

ACOUSTICAL LETTER

Word identification of New Zealand English by native Japanese listeners with and without exposure to New Zealand English

C. T. Justine Hui[1,*], Hinako Masuda[2,†], Yusuke Hioka[1] and Catherine I. Watson[3]

[1]*Acoustics Research Centre, Department of Mechanical and Mechatronics Engineering, University of Auckland, Auckland 1142, New Zealand*
[2]*Faculty of Science and Technology, Seikei University, Tokyo, 180–8633 Japan*
[3]*Department of Electrical, Computer, and Software Engineering, University of Auckland, Auckland 1142, New Zealand*

(*Received 9 July 2022, Accepted for publication 22 September 2022*)

Keywords: Speech intelligibility, Error confusion, Second language acquisition

1. Introduction

With English being one of the most spoken languages in the world and has been used as a lingua franca for many years, there are many studies examining non-native English listeners' speech intelligibility in various conditions, such as noisy and reverberant environments, e.g., [1,2]. The stimuli used usually come from pre-recorded corpora in a variety of English at the place of the study. In a recent study, we examined the benefit listeners with different English familiarity get from spatial cues when listening to speech in noise from different directions [3]. As the study was mainly conducted in New Zealand, we used a New Zealand English (NZE) corpus so that participants in the control group were listening to an English variety they were native in. To vary familiarity in New Zealand English as well as exposure to English in general, we also recruited a group of listeners in Japan who has not been immersed in an English speaking country (unpublished data). In order for such an experiment to have valid conclusions that differences in performance were not affected by the corpus in terms of the difficulty of the words and the variety of English, we needed to obtain a baseline of how well non-native listeners can understand the corpus without any distortion. The current paper reports on the findings from this baseline test.

English education in Japan has been taught through mostly General American English (GenAm). The British Council estimated in 2013 that there are 1.75 billion people worldwide using the language. Among English speakers, there are native and non-native speakers, and even within the native speaker group, there are many varieties including General American, British, Canadian, Australian, and New Zealand English, not to mention the regional dialects within these countries [4]. Needless to say, learning to accurately perceive different varieties of English is essential in being able to communicate with both native and non-native English speakers in the current globalised society. Unlike GenAm, NZE is a relatively unfamiliar English variety to students who have studied English via the Japanese education system [5], and anecdotally difficult to follow [6].

Wells [7] proposed a system of lexical sets which enables easy comparison between English accents. Each word in the lexical set represented a collection of words in which the vowel is pronounced the same way for a particular accent, or family of accents. We will adopt this system in this study. NZE is considered to be a non-rhotic variety of English, characterised by its vowel system, especially among the FLEECE, KIT, DRESS and TRAP vowels [8]. Listeners of other English varieties have been reported anecdotally to be confused by the raised NZE DRESS and TRAP vowels [9]. There are also other notable differences between GenAm and NZE. For example, NZE distinguishes between the THOUGHT and LOT vowels, and between the TRAP-BATH vowels, but the SQUARE and NEAR vowels merge. There is also lip-rounding that accompanies the NURSE vowel. This is in contrast to GenAm, a rhotic variety of English, where LOT splits from CLOTH, but PALM and LOT merge [7,10]. NZE also has a fronted GOOSE vowel, and unlike GenAm the back point vowel is THOUGHT, not GOOSE.

On the other hand, there are well-reported confusions for Japanese learners of English stemming from the differences between the sound systems of the two languages. These include vowel confusions in [STRUT-TRAP], [NURSE-START], [GOAT-THOUGHT], [FOOT-GOOSE], and consonant confusions in /s/-/θ/, /r/-/l/ and substitution of /b/ for /v/, due to the lack of /θ, v/ and distinction between /r/ and /l/ in Japanese [11–15].

While non-native listeners' attitudes towards NZE have been investigated previously [5,16,17], there is little investigation on the intelligibility of NZE by Japanese learners of English. The current small-scale study aims to examine how well Japanese listeners with little exposure to NZE understand speech spoken in NZE and the types of errors they may make compared to Japanese listeners with exposure to NZE. It is motivated by the need to control any speech intelligibility performance due to corpus-related effects, with Japanese listeners as our target population. This study contains an analysis of the results from a baseline test to ensure the NZE corpus is suitable for testing Japanese listeners in different noise conditions. Specifically we analysed the errors due to the lack of exposure to the variety of English, NZE, comparing to the errors stemmed from the different language sounds system between English and Japanese. This paper is by no means an exhaustive study for all types of errors when

*e-mail: justine.hui@auckland.ac.nz
†e-mail: h-masuda@st.seikei.ac.jp
[doi:10.1250/ast.44.29]

listening to NZE, but rather the aim is to firstly establish whether the NZE corpus can be used to examine intelligibility for Japanese listeners with no prior overseas English exposure, and secondly, understand the impact of a particular variety of English when conducting speech intelligibility tests on non-native listeners.

2. Methodology

A web-based experiment was carried out to examine how well Japanese listeners without exposure to NZE can perceive NZE sentences. A control group of native Japanese listeners who have resided in New Zealand for more than two years also participated in the experiment.

2.1. Stimuli

The Bamford-Kowal-Bench (BKB) sentences spoken in NZE from the Speech Perception Assessments New Zealand (SPANZ) corpus [18] were used. The BKB sentences are designed to assess speech recognition of partially hearing impaired children. They have been used for non-native English speaker studies since the sentences are considered to be semantically meaningful, syntactically simple and consist of words that are highly familiar to non-native English speakers [19–21]. The BKB sentences in the SPANZ corpus have been modified to suit the New Zealand population [19]. Altogether, 288 sentences were tested, where each sentence consisted of three to four keywords. An example sentence is "The clown had a funny face," where the words "clown," "funny" and "face" are the keywords to be analysed. The scoring of the results is based on the recommendation according to the BKB corpus. Only the root of the keyword is required, e.g., run, ran, running would be marked as correct for the word "running."

2.2. Participants

Twenty two native Japanese listeners (JPN) were recruited. None had lived in an English speaking country for more than one year. Three participants went on short-term exchange programmes from 6–10 months in the U.S., U.K. and Australia. None had resided in New Zealand. Their mean age was 24.3 years old (sd = 5.5) and their mean Test of English for International Communication (TOEIC) scores was 638 out of 990. They all either had gone or were going through tertiary level education. The control group (NZJ) consisted of seven native Japanese listeners who had been living in New Zealand for more than two years (mean: 12.6 years, range: 4–21 years). Their mean age was 41.4 years old (sd = 8.9). All carried out their schooling in Japan and spoke English in their daily life. Their mean age of immersion into NZE was 28.4 years old (sd = 5.8).

2.3. Test procedure

In order to keep the test within reasonable time constraints, the 288 sentences from the corpus were divided into three groups for the JPN participants, where each JPN participant heard 96 sentences. As the NZJ group was more familiarised with NZE, they were divided into two groups, where each NZJ participant heard 144 sentences. The experiment was a web-based sentence identification test implemented using psytoolkit [22]. The participants were asked to transcribe the sentence they heard and were instructed to wear headphones.

3. Results

The NZJ group scored a mean of 97.3% correct, and the JPN group 83.2%. The two groups' scores were significantly different ($t(21.5) = 7.38$, $p < 0.0001$). Error analyses were carried out on the types of confusions the participants experienced. Only words that were scored less than 50% correct were analysed. There were five words that scored less than 50% correct in the NZJ group, compared to 53 words in the JPN group. As the keywords were marked according to the guidelines from the BKB corpus, vowels and consonants in the keywords were assumed to be equally represented. We identified and grouped the error patterns based on the characteristics of NZE, GenAm and Japanese.

The errors were grouped depending if (a) they stemmed from pronunciation features specific to NZE (e.g., TRAP/DRESS raising, a central vowel for KIT, NEAR-SQAURE merger) [23] and (b) they stemmed from the differences between the Japanese and English sound systems. Type (a) errors include differences in NZE and GenAm sounds such as vowel quality confusion and rhoticisation differences. Type (b) errors include well-documented errors typical of Japanese learners such as confusions due to lack of certain vowel and consonant distinctions in Japanese.

The analysis of the errors was not balanced and may include conflation of type (a) and (b) errors due to the nature of the test design and the purpose for testing speech intelligibility using the set corpus. In addition, only errors with less than 50% correct are reported. As the listeners were asked to type out the sentences, we could not always confidently say that the errors were either type (a) or (b), due to possible spelling mistakes, e.g., "knocked" being confused with "know." In those cases, we discarded these errors and only concentrated on the ones which we were confident that they were caused by either NZE pronunciation features or differences between the language sound systems.

3.1. Analysis of NZJ group

Table 1 shows the error patterns of the NZJ group on words where they scored less than 50% correct. Apart from the word "across" being reduced to "cross," we grouped the remaining errors into type (b) errors due to the differences in sound systems between English and Japanese, namely, /r/-/l/, [GOAT-THOUGHT] ("bowl" to "ball") and [NURSE-START] ("stirs" to "stars") confusions. As "stare" has a SQUARE vowel, this may be a case of a spelling mistake.

3.2. Analysis of JPN group

Table 2 shows the error patterns of JPN group on words with less than 50% correct that had consistent type (a) or (b)

Table 1 Error patterns of NZJ group on words scored less than 50% correct.

Word	Responses	Error pattern
across	cross	—
bowl	ball	(b)
broom	bloom	(b)
jelly	jerry	(b)
stirs	stare, stars	(b)

Table 2 Error patterns of JPN group on words scored less than 50% correct.

Word	Responses	Error pattern
bed	bad, beard, beef, beat, beach, bit	(a)
fence	fin(s), sins, things	(a)
hen	him	(a)
heard	hood, food	(a)
mat	met, meat	(a)
men	min, mean, need, minutes	(a)
park's	pack's	(a)
pears	pen, pigs, peas, piece, bee	(a)
set	sit, sip	(a)
sharp	shot, chop, sure	(a)
shirt(s)	shoes	(a)
stirs	stuuze, stood, stoose, stew, stood	(a)
taps	test, teds, teaps	(a)
bowl	ball	(b)
bull	boots, ball, book, boo, bow, boat	(b)
broom	bloom	(b)
clown	crown	(b)
grass	glass	(b)
grows	glows, cross, closed	(b)
ladder	latter, letter	(b)
lies	rise, ride	(b)
mud	mad	(b)
silly	three, thrilling	(b)
thumb	tangue, tongue, stamp, son, sum	(b)
van's	bens, buns, ban	(b)
wore	bow, bought, boa	(a)/(b)

errors. We identified 26 words that had consistent responses that we could deduce from potential type (a) or (b) errors, with 13 words with type (a) errors and 13 words with type (b) errors. Due to limit of space, only these 26 errors are reported. Within the type (a) errors stemming from pronunciation features specific to NZE, we have the KIT-DRESS-TRAP vowel confusions ("bed," "fence," "hen," "mat," "men," "set," "tap"), but also a possible SQUARE-NEAR merger in "pears" to "pigs, peas, piece, bee," and the influence of lip-rounding accompanying NURSE in "stir" to "stuuze, stood, stoose, stew, stood" and "shirt(s)" to "shoes." The absence of rhoticisation can be observed in "park's" as "pack's." In addition, the NZE centralised non-rhotic START vowel is close in position to a GenAm TRAP vowel. The confusion of "sharp" to "shot, chop, shot" can be attributed to both the absence of rhoticisation and PALM-LOT merging in NZE.

For type (b) errors that may stem from the differences between the Japanese and English sound systems: the JPN group exhibited /r/-/l/ confusions for "broom," "clown," "grass," "grows" and "lies." Vowel confusion between [FOOT-GOOSE] can be observed in "bull" to "boots, boo," [GOAT-THOUGHT] in "bowl" to "ball" and [STRUT-TRAP] in "mud" to "mad." The use of tapping similar to GenAm in /t/ can be seen in the confusion between "ladder" as "letter, latter" [24]. Lack of certain consonants in the Japanese sound system such as /θ/ and /v/ can be seen in the confusions for "thumb" to "stamp, son, sum" and "van's" to "bens, buns, ban." The error pattern "silly" to "three" can be attributed to /r/-/l/ and /θ/-/s/ confusions, in addition to lack of consonant clusters in Japanese, where the word "three" can be transcribed and pronounced as "surii" in Japanese as a loan word. While both English and Japanese have initial /w/, the confusion from "wore" to "bow, bought, boa" may be attributed to Japanese /w/ having less lip-rounding than English /w/ [15,25], in addition to the lack of rhoticisation in NZE.

4. General discussion and conclusions

Anecdotally, NZE has been reported to be difficult to perceive accurately for Japanese learners without exposure. The current paper provides evidence of confusions for this lesser spoken English variety, but also, data to confirm the suitability of the BKB corpus for assessing non-native listeners' speech intelligibility, where the JPN group scored 87% correct with clean signals. As a reference, listeners with similar language background scored a mean of 40% correct for the Harvard sentences [26]. With a near 90% accuracy, we would recommend using BKB sentences in studies where there is a need to control for possible effect from the non-native listeners not knowing the words in the corpora.

The analysis in this paper was intended to serve as a report of possible error confusions and not an indication of the distribution of the different error patterns. The experiment was designed as a speech intelligibility test where the participants transcribed what they heard and we only examined the errors that were less than 50% correct. The corpus was designed for children age 8–15, where previous studies deemed the words to be familiar to non-native listeners. This means that there may be other possible confusions arising from the differences between NZE and GenAm that were not captured. To separate such other confusions, further experiments by recruiting Japanese listeners with extensive exposure to GenAm but not NZE would be needed. In addition, some errors could not be confidently analysed due to possible spelling errors and therefore the errors may not be balanced, hence more balanced error pattern analysis by focusing on specific phonemes remains for future study. We should not consider this a comprehensive view to the differences between NZE and GenAm perceived by non-native Japanese listeners, but rather, a snapshot of the different errors exhibited by the two groups with different exposure to NZE.

Vowel confusions due to the lack of vowel distinction in Japanese and /r/-/l/ confusions being observed in both groups suggest that phonemic system differences between English and Japanese were difficult to overcome, in corroboration with studies on /r/-/l/ distinction in Japanese listeners of different language experiences [12]. The JPN group also showed confusions from type (a) errors (due to NZE characteristics), while the NZJ group only exhibited type (b) errors (differences in language sound systems), showing that confusions due to an unfamiliar variety of English can be adapted after exposure to the particular variety. This suggests the importance of English education in Japan to include other varieties of English, especially considering the lingua franca status of English being spoken by various speakers around the world.

The current study provides preliminary understanding on second language acquisition differences for English between fundamental phoneme system differences, which is more difficult to overcome, compared to differences between varieties of English, which can be adapted. A thorough investigation with a corpus and test design dedicated to examine errors caused by vowel and consonant differences between English and Japanese; particular vowels of different English varieties; and rhotacised/non-rhotacised words, are needed to gain more granular level understanding. The results of such studies would potentially contribute to further understanding of second language speech perception.

Acknowledgement

Our gratitude goes to Prof. Suzanne Purdy for access to the SPANZ corpus, and our participants. This work was supported by the Engineering Faculty Research Development Fund at the University of Auckland.

References

[1] A. K. Nábělek and A. M. Donahue, "Perception of consonants in reverberation by native and non-native listeners," *J. Acoust. Soc. Am.*, **75**, 632–634 (1984).
[2] M. L. G. Lecumberri, M. Cooke and A. Cutler, "Non-native speech perception in adverse conditions: A review," *Speech Commun.*, **52**, 864–886 (2010).
[3] C. T. J. Hui, Y. Hioka, H. Masuda and C. I. Watson, "Differences between listeners with early and late immersion age in spatial release from masking in various acoustic environments," *Speech Commun.*, **139**, 51–61 (2022).
[4] British Council, The English Effect — The impact of English, what it's worth to the UK and why it matters to the world —, Tech. Rep. (2013).
[5] B. E. Evans and T. Imai, "'If we say English, that means America': Japanese students' perceptions of varieties of English," *Lang. Aware.*, **20**, 315–326 (2011).
[6] K. Tanaka, "Japanese students' contact with English outside the classroom during study abroad," *New Zeal. Stud. Appl. Linguist.*, **13**, 36–54 (2007).
[7] J. C. Wells, *Accents of English* (Cambridge University Press, Cambridge, 1982).
[8] M. Maclagan and J. Hay, "Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift," *Lang. Var. Change*, **19**, 1–25 (2007).
[9] M. Maclagan, "The English(es) of New Zealand," in *Routledge Handbook of World Englishes* (Routledge, 2010).
[10] Z. Evans and C. Watson, "Flapping in three varieties of English," *Proc. 9th Aust. Int. Conf. Speech Sci. Technol.*, pp. 509–514 (2002).
[11] A. Sheldon and W. Strange, "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," *Appl. Psycholinguist.*, **3**, 243–261 (1982).
[12] K. Miyawaki, J. J. Jenkins, W. Strange, A. M. Liberman, R. Verbrugge and O. Fujimura, "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Percept. Psychophys.*, **18**, 331–340 (1975).
[13] K. Tsukada, *An Acoustic Phonetic Analysis of Japanese-Accented English* (Ph.D., Macquarie University, 1999).
[14] W. Strange, R. Akahane-Yamada, R. Kubo, S. A. Trent, K. Nishi and J. J. Jenkins, "Perceptual assimilation of American English vowels by Japanese listeners," *J. Phon.*, **26**, 311–344 (1998).
[15] K. Aoyama, S. Guion, J. E. Flege, T. Yamada and R. Akahane-Yamada, "Perception and production of English consonants by Japanese adults and children," *J. Acoust. Soc. Am.*, **110**, 2685–2685 (2001).
[16] R. M. McKenzie, "The role of variety recognition in Japanese university students' attitudes towards English speech varieties," *J. Multiling. Multicult. Dev.*, **29**, 139–153 (2008).
[17] G. B. Ray and C. J. Zahn, "Language attitudes and speech behavior: New Zealand English and standard American English," *J. Lang. Soc. Psychol.*, **18**, 310–319 (1999).
[18] J.-h. Kim and S. C. Purdy, "Speech Perception Assessments New Zealand (SPANZ)," *New Zeal. Audiol. Soc. Bull.*, **24**, 9–16 (2015).
[19] O. M. Cañete and S. C. Purdy, "Spatial speech recognition in noise: Normative data for sound field presentation of the New Zealand recording of the Bamford-Kowal-Bench (BKB) sentences and Consonant-Nucleus-Consonant (CNC) monosyllabic words," *Bull. New Zeal. Audiol. Soc.*, **25**(2) (2015).
[20] K. J. V. Engen, "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble," *Speech Commun.*, **30**, 943–953 (2010).
[21] A. R. Bradlow and T. Bent, "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.*, **112**, 272–284 (2002).
[22] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behav. Res. Methods*, **42**, 1096–1104 (2010).
[23] L. Bauer and P. Warren, "New Zealand English: Phonology," in *A Handbook of Varieties of English A Multimedia Reference Tool*, Vol. 1, pp. 580–602 (2008).
[24] J. Hay, M. Maclagan and E. Gordon, *New Zealand English* (Edinburgh University Press, Edinburgh, 2008).
[25] T. Vance, *An Introduction to Japanese Phonology* (State University of New York Press, 1987).
[26] H. Masuda, Y. Hioka, J. James and C. I. Watson, "Protecting speech privacy from native/non-native listeners — effect of masker type," in *Int. Congr. Phonetic Sci.*, pp. 3070–3074 (2019).