure which results in turbulent airflow, whereas nasal stops vent oral pressure through the nasal cavity and are characterized by low oral pressure and laminar airflow.

/t/ vs. those derived from /d/, but this distinction is not represented in the dataset.

Glottalization of the sort [tʔ] is present in the recordings, but it is not encoded in the transcriptions as such. Instead these types of segments are transcribed as either [ʔ] or [t͡ˀ]. The glottal stop is used when a glottal pulse is audible and detected in the spectrogram where an oral consonant would otherwise be. In this respect, glottalized stops like [tʔ] are not represented as [ʔ] unless the glottal portion is released and the coronal portion is not released. This most frequently occurs word-finally although there are some cases of this word-medially with concomitant creaking in words like in *'curtain'*.

Unreleased stops most often occur word-medially in codas before nasal onsets. Plosives in this position are generally unrelased because the pressure built up for the plosive is usually released through the nasal cavity as the airflow transitions from oral to nasal. Thus a plosive release burst is not attained. There are some instances of unreleased stops word finally in Khan's and Burns' recordings and many in Preseau's. For Khan, this occurs due to frequent syllabification of the final C of the target word with the initial unstressed V of the carrier phrase which triggers tapping. If the tap is not very audible, it is transcribed as an unreleased /t/. For Burns, the unreleased /t/ is a variant of a word final glottalized /t/ where neither constriction point is released. In the case that neither constriction is released, it is transcribed as [t͡ˀ], not [t͡ˀʔ͡ˀ]. For Preseau's speech it is just common to have unrealeased stops (she seldom syllabified the target word with the carrier phrase).

Although there are syllabic consonants in the dataset which can be detected acoustically, the difference between a syllabic consonant and the combination of /ə+$C_{[son]}$/ is very difficult to hear for many language users. For this reason, they are all encoded as a vowel and a sonorant together and the options described in Diacritics handle this feature.

Velarization of laterals was assessed aurally and transcribed based on known behaviors of American English. All coda laterals are assumed to be velarized as are onset laterals in clusters with velars. Some speakers, such as Burns, always produced velarized laterals in onset clusters regardlss of the place of the neighboring consonant. Some speakers produced velarized laterals in initial onset position based on how they syllabified the final V of the carrier phrase with the initial C of the target word.

5.2 Vowels

The plain text files containing formant measurements described in §4.3 were combined into a single file. F1, F2, and F3 measurements from times 3 and 5 were analyzed as the vowel nucleus and the vowel off-glide respectively. The file was then subset to only vowels carrying primary stress and imported into R. Vowels were normalized using the Labov method method in the R package VOWELS. The F1 F2 trajectories were first analyzed using the R package VOWELS and then the lexical label of the nuclei were plotted to get a sense of both individual and comparative vowel spaces. Some notable behaviors include the following.

Speakers from the western coast of North America (e.g., Burns, Khan,

Meyer) consistently exhibited merger of CAUGHT /ɔ/ and COT /ɑ/ in the direction of COT /ɑ/. This is a known-feature reported by all individuals who submitted speech samples from this region. As a speaker of the Eastern New England dialect, Browne reported being CAUGHT - COT merged, but in the direction of CAUGHT /ɔ/.

For both types of CAUGHT -COT merger, despite the fact that speakers reported being merged, they were able to physically produce the difference between these sounds in some, but not all, cases. The author frequently relied on the category assignment from the Montreal Forced Aligner to assist in determining which transcription to use.[8] It is common for people who are merged towards /ɑ/ to have [ɔ] in contexts which naturally trigger rounding or velarization via co-articulation (e.g., *'swallow'*, *'salt'*, *'water'*).

Speakers often exhibited vowel raising before rhotics and the common mergers which develop from raising. For speakers from the west coast and the upper midwest (i.e., Burns, Khan, Mayer, Vais, Preseau), the vowels MARRY /æ/ and MERRY /ɛ/ were often merged to MERRY as /ɛ̝/. As there were no appropriate samples of MARY /e͡ɪ/ before rhotics in the noun/verb word list, this class could not be directly compared in the known three-way merger set. Speakers from the north east (i.e., Browne) are unmerged.

Across the current sample, no one exhibited merger of LARD /ɑ/ and LORD /ɔ/. Although it is common to transcribe the unmerged vowels as /ɑ/ and /ɔ/ respectively, the LORD vowel was often tense and closer to the quality of /o/.

Although the CMU dictionary consistently differentiates between /ʊ/ and /ə/ before rhotics in stressed syllables,[9] most speakers failed to make this distinction with the exception of Browne. In his speech, [ʊ̞ɹ] is found in many words like *'curtain'*, *'skirt'*, and *'version'* and [əɹ] is found in words like *'inquiry'* and *'wonder'*. For all other speakers, the only exception to UH1R and ER1 merger is if /ʊ/ tensed to /u/ (or in the case of these speakers, /ʉ/) like in the word *'cure'*.

Most speakers also exhibited merger of /ɪ/ before rhotics with /i/ with the occasional exceptions found in Browne and Khan's speech the most notable exception being the word *'lyrics'*.

Speakers exhibited a variety of processes before nasals. Only one speaker, Burns, exhibits merger between PIN /ɪ/ and PEN /ɛ/ in the direction of PIN. All other speakers are unmerged.

All speakers exhibited tensing of /æ/ > [ɛ̝] before /ŋ/, with some upper midwestern speakers exhibiting more extreme tensing to [e͡ɪ] (e.g., Vais, Preseau). There was variation as to whether or not additional related tensing processes were observed. Mayer exhibited a vowel chain shift with two types of extensions of /æ/ before /ŋ/ tensing. First, /ŋ/ exhibited an extended set of raising targets (all other non-high lax vowels). Second, other consonants (notably [g]) could trigger raising across the extended set of targets. Burns and Khan do not exhibit these extensions.

For speakers from the western coast of North America, /æ/ tenses

[8]The CMU pronouncing dictionary comes by default with pronunciation variants for this specific merger.

[9]ARPABET UH1R and ER1 respectively.

before the coronal nasal, but this tensing is often not as high and as front as tensed /æ/ before /ŋ/. The resulting tensed vowel before the coronal nasal is transcribed as [æ̝] as it has a quality that is different from the diphthongs characteristic of varieties like New York City English or the Northern Cities region. While the tensing is most noticeable before the coronal nasal, the bilabial nasal /m/ participates to varying degrees across speakers.

Speakers from the upper midwest often exhibit diphthongs that result from /æ/ tensing before nasals which I transcribe as [e͡a]. Although a lot of Vais' tensed /æ/ diphthongs occur before nasals, Preseau exhibits a more generalized pattern of tensing across all instances of /æ/. While many of Preseau's tensed /æ/ vowels are diphthongs, a hand full are transcribed as [æ̝] as in *'broadcast'* and *'photograph'*.

Laterals exhibited a variety of effects on lax vowels most commonly involving either backing or raising. In Browne's speech, /ɔ/ tenses to [o] before laterals in words like *'golf'*. The exception to the backing and raising effect is most readily observed with the high back lax vowel /ʊ/ which can exhibit dissimilatory lowering in the context of a post-vocalic lateral thus leading to merger with /ʌ/ as often is found in the speech of Burns and Preseau. In their speech, the vowel in *'pull'* is similar to the vowel in *'pulse'* whereas other speakers differentiate two. The words *'woman'* an *'wolf'* were exceptions due to the heavy labial co-articulation of the initial /w/.

In the speech of Preseau and Vais, laterals triggered retraction and lowering of /ɪ/ > [ɛ] in the word *'pillow'* but not *'fill'* or *'will'*.

The Low Back Merger Shift, common of the Western US, was observed in the speech of Khan. Structurally, this shift begins with the merger of CAUGHT and COT in the back vowel space and then lax/short front vowels rotate counter-clockwise in response (except for low front vowels which undergo tensing in the environments described above).[10] In Khan's speech, the shift was more prominent in lower front vowels where /æ/ > [a] in some cases. All speakers from the west coast exhibited fronting of the high back vowel /u/ > /ʉ/.

Speakers from the upper midwest exhibited the Northern Cities Shift common of the Great Lakes region (found as far east as Upstate New York). Structurally, the shift begins with tensing of /æ/ and then continues with the clockwise rotation of lax high vowels in response.[11] Preseau exhibits the reflexes of the early structural changes in the shift with her generalized /æ/ tensing discussed above. In Presau's and Vais' speech, the vowel /ɑ/ > [a], but the fronting is more pronounced for Preseau and occurs across a wider range of words. Following the normal clockwise direction of the shift, Vais exhibits /ʌ/ > [ɑ] in the word *'wonder'* which sounds similar to *'wander'*.[12] Preseau exhibits reversal of the clockwise rotation where /ɑ/ > [ʌ] the word *'golf'* which is treated as near-merged

[10] For an animated schema of this innovation, see the Sound Change Visualization Project's Low Back Merger Shift (Burns 2020a).

[11] For an animated schema of this innovation, see the Sound Change Visualization Project's Northern Cities Shift (Burns 2020b).

[12] Although the word *'wander'* is not in the sampled set, Vais did confirm that she often says these words similarly and it is common where she is from.

with *'gulf'* (a potential case of hypercorrection).

All speakers, to some extent, exhibited raising of the central nucleus of the diphthongs /a͡ɪ/ and /a͡ʊ/. In order to be assigned a transcription consistent with Canadian Raising, [ʌ͡ɪ] and [ʌ͡ʊ] respectively, the F1 of the nucleus had to be at or less than the following benchmarks that I have defined: [ʌ͡ɪ] 660 Hz, [ʌ͡ʊ] 780 Hz. Browne, Khan, Mayer, Preseau, and Vais all exhibit Canadian Raising. The raising is most audible in the back diphthong space but the raising is especially audible in the front diphthong space for Mayer and Khan. While Browne lacks raising that crossed the threshold for inclusion in the front diphthong acoustic space, some of Preseau's raising in the front diphthong acoustic space includes non-traditional environments in words like *'hire'*.

Finally, one speaker, Browne, exhibited a characteristic New England speech pattern which follows the behavior of the British English /æ/-split. In Browne's speech, the BATH vowel is realized as monophthong [aː] in *'bath'* and *'laugh'* but most other forms of /æ/ are front instead of central.

5.3 Prosody

Syllable structure was determined based on the principle of onset maximization (i.e., as many things go into the onset as possible), known phonotactic behaviors of English onsets, and aided by identifying phonetic cues to syllable membership (e.g., presence or absence of aspiration on voiceless plosives and affricates).

Primary stress was determined based on aural assessment of pitch, intensity, and duration and by viewing dB intensity contours in Praat. Secondary stress was determined based on aural assessment, cross-checking with ARPABET codes, and looking for other cues to membership in a stressed syllable (i.e., aspiration on voiceless plosives and affricates, labialization on rhotics).

6 The Tool

The American English Transcription Trainer tool was created in Adobe Animate 2023 using an HTML5 Canvas document. The tool is intended to be used on either desktop or laptop computers. Certain functions may not be available on mobile devices.

The first page is the main menu with the ABOUT and KEYBOARD buttons.

6.1 ABOUT button

Clicking the ABOUT button from the main menu takes the user to the about section. This section provides the user with basic information about the DiversiPHy project and the Transcription trainer itself. The user may learn basic demographic information about the speakers that is provided under the SPEAKERS navigation button or return to the main page with the MENU button at any time.

6.2 KEYBOARD button

Clicking the KEYBOARD button from the main menu takes the user to the main tool. Here the user can interact with three basic modules.

Keyboard: Type, paste, and copy characters on a virtual IPA keyboard

Audio: Navigate and listen to different native speakers of American English pronounce the word list described in §3

Answer Check: Check IPA transcriptions and set the settings for how narrow the transcription should be

The Keyboard and Audio modules may be used independently of each other and independently of the Answer Check module. The behaviors which are programmed into the keyboard were chosen based on the results of a survey of instructors of American English IPA instructors that was administered in Fall of 2023 (e.g., tie bars on affricates, but not on diphthongs was more common in the responses to the survey).

The keyboard allows the user to either insert or delete characters in the main text entry box. Characters may be inserted by pasting from another source (e.g., a website or text document), typing on your own keyboard, or by using the tool's virtual keyboard. The advantage of using the virtual keyboard is that it is programmed with UTF-8 IPA characters. This is advantageous because there are non-IPA characters which students frequently mistake for IPA characters, such as <g> instead of IPA [ɡ], and characters which students mistakenly enter on a keyboard which have a different IPA value than what they intend, such as <a> when they mean to use IPA [ɑ].[13]

To insert characters from the virtual keyboard, hover your mouse over a button in either the consonant, vowel, or diacritic chart. When the button changes to a pointing finger icon, click the desired button and the symbol will be inserted into the main text entry box on screen. For combining diacritics, write the symbol first and then the diacritic. For nasals, first type the segment, e.g., [ɑ], and then the nasalization button to get [ɑ̃]. For affricates and diphthongs, type the first character, the tie bar, and then the second character as in [t],[t͡], and finally [t͡ʃ]. For diphthongs which are encoded in the keyboard, you can click the diphthong to get [aɪ], move the cursor between the two symbols, and then click the tie bar to get [a͡ɪ].

Be aware that some characters may have been entered, even if they do not automatically appear in the text field. This is especially true of combining diacritics which go below the character (e.g., the voicelessness, syllabic, advanced, etc) when applied to characters with long descenders (i.e., j, ɡ, ʃ, ʒ, etc.). If the characters have been entered properly, they will appear as normal when copy and pasted into a separate document (e.g., Microsoft Word, Mac Pages, Open Office Writer, etc.).

To remove characters, click on the BACKSPACE button on the virtual keyboard, your own keyboard, or press the CLEAR button on the virtual

[13] While <g> is not an IPA character, <a> is used in IPA, but specifically for central vowels. If someone wants to type IPA [a] for the MONOPHTHONG /A/ option, they may either type it directly from their own keyboard or they can access it on the IPA keyboard by clicking one of the central diphthongs and deleting the off-glide.

keyboard. Pressing BACKSPACE will delete one character at a time as on a normal keyboard. Pressing the CLEAR button will delete all characters in the main text entry box.

6.3 PLAY button

To hear audio, press the PLAY button indicated with the rightward pointing arrow. The PLAY button will appear when you start navigating through the word list. The first way is to navigate with the black buttons above the main text entry box. Although the buttons are black, you will know you are in the correct area when the cursor changes to a pointing finger. The PREVIOUS and NEXT buttons will lead the user through the word list sequentially in alphabetical order. The JUMP button will take the user to a random word in the list.

The second way to navigate the word list is by text in the page text box above the main text entry box. When your cursor is in the page text box, type either the word list index number (1 through 300) or the desired word and press ENTER on your keyboard. While this navigation option is not case sensitive, all spelling must conform to U.S. American English standards (i.e. *'utilize'* instead of *'utilise'*). Note that the text-based navigation feature may not work on mobile devices if your device reserves the ENTER key for use in the main text entry box only.

To hear the audio of different speakers, go to the OPTIONS button on the right hand side menu. The first page of options will have a list of the different speakers who are currently available and buttons the left of each speaker. Press a button to select your desired speaker and you will receive a browser notification and the button next to the speaker will illuminate. To exit the options menu, click the OPTIONS button again. If you have selected a new speaker, their audio should play when you hit the PLAY button.

6.4 Answer Checking

Answer checking is available only when there is audio. To check an answer, either press ENTER on your keyboard while the cursor is in the main text entry box or click the SUBMIT button on the left hand side menu. You will receive feedback on the page below the main text entry box that says either *"Correct"* or *"Try Again"*. This response is dependent on three different factors: what you have written in the main text entry box, the speaker that you have selected, and the transcription options that are activated. As we have already covered entering text in the main text box and speaker selection, we will focus on transcription options.

Whether you are checking answers for a *broad* (phonemic) vs. *narrow* (phonetic detail) transcription depends on your selected transcription options. Transcription options may be accessed in the options menu by clicking the OPTIONS button and navigating away from the speaker option page. To exit the options menu, click the OPTIONS button again. Active options are indicated by an illuminated green button to the left of the option whereas inactive options are indicated by a dim green button.

Every time you change an option, you will receive a notification from your browser and the lighting of the relevant button will change.

Be aware that you cannot have some options activated without other options also being activated. For example, although you can check the transcription of primary stress without checking the transcription of secondary stress, you cannot check the transcription of secondary stress without also checking the transcription of primary stress. The tool will automatically make these dependency-based option adjustments for you and your browser notification will indicate that two features have been changed.

The transcription options are divided into three major categories: diacritics, full symbols, and syllable options.

6.4.1 Diacritics

The diacritic options are:

- Aspiration
- Nasalization
- Voiceless Obstruents
- Voiceless Sonorants
- Tie Bar Consonants
- Tie Bar Vowels
- Rhoticity
- Dental
- No Audible Release
- Velarization
- Labialization
- Advanced
- Raised
- Long
- Syllabic

The ASPIRATION option enables the transcription of aspiration. Aspiration is written following the whole segment which carries aspiration (i.e., [tʰ] and [t͡ʃʰ] are correct; [tʰ͡ʃ] is incorrect).

The NASALIZATION option turns on the transcription of nasalization for the nasalized tap and vowels. Nasalization is written on nasalized taps as [ɾ̃] only if the TAP option is also on, otherwise nasalized taps are written as [n] even if NASALIZATION is on. In this tool, nasalization is written only on the vowel nucleus (i.e., [ãʊ] is correct; [aʊ̃] and [ãʊ̃] are incorrect).

The VOICELESS OBSTRUENT option turns on the voiceless ring for obstruents only. In this tool, the ring goes under the first part of the obstruent (i.e., [d̥͡ʒ] is correct; [d͡ʒ̥] and [d̥͡ʒ̥] are incorrect). This option only applies to segments which are naturally voiced, not ones which are already voiceless (i.e., [t̥͡ʃ], [t͡ʃ̥], and [t̥͡ʃ̥] are all incorrect).

The VOICELESS SONORANTS option turns on the transcription of voiceless sonorants which are transcribed one of two ways. If the sonorant is a non-glide approximant, it is written with a ring under the segment (i.e., [ɹ̥], [ɬ̥]). If it is a voiceless glide approximant (i.e., [j], [w]), it is written as its voiceless fricative counterpart (i.e., [ç], [ʍ]) for reasons discussed in §5.1. As voiceless sonorants in English are only partially devoiced, they can be optionally written with a tie bar if the TIE BAR CONSONANT option is turned on (e.g., [ɹ̥͡ɹ], [ʍ͡w]), otherwise they are written without the tie bar (e.g., [ɹ̥ɹ], [ʍw]). As partial devoicing of

sonorants is derived from the process of aspiration, ASPIRATION will be turned on if it is not already on when VOICELESS SONORANTS turned on. If ASPIRATION is turned off, VOICELESS SONORANTS will also be turned off.

Tie bars are used for single segments with two articulatory targets such as affricates, partially devoiced sonorants, and diphthongs.[14] When the TIE BAR CONSONANTS option is turned on, the tie bar should be written with affricates (e.g., [t͡ʃ]) and, if VOICELESS SONORANTS is turned on, with partially devoiced sonorants (e.g., [ç͡j]). If the TIE BAR VOWELS option is turned on, the tie bar should be used with diphthongs regardless of whether the off-glide is represented a vowel or a glide (i.e., [a͡i], [a͡j]).

The RHOTICITY option turns on the rhoticized schwa [ə˞] regardless of whether it is in a stressed or unstressed syllable. Currently, other vowels are not programmed for rhoticity. When RHOTICITY and SYLLABIC are both turned on, syllabic consonants are written with the pipe underneath the syllable nucleus except for the syllabic rhotic which is written as [ə˞̩].

The DENTAL option enables the transcription of the dental diacritic on coronals which are otherwise alveolar.

The NO AUDIBLE RELEASE option enables the transcription of the open upper right corner on plosive segments which have no audible release.

The VELARIZATION option enables the transcription of the velarized lateral.

The ADVANCED option enables the transcription of the plus diacritic on dorsal consonants and low vowels and some forms of tensed /æ/. Note that /u/ fronting is handled with the /U/ ALLOPHONE option discussed in §6.4.2 and not the ADVANCED option.

The RAISED option enables the transcription of the up tack diacritic on vowel nuclei. Note that /æ/ tensing is handled by the ADVANCED option.

The LONG option enables the transcription of length on phonemically long vowels that are not diphthongs (i.e., /i/ and /u/). This feature is not for derived length.

The SYLLABIC option enables the transcription of syllabic consonants. When it is used, all /ə + C_{SON}/ sequences are written as [C̩]. The only exception is if RHOTICITY is also turned on in which case the rhoticized schwa is used for syllabic rhotics.

6.4.2 Full Symbols

The full symbol options are

[14] While some people transcribe partial devoicing of obstruents with tie bars, I have decided not to include this as to not complicate the transcription of partially devoiced affricates. All partially devoiced obstruents do not involve tie bars to link the devoiced portion to the voiced one.

- Tap
- Glottal Stop
- Diphthong Off-Glide
- Retroflex
- /h/ Allophones
- /u/ Allophones
- Canadian Raising
- Monophthong /a/
- Diphthong /æ/
- Tense /ɔ/

The TAP option enables the transcription of taps. If the NASALIZATION option is off, the nasalized tap is transcribed as [n] instead of [ɾ̃] even if the TAP option is turned on. When the TAP option is turned off, tapped rhotics are transcribed as either bunched [ɹ] or retroflex [ɻ] depending on the state of the RETROFLEX option.

The GLOTTAL STOP option enables the transcriptions of the glottal stop in words where either an onset or coda consonant debuccalizes (e.g. /t/ > [ʔ]). It is not written in cases of double articulation (e.g., [tʔ]). In this tool, glottal stops are not written in cases of V-initial onset that have an epenthetic glottal stop.[15] Although there my be concomitant creaking on vowels adjacent to a glottal stop, creak itself is not transcribed with the glottal stop.

The DIPTHONG OFF-GLIDE option enables the transcription of off-glides in the diphthongs /aɪ/, /aʊ/, and /ɔɪ/ as [j] and [w] instead of [ɪ] and [ʊ].

The RETROFLEX option enables the transcription of the retroflex rhotic [ɻ] instead of the bunched-r rhotic [ɹ].

The /H/ ALLOPHONES option enables the transcription of non-basic allophones of /h/. The allophone [ɦ̟] occurs for speakers who have both a general breathy quality to their voice and who syllabify the carrier phrase with the target word thus allowing voicing to carry over (e.g. Khan). The allophone [ç] occurs when the oral cavity forms a narrow channel that causes turbulent airflow in anticipation of close segments with free-flowing air (i.e., /i/ and /j/).

The /U/ ALLOPHONES option enables transcription of multiple allophones of /u/. While the basic allophone is [u] in some varieties, the allophone [ʉ] normally occurs in anticipation of a following coronal segment which pulls the tongue forward. In some cases, due to vowel chain shifting, [ʉ] becomes the basic allophone and [u] is derived in environments which trigger backing (i.e., velar segments and velarized segments).

The CANADIAN RAISING option enables the transcription of the nucleus [ʌ] for the diphthongs /aɪ/ and /aʊ/. This transcription is used for speakers who meet the benchmarks described in §5.2.

The MONOPHTHONG /A/ option enables the transcription of [a] for speakers whose vowel systems have shifted to include this segment (see §5.2). As the vowel [a] derives from different contrastive vowels in different regions, when MONOPHTHONG /A/ is turned off, the transcription matches the source contrastive vowels (e.g., /æ/ in the Low Back Merger Shift, but /ɑ/ in the Northern Cities Shift).

[15] This is because in many cases it is difficult to hear and it is not a reliable enough of a feature to transcribe by default among the speakers.

The DIPHTHONG /AE/ option enables the transcription of [e͡æ] for speakers whose vowel systems have shifted to include this segment (see §5.2). When DIPHTHONG /AE/ is turned off, these segments are transcribed as [æ].

The TENSE /O/ option enables the transcription of [o] for speakers who have undergone tensing of this vowel. When TENSE /O/ is turned off, these segments are transcribed as [ɔ].

6.4.3 Syllable Options

The syllable options are as follows:

- Syllable Breaks
- Primary Stress
- Secondary Stress

The SYLLABLE BREAKS option enables the transcription of syllable breaks. Syllable breaks are only written in words that are larger than two syllables. For the most part, ambisyllabicity is not encoded in the responses. This means that in a word like *'forest'* there is only one rhotic and not two. The membership of the rhotic in either the first or second syllable is determined based on whether or not any syllable would be left without an onset. Exceptions where ambisyllabicity is encoded include words with syllabic consonants and homo-organic onsets like *'tomato'* and *'worry'*. In these cases, the segment is written twice, once in the first syllable and once in the second. Note that the double transcription of a syllabic consonant and homo-organic onset applies even in the case that either SYLLABLE BREAKS or SYLLABIC is turned off.

The PRIMARY STRESS feature enables the transcription of primary stress. Primary stress is always written before the first segment of the stressed syllable (e.g. [ˈt^{h}a͡im] is correct; [t^{hˈ}a͡im] is incorrect).

The SECONDARY STRESS option enables the transcription of secondary stress. Similar to primary stress, secondary stress is always written before the first segment of the stressed syllable. When SECONDARY STRESS is on, PRIMARY STRESS must also be on. If PRIMARY STRESS is turned off, then SECONDARY STRESS will also turn off.

6.5 Troubleshooting Answer Checking

If you are using the answer checking function and keep getting the wrong answer there are certain things that you can do to troubleshoot.

1. Check the features that are active in the OPTIONS menu.
2. Make sure you understand the description of different speakers' pronunciations.
3. Try alternative vowels especially if the vowel is either unstressed or in a merger environment.
4. Try alternative transcriptions with consonants.

It is common to get the correct transcription but not for the feature set that is selected (e.g., forgetting to write aspiration even though the option is selected). Similar to getting an incorrect response for writing something that is too broad, you will get an incorrect response if you type something that is too narrow for the current set of feature options.

The description of assigning transcriptions often references individual speakers, specific words, and examples of transcriptions. Using this as a guide can help you understand what transcription to write.

Vowels can be difficult to discern in some parts of the vowel space especially if they are frequent targets of merger in specific dialects (e.g., CAUGHT - COT merger, pre-rhotic merger, pre-lateral merger, pre-nasal merger). Vowels may also sound similar to other vowels due to reduction processes where [ə] might sound more like [ɪ] depending on the neighboring consonants.

Finally trying different consonants is useful if there are two different outcomes to the same process such as partial affrication and affrication.

References

Boersma, Paul & David Weenink. 2019. *Praat: Doing Phonetics by Computer*. v. 6.0.46. praat.org

Burns, Roslyn. 2020a. *The Low Back Merger Shift*. Sound Change Visualization Project. Sound Change Visualization Project Google Drive

Burns, Roslyn. 2020b. *The Northern Cities Shift*. Sound Change Visualization Project. Sound Change Visualization Project Google Drive.

Burns, Roslyn. 2023. *The American English Sound Bar*. v. 1.0. DiversiPHy. http://www.linguisticking.com/SoundBar/html

Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

Carnegie Mellon Speech Group. *Carnegie Mellon University Pronouncing Dictionary*.http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Labov, William. 1994. *Principles of Linguistic Change*. vols. 1–3. Malden: Wiley-Blackwell.

Labov, William, Sharon Ash, & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology and sound change*. Mouton de Gruyter.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal forced aligner: trainable text-speech alignment using kaldi. In *Interspeech*, vol. 2017, 498–502.

Sprouse, Ronald & Kieth Johnson. 2016. The Berkeley Phonetics Machine. *UC Berkeley PhonLab Annual Report* 12(1).

Ueda, Y., Hamakawa, T., Sakata, T., Hario, S., & Watanabe, A. 2007. A real-time formant tracker based on the inverse filter control method. *Acoustical science and technology* 28(4): 271–274.

Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Appendix A: Word List

This appendix contains the full word list used in the tool.

acquire	clothes	exchange	identification	oxygen
adult	cloud	excitement	identity	pal
advertisement	club	exit	ignore	palm
agriculture	clue	explode	illusion	peel
aisle	coin	exploit	imply	pen
anchor	colonel	expression	income	photograph
anniversary	component	feel	input	photography
answer	contrast	fighter	inquiry	piano
apple	cooking	fill	instruct	pillow
appoint	cord	fishing	insurance	pin
array	couch	flaw	integrity	plot
arrow	counter	flood	interval	point
ask	county	flower	join	police
attempt	courtesy	fluid	joy	poll
aunt	cousin	focus	jump	pool
avoid	crystal	food	knife	porch
bankruptcy	cure	fool	know	portray
barrel	curtain	footage	laugh	powder
bath	dark	forest	lawyer	prayer
bear	dawn	fuel	length	prescription
beg	deploy	garage	life	privacy
being	depth	genius	lightning	production
black	derive	genre	limit	prosecutor
blade	destroy	ghost	lyrics	publication
board	detail	globe	marry	pull
boil	disappear	glory	meal	pulse
book	discuss	goal	media	push
borrow	diversity	golf	meeting	put
boundary	doubt	goodness	milk	queen
bowl	dough	gospel	mistake	question
breathing	dragon	grab	monkey	radio
broadcast	drum	grocery	mountain	reject
brother	dust	growth	mouse	response
bull	duty	gulf	mouth	roof
bush	earthquake	hearing	music	root
butter	economics	herb	news	rose
campus	electricity	hire	noise	route
card	emerge	hook	nominee	sail
choice	employee	horizon	observe	salad
chuckle	engagement	horror	occupy	salmon
circuit	enjoy	hour	occur	salt
claim	enter	house	ocean	sauce
clear	enthusiasm	housing	oil	scout
clerk	equation	human	onion	section
closet	equipment	hurricane	operator	security
cloth	escape	idea	orange	sell

shine	sphere	thief	universe	wheat
should	spouse	thought	utilize	wheel
shout	square	thousand	value	whip
skirt	squeeze	thrive	variation	white
skull	stir	throat	vehicle	will
slice	stop	thumb	verify	withdraw
slope	strengthen	toilet	version	wolf
smile	strive	tomato	virtue	woman
smoke	sugar	top	vision	wonder
snake	swallow	toy	vitamin	wood
sodium	tag	trainer	voice	word
solve	tank	tube	water	world
sort	tattoo	twin	wealth	worry
south	tell	unemployment	whale	youth