ENGLISH READ BY JAPANESE PHONETIC CORPUS

Takehiko Makino

Chuo University, Tokyo, Japan
mackinaw@tamacc.chuo-u.ac.jp

Keywords: pronunciation, corpus, English Read by Japanese, Intonation Variation Transcription System

1. INTRODUCTION

English Read by Japanese (ERJ) Phonetic Corpus is a phonetically transcribed electronic corpus of the pronunciation of English by learners whose L1 is Japanese. It is intended to be a source of all the phonetic characteristics of Japanese speakers' English speech, which has tended to be described rather informally [3]. So far the transcription of individual phones is completed and prosodic transcription is underway.

2. ENGLISH READ BY JAPANESE SPEECH DATABASE

The corpus uses as its basis the 800 out of more than 70,000 recordings in English Read by Japanese (ERJ) speech database [5], which was originally collected mainly in order to help CALL system development but has not been much utilized for linguistic research.

Some of the characteristics of ERJ speech database is given below:

- 807 different sentences and 1,009 different words/word sets were used for recording. A majority of the sentences were from those used for TIMIT speech database.
- They were read aloud by 200 (100 male and 100 female) university students in 20 different recording sites all over Japan.
- The speakers did not record all of the materials; each sentence was read by about 24 speakers and each word (set) by 40.
- Phonemic transcriptions in ARPAbet were given for practice; in the recording session, the speakers only looked at orthographic sentences.

3. BUILDING PROCEDURE OF THE CORPUS

The 800 recorded sentences selected for the corpus building were the same set used in [6]. In that study, the sentences were played back on the phone to Americans who were then asked to repeat what they thought they heard, and the correlations between "errors" and the speakers' general pronunciation scores (which were evaluated independently of this study) were calculated. The orthographic transcription of what the Americans repeated could be utilized for future study about the correspondence between pronunciation deviations and intelligibility.

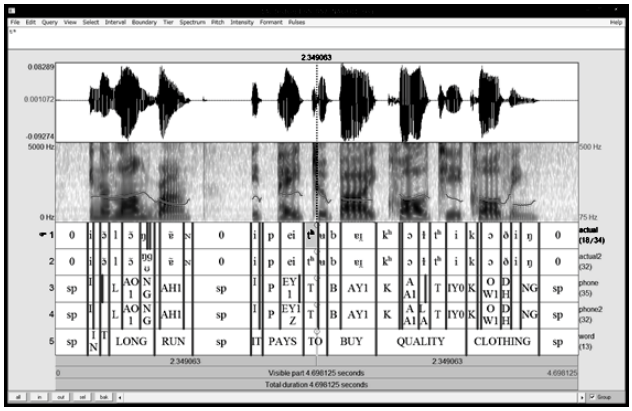
3.1. Segmental annotations

In order to facilitate the transcribing process, the recording was pre-processed by the Penn Phonetics Lab Forced Aligner (p2fa) [9], which produced forced aligned transcriptions of English words and phonemes for each file in the Praat [1] TextGrid format.

The output from p2fa was full or errors, which is not surprising at all as it was developed for native-speaker English. Thus most of the corpus-building work involved manual transcription of "actual phones," which represent what was actually pronounced in narrow phonetic IPA, and manual alignment of both "target phones," a phonemic transcription in ARPAbet which the speaker should have aimed at, and "words," to actual phones, all using Praat.

The TextGrid has two more tiers, namely "actual2" and "phone2" (Figure 1). They are there to collapse into one two or more consecutive actual phones or target phones corresponding to a single target/actual phones. They are necessary because ELAN [7], the software which is used for searching, cannot currently handle one-to-many correspondences.

Figure 1: Segmental annotations.



3.2. Prosodic annotations

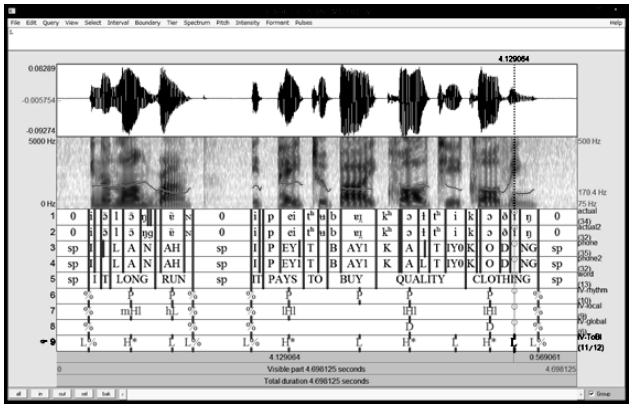
Because the prosodic system in question is part of an interlanguage [2], it cannot be transcribed with the notational system for Japanese or English. Thus it was necessary to devise a new notational system for this corpus.

Actually, this should also have been a problem for segmental transcription. But as far as individual phones are concerned, there is a framework of narrow phonetic transcription, and it was possible to use this to transcribe actual phones. Its actual implementation was not at all straightforward, though, since the truly narrow phonetic transcription independent of any language is an ideal which could not be reached. But still, it is doable.

Not so with prosody. There is no ready-to-use framework for its narrow phonetic transcription. The only framework which has been found for possible use for L2 (interlanguage) prosody is Intonation Variation Transcription System [8], abbreviated to IVTS, which was originally devised to transcribe dialect differences.

There are four tiers in IVTS: (1) rhythmic beat, (2) local phonetic pitch, (3) global pitch change such as downstep, and (4) tentative phonological pitch targets (Figure 2).

Figure 2: Segmental and IVTS annotations



The specific adaptation of IVTS to accommodate Japanese speakers' prosody of English is inevitably tentative.

Since the interlanguage system is fluid (i.e., proportions of influences from L1 and L2 should be different in one speaker/situation from another), it is not possible to devise a watertight notational system. Adjustments will be made as we transcribe more sentences, until we finish with all the 800 files. Even then the system is not complete, since there may be other utterances which could require further adjustments.

4. SOME PRELIMINARY FINDINGS

Some of the preliminary findings from the segment-only corpus survey [4] are given below:

- Voiceless plosive targets are frequently realized as fricatives, especially for /p/ (Tables 1, 2 and 3). This occurs not only in "weakening" positions (e.g., between vowels) but also in "strong" positions. Why /p/ is spirantized much more often than /t, k/ remains to be explored.

Table 1: Actual phones for target /p/ (n=549).

p	pʰ	p̄	ɸ	pɸ	pɨ	others
259	84	14	123	20	15	34
			Spirantized (26.0%)			

Table 2: Actual phones for target /k/ (n=770).

k	kʰ	k̟	k̟ʰ	k̄	k'	x	kx	kɨ	others
330	204	53	21	16	14	44	12	23	53
						Spirantized (7.2%)			

Table 3: Actual phones for target /t/ (n=1,280)

t	tʰ	t̄	t'	ts	tʃ	tɕ	θ	s	tɨ	others
579	325	58	13	28	18	16	15	10	64	154
				Spirantized (6.8%)						

- Target /r/ and /l/ are correctly pronounced in half of their tokens (Tables 4 and 5). This should mean that the distinction between them is far from absent, as is often pointed out.

Table 4: Actual phones for target /r/ (n=879).

ɹ	ɚ	ɾ	ɻ	ə	ɐ	l	ɯ	others
389	40	162	27	69	36	57	29	70
(48.8%)		Japanese /r/ (21.5%)		Vocalized				

Table 5: Actual phones for target /l/ (n=1,037).

l	ɬ	ɾ	ɻ	ɹ	lɨ	ɾɨ	others
385	203	139	20	102	37	36	115
(56.7%)		Japanese /r/ (15.3%)		Hyper-Corrected?	Vowel insertion		

4. REFERENCES

[1] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer (version 5.4.08) [Computer software]. Available: http://www.praat.org/

[2] Ioup, G., Weinberger, S. H. 1987. *Interlanguage Phonology*. Cambridge, MA: Newbury House.

[3] Makino, T. 2013. Pronunciation characteristics of Japanese speakers' English: a preliminary corpus-based study. *Pronunciation in Second Language Learning and Teaching: 5th Annual Proc.*, Ames, IA, 120-136.

[4] Makino, T. 2014. Vowel and consonant patterns of Japanese speakers' English: a study based on English Read by Japanese Phonetic Corpus. Presentation at the 3rd Conference of the International Society for the Linguistics of English, Zurich.

[5] Minematsu, N., et al. 2002. English speech database read by Japanese learners for CALL system development. *Proc. LREC 2002*, Las Palmas, Spain, 896-903.

[6] Minematsu, N., et al. 2011. Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database. *Proc. Interspeech 2011*, Florence, Italy, 1481-84.

[7] Wittenburg, P., et al. 2006. ELAN: A professional framework for multimodality Research. *Proc. LREC 2006*, Genoa, 1556-59.

[8] Post, B., Delais-Roussarie, E. 2006. Transcribing intonational variation at different levels of analysis. *Proc. Speech Prosody 2006*, Dresden, paper 190.

[9] Yuan, J., Liberman, M. 2008. Speaker identification on the SCOTUS Corpus. *Proc. Acoustics '08*, Paris.