

ENGLISH READ BY JAPANESE PHONETIC CORPUS: AN INTERIM REPORT

*TAKEHIKO MAKINO*¹

*RIKA AOKI*²

Chuo University¹

The University of Tokyo²

Abstract

The primary purpose of this paper is to explain the procedure of developing the English Read by Japanese Phonetic Corpus. A series of preliminary studies (Makino 2007, 2008, 2009) made it clear that a phonetically-transcribed computerized corpus of Japanese speakers' English speech was worth making. Because corpus studies on L2 pronunciation have been very rare, we intend to fill this gap. For the corpus building, the 1,902 sentence files in the English Read by Japanese speech database scored for their individual sounds by American English teachers trained in phonetics in Minematsu, et al. (2002b) have been chosen. The files were pre-processed with the Penn Phonetics Lab Forced Aligner to generate Praat TextGrids where target English words and phonemes were forced-aligned to the speech files. Two additional tiers (actual phones and substitutions) were added to those TextGrids, the actual phones were manually transcribed and the other tiers were aligned to that tier. Then the TextGrids were imported to ELAN, which has a much better searching functionality. So far, fewer than 10% of the files have been completed and the corpus-building is still in its initial stage. The secondary purpose of this paper is to report on some findings from the small part of the corpus that has been completed. Although it is still premature to talk of any tendency in the corpus, it is worth noting that we have found evidence of phenomena which are not readily predicted from L1 phonological transfer, such as the spirantization of voiceless plosives, which is not considered normal in the pronunciation of Japanese.

1. Introduction

The purpose of this paper is to explain the procedures in developing the English Read by Japanese (henceforth ERJ) Phonetic Corpus and to report on some findings from the small part of the corpus that has been completed.

A series of preliminary studies (Makino 2007, 2008, 2009) made it clear that a phonetically-transcribed computerized corpus of Japanese speakers' English speech was worth making. So the first author began building the ERJ Phonetic Corpus by making use of ERJ speech database (Minematsu, et al. 2002a), which he also used in the preliminary studies.

Corpus studies on L2 pronunciation have been very rare (cf. Gut 2009, Meng, et al. 2009). We intend to fill this gap with this study.

2. The ERJ speech database

The ERJ speech database was collected mainly in order to help CALL system development (Minematsu, et al. 2002a). 807 different sentences and 1,009 different word sets were read aloud by 100 male and 100 female speakers in 20 different recording sites in Japan. All of the sites were universities and all the speakers were students in those universities.

Each sentence and each word were read by approximately 12 speakers and 20 speakers respectively of each sex. In total, the ERJ speech database consists of more than 70,000 speech files: 24,744 sentence files and 45,495 word-set files.

2.1 ERJ recording procedure

The following explanation of the recording procedure of the ERJ speech database is based on Minematsu, et al. (2002). Before recording, speakers were asked to practice pronouncing the sentences and words on the given sheets. In the practice, they were permitted to refer to the reading sheets with phonemic and prosodic notation.

The phonemic symbols used in the training sheets are based on those of the TIMIT database and the CMU pronunciation dictionary. The model of the pronunciation is therefore Mainstream American English. The actual symbols used are shown below with their IPA equivalents:

Consonants: P /p/, T /t/, K /k/, B /b/, D /d/, G /g/, CH /tʃ/, JH /dʒ/, F /f/, TH /θ/, S /s/, SH /ʃ/, HH /h/, V /v/, DH /ð/, Z /z/, ZH /ʒ/, M /m/, N /n/, NG /ŋ/, L /l/, R /r/, W /w/, Y /j/
 Vowels: IY /i/, IH /ɪ/, EH /ɛ/, EY /eɪ/, AE /æ/, AA /ɑ/, AW /aʊ/, AY /aɪ/, AH /ʌ/, AO /ɔ/, OY /ɔɪ/, OW /oʊ/, UH /ʊ/, UW /u/, ER /ɜ/, AXR /ɝ/, AX /ə/

Each vowel was specified for degrees of stress: “1” for primary, “2” for secondary and “0” for unstressed.

Since the IPA is used for transcribing pronunciations in English dictionaries in Japan, the above set of symbols was unfamiliar to the Japanese subjects. In order to ensure that the speakers understood these symbols correctly, a website was prepared where they could listen to word examples for each phonemic symbol. On that website, they also could listen to sample sentences with prosodic notations (explained below) so that they could understand what those notations meant.

However, the degree to which the speakers made use of the learning materials was entirely up to them; it is possible that some of the speakers were more influenced by spelling rather than phonemic notation.

Examples of sentences in the training sheets are shown below:¹

- S1_0001 This was easy for us.
[DH IH1 S] [W AA1 Z] [IY1 Z IY0] [F AO1 R] [AH1 S]
- S1_0002 Is this seesaw safe ?
[IH1 Z] [DH IH1 S] [S IY1 S AO2] [S EY1 F]
- S1_0003 Those thieves stole thirty jewels.
[DH OW1 Z] [TH IY1 V Z] [S T OW1 L] [TH ER1 T IY0] [JH UW1 AX0 L Z]

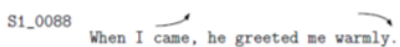

The phonemic notations were removed in the sheets used in the recording sessions, because it was inferred that reading sentences with phonemic notation could induce unnatural pronunciation.

Examples of the sentences with rhythmic specifications are shown below:

- S1_0106 Come to tea with John.
/+ - + - @ /
[K AH1 M] [T UW1] [T IY1] [W IH1 DH] [JH AA1 N]
- S1_0107 Come to tea with John and Mary.
/+ - @ / - + - @ - /
[K AH1 M] [T UW1] [T IY1] [W IH1 DH] [JH AA1 N] [AE1 N D] [MEH1 R IY0]

“@” stands for nuclear stress, “+” for non-nuclear primary/secondary stress, and “-” for unstressed syllables. Here again, the phonemic notations were removed from the reading sheets for the recording sessions, while the rhythmic specifications were retained.

Examples of the sentences specified for their intonation are shown below.

- S1_0088 
When I came, he greeted me warmly.
[W EH1 N] [AY1] [K EY1 M] [HH IY1] [G R IY1 T AX0 D] [M IY1] [W AO1 R M L IY0]
意図：特に指定無し
- S1_0089 
What difference does it make ?
[W AH1 T] [D IH1 F AXRO AX0 N S] [D AH1 Z] [IH1 T] [M EY1 K]
意図：「どちらにしても変わらない」という気持ちを込めて。

Note that the intonation curves are not based on any particular theoretical frameworks but only indicate impressionistically what was decided important. The last line of each sentence is an instruction in Japanese about the meanings/attitudes which were (supposedly) conveyed by the intonation.

¹ It is evident from these examples that different degrees of “sentence accents” and “weak form” pronunciations of function words were not taken into consideration when preparing the phonemic notation. The same is true for the “rhythm-specified” and “intonation-specified” sentences discussed below. This could have led the speakers to pronounce the sentences using “citation form” for every word.

Again, the phonemic notations were removed from the reading sheet used in the recording sessions, while the intonation curves were retained..

In the recording sessions, speakers were asked to read aloud sentences and words on the given sheets repeatedly until they were sure that they pronounced them correctly. If they made errors on the same sentences three times, they were allowed to skip them and go on to the next one.

After recordings, each speech file was checked by the technical staff of the recording site. If they found any technical errors in sentences or words, they were recorded again.

Minematsu, et al. (2002a) claims that with this procedure, the pronunciation errors in the database are supposed to have been made purely because of the speakers' lack of skills in English pronunciation and not because of their lack of knowledge about phonological forms of individual words or spelling-to-pronunciation correspondences.

3. Corpus building procedure

3.1 Selection of speech files

Obviously, it was unpractical to use the whole database for the corpus building because of its sheer size. Fortunately, however, 9,494 speech files have been selected and given pronunciation proficiency scores by American teachers trained in phonetics in another study (Minematsu, et al. 2002b). The selected files are grouped into five sets:

- Sentence files scored for their individual sounds: 1,902
- Sentence files scored for their rhythm: 952
- Sentence files scored for their intonation: 952
- Word files scored for their individual sounds: 3,786
- Word files scored for their stress pattern: 1,902

In the ERJ Phonetic Corpus, we have chosen to use only the first group, i.e., the 1,902 sentence files scored for their individual sounds for transcription. The reason for this choice is that the other sentence groups were specified for their rhythm or intonation, which could have distorted what Japanese speakers normally do when they read English aloud.

Word sets have not been chosen because we are not interested in the pronunciation of individual words.

3.2 Transcription

To reduce the effort of manual transcription, the files were pre-processed by the Penn Phonetics Lab Forced Aligner (Yuan and Liberman 2008;

<http://www.ling.upenn.edu/phonetics/p2fa/>), which produced forced aligned Praat (Boersma and Weenink 2011) TextGrids for each speech file with two tiers: the “word” tier and “phone” (=phoneme) tier.

The p2fa is designed for Mainstream American English speech, so it was inevitable that the Japanese speakers’ speech resulted in transcriptions with numerous errors.

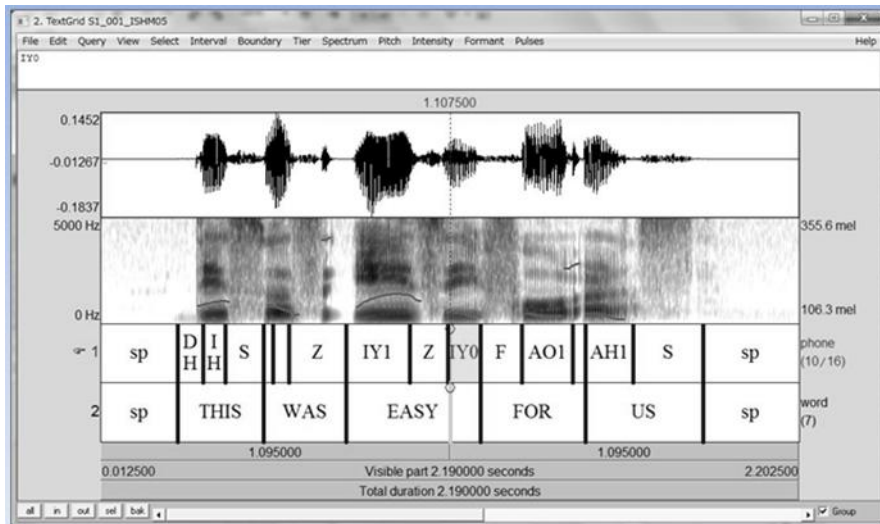


Figure 1: An example of a TextGrid output from the Penn Phonetics Lab Forced Aligner shown on Praat.

Then, using Praat software, two tiers (actual phones > “actual” and substitutions > “subst”) were added to the TextGrids and “word” and “phone” tiers were re-interpreted as target words and target phonemes respectively.

The actual phones were manually transcribed, and boundaries of target phones and target words were manually aligned with the actual phones. The second author of this paper was involved at this very important stage.

The substitution tier is the same as the actual phone tier, except that consecutive actual phones were merged into one unit if they corresponded to a single target phoneme. This tier is only necessary for searching purposes; the duration information of each phone is retained in the actual phone tier.

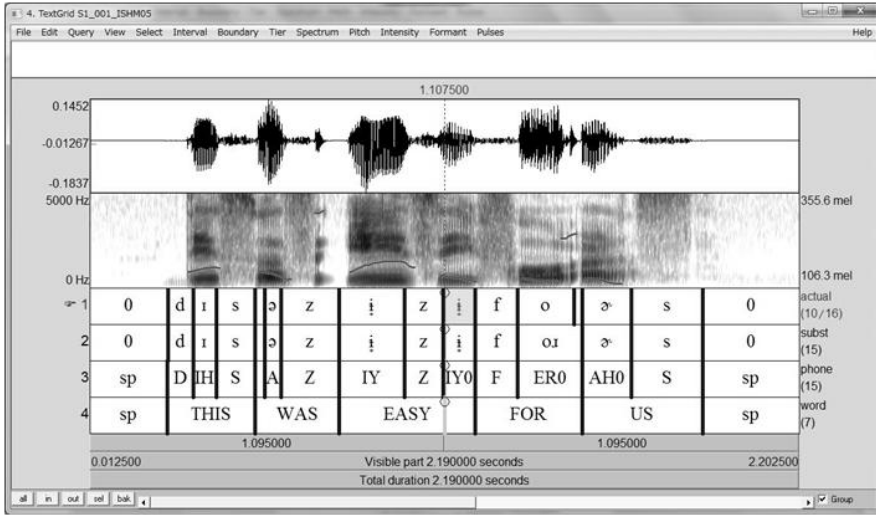


Figure 2: An example of a re-formatted and corrected TextGrid shown on Praat.

The re-formatted and corrected TextGrids were then imported to ELAN software (Sloetjes and Wittenburg 2008; <http://www.lat-mpi.eu/tools/elan/>), which has a much better searching functionality than Praat. The resulting .eaf files and the original .wav files are the complete individual data of the corpus. So far, fewer than 10% of the files have been completed and the corpus-building is still in its initial stage.

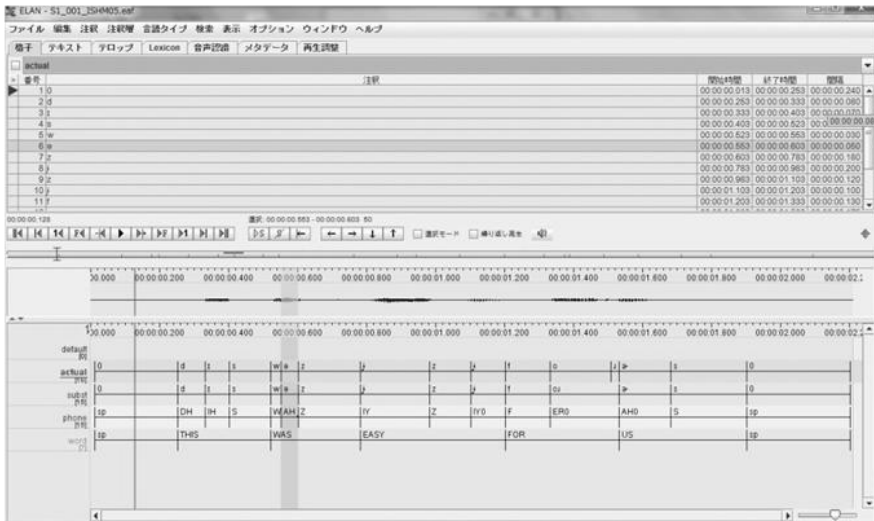


Figure 3: An example of the Corpus data shown on ELAN.

4. Preliminary findings

In this tiny micro-corpus, the following consonantal tendencies, among others, have been found.

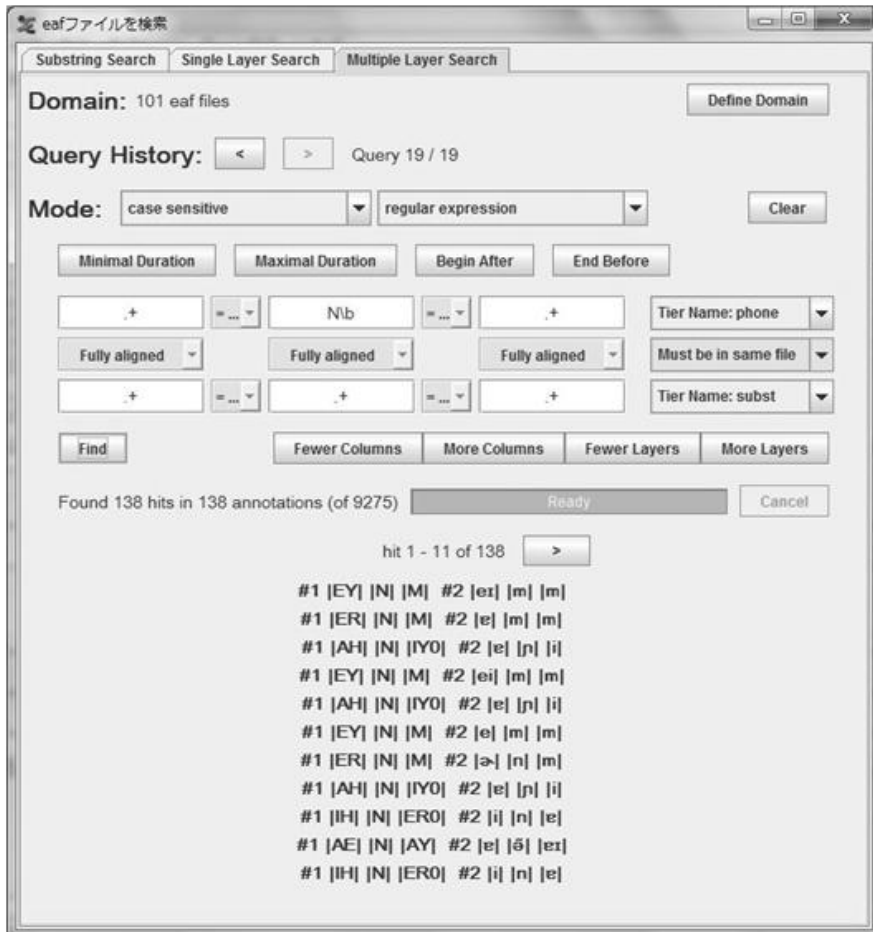


Figure 4: An example of a search result by ELAN.

4.1 Voiced plosives

The voiced plosive phonemes are frequently spirantized (realized as fricatives): 32% for /b/, 15% for /d/ and 8% for /g/. The equivalent phonemes are often (but not obligatorily like, for example, in Spanish) spirantized between vowels in Japanese, so this distribution seems entirely natural. But the situation is quite not so simple. Let's look at the individual cases below.

4.1.1 /b/

Realization	Percentage	Count (n=41)	Contexts
/b/ --> [b]	56.1	23	0_V 9
			0_[approx] 1
			[nas]_V 4
			[obstr]_V 2
			[obstr]_[approx] 2
			V_V 4
			V_[approx] 1
/b/ --> [β]	26.8	11	V_V 5
			V_[approx] 3
			V_[obstr] 1
			[nas]_V 1
			0_[approx] 1
/b/ --> [bi]	2.4	1	V_[approx]
/b/ --> [biu]	2.4	1	[obstr]_[approx]
/b/ --> [b̥]	2.4	1	[obstr]_[approx]
/b/ --> [ḃ]	2.4	1	V_[obstr]
/b/ --> [p̥]	2.4	1	V_[obstr]
/b/ --> [v]	2.4	1	0_V
/b/ --> [ϕ]	2.4	1	V_[obstr]

Table 1: ERJ realizations of /b/ and their phonetic contexts

In the above table, shaded cells in the “Realization” column represent spirantized realizations, and those in the “Contexts” column represent possible spirantizing conditions. “V” represents target vowels, [approx] approximants (liquids and semivowels), [nas] nasals, and [obstr] obstruents (plosives, fricatives and affricates). “0” represents a pause, so “0_” and “_0” correspond to syllable-initial and syllable final positions respectively.

The spirantizing condition for Japanese voiced plosives is “between vowels,” but this does not necessarily result in the spirantization of /b/, as shown in the table. This reflects the fact that spirantization is a variable process in Japanese.

Other possible spirantizing conditions, from a universal phonetic point of view, which do not appear in Japanese but do so in English include syllable-final (or “weak”) positions. “V_[obstr]” (after a vowel and before an obstruent) is a possible context where the following obstruent is very likely to be the onset of the following syllable. This is basically an impossible consonantal sequence in Japanese, and the difficulty in pronunciation can also be resolved by other means than spirantization such as vowel insertion, which does not occur in the current data. The devoiced ([ḃ]) and unreleased ([p̚]) realizations seem to be more English-like resolutions in this condition.

4.1.2 /d/

Realization	Percentage	Count (n=54)	Contexts
/d/ --> [d]	66.7	36	0_V 5
			[nas]_V 6
			[obstr]_V 5
			[nas]_[nas] 1
			[obstr]_V 1
			[nas]_[obstr] 1
			V_[obstr] 2
			V_0 2
V_V 13			
/d/ --> [ɖ]	9.3	5	V_0 2
			V_h 1
			[approx]_0 1
			[nas]_[obstr] 1
/d/ --> [tʰ]	5.6	3	0_V 1
			V_V 1
			V_0 1
/d/ --> [ð]	5.6	3	[obstr]_V 1
			V_V 1
			[approx]_V 1
/d/ --> [ɖʰ]	1.9	1	V_0
/d/ --> [t]	1.9	1	[nas]_[nas]
/d/ --> [tʰ]	1.9	1	V_0
/d/ --> [z]	1.9	1	V_V
/d/ --> [r]	1.9	1	V_V
/d/ --> [ʃ]	1.9	1	[approx]_[obstr]
/d/ --> [θi]	1.9	1	[approx]_[obstr]

Table 2: ERJ realizations of /d/ and their phonetic contexts

/d/ is spirantized rather infrequently in Japanese, much less often than the other voiced plosives. This is reflected in the table, where shaded conditions correspond to many cases of non-spirantized realizations.

4.1.3 /g/

Realization	Percentage	Count (n=25)	Contexts
/g/ --> [g]	80.0	20	0_V 2
			0_[approx] 3
			[nas]_V 2
			[obstr]_V 2
			[obstr]_[approx] 3
			V_V 2
			V_[approx] 2
			[approx]_[approx] 1
V_[obstr] 3			
/g/ --> [gi]	4.0	1	V_[obstr]
/g/ --> [ġ]	4.0	1	V_[obstr]
/g/ --> [ŋg]	4.0	1	V_[obstr]
/g/ --> [ɣ]	4.0	1	V_[obstr]
/g/ --> [x]	4.0	1	V_0

Table 3: ERJ realizations of /g/ and their phonetic contexts

Here again, the fricative realizations are infrequent. More cases of non-spirantized [g] appear in spirantizing conditions than fricative realizations.

The /g/ in Japanese can be realized as a velar nasal [ŋ] as well as a [g] or spirantized [ɣ] between vowels, but this variant does not appear in the current data.

4.2 Voiceless plosives

The voiceless plosive phonemes are also sometimes spirantized: 14% for /p/, 7% for /t/ and 6% for /k/. This cannot be the case of L1 transfer because this sort of “weakening” is not considered normal for Japanese speech.

4.2.1 /p/

Realization	Percentage	Count (n=50)	Contexts
/p/ --> [p]	48.0	24	0_V 2
			[nas]_V 2
			[nas]_[approx] 5
			[obstr]_V 1
			V_[nas] 1
			V_V 9
			V_[approx] 2

Realization	Percentage	Count (n=50)	Contexts
			V_[obstr] 2
/p/ --> [p ^h]	30.0	15	0_V 4 [nas]_V 3 [obstr]_V 2 V_V 2 V_[obstr] 1 [obstr]_0 2
/p/ --> [ϕ]	16.0	8	V_[obstr] 2 [approx]_[obstr] 3 V_0 1 [nas]_[approx] 1 0_[approx] 1
/p/ --> [p ⁱ]	2	1	V_[obstr]
/p/ --> [p ^w]	2	1	[approx]_0
/p/ --> [p ^ʰ]	2	1	V_[obstr]

Table 4: ERJ realizations of /p/ and their phonetic contexts

Here, we are only concerned with spirantized cases; the phonetic conditions in the non-spirantized cases ([p, p^h, p^ʰ]) are too varied, and in any case released [p]s are what is generally found for this sound in Japanese in the phonetics literature.

The fact that a spirantized realization [ϕ] does appear (though infrequently) is in itself notable. It is possible that /p/ is sometimes spirantized in spontaneous Japanese speech under some conditions, but we do not possess the data necessary to confirm this. All the conditions where it appears are spirantizing conditions for voiced plosives. There might be some universal phonetic process at work which can spirantize voiceless plosives in these conditions.

4.2.2 /t/ and /k/

Realization	Percentage	Count (n=112)	Contexts
/t/ --> [t]	65.2	73	V_V 28 V_[obstr] 10 0_[approx] 9 [obstr]_V 9 [obstr]_[approx] 4 [nas]_V 3 [nas]_[nas] 3 V_0 3 0_V 1 [obstr]_[nas] 1 [nas]_[obstr] 1

Realization	Percentage	Count (n=112)	Contexts
/t/ --> [tʰ]	18.8	21	0_V 2
			V_0 2
			[obstr]_V 5
			V_V 9
			V_[approx] 1
			[obstr]_[approx] 1
			[nas]_0 1
/t/ --> [tˀ]	3.6	4	V_[obstr] 2
			V_0 1
			[nas]_[obstr] 1
/t/ --> [ti]	2.7	3	[obstr]_[approx]
/t/ --> [ts]	1.8	2	V_[obstr]
/t/ --> [tʲ]	0.9	1	V_[nas]
/t/ --> [tθ]	0.9	1	V_[nas]
/t/ --> [d]	0.9	1	[obstr]_[nas]
/t/ --> [s]	0.9	1	V_[approx]
/t/ --> [tʷ]	0.9	1	[obstr]_[approx]
/t/ --> [tʃ]	0.9	1	[obstr]_V
/t/ --> [θ]	0.9	1	[nas]_[obstr]
/t/ --> [ð]	0.9	1	V_0
/t/ --> [r]	0.9	1	V_V

Table 5: ERJ realizations of /t/ and their phonetic contexts

Again, we are only concerned with spirantized cases. It is to be noted that spirantized realizations are found even in “non-spirantizing” conditions. Much the same can be said of the spirantization of /k/.

Realization	Percentage	Count (n=73)	Contexts
/k/ --> [k]	53.4	39	V_V 16
			V_[obstr] 8
			[obstr]_V 6
			0_V 2
			V_[approx] 1
			[obstr]_[approx] 1
			0_[approx] 1
/k/ --> [kʰ]	37.0	27	V_V 11
			0_V 4
			[obstr]_V 3
			[approx]_V 2
			[nas]_V 2

Realization	Percentage	Count (n=73)	Contexts
			[obstr]_[approx] 1 V_[approx] 1 0_[approx] 1 V_[obstr] 1 V_0 1
/k/ --> [x]	5.5	4	0_V 1 V_V 2 V_0 1
/k/ --> [ki]	1.4	1	[obstr]_[approx]
/k/ --> [kʰ]	1.4	1	V_[obstr]
/k/ --> [xk]	1.4	1	0_V

Table 6: ERJ realizations of /k/ and their phonetic contexts

4.3 Voiced (inter)dental fricatives

/ð/ is very frequently mispronounced: only 13.5% were canonical [ð]. The most frequent pronunciation was [d], which accounts for 32.4%, and the next most frequent were [dz] (27%) and [z] (21.6%).

Realization	Percentage	Count (n=37)	Contexts
/ð/ --> [d]	32.4	12	0_V 7 [obstr]_V 3 V_V 2
/ð/ --> [dz]	27.0	10	0_V 4 V_V 3 [nas]_V 1 [approx]_V 1
/ð/ --> [z]	21.6	8	0_V 2 V_V 2 [approx]_V 2 [nas]_V 1 [obstr]_V 1
/ð/ --> [ð]	13.5	5	V_V 3 [approx]_V 1 [obstr]_V 1
/ð/ --> [dʰ]	5.4	2	0_V 1 [obstr]_V 1

Table 7: ERJ realizations of /ð/ and their phonetic contexts

The different realizations are more or less evenly distributed, and we should not comment about the conditions where they are found with such small data, although plosive realizations [d, d^h, dz] seem to be preferred in the syllable-initial positions.

4.4 /n/

/n/ was found to be pronounced as some sort of nasalized vowel in more than 30% of the cases. This can be predicted from Japanese phonology, whose moraic nasal /N/ is regularly realized as a nasalized vowel before a vowel, semivowel, sibilant fricative /s/ (which is usually realized either as [s] or [ʃ]) or /h/.

In the table below, [sib] means “sibilant fricative” and specific sounds in their contexts are also transcribed where appropriate.

It is to be noted that nasalized vowel realizations appear even before obstruents in some cases. This again is not predictable from the phonology of Japanese, and cannot be the case of L1 transfer.

Realization	Percentage	Count (n=138)	Contexts
/n/ --> [n]	45.7	63	V_[stop] 17 V_V 25 V_[approx] 4 V_[obstr] 4 V_[nas] 3 [nas]_V 2 0_V 2 [obstr]_V 3 V_0 2 V_[approx] 1
/n/ --> [ɛ̃]	18.8	26	V_[obstr] 16 V_V 7 V_0 3
/n/ --> [m]	10.9	15	V_[p, b] 10 V_[m] 5
/n/ --> [ĩ]	4.3	6	V_[sib] 2 V_[approx] 1 V_0 1
/n/ --> [ŋ]	4.3	5	V_[i~i] 4 [obstr]_V 1
/n/ --> [õ]	2.2	3	V_V 1 V_0 1 V_[sib] 1
/n/ --> [ŋ]	2.2	3	V_/g/ 2 V_0 1

Realization	Percentage	Count (n=138)	Contexts
/n/ --> [ẽ]	2.2	3	V_[stop] 1 V_[sib] 2
/n/ --> [n <silence> n]	0.7	1	V_V
/n/ --> [ŋ]	0.7	1	V_[obstr]
/n/ --> [ã]	0.7	1	V_0
/n/ --> [ẽ]	0.7	1	V_V
/n/ --> [ũ]	0.7	1	V_[approx]
/n/ --> [j̃ɲ]	0.7	1	[obstr]_V
/n/ --> [ø]	0.7	1	V_V
/n/ --> [õ]	0.7	1	V_0

Table 8: ERJ realizations of /n/ and their phonetic contexts

5. Remaining problems

5.1 Lack of prosodic notation

The corpus is intended to be a source of all the phonetic characteristics of Japanese speakers' English speech. Therefore, prosodic notation is also necessary.

However, L2 prosody is very difficult to describe. Studies such as Gut (2009) and Li, et al. (2011) use English ToBI (Beckman, et al. 2005) for L2 English, which I believe is a mistake. L2 prosodic system is neither that of L1 nor of the target language, but something of a mixture of the two.

The first author of this paper will be addressing this problem and proposing a notational system of Japanese speakers' English prosody in Makino (forthcoming).

5.2 Inefficiency of manual transcription

Development of spoken corpora lags far behind that of written corpora for obvious reasons; that is, transcribed texts are not readily available, although making such texts can be facilitated by using automatic speech recognition (ASR) technologies.

The development of L2 spoken corpora is even more difficult, because ASR technologies have not been developed for non-native speech. Even more difficult than this is an L2 phonetically-transcribed corpus like what we are doing, because narrow phonetic transcription (independent of any language) is required.

Tsubaki and Kondo (2011) tried using ASR technologies in the development of their Japanese speakers' L2 English corpus, with reasonably good results, but this entailed an enrichment of the dictionary with all the possible pronunciations for each entry that

could be conceived of in terms of contrastive phonetics of the two languages. Unless the size of the dictionary necessary is very small like theirs (the text they used was “The North Wind and the Sun”), I do not think it practical.

6. Further work

We have decided that a different set of files (800 in total) are to be included in ERJ Phonetic Corpus. Those files were selected independently of the study discussed in §3.1 from the ERJ database for another study (Minematsu, et al. 2011), where the recordings were played over the telephone to Americans who were not familiar with Japanese speakers’ English. The subjects were asked to repeat the sentences they heard and the responses were written down orthographically.

With this data, we will be able to explore what sort of actual phones tend to be misheard or not understood at all. This can be a basis for the study of intelligibility.

Acknowledgements

The research for this study was supported by Grant-in-Aid for Scientific Research (B) No.23300067 (project leader: Nobuaki Minematsu) from the Japan Society for the Promotion of Science, and a Chuo University Grant for Special Research.

References

- Beckman, M., J. Hirschberg and S. S. Shattuk-Hufnagel. (2005) The original ToBI system and the evolution of the ToBI framework. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun. Oxford: Oxford University Press.
- Boersma, P. and D. Weenink. (2011) Praat: doing phonetics by computer. <http://www.praat.org/> [Accessed 30 September, 2011]
- Gut, U. (2009) *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang.
- Li, M., S. Zhang, K. Li, A. Harrison, W.-K. Lo and H. Meng. (2011) Design and collection of an L2 English corpus with a suprasegmental focus for Chinese learners of English. *Proceedings from the 17th International Congress of the Phonetic Sciences*, Hong Kong.
- Makino, T. (2007) A corpus of Japanese speakers’ pronunciation of American English: preliminary research. *Proceedings from Phonetics Teaching and Learning Conference 2007 (PTLC2007)*, University College London, UK.
- Makino, T. (2008) Consonant substitution patterns in Japanese speakers’ English. Paper read at Accents 2008: II International conference on native and non-native accents of English, University of Łódź, Poland.

- Makino, T. (2009) Vowel substitution patterns in Japanese speakers' English. In *Ta(l)king English Phonetics Across Frontiers*, eds. Biljana Čubrović and Tatjana Paunović, 19-31. Newcastle: Cambridge Scholars Publishing.
- Makino, T. (forthcoming) Devising a notational system for the interlanguage prosody of Japanese speakers' English speech: a pilot study for corpus building.
- Meng, H., C. Tseng, M. Kondo, A. Harrison and T. Viselgia. (2009) Studying L2 suprasegmental features in Asian Englishes: a position paper. *Proceedings from Interspeech 2009*, Brighton, UK.
- Minematsu, N., et al. (2002a) English speech database read by Japanese learners for CALL system development. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2002)*: pp.896-903. Las Palmas, Spain.
- Minematsu, N., et al. (2002b) Rating of Japanese students' utterances of English by native English teachers. (In Japanese) *Proceedings from the Fall 2002 General Meeting of the Acoustical Society of Japan*.
- Minematsu, N., et al. (2011) Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database. *Proceedings from Interspeech 2011*, Florence, Italy.
- Sloetjes, H. and P. Wittenburg. (2008) Realization by category - ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.
- Tsubaki, H. and M. Kondo. (2011) Analysis of L2 English speech corpus by automatic phoneme alignment. *Proceedings from SLATE 2011*, Venice, Italy.
- Yuan, J. and M. Liberman. (2008) Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*, Paris, France.