

# INTELLIGIBILITY AND RELATED FEATURES OF ENGLISH SOUNDS PRODUCED BY YOUNG JAPANESE LEARNERS

Akiyo JOTO, Yuri NISHIO

Prefectural University of Hiroshima, Meijo University  
joto@pu-hiroshima.ac.jp, ynishio@meijo-u.ac.jp

## ABSTRACT

This study investigated the intelligibility of young Japanese learners' pronunciations of English alphabet letter names and clarified acoustic features related to the varying intelligibility levels of their productions. Eight native English speakers of four different nationalities evaluated productions made by 10 Japanese third-graders. We performed acoustic analysis to elucidate the features of their productions as well as those of native English speakers and Japanese sounds for comparison.

Young learners' pronunciations of L, P, T, V, and Z were less intelligible. We attributed lower intelligibility to the shorter voice onset time (VOT) of P and T, the shorter duration of the fricative consonants V and Z, and the different formant patterns of L compared to the productions of native English speakers, along with other features affected by Japanese sounds. We offer some pedagogic suggestions based on the results to prevent the stabilization of English pronunciation with a Japanese accent.

**Keywords:** intelligibility, young Japanese learners, English pronunciation, English alphabet letter names

## 1. INTRODUCTION

A new system of early English education was started in public elementary schools in 2020 after a 2-year transitional period in Japan, under which the alphabet, an important part of learning English, is first taught to children in third grade: they learn to pronounce the name of, and identify, each letter. Teaching and learning the English alphabet is crucial not only orthographically but also phonetically, especially in early English education in Japan, because many of the sounds in the English alphabet letter names, which account for about half of English phonemes, do not exist in the Japanese sound system; it is presumed that young Japanese learners (JLs) will find it difficult to produce them. To effectively teach young learners how to pronounce each alphabet letter name, the features of JLs' English pronunciations need to be studied. However, compared with a large number of studies on teaching and learning the orthography of the English alphabet in early English education in

Japan, there are few objective studies of the features of pronunciations produced by JLs [1], [2], and very few focus on the productions of alphabet letter names [3].

To better understand the English pronunciation of young JLs, particularly aspects that have yet to be fully explored, we conducted a study to assess the intelligibility of the English alphabet letter names spoken by elementary school students in Japan (investigation 1), and attempted to determine acoustic features of their productions related to the intelligibility levels in comparison with the productions of native English speakers and Japanese sounds (investigation 2). The findings provide suggestions to improve the teaching of English pronunciation, in which many elementary school teachers lack confidence.

## 2. INVESTIGATION 1

Native speakers of English evaluated the intelligibility levels of the pronunciations of English alphabet letter names produced by JLs.

### 2.1. Materials

The stimulus corpus for the evaluation comprised 260 English productions; 10 Japanese third-graders individually pronounced each alphabet letter name. Each production was sampled at 44.1 kHz, randomly ordered, and filed on a computer.

#### 2.1.1. Speakers

The speakers were 10 Japanese third-graders at a public elementary school (5 boys and 5 girls) who had no experience living outside Japan. They were asked to pronounce each English alphabet letter name in order after receiving four classroom lessons on the alphabet taught by a Japanese teacher.

#### 2.1.2. Evaluators

Eight native English speakers of four different nationalities participated in the experiment: two Americans, two British, two Canadians, and two Australians, chosen to reflect the present situation in Japanese public elementary schools where students

have the opportunity to learn English from assistant language teachers (ALT) of these nationalities.

All evaluators teach English in Japan at the preschool, elementary school, or university level. The length of time they have lived in Japan ranges from 16 to 37 years; all of them can identify English with a Japanese accent because of their daily experiences speaking with Japanese people in English. None of the evaluators had received phonetic training.

## 2.2. Procedure

Each evaluator listened to 260 productions by JLs on a computer and rated the intelligibility level of the alphabet letter names on a 5-point Likert scale, from 1 *very poor* to 5 *excellent*. They were provided with the alphabet letters whose names the students intended to pronounce when rating each JL production. They could listen to each sound as many times as they wanted and rated the productions at their own pace. They could also write comments about each pronunciation if they desired.

After giving a rating, they listened to the same productions by the JLs just one or two times each and typed the letter whose name they heard in an Excel spreadsheet. If they were unable to identify a production as any letter name, they typed a question mark. They could write comments about each production if they wished.

## 2.3. Results

We averaged the rating scores assigned by the eight different evaluators to each production. Table 1 shows the mean rating scores and identification rates for each letter. The inter-rater reliability was high according to Cronbach's alpha:  $\alpha=0.919$  for the intelligibility rating and  $\alpha=0.990$  for letter identification. A strong positive correlation was observed between the rating scores and the identification rates ( $r=0.878$ ); productions with higher intelligibility tended to be more identifiable.

Table 1 indicates that the overall intelligibility level of the productions by the JLs was "fair" (3.74) but that the levels differed across letters: E was the most intelligible (4.66), whereas V, Z, P, and L were less intelligible; for these letters, the rating scores were below 3 (*fair*), with the lowest being L at 2.14.

As outlined in Table 1, the productions for E, H, U, and X were highly intelligible and identified by all evaluators. On the other hand, V, Z, P, L, and T were less identifiable. The intelligibility of T was "fair" (3.01); its identification rate was the third lowest of all (57.5%) and lower than L (58.8%).

To investigate acoustic features related to the various intelligibility levels, we focused on the JLs' productions of V, Z, P, T and L whose intelligibility

levels were below 3.0 and/or identification rates were below the rate for V (65.0%).

**Table 1:** The rating scores of intelligibility levels and the identification rates for the pronunciations of English alphabet letter names produced by the young Japanese learners.

Letter name	Mean of rating scores	Identification rate (%)	Letter name	Mean of rating scores	Identification rate (%)
E	4.66	100.0	Q	3.86	95.0
X	4.51	100.0	A	3.79	93.8
U	4.49	100.0	F	3.69	92.5
I	4.38	97.5	K	3.54	87.5
D	4.36	98.8	M	3.54	78.8
G	4.35	97.5	W	3.54	97.5
H	4.19	100.0	N	3.45	85.0
Y	4.18	98.8	T	3.01	57.5
J	4.14	98.8	R	3.00	80.0
S	4.11	96.3	V	2.85	65.0
B	4.08	80.0	Z	2.73	51.3
O	4.08	97.5	P	2.70	42.5
C	4.01	97.5	L	2.14	58.8
Mean				3.74	86.4

## 3. INVESTIGATION 2

We performed acoustic analyses of the JLs' and NSs' productions of V, Z, P, T, and L to determine the features of individual JLs' productions with varying levels of intelligibility. To compare the JLs' and NSs' productions, we also acoustically analyzed the Japanese equivalents of the English letter names.

### 3.1. Materials

The JLs' productions were the same as in investigation 1. The English materials were English letter names produced by eight native NSs: two male Americans (General American), two female British (one from the Midlands and the other from the middle north), two female Canadians, and two Australians (General Australian), one male and one female. They were the same speakers as the evaluators for investigation 1 except for one Canadian and one American speaker.

The Japanese sound materials were produced by 8 Japanese third-graders (5 males and 3 females) of the 10 participants in investigation 1. They were asked to read Japanese words written in Katakana characters, specifically the Japanese equivalents of English alphabet letter names transcribed into Katakana in the

way they are used and pronounced in daily life in Japan. Each production was sampled at 44.1 kHz and filed on a computer.

### 3.2. Procedure

We conducted acoustic measurements for all the productions of JLs, NSs, and Japanese sound materials using Praat [4]. We measured the VOT and intensity of stop consonants for P and T, and the duration of the fricatives /v/ and /z/ (V and Z). As for L, we measured three formants of /l/.

### 3.3. Results

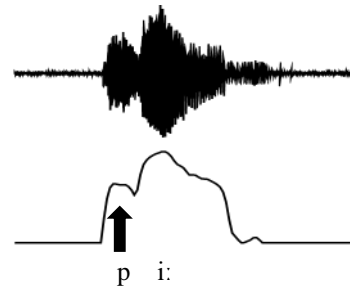
Table 2 shows the mean values of the VOTs of /p/ and /t/ for P and T produced by the 10 JLs and 8 NSs, as well as Japanese stops. We observed a temporary rapid rise of intensity during the VOT of /p/ and /t/ for almost all NSs' productions. We call this the "intensity boost" (henceforth, IB). The temporal point of the IB peak is also shown in Table 2. An example of the IB, due to strong air pressure of aspirated voiceless stop consonants, is indicated by the arrow in Figure 1. Table 2 presents the alphabet letters identified by the evaluators for the JLs' productions of P and T: the left side shows the most frequently identified letters, while the right side displays the second most identified letters. The numbers in brackets indicate the percentage of identification.

**Table 2:** VOT and the temporal point of the IB of stops for P and T produced by JLs and NSs, as well as Japanese stops, and letters identified for P and T.

Stops	Productions (number of JLs)	VOT (ms)	Temporal point of IB (ms)	Identified letter and identification rate (%)
/p/	NS	65.7	24.0	–
	JL (HI) (4)	37.2	–	P (87.5), B (6.3)
	JL (LI) (6)	15.4	–	B (68.8), P (12.5)
	Japanese	21.5	6.3	–
/t/	NS	81.2	28.1	–
	JL (HI) (5)	40.5	43.3	T (92.5), D (7.5)
	JL (LI) (5)	22.3	–	D (70.0), T (22.5)
	Japanese	36.5	12.0	–

For both stops, a positive correlation was found between the intelligibility rating scores and the VOT:  $r = .509$  for P and  $r = .819$  for T. The VOT was shorter for the JLs' productions than for the NSs' ones, even for the productions with high intelligibility (HI) rated as 3.0 or higher, but much shorter for productions with low intelligibility (LI) rated lower than 3.0 and the Japanese equivalents. In addition, the IB did not appear in the JLs' stops, but did occur in all NSs'

productions except those of two British speakers for P and one British speaker for T. The weak IB appeared in only one JL's production with HI rated as 5 for T and only one Japanese equivalent of each stop consonant. The IB occurred within an average of 30 ms from the release of the stop consonants /p/ and /t/ for most of the NSs' productions.



**Figure 1:** The intensity boost for /p/ produced by a native speaker of American English.

Table 3 displays the mean values of the duration of the fricative consonants and the identified letters for V and Z as pronounced by the JLs and the NSs, as well as the Japanese equivalents. The Japanese equivalent of V is /bui/ and that of Z is /dzi:/ or /dsetto/. Since there is no /v/ sound in Japanese, the duration of the Japanese equivalent of V was not measured.

**Table 3:** The duration of the fricatives and identified letters for the V and Z productions of the JLs and NSs, as well as Japanese fricatives.

Fricatives	Productions (number of JLs)	Duration (ms)	Identified letter and identification rate (%)
/v/	NS	139.1	–
	JL (HI) (4)	127.7	V (100)
	JL (LI) (6)	44.1	V (41.7), B (35.4)
/z/	NS	158.1	–
	JL (HI) (5)	97.4	Z (77.5), G (12.5)
	JL (LI) (5)	–	? (27.5), Z/G (25.0)
	Japanese	41.3	–

All the JLs' productions with high intelligibility of V had a /v/ sound; the mean duration was almost as long as that of the NSs' fricative and longer than one JL's production with low intelligibility. The other five with low intelligibility productions had /b/ for /v/.

The acoustic analysis for Z showed a release-burst of /dz/ or /dʒ/ in all JLs' productions and all Japanese equivalents except for one JL production with high intelligibility and one Japanese sound, both without a release-burst.

The correlation coefficient was measured between the intelligibility rating scores and the durations of JLs' /v/ and /z/. JLs' pronunciations of V and Z with higher intelligibility were correlated with longer duration of /v/ and /z/, respectively:  $r = 0.96$  for V and  $r = 0.766$  for Z.

Table 4 presents the mean values of F1, F2, and F3 of /l/ for L produced by the JLs and NSs. The Japanese equivalent of L is /eru/: instead of the approximant /l/, the tap or flap sound is used in Japanese, in which formants cannot be measured. We could measure three formants for L in three male productions.

**Table 4:** The mean frequencies (Hz) of three formants of /l/ in the JLs' and NSs' productions and the Japanese equivalent, and letters identified for L.

Productions (number of JLs)	F1	F2	F3	Identified letter and identification rate (%)
NS	594.6	1072.5	3022.0	–
JL (LD) (10)	507.3	1886.4	2515.9	L (58.8), ? (18.8)

All JLs' pronunciations were of low intelligibility rated below 3.0. F2 is lower in English dark /l/ than light /l/ due to velarization [5], [6]. Notably, regardless of the speaker's age or gender, the JLs' productions have higher F2 and lower F3 than NSs'. The continuous descent of F2 from the vowel observed in NSs' dark /l/ [5] was not present in all three JLs' /l/ sounds. Due to the limited number of JLs' productions available for statistical analysis, we did not measure the correlation coefficient between intelligibility and formant patterns.

There was some variation in the JLs' pronunciations. Seven of the ten students added the Japanese vowel /u/ at the end, and half of them pronounced /r/ for /l/. We also observed the slightly retroflexed lateral /l/, the weak intervocalic /l/, and the weak light /l/; for all of these sounds, a higher F2 or descending F3 is characteristic [6], which affected the acoustic features of JLs' productions for L.

#### 4. DISCUSSION

As presented in Table 1, there were differences in intelligibility across the 10 JLs' productions for each of the five letter names. Most evaluators heard low-intelligibility pronunciations of P or T as B or D, respectively, attributable to the shorter VOT of the voiceless consonants /p/ and /t/. Table 2 shows that the VOTs of JLs' low intelligible productions are less than 25 ms, which suggests English voiced plosives [7]. The VOTs of the Japanese voiceless stops are also shorter than the English ones. This means that

negative transfer of Japanese sounds affected the JLs' pronunciations of P and T.

Negative language transfer also occurred in the JLs' pronunciations of V and Z: /b/ was used instead of /v/ and /dz/ or /dʒ/ instead of /z/. In Japanese, the affricate /dz/ and the fricative /z/ occur before /i/; /dz/ or /dʒ/ usually appear word-initially and /z/ or /ʒ/ word-medially [8]. Thus, it is difficult for Japanese speakers to pronounce /z/ [9], especially in word-initial position. In their comments, some of the evaluators described Z as a "strong /d/ sound mixed in," "hard to decide G or Z," and "the sound in between G and Z." This confusing pronunciation is reflected in the results shown in Table 3; the native English speakers heard the JLs' productions of Z with low intelligibility as Z or G at the same percentage.

The dark /l/ was the most difficult sound for the JLs to pronounce. A strong Japanese accent in the JLs' pronunciation was mentioned in most comments, partly due to the additional vowel at the end and partly because of /l/ sounds different from the dark /l/. In order to gain a deeper comprehension of the dark /l/ pronunciation among young JLs, it is imperative to undertake more detailed and extensive research (quantitative and qualitative).

The JLs' difficulties in pronouncing the above five English consonants have also been observed among adult Japanese speakers. Nishio and Joto [10] reported that over 50% of university students participating in their experiment made errors in pronouncing the letter names of L, P, T, V, and Z. The VOT was shorter in their productions of P and T than in native English speakers' ones. Aspiration of voiceless stops is listed as one of the Lingua Franca Core crucial for international communication [11].

There are some pairs of phonemic contrast among the English letters: B-P, B-V, T-D, and Z-G (in American English). These contrasts carry a higher functional load [12] and need to be taught to young Japanese learners with careful attention to avoid the stabilization of pronunciation. From a pedagogic point of view, we suggest that P and T be pronounced with strong aspiration, and V and Z with longer friction in initial position. As for L, the tongue tip should touch the alveolar ridge while the lips make the gesture of the Japanese vowel /o/ in order to realize velarization of dark /l/, and should be pronounced without an additional vowel at the end.

#### 5. CONCLUSIONS

The intelligibility levels of young JLs' pronunciations of English alphabet letter names differed based on the letter in question as well as the student. Acoustic features related to intelligibility provide suggestions for teaching English pronunciation to young JLs.

## 6. ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant numbers JP18K00787, JP20H01295. We would like to extend our gratitude to Mr Kazuhiro Toi, currently vice-principal at an elementary school in Hiroshima City, for collaborating with elementary school students on their recordings.

## 7. REFERENCES

- [1] Akatsuka, M. and Yamami, Y. 2016. Nihonjin shogakusei niokeru eigo no hatsuon to kokusaiteki rikaido no ichi-kosatsu [English pronunciation of elementary school students and their international intelligibility: focusing on word front consonants]. *Journal of the School of Contemporary International Studies, Nagoya University of Foreign Studies*, 12, 51-61.
- [2] Kato, T. and Yamamoto, S. 2020. Nihonjin shogakusei no eigo onsei no jizokuteki shuroku to eigo tanboin no onkyoteki jikanteki bunseki [Longitudinal collection of English speech produced by Japanese elementary school children and acoustic and temporal analysis of their vowel production]. *Journal of the Phonetic Society of Japan*, Vol. 24, 71-83.
- [3] Joto, A. and Nishio, Y. 2021. The phonetic features of English diphthongs produced by Japanese elementary school students [Online poster presentation], *The 19th World Congress of Applied Linguistics (AILA2020)*, Groningen, the Netherlands.
- [4] Boersma, P. and Weenink, D. 2022. Praat, ver. 6.3.03, <https://www.fon.hum.uva.nl/praat/>
- [5] Olive, J. P., Greenwood, A., and Coleman, J. 1993. *Acoustics of American English Speech: A Dynamic Approach*. Springer.
- [6] Ladefoged, P. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Blackwell.
- [7] Baken, R, J. 1987. *Clinical Measurement of Speech and Voice*. Taylor & Francis Ltd.
- [8] The Japan Foundation. 2009. *Onsei o Oshieru* [Teaching Pronunciation], The Japan Foundation Teaching Japanese Series 2. Hitsuji Shobo.
- [9] Shirota, S. 1995. *Nihongo no Oto* [The Sounds of the Japanese Language]. Hitsuji Shobo.
- [10] Nishio, Y. and Joto, A. 2023. Improving fossilized English pronunciation by simultaneously viewing a video footage of oneself on an ICT self-learning system. In: Albuquerque, J. I. A. and Alves, U. K. (eds), *Second Language Pronunciation: Different Approaches to Teaching and Training*. De Gruyter/Mouton, 249-286.
- [11] Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford University Press.
- [12] Cruttenden, A. 2014. *Gimson's Pronunciation of English*, 8th ed. Routledge.