

Speech Assessment via Read-Alouds: A Critical Analysis of Diagnostic Passages



Takehiko Makino

Abstract My previous studies on Japanese speakers' pronunciation of English were based on recordings of TIMIT—a phonetic corpus that consists of isolated sentences with words chosen for the sake of phonological balance. The scope of possible analyses with TIMIT is limited because the recorded utterances are prosodically monotonous and the pronunciation of difficult words is unstable. To conduct more reliable speech assessments via read-alouds, I embarked on a critical analysis of diagnostic passages. This chapter critically reviews eight different diagnostic passages that have been used for data collection mainly in terms of phonetic coverage. My goal was to find one that meets the following criteria: contains every English phoneme; includes as many diphones as possible, especially those found to be difficult for Japanese speakers; does not contain words that are infrequent or too difficult for learners; constitutes a coherent passage which can elicit different prosodic patterns according to context; and illustrates a variety of speech acts that can elicit different intonation choices. Based on the critical review, the chapter argues for the selection of one diagnostic passage as the most suitable (after a few suggested adaptations) to collect English speech data from Japanese learners of English.

Keywords Learners' English pronunciation · Japanese speakers · Diagnostic passage · Phonetic coverage · Sentence types

1 Introduction

A few years ago, I developed *English Read by Japanese (ERJ) Phonetic Corpus* (Makino, 2013; Makino & Aoki, 2012) by using a small part of ERJ speech database (Minematsu et al., 2002)—a large collection (more than 70,000 files) of English sentences and words read aloud by 200 university students in Japan. Most of the ERJ sentences are based on TIMIT phonemically balanced set of sentences (Garofolo et al., 1993). The set of 800 files used for ERJ Phonetic Corpus was the same as

T. Makino (✉)
Faculty of Economics, Chuo University, Tokyo, Japan
e-mail: mackinaw@tamacc.chuo-u.ac.jp

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
V. G. Sardegna and A. Jarosz (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*, Second Language Learning and Teaching, https://doi.org/10.1007/978-3-030-98218-8_6

that used in Minematsu et al. (2011), which studied the intelligibility of Japanese-accented pronunciation. I have used the ERJ Phonetic Corpus mainly to look at segmental patterns. However, while conducting my analyses, I found some significant drawbacks in the sentence set and the selected recordings in the Corpus. First, the phonemically balanced sentence set was originally devised for speech engineering. As a result, the set includes quite difficult words. According to the computation by Wordcounter.net (<https://wordcounter.net/>), the lexical items in the Corpus have a 11-12th grade reading level. This reading level can be very difficult for the average learner of English in Japan, and has probably led to unstable or erroneous pronunciations. Second, the sentence set consists of 420 isolated sentences (total 3,167 words) and many of them are short (8 words average). Because of this, most of the recorded utterances are prosodically monotonous, which has made it difficult to study different prosodic possibilities. Finally, the Corpus is not phonemically balanced for any particular speaker because only four out of the 120 sentences that comprise the balanced set for each speaker were randomly chosen. These limitations have led me to consider doing new recordings using a short passage.

The purpose of this chapter is to critically review different passages that have been used to collect learners' pronunciation of English so that I can find a more adequate data collection instrument for my study of Japanese speakers' English pronunciation. The chapter starts with an overview of the possible and attested problems in Japanese speakers' pronunciation of English. Then, it proposes the requirements for an ideal diagnostic passage. After these background discussions, it analyzes commonly used individual passages in terms of those requirements and identifies the most preferable choice. The chapter concludes with a discussion of the implications of this investigation.

2 Literature Review

2.1 *Japanese Learners of English*

The Japanese phonological system differs from that of English (Vance, 2008). It has a five-vowel system, and 14 phonemic consonants. Hence, its segmental inventory is less rich than that in English. It only has open syllables and virtually no initial consonant clusters. As for its prosody, lexical stress and sentence stress are always realized as falling pitch, but about half of the vocabulary does not have a lexical stress. Sentence intonation is realized as pitch movements at the end of intonation phrases (Venditti, 2005).

When pronouncing English vowels, Japanese speakers conflate /i/ and /ɪ/, /ɑ:/ and /ɑə/, and /æ:/, and /ʌ/ and /ə/, and several other pairs or sets. They also struggle with pronouncing English consonants /l, r, f, v, θ, ð/, and other consonants in certain phonetic contexts. For example, they tend to pronounce syllable-initial English /z/ as an affricate [dz], intervocalic English /dʒ/ as a fricative [ʒ], and intervocalic voiced

plosives /b, d, g/ as fricatives [β, ð, ɣ]. They also do not distinguish word-final /z/ and /dz/. These are all cases of negative transfer from the Japanese pattern where voiced obstruents are usually realized as fricatives between vowels but as plosives or affricates word-initially. In my previous study on Japanese speakers' segmental patterns of English (Makino, 2013), voiceless plosives were also found to be spirantized in many cases, even though such patterns are not documented for spoken Japanese. I expect to find other "unexpected" patterns in new recordings.

Consonant clusters are also a major problem for Japanese speakers because they often insert a vowel between the consonants. *Strike* /straɪk/ becoming [sutoraikɯ] is one of their notorious mispronunciations, although in reality, the [ɯ] between voiceless consonants is usually dropped in spoken Japanese, so the more plausible mispronunciation is [storaikɯ]. They also find difficult the pronunciation of word-to-word linking, especially between a final consonant and an initial vowel.

Japanese speakers are not good at placing English nuclear stresses on the appropriate syllables. They tend to use a rising pitch at the very end of yes–no questions even if the last syllable is unstressed and a continuous rise to the end from the earlier narrow focus is desirable (Ueyama, 1997), or place the nuclear stress on the last syllable even if it should be placed earlier in the intonation phrase. These errors may result from negative transfer from the already mentioned prosodic characteristics of Japanese. Surprisingly, such prosodic deviations have not been adequately documented in the literature even though they are sometimes discussed in informal observations. One of my motivations to collect Japanese speakers' reading of a passage in English is to be able to objectively describe the accented speech of Japanese learners of English.

2.2 *Requirements for an Ideal Passage for a Read-Aloud Assessment*

Ideally, a diagnostic passage should include the following:

- every phoneme of the target language, preferably in the same proportion as what occurs in authentic speech;
- as many diphones (types of two phoneme sequences) as possible, especially those found to be difficult for speakers of particular L1s (Japanese, in my case);
- a variety of sentence types and speech acts that elicit a range of prosodic and intonation patterns.

Additionally, the passage should *not*:

- contain words that are infrequent or too difficult for learners;
- be long—long passages may impose heavy burdens on the informants in recording sessions.

To my knowledge, only some of these requirements (most notably, the phoneme and diphone coverage) have been discussed in the construction of currently available passages (e.g., Hiki & Kakita, 2013; Kominek & Black, 2003). Although different sentence types may have been considered in the construction of the passages, I could not find a systematic review that discussed requirements for sentence types.

3 My Investigation

Before attempting to construct a passage that follows the ideal requirements just outlined, I set out to analyze the characteristics, advantages, and shortcomings of commonly used diagnostic passages in order to determine which could serve my data collection purposes most satisfactorily. This section first identifies the passages and criteria used for analysis, and then reports and discusses the findings.

3.1 Passage Selection

I have chosen the following passages for this survey:

- the “Stella” passage from the *Speech Accent Archive* (Weinberger, 2015);
- “The North Wind and the Sun” passage used for “Illustrations of the IPA” in the *Journal of the International Phonetic Association* (International Phonetic Association, 1999);
- “The Boy who Cried Wolf” passage, which was developed to improve upon “The North Wind and the Sun” (Deterding, 2006);
- the diagnostic passage in the *Manual of American English Pronunciation* (Prator & Robinett, 1984);
- the diagnostic passage in *Teaching Pronunciation* (Celce-Murcia et al., 2010);
- the diagnostic passage in *Well Said* (Grant, 2017)—a widely used textbook;
- the short version of “Arthur the Rat,” reproduced in *A Course in Phonetics* (Ladefoged & Johnson, 2015);
- “Text for phonemic contrasts” in William Labov’s study of New York speech (Labov, 2006, originally in 1966).

The passages I have chosen are by no means all the ones that have been used to collect learners’ pronunciation of English. There are at least three other important sentence sets that I could have included: “The Rainbow passage” (formerly used) and the “Comma Gets a Cure” (currently used) by the online International Dialects of English Archive (Maier, 2019), and the Arctic sentence set (Kominek & Black, 2003).

“The Rainbow passage” consists of 331 words of 9th-10th grade level, with an average sentence length of 18 words. “Comma Gets a Cure” has 372 words in 9th-10th grade level, with an average sentence length of 17 words. Critics have found these

two passages rather “unnatural,” perhaps because words were chosen to represent all the English phonemes in terms of “standard lexical sets” (Wells, 1982) that could cover all the possible (vowel) contrasts in different standard native-speaker varieties. Because unnatural passages can produce unnatural utterances, especially as regards prosody, I have chosen not to analyze these two passages. Another major reason is their lack of interrogative sentences, which are necessary for collecting different uses of question intonations.

The Arctic sentence set was originally produced for use in the development of speech synthesis. Importantly for us, it has been used for the collection of L2-Arctic Corpus (Zhao et al., 2018), a collection of L2 English speech by (currently) 24 speakers whose L1s are Arabic, Mandarin Chinese, Hindi, Korean, Spanish, and Vietnamese. This set consists of 1,132 “sentence prompts” (whose average length is 9 words) with 9,998 words (in 9th-10th grade level). One of its major merits is the diphone coverage at 79.6%. Although I have not computed the diphone coverages for the passages assessed in this chapter, they must be much smaller because the number of possible diphones in English is 1,610, according to the developers of the Arctic set, and the numbers of patterns I present below in the analysis are in lower hundreds at most. However, the set’s major merit is also one of its drawbacks. This large set is probably too burdensome for speakers to comfortably read aloud in the recording session. Another problem derives from the way the developers chose sentences for their prompts. Although the set draws on a running literary text, they “pruned” it automatically and manually so that the prompts meet their requirements about length, pronounceability, and types of words and grammar. As a result, the set is largely a collection of isolated sentences rather than a coherent text. As I argue in this chapter, this is not desirable for collecting uses of different prosodic patterns according to contexts.

3.2 *Data Analysis*

The analysis of each passage starts with a general description in terms of its author, number of words, reading level according to Wordcounter.net (<https://wordcounter.net/>), and average sentence length. What follows is a descriptive linguistic analysis of the range of phonemes (based on the General American phonemic inventory), consonant clusters, word-to-word sound combinations (excluding vowel-to-consonant combinations¹), and sentence types embedded in each text. The identified sounds and sentence types constitute potential candidates for speech analyses of read-aloud speech. The findings from the linguistic analysis are then critically discussed in view of the advantages and shortcomings they pose for speech data collection and analysis.

¹ Vowel-to-consonant combinations are not considered because they are not candidates for connected speech phenomena, such as linking and assimilation, and therefore do not generally pose pronunciation problems for language learners.

3.3 Results and Discussion

3.3.1 “Stella”

The passage “Stella” has been used to collect speech at the website *Speech Accent Archive* (Weinberger, 2015). The collection includes recordings from speakers of as many as 386 different first languages. The text is only 69 words long (with 55 unique word forms). According to Wordcounter.net, the text has a 5th-6th grade reading level. The average sentence length is 18 words. The linguistic analysis revealed that this passage included the following candidates for speech analysis:

- All phonemes except /ʒ, dʒ, j, aʊ, ɪə, ɛə, ɑə, ʊə/.
- 23 types of word-internal consonant clusters (“#” denoting word boundaries):
 - initial (12): #bl, #br, #fr, #pl, #sk, #sl, #sm, #sn, #sp, #st, #tr, #θr;
 - final (8): bz#, ts#, gz#, ks#, nd#, nz#, ɲz#, sk#;
 - intervocalic medial (3): ls, nt, nzd.
- 33 types of word-to-word sound combinations:
 - vowel to vowel (3): i#ə, i#ə, a#ə;
 - consonant to vowel (3): k#ə, z#ɪ, z#ə;
 - possible t/d-flapping (1): d#ə;
 - possible place assimilation (2): d#b, d#m;
 - other consonant to consonant (24): p#ð, t#ð, t#h, d#w, k#f, k#s, k#h, g#t, g#f, v#b, v#f, v#θ, ð#h, s#s, z#k, z#θ, z#w, ʃ#s, m#ð, n#s, ɲ#ð, l#p, l#g, l#s.
- 4 sentence types:
 - statements (with falling tones): We also need a small plastic ↘snake...;
 - commands (with falling tones): Please call ↘Stella;
 - non-final intonation phrase (rises): She can scoop these things into three red ↘bags↗...;
 - a list (with non-final rises and a final fall): Six spoons of fresh ↗snow peas || five thick slabs of ↗blue cheese || and maybe a ↘snack.

The short length of this passage makes it handy for collecting samples from a large number of people from different language backgrounds, and that is the main purpose of the *Speech Accent Archive*. However, this passage is far from adequate for collecting speech data for a phonetic study because most of the requirements identified in Sect. 2.2 for an ideal diagnostic text are not fulfilled. Most importantly, some phonemes are not included, and the coverage of consonant clusters and word-to-word combinations is very limited. Also, the majority of its sentences is commands and statements, which will only elicit falling intonation. Although the list included in the text will probably elicit rising tones, the major use of rises in questions is not represented. Hence, the reading will be monotonous, to say the least, and the passage will not be useful for assessing prosody.

3.3.2 “The North Wind and the Sun”

The passage “The North Wind and the Sun” was first used to demonstrate the uses of IPA symbols in “Illustrations of the IPA,” a major section in the *Journal of the International Phonetic Association*. The text has been translated into different languages to demonstrate the uses of the alphabet in different languages. As such, it was not designed for collecting speech samples with adequate phonetic balance. The English version consists of 113 words (66 unique word forms) in the 9th-10th grade level. The average sentence length is 23 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ɑ, ɔɪ, ɪə, εə, ʊə/.
- 20 types of word-internal consonant clusters:
 - initial (4): #bl, #kl, #str, #tr;
 - final (6): dʒd#, ld#, mpt#, nd#, pt#, st#;
 - intervocalic medial (10): bl, gr, ks, ml, nf, ns, ŋg, sl, sp, tl.
- 45 types of word-to-word sound combinations:
 - vowel to vowel (4): i#ʌ, i#ə, u#ə, eɪ#ə;
 - consonant to vowel (7): k#ɔ, k#ə, t#ɪ, d#ə, z#ə, v#ʌ, m#ə;
 - n- and r-linking (3): n#aʊ, n#ə, ə#ə;
 - possible t/d-flapping (1): ɫ#ɪ;
 - lateral release (1): t#l;
 - possible place assimilation (2): d#b, d#g;
 - other consonant to consonant (27): p#ð, t#ð, t#s, t#w, d#ð, d#h, d#s, d#t, d#w, k#h, f#h, f#ʃ, v#ð, θ#w, s#ð, z#ð, z#h, z#k, tʃ#w, m#k, n#ð, n#ʃ, n#h, n#w, ŋ#ð, ŋ#r, ŋ#w.
- 2 sentence types:
 - statements (falls): the North Wind gave up the at ↘temp;
 - non-final intonation phrases (rises): ↘Then ↗ || the North Wind...

I know of at least one L2 speech corpus project (AESOP corpus of Asian Englishes by Meng et al., 2009) which makes use of this passage for collecting speech samples, but the nature of the text (especially the lack of some phonemes and monotonous intonation) can limit its usefulness for speech data analysis. Importantly, the text is far from ideal for obtaining a variety of prosodic patterns given that it only includes statements, which will mostly elicit falling intonation patterns apart from possible rises in sentence-medial phrasings.

3.3.3 “The Boy Who Cried Wolf”

“The Boy who Cried Wolf” (Deterding, 2006) was developed to eliminate deficiencies of “The North Wind and the Sun” passage by collecting phonetically balanced speech samples with fewer repetition of the same words (Hiki & Kakita, 2013). It consists of 216 words (134 unique word forms) in the 7th-8th grade level, and the average sentence length is 27 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes (despite being originally devised for standard British pronunciation).
- 36 types of word-internal consonant clusters:
 - initial (7): #fl, #fr, #kr, #pl, #st, #tr, #θr;
 - final (17): kst#, ldz#, lf#, mz#, nd#, ndʒ#, ns#, nst#, nt#, ntn#, pt#, st#, ʃt#, tʃl#, tn#, vn#, znz#;
 - intervocalic medial (12): ft, gz, ks, ktl, ktʃ, ls, mp, ms, ns, sf, st, tl.
- 97 types of word-to-word sound combinations:
 - vowel to vowel (6): i#eə, eɪ#ə, ɔɪ#ə, oʊ#oʊ, oʊ#ə, u#ə;
 - consonant to vowel (15): p#ɪ, t#ɪ, d#ɔ, k#ə, v#ə, s#ɪ, s#ə, z#ɪ, z#ɔ, m#ɪ, m#aʊ, m#ə, ɪ#æ, ɪ#ɪ, l#ə;
 - possible t/d-flapping (7): t#ɪ, t#æ, t#ə, t#ʌ, d#ə, d#ɪ, d#aʊ;
 - n- and r-linking (3): n#ə, ɪə#ə, ə#ə;
 - lateral release (2): t#l, n#l;
 - possible place assimilation (5): t#b, d#p, d#b, d#k, n#b;
 - other consonant-to-consonant combinations (59): t#f, t#h, t#n, t#s, t#t, d#t, d#d, d#dʒ, d#h, d#s, d#w, k#f, f#ð, f#h, f#k, f#w, v#b, v#ð, v#h, v#k, v#tʃ, ð#h, s#g, s#k, s#m, z#ð, z#f, z#k, z#l, z#s, z#t, z#v, z#w, ts#f, ts#j, tʃ#h, tʃ#p, dʒ#f, dʒ#ʃ, m#t, m#k, m#f, m#ð, n#t, n#d, n#f, n#ð, n#s, n#h, ɪ#t, ɪ#d, ɪ#f, ɪ#ʃ, ɪ#h, ɪ#w, l#d, l#f, l#ð, l#r.
- 4 sentence types:
 - statements (falls): the wolf had a ↘feast;
 - commands (falls): Go a ↘way;
 - non-final intonation phrases (rises): As soon as they ↘heard ↗him;
 - calls (falls): ↘Wolf || ↘wolf.

This passage was constructed for use in phonetic studies. Hence, it has some clear advantages for phonetic analyses: all the phonemes are represented, and there are more word-to-word combinations for its relatively short text. Also, the sentences in the text are relatively long (the longest among the passages discussed in this chapter), so they are more likely to be divided into intonation phrases, which can carry non-final (rising) tones. However, the passage does not seem to have been designed with prosody in mind as it does not contain questions, which would elicit sentence-final rising intonation. This is a major omission. Although the passage has reasonably

animated content which could elicit more expressive performance (i.e., wider pitch ranges), it is clearly more adequate for segmental studies than for collecting data on different prosodic patterns.

3.3.4 *Manual of American English Pronunciation Diagnostic Passage*

The Manual of American English Pronunciation (Prator & Robinett, 1984) is a textbook which was widely used during the latter half of the twentieth century in programs where the target accent was American. It contains a diagnostic text that can be used by the instructor to assess English learners' pronunciation. The passage has 165 words in the 9th-10th grade level, and the average sentence length is 15 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ɪə, əə, ɔɪ/.
- 32 types of word-internal consonant clusters:
 - initial (7): #pr, #tr, #dr, #fr, #kw, #sp, #st;
 - final (10): dz#, dnt#, ks#, kt#, mz#, nt#, nd#, ŋk#, nz#, lf#;
 - intervocalic medial (15): pr, bl, mp, nt, ntr, ms, nf, ns, ŋgw, st, stj, dv, kt, lfr, dnl.
- 45 types of word-to-word sound combinations:
 - vowel to vowel (1): i#i;
 - consonant to vowel (10): t#i, d#ɔ, d#aʊ, k#æ, k#ə, z#i, dʒ#ə, l#æ, l#ə, l#i;
 - possible t/d-flapping (4): t#i, t#i, t#ə, d#i;
 - lateral release (1): t#l;
 - r-linking (4): ə#ɔ, ə#ə, ə#i, ɔə#i;
 - other consonant-to-consonant combinations (25): p#s, t#d, t#f, t#s, t#h, t#r, d#f, d#s, d#h; f#k, f#ð, v#h, s#t, s#l, z#t, z#ð, z#m, z#w, m#t, m#dʒ, ŋ#d, l#p, l#b, l#s, l#ʃ.
- 8 sentence types:
 - statements (falls): All of this will take will \power;
 - yes-no questions (rises): Should he spend all of his time ↗studying?;
 - wh-questions (falls): Where should he \live?;
 - alternative questions (a medial rise and a terminal fall): Would it be better if he looked for a private room ↗off campus, || or if he stayed in a \dormitory?;
 - a yes-no question spanning some intonation phrases (a sequence of rises): Shouldn't he try to take ad↗vantage || of the many social ac↗tivities || which are ↗offered?;
 - non-final intonation phrases (rises): When a student from another country comes to study in the United ↗States...;
 - tag questions (falls or rises): ...doesn't develop \suddenly, || \does it?;
 - vocatives (rises): But let me \tell you, || my ↗friend.

This passage contains a wide range of sentence types. However, in some cases, the plausible intonation patterns are rather unusual and difficult. For example, the sequence of rises in the question spanning three intonation phrases is hard to pronounce for some, and probably not common enough. One may want to collect simpler patterns before delving into that sort of prosody. Another possible objection is that the sentences sound dated. For instance, using “he” for a person unspecified for gender does not feel correct anymore. Because the omission of some phonemes is also a problem, the text needs to be revised before it can be chosen for data collection.

3.3.5 Teaching Pronunciation Diagnostic Passage

Teaching Pronunciation (Celce-Murcia et al., 2010) is one of the major pronunciation textbooks aiming at prospective teachers. Its diagnostic passage has 226 words in the 9th-10th grade level, and the average sentence length is 12 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ɔɪ/.
- 42 types of word-internal consonant clusters:
 - initial (6): #pr, #fr, #θr, #str, #kw, #sp;
 - final (13): kt#, st#, ns#, nt#, nts#, ntʃ#, nd#, ndz#, ndʒd#, ɪz#, lt#, lz#, pl#;
 - intervocalic medial (23): ks, ksp, kt; gz, gn, gr; stʃ, sn; mp, mpr, mb; ntr, nd, nt, nf, nfl, ns, nl; ɪgl, ɪgw; ls, ld, ldr.
- 75 types of word-to-word sound combinations:
 - vowel to vowel (4): i#ɔ, oʊ#ai, u#ɔ, aɪ#i;
 - consonant to vowel (17): d#i, t#i, t#oʊ, t#ə, v#ʌ, s#ɔə, z#i, z#ɔ, z#oʊ, ʃ#ə, ts#ə, tʃ#æ, dʒ#i, m#ə, ɪ#æ, ɪ#ə, l#ə;
 - n- and r-linking: (6): n#æ, n#ə, n#i, ə#ə, ə#i, eə#ə;
 - possible t-flapping (2): t#ə, t#i;
 - possible coalescence (1): t#j;
 - possible place assimilation (3): t#p, t#b, d#p;
 - lateral release (1): d#l;
 - other consonant-to-consonant combinations (41): t#t, t#d, t#f, t#tʃ, t#m; d#t, d#d, d#h, d#w; k#p, k#s, k#h, k#w; f#n; v#l, v#s, ð#p; s#b, s#k; z#p, z#t, z#k, z#f, z#ð, z#j; ʃ#s, ʃ#j; dʒ#w, dʒ#j; n#t, n#k, n#ð, n#l, n#w; ɪ#k, ɪ#t; l#t, l#j; l#f, l#ð, l#r.
- 9 sentence types:
 - statements (falls): There are a couple of \answers to this question;
 - yes-no questions (rises): Is English your native ↗language?;
 - wh-questions (falls): Why is it difficult to speak a foreign \language without an accent?;
 - non-final intonation phrases (rises): If \not ↗...;

- alternative questions (a medial rise and a final fall): Will you make ↗progress, || or will you give ↘up?;
- final comment clauses (low levels): Only ↘time will tell, || I'm a → afraid;
- parentheticals (rises): for e ↘xample ↗;
- lists (non-final rises and a final fall): concentrated hard ↗work || a good ↗ear || and a strong ↘ambition;
- strong assertions (a wider pitch range): You can im ↗ ↘prove!

This passage contains a wide range of sentence types. The sentences are shorter and less complex than those by Prator and Robinett (1984), and this is clearly an advantage in eliciting straightforward intonation patterns that the speaker has acquired. The incomplete phonemic coverage could be a problem, but the omission of only /ɔɪ/ may not be a major concern because it is not one of the difficult vowels of English, and a word containing /ɔɪ/ could easily be added to the text.

3.3.6 Well Said Diagnostic Passage

Well Said (Grant, 2017) is probably one of the most widely used American pronunciation textbooks currently in print. The diagnostic passage has 138 words (89 unique word forms) in the 11th-12th grade level, and the average sentence length is 12 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ɔɪ, εə, əə, ʊə/.
- 32 types of word-internal consonant clusters:
 - initial (4): #kl, #pl, #pr, #sp;
 - final (14): ks#, kt#, lt#, lts#, mz#, nd#, ndz#, nt#, nz#, ɲz#, tnt#, tʃt#, zn#, znz#;
 - intervocalic medial (14): gr, ks, ksp, ldr, ltʃr, mb, mp, mpl, mpr, ns, nstr, ɲgl, ɲgw, sp.
- 66 types of word-to-word sound combinations:
 - vowel to vowel (6): i#ɪ, u#ə, u#ε, eɪ#ə, eɪ#ɪ, aɪ#ɪ;
 - consonant to vowel (15): t#ɪ, d#ɪ, k#ə, v#ə, s#ɪ, z#ɪ, z#ɔ, z#ə, z#ə, dz#ə, dʒ#ɪ, m#ε, ɲ#ɪ, l#ɪ, l#ə;
 - n- and r-linking (5): n#ə, n#ɪ, ɔə#ε, ə#æ, ə#ə;
 - possible t/d-flapping (5) t#ɪ, t#ε, t#ɔ, d#ɪ, d#ə;
 - nasal release (1): t#n;
 - lateral release (2): t#l, d#l;
 - possible coalescence (1): z#j;
 - other consonant-to-consonant combinations (31): t#t, t#k, t#f, t#ð, t#j, t#r, d#k, d#s, d#h, f#t, f#j, v#g, v#ð, v#s, v#j, θ#m, ð#ð, s#l, s#m, z#p, z#s, z#h, z#n, z#w, ts#s, dʒ#ð, dʒ#s, n#w, ɲ#t, ɲ#s, ɲ#tʃ.
- 5 sentence types:
 - statements (falls): pronunciation of a new language is not auto ↘matic;

- yes-no questions (rises): Have you ever watched young ↗children || play with the sounds of the languages they are ↗learning;
- wh-questions (falls): Why is progress in adults more ↘limited?;
- non-final intonation phrases (rises): For young ↘child ↗ren;
- lists (rises): They ↘imi ↗tate, || re ↘peat ↗, || and sing sound combi ↘nations without effort.

The phonemic coverage is incomplete in this passage. Supplying the missing five phonemes may not be an easy task. One might think that the vowels /εə, əə, uə/ are combinations of /ε, ə, u/ plus /r/ phonologically and, hence, not absolutely necessary for pronunciation assessment. However, Japanese speakers perceive postvocalic r's as vowels which sound quite different from prevocalic r's, so it is necessary to collect how Japanese learners of English pronounce them in these combinations.

The coverage of sentence types is fair but not good enough, either. Having yes-no and wh- questions barely fulfills the minimum requirement. The lack of alternative questions is a major omission, and I would like to collect more types of speech acts such as those found in the passage from Celce-Murcia et al. (2010) (see Sect. 3.3.5). This passage is probably too short to include other types of speech acts.

3.3.7 “Arthur the Rat” Short Version

The original version of “Arthur the Rat” (consisting of 594 words) was devised by Henry Sweet and used extensively in the fieldwork for the *Dictionary of American Regional English* (Cassidy, 1985). The shortened version of the original, which is reproduced in *A Course in Phonetics* (Ladefoged & Johnson, 2015), has 339 words (197 unique word forms). Its vocabulary is in the 5th-6th grade level, and the average sentence length is 22 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes except /ʒ, ʊə/, which are the most infrequent consonant and vowel in English.
- 40 types of consonant clusters:
 - initial (8): #bl, #fl, #fr, #gr, #kr, #sk, #st, #tr;
 - final (21): dn#, dnt#, ft#, kt#, ld#, ld#, lf#, lm#, lz#, md#, nd#, ns#, nt#, ŋk#, skt#, st#, sts#, tʃ#, tn#, tʃt#, vd#;
 - intervocalic medial (11): ft, lw, ml, ms, nd, ndl, nl, ns, nʃ, ŋgr, sl.
- 119 types of word-to-word sound combinations:
 - vowel to vowel (10): oʊ#aʊ, i#æ, oʊ#i, u#i, i#a, eɪ#ə, ə#a, u#ə, i#oʊ, i#ə;
 - consonant to vowel (23): p#ə, t#ə, d#ə, d#ə, d#aʊ, d#ə, d#ð, k#l, k#ə, k#ai, f#aʊ, f#i, v#aʊ, ð#ə, s#ə, s#ə, z#i, z#æ, z#ə, dz#æ, m#ə, ŋ#ə;
 - possible t/d-flapping (5): t#ə, d#ə, d#l, d#oʊ, d#ə;
 - n- and r-linking (11): n#ai, n#ə, n#ε, n#i, n#j, ə#i, iə#ə, ə#ɒ, ə#ə, ə#i, ə#l;
 - nasal release (2): t#n, d#n;

- lateral release (3): t#l, d#l, n#l;
- possible assimilation (4): t#k, d#b, d#k, d#g;
- possible coalescence (1): n#ð;
- other consonant-to-consonant combinations (60): p#h, t#d, t#ð, t#f, t#s, t#h, t#w, d#t, d#f, d#h, d#w, k#t, k#ð, k#h, k#m, v#g, v#ð, v#n, v#w, ð#f, ð#s, ð#h, d#r, s#ð, s#m, s#n, s#w, z#d, z#g, z#f, z#f, z#h, z#m, z#r, z#w, ts#k, ts#g, ts#s, ts#h, t#f, m#b, m#t, m#d, m#h, m#m, m#n, n#f, n#s, n#h, n#r, n#w, ɲ#t, ɲ#l, ɲ#r, l#t, l#g, l#ð, l#s, l#h, l#r.
- 5 sentence types:
 - statements (rises): there was a young rat ↘Arthur...;
 - commands (falls): Now look ↘here;
 - tag questions (falls or rises): You’re ↘coming, ll of ↗course?;
 - non-final intonation phrases (rises): One rainy ↘day↗;
 - calls (falls): Right about ↘face.

This passage lacks the most infrequent phonemes /ɜ, ʊə/. In its original longer version, the only missing phoneme is /ʊə/. It is surprising that a passage extensively used for dialect fieldwork does not include all phonemes. However, it is possible to add words like *usual*(ly) and *poor, pure* or *cure* to cover all phonemes. The coverage of consonant clusters and word-to-word combinations is the best of all the texts reviewed in this chapter probably because “Arthur the Rat” is longer than most of them, except the one by Labov (2006) analyzed in the next section. Perhaps the biggest weakness of this passage is the coverage of sentence types: wh-questions and alternative questions are missing. These omissions are problematic because it is impossible to include other sentence types without revising the text extensively. Finally, the story itself does not seem to be exciting enough.

3.3.8 Labov’s “text for Phonemic Contrasts”

Bill Labov used this passage for his study of New York speech (Labov, 2006) in the 1960s, but I do not know of any other study utilizing it. It consists of 23 sentences with 349 words (212 unique word forms) in the 9th-10th grade level. The average sentence length is 15 words. The linguistic analysis revealed the following regarding representative features:

- All phonemes.
- 44 types of word-internal consonant clusters:
 - initial (10): #br, #fr, #gr, #pl, #sl, #sm, #st, #str, #sw, #θr;
 - final (18): bl#, dɜd#, fθ#, kst#, ld#, lf#, lk#, lm#, lz#, nd#, nt#, nz#, sk#, skt#, st#, t#l#, zd#, znt#;
 - intervocalic medial (16): bm, bw, dtw, ktʃ, mw, nd, ndr, nf, nh, nt, ɲg, sk, st, tnl, tns, tʃr.
- 110 types of word-to-word sound combinations:

- vowel to vowel (6): oʊ#ə, i#ə, aɪ#ɑ, i#e, i#æ, u#æ;
 - consonant to vowel (20): p#ə, t#ɒ, t#ɪ, tʃ#ə, f#ə, v#ə, θ#ɪ, s#ɑɪ, s#ʌ, s#ə, s#ə-, z#ə, ʒ#ɑ, m#ʌ, ɪ#ɑ, ɪ#ə, ɪ#ə, l#ɪ, l#oʊ, l#ə;
 - n-linking (4): n#ɑɪ, n#ə, n#eɪ, n#e, n#j;
 - r-linking (2): ə#ɪ, ɛə#æ;
 - nasal release (2): p#m, d#n;
 - lateral release (3): t#l, d#l, n#l;
 - place assimilation possible (11): n#b, n#d, n#k, n#m, t#p, t#k, t#g, t#m, d#b, d#g, d#m;
 - coalescence possible (1): n#ð;
 - other consonant-to-consonant patterns (63): p#ʃ, p#θ, t#d, t#ð, t#f, t#h, t#s, t#ʃ, t#t, t#tʃ, t#w, t#θ, d#ð, d#h, d#s, d#w, k#ð, k#f, k#k, k#l, k#m, k#s, k#t, v#b, v#ð, v#j, v#v, θ#m, θ#s, ð#k, ð#t, s#f, s#h, s#m, z#m, z#s, z#ʃ, z#t, z#θ, t#æ, ts#b, ts#ð, ts#dʒ, ts#t, ts#w, tʃ#w, dʒ#ʃ, m#ð, m#f, m#k, m#t, m#w, n#n, n#r, n#s, n#ʃ, ɪ#f, ɪ#s, ɪ#w, l#ð, l#l, l#r, l#w.
- 10 sentence types:
 - statements (falls): Mary got her \finger in the pie;
 - yes-no questions (rises): Are they running ↗submarines to the Jersey shore?;
 - wh-questions (falls): And what’s the source of \your information;
 - commands (falls): Don’t tell this man any fairy \tales about a ferry;
 - non-final intonation phrases (rises): When Mary starts to sound \humor↗ous,...;
 - vocatives (rises): “And what’s the source of your infor\mation, ↗Joseph?”;
 - calls (wide falls): My \God!;
 - strong assertion (wide falls): Oh yes he \can!;
 - reported speech: “You’re certainly in the \dark,” I \told her;
 - irony: They tore down that dock ten \years ago, when you were in \diapers.

This passage has a complete phoneme coverage, the largest number of consonant clusters and the second most types of word-to-word combinations among all the texts reviewed in this chapter. It also has the most variety of sentence types. Its content is representative of mid-twentieth century New York City (with true proper nouns for people and places which no longer exist after more than half a century), but it has very animated content which could help speakers produce a variety of possible prosodic patterns.

One drawback of this passage is that it is long. Individual sentences also are long, and some of the words are rather difficult, although they are generally much easier than those in ERJ. It will probably place heavier burdens on the people reading it aloud in the recording session. Future research should investigate if its length may negatively affect the results.

3.3.9 Summary of Findings and Passage Selection

Naturally, the longer the passages are, the more phonetic coverage they have. Shorter passages are handy when it comes to recording, but they also have significant omissions. A major exception is the passage “The Boy who Cried Wolf” which has a relatively good phonetic coverage despite its shortness, but its sentence type coverage is not good enough probably because that was not a main concern when it was constructed. The longer passages under review can be tolerable. In fact, they have much fewer words than the sentence set used in my previous studies with *English Read by Japanese* database. Taking this into consideration, Labov’s (2006) passage, which has the best score in each column, seems the strongest candidate for use in my study to collect read-aloud recordings of Japanese-accented English.

Table 1 displays a summary of the main characteristics of the passages surveyed in this chapter.

One possible objection to Labov’s original text is that it contains only one instance of target /ʒ/. Also, I find it desirable that the text contain some instances of intervocalic voiced affricates which tend to be neutralized with fricatives in Japanese. So, I have decided to make the following minor adaptations:

Table 1 Summary of main characteristics in selected passages

	<i>N</i> of words	Phoneme coverage	<i>N</i> of consonant cluster types	<i>N</i> of word-to-word combination types	<i>N</i> of sentence types
“Stella”	69	Missing /ʒ, dʒ, j, aʊ, ɪə, ɛə, ɑə, ʊə/	23	33	4
“The North Wind and the Sun”	113	Missing /ʒ, ɑ, ɔɪ, ɪə, ɛə, ʊə/	20	45	2
“The Boy who Cried Wolf”	216	All	36	97	4
Prator and Robinett (1984)	165	Missing /ɪə, ɑə, ɔɪ/	32	45	8
Celce-Murcia et al. (2010)	226	Missing /ɔɪ/	42	75	9
Grant (2017)	138	Missing /ʒ, ɔɪ, ɛə, ɑə, ʊə/	32	66	5
“Arthur the Rat”	339	Missing /ʒ, ʊə/	40	119	5
Labov (2006)	349	All	44	110	10

- Well, we were waiting in (<— on) line about half an hour.
- “And what’s the source of your information, Roger (<— Joseph)?”
- She suggested that he (<— told him to) ask a subway guard.
- Well, I managed to sleep through the worst part of the picture, and the stage show wasn’t too hard to bear, which was a pleasure for me [inserted].

Although some of the lexical items are still a little difficult (e.g., Palisades, Paramount, rubies, etc.) and there are several words chosen specifically to elicit different characteristics of New York speech (e.g., “carry” vs. “Carey,” “Mary” vs. “merry,” “guard” vs. “God,” “Chock” vs. “chalk”), I have made no effort to replace them with easier and more general ones. The adapted text (see Appendix) has 353 words (226 unique word forms) in the 9th-10th grade level.

4 Implications

Before conducting research for this chapter, Labov’s (2006) passage was already my favorite candidate because of its animated content. Nevertheless, I was also wondering if it would be hard for speakers to handle because of some difficult lexical items and longer individual sentences. These could make the recording sessions more demanding. I now know that it is slightly better than others as far as the phonetic coverage is concerned.

The survey in this chapter has revealed that the phonetic coverage of the diagnostic passage is roughly in proportion to its length. While it is possible to make a passage that is more “efficient” for its shortness (like the “Wolf” passage), actually constructing one is quite another matter. The fact that many of the texts reviewed in this survey do not have a complete phonemic coverage reflects its sheer difficulty. All we can do is to look for some available materials before setting out to make a new passage, and to modify the one we have chosen so that it suits our own purposes more efficiently. I might want to refine Labov’s text in some additional aspects, especially with respect to the coverage of word-to-word combinations, but then again, there may be no “ideal” passage fulfilling all the requirements. If I find it desirable to cover more patterns, the better idea will be to supplement the passage with a word list.

Another point to discuss is the uses of speech recordings collected with the passage. Fundamentally, we are not aware of all the pronunciation characteristics of Japanese speakers of English (or speakers of any L1, for that matter). The recordings help us not just to objectively confirm those that we know, but also to discover what has not been observed. Each of the passages analyzed in this chapter contains materials which can elicit the latter as well as many of the former. It is true that these short passages cannot uncover all the unknown pronunciation difficulties Japanese speakers of English (and other L1 speakers) may experience. That would require a passage or sentence set that contains all the patterns that can elicit them, and preparing such materials is unrealistic even if we do not set the upper limit to the

number of words in it. As far as we do not require such an unrealistic comprehensiveness, however, collecting speech with passages is useful for phonetic studies. Also, these passages are useful in the teaching of pronunciation. They can be utilized not only for diagnostic purposes, as some of them were originally crafted for, but also as target models for learners.

5 Conclusion

In this chapter, I presented a linguistic analysis of eight different passages that can be used to assess L2 English speakers' read-aloud pronunciation. I proposed a set of requirements for the ideal passage, and analyzed the texts based on those requirements and with a focus on Japanese speakers of English. The passage that I have found meets most of the selection criteria is Labov's "text for phonemic contrast" (Labov, 2006). I slightly adapted the original text to make it better suited to elicit phonetic characteristics specific to Japanese speakers.

Acknowledgements The research for this study was supported by KAKEN Grant-in-Aid for Scientific Research (C) 18K00663 and (A) 18H04107 from the Japan Society for the Promotion of Science, and Chuo University Grant for Special Research.

Appendix

Labov's "text for phonemic contrasts" (Labov, 2006), with adapted part underlined.

Last Saturday night I took Mary Parker to the Paramount Theatre. I wanted to go and see *The Jazz Singer*, but Mary got her finger in the pie. She hates jazz, because she can't carry a tune, and besides, she never misses a new film with Cary Grant. Well, we were waiting in line about half an hour, when some farmer from Kansas or somewhere asked us how to get to Palisades Amusement Park.

Naturally, I told him to take a bus at the Port Authority Garage on 8th Avenue, but Mary right away said no, he should take the I.R.T. to 125th Street, and go down the escalator. She actually thought the ferry was still running.

"You're certainly in the dark," I told her. "They tore down that dock ten years ago, when you were in diapers."

"And what's the source of your information, Roger?" She used her sweet-and-sour tone of voice, like ketchup mixed with tomato sauce. "Are they running submarines to the Jersey shore?"

When Mary starts to sound humorous, that's bad: merry hell is sure to break loose. I remembered the verse from the Bible about a good woman being worth more than rubies, and I bared my teeth in some kind of a smile. "Don't tell this man any fairy tales about a ferry. He can't go that way."

“Oh yes he can!” she said. Just then a little old lady, as thin as my grandmother, came up shaking a tin can, and this farmer asked her the same question. She suggested that he ask a subway guard. My god! I thought, that’s one sure way to get lost in New York City.

Well, I managed to sleep through the worst part of the picture, and the stage show wasn’t too hard to bear, which was a pleasure for me. Then I wanted to go and have a bottle of beer, but she had to have a chocolate milk at Chock Full O’ Nuts. Chalk this up as a total loss, I told myself. I bet that farmer is still wandering around looking for the 125th Street Ferry.

References

- Cassidy, F. G. (Ed.). (1985). *Dictionary of American regional English* (Vol. 1). Harvard University Press.
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.
- Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *JIPA*, 36(2), 187–196.
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., Dahlgren, N., & Zue, V. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1* (Electronic database). Linguistic Data Consortium. Retrieved January 29, 2022, from <https://catalog.ldc.upenn.edu/LDC93S1>
- Grant, L. (2017). *Well Said: Pronunciation for clear communication* (4th ed.). National Geographic Learning.
- Hiki, S., & Kakita, K. (2013). The ‘panphonic’ text of ‘The North Wind and the Sun’ for the illustration of the International Phonetic Alphabet of Japanese consonants and its use in the phonetic analysis of Japanese speech. *Proceedings of Meetings on Acoustics*, 19, 060285. ASA. <https://doi.org/10.1121/1.4800255>
- The International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- Kominek, J., & Black, A. W. (2003). CMU ARCTIC: Databases for speech synthesis. Online document. *Carnegie Mellon University Language Technologies Institute Tech Report CMU-LTI-03-177*. Retrieved January 29, 2022, from http://festvox.org/cmu_arctic/cmu_arctic_report.pdf
- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press.
- Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Cengage Learning.
- Maier, P. (2019). *IDEA international dialects of English archive*. Retrieved January 29, 2022, from <https://www.dialectsarchive.com/>
- Makino, T. (2013). Pronunciation characteristics of Japanese speakers’ English: A preliminary corpus-based study. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 5th Pronunciation in Second Language Learning and Teaching Conference* (pp. 120–136). Iowa State University. Retrieved January 29, 2022, from https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2015/05/PSLLT_5th_Proceedings_2013.pdf
- Makino, T., & Aoki, R. (2012). English read by Japanese Phonetic Corpus: An interim report. *Research in Language*, 10(1), 79–95. <https://doi.org/10.2478/v10015-011-0046-5>
- Meng, H., Tseng, C.-Y., Kondo, M., Harrison A., & Viscelgia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: A position paper. In R. Moore (Ed.), *Proceedings of Interspeech 2009* (pp. 1715–1718). ISCA. <https://doi.org/10.21437/Interspeech.2009-517>

- Minematsu, N., Okabe, K., Ogaki, K., & Hirose, K. (2011). Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database. In P. Cosi & R. De Mori (Eds.), *Proceedings of Interspeech 2011* (pp. 1481–1484). ISCA. <https://doi.org/10.21437/Interspeech.2011-310>
- Minematsu, N., Tomiyama, Y., Yoshimoto, K., Shimizu, K., Nakagawa, S., Dantsuji, M., & Makino, S. (2002). English speech database read by Japanese learners for CALL system development. *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)* (pp. 896–903). European Language Resources Association. Retrieved January 29, 2022, from <http://www.lrec-conf.org/proceedings/lrec2002/pdf/155.pdf>
- Prator, C. H., & Robinett, B. W. (1984). *Manual of American English pronunciation* (4th ed.). Harcourt College Publishers.
- Ueyama, M. (1997). The phonology and phonetics of second language intonation: the case of “Japanese English”. In G. Kokkinakis (Ed.), *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)* (pp. 2411–2414). ISCA. Retrieved January 29, 2022, from https://www.isca-speech.org/archive/pdfs/eurospeech_1997/ueyama97_eurospeech.pdf
- Vance, T. (2008). *The sounds of Japanese*. Cambridge University Press.
- Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199249633.003.0007>
- Weinberger, S. (2015). *Speech accent archive*. George Mason University. Retrieved January 29, 2022, from <http://accent.gmu.edu>
- Wells, J. C. (1982). *Accents of English*. Cambridge University Press.
- Zhao, G., Sonsaat, S., Silpachai, A., Lucic, I., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2018). L2-ARCTIC: A non-native English speech corpus. In B. Yegnanarayana (Ed.), *Proceedings of Interspeech 2018* (pp. 2783–87). ISCA. <https://doi.org/10.21437/Interspeech.2018-1110>

Takehiko Makino is Professor of English as a Foreign Language at Chuo University in Tokyo, Japan. He was awarded a Certificate of Proficiency in the Phonetics of English by IPA in 2004. His publications include the Japanese translation of Peter Ladefoged’s *A Course in Phonetics* (1999) and an original textbook (in Japanese) *English Phonetics for Japanese Speakers* (2005).