HOW JAPANESE L2 LISTENERS PERCEIVE ENGLISH SILENT-CENTRE FRONT VOWELS

Shinichi Tokuma

Chuo University
tokuma@tamacc.chuo-u.ac.jp

ABSTRACT

This study examines the perception of three American English front vowels, /iː/, /ɪ/, /e/, by Japanese listeners in ‘Silent-Centre’ (SC) paradigm. A perceptual experiment was conducted to investigate whether English L2 vowel perception by Japanese listeners is affected by the SC paradigm, using the set-up utilised by previous research. The results demonstrate that: (1) the Japanese sound system, in particular the importance of duration on vowels and consonantal geminates as a perceptual cue, plays a significant role in the perception of English SC /iː/ and /ɪ/; and (2) certainly for some vowels, duration-neutralised tokens create some perceptual confusions, but the influence of Japanese sound system outweighs the perceptual effect of neutralised duration.

Keywords: L2 vowel perception, Silent-Centre vowels

1. INTRODUCTION

Anybody with good knowledge of acoustic phonetics knows that the F1/F2 formant patterns of the vowels flanked by consonants show great variability in comparison with the values when the vowels are produced in isolation. Listeners somehow cope with this variability and find no difficulty in perception.

This phenomenon, vowel formant undershoot [1] and its perpetual overshoot [2], had been a central issue in early days. However, it was discovered later that vowel inherent duration and vowel dynamic spectral change carries more weight in perception. One kind of strong evidence comes from what is called a ‘silent-centre’ (SC) vowel paradigm, suggested by Strange and her colleagues (e.g. [3]), where listeners are asked to identify the vowel in /CVC/ stimuli whose entire central portion was silenced while durational information is kept intact. Their results demonstrate that “vowels in SC syllables were identified remarkably well, despite the total absence of vocalic nuclei. In fact, when only a single speaker’s utterances were represented, errors on SC syllables were no higher than on unmodified control stimuli.” ([3] p.2186)

This perceptual independence on static vowel information proposes one intriguing question: do L2 listeners show the similar perceptual behaviour when they listen to SC vowels, because it is well known that L2 listeners sometimes rely on different acoustic cues for perception (e.g. perception of English /l/-/ɹ/ by Japanese listeners in [4])?

This L2 perception of SC vowels has been studied previously. For example, Rogers & Lopez [5] investigated the perception of American English vowels in SC paradigm by L1 and Spanish-speaking L2 listeners. Their results demonstrate that the L2 listeners who started learning English after the age of 18 showed the identification rate consistently lower than the L1 listeners, while the L2 listeners with the age of onset of immersion of 12 years or earlier identified the whole-word and 40ms duration-preserved syllables as accurately as the L1 listeners, but identified the silent-centre syllables significantly with less accuracy overall. Also, Schwartz and his colleagues ([6] [7]) studied how some SC vowels of British English were perceived by L1 and L2 Polish listeners, the latter of whom were at different English proficiency levels. They discovered that depending on proficiency levels, the listeners adopted a more dynamic approach to vowel identification and showed higher accuracy rates on the SC vowels: they heed more attention to dynamic formant cues, or vowel inherent spectral change.

However, no studies were found which deal with perception of SC English vowels by Japanese L2 listeners, and it would be of particular interest to investigate the perception of English vowels which cause a great confusion among Japanese listeners. such as /iː/-/ɪ/, and compare the results with those of /e/ that causes less confusion (see [8] [9]). As in other SC studies, the effect of neutralised durational difference is also examined: /iː/-/ɪ/, in this study.

Hence, this study investigates the perception of English SC vowels /iː/, /ɪ/, /e/ by Japanese listeners. The experimental set-up used in [5] is followed because their L1 and L2 data can be compared with the data obtained in this study. The main objectives of this study are: (1) to compare the perceptual data with those obtained in the previous studies and see if there is any effect by Japanese sound system (2) to compare perceptual patterns, particularly in difficult conditions like neutralised or /iː/-/ɪ/ vowel pairs that are likely to cause confusions.

2. EXPERIMENT

2.1. Materials

Six American English /bVb/ and /dVd/ words were used as stimuli, where /V/ is /iː/, /ɪ/ or /e/: in the English writing system, they are written as ‘Beeb’, ‘bib’, ‘bebb’ (a nonsense word), ‘deed’, ‘did’ and ‘dead’.

The vowels /iː/ and /ɪ/ were selected since it has been reported (e.g. [9]) that Japanese listeners of English show great perceptual confusion between these two vowels even in ordinary listening environments. The rest of the vowels, /e/, which shows less perceptual confusion among Japanese listeners, was selected for comparison.

Rogers & Lopez [5] explored three more American English vowels, but this study focuses on the vowel pair /iː/-/ɪ/ which demonstrates more significant confusions by Japanese listeners than other vowels. In contrast, although Rogers & Lopez [5], without no apparent reason, investigated only the /bVb/ consonantal frame, two consonants, /b/ and /d/, were examined in this paper since, as stated in [8], the perception of of /ɪ/ or /e/ by Japanese listeners is affected by the types of flanking consonants.

The target /CVC/s for editing were recorded as follows. A native speaker of General American (male in his 50s, with no speech impairment) was asked to read, at a comfortable speed, the target /CVC/s in the frame sentence used in [5], “I say _____ on the tape.” The sentences were printed on a sheet, in the English orthography, described above. He was asked to read the sheet twice, and this session was repeated. Therefore, four tokens per each /CVC/ type were recorded.

The left-channel monaural recording was made in a quiet room with Sony Electret Condenser Microphone ECM-PCV80U attached to a Windows laptop PC, and the sound was transferred to the PC by Praat (ver. 6.2.10) [10], at a sampling frequency of 44100 Hz.

Out of four tokens for each /CVC/ types recorded as above, one optimal token for stimulus manipulation was selected based on the spectral analyses and the auditory impression. The target /CVC/s were excised from the frame sentences and they were manipulated, by Praat, to create SC vowels.

The procedures were as follows: first, the onset and offset of the vowel in the /CVC/ token were determined by waveform and spectrograph analysis. Then, four types of /CV/ and /VC/ transition, whose vowel duration is 10/20/30/40ms from the onset/offset, were created and the central parts of the vowel were silenced, by manipulating Intensity tier on Praat. After that, the edges of the initial /CV/ and final /VC/ (i.e. just before and after the central silence part) were smoothed by applying a 2ms linear filter ramp, as in [5], in order to minimise the auditory impression of clipping. Also, as a control group, the non-manipulated (i.e. no SC) /CVC/ tokens were presented to the listeners.

Furthermore, as in the previous studies, duration-neutralised (DN) tokens of /iː/ and /ɪ/ were also created and presented to participants since the importance of vowel intrinsic duration in SC paradigm has been demonstrated in the previous studies ([3] [5], amongst others), and in particular, it is demonstrated ([11], also amongst others) that the duration is a main factor that affects the substantial perceptual confusion in non-SC English /iː/ and /ɪ/ by Japanese listeners. Henceforth, for clarification, tokens without durational manipulation are called duration-preserved (DP).

The duration of DN tokens was calculated as follows. First, the mean duration was obtained across four tokens for each /CVC/ types. Then the mean of the two was calculated for the duration of DN tokens. The actual values are shown in Table 1 below.

	/biːb/	/bɪb/	/diːd/	/dɪd/
Mean duration of 4 /CVC/ tokens (ms)	195.8	130.4	228.2	168.0
Duration of DN tokens (ms)	163.1		198.1	

Table 1: Mean duration of 4 /CVC/ tokens and the DN duration.

Manipulating Duration tier on Praat, the DN tokens were created by editing the central silent part, or the central steady part in the case of non-SC vowels, without changing the /CV/ and /VC/ transitions, as in the previous studies. Note that this process created DN tokens for each consonant and vowel type (i.e. one for /biːb/, another for /bɪb/) .

This process created 30 token types for DP (2 consonants x 3 vowels x 5 duration patterns of transition), and 20 for DN (2 consonants x 2 vowels x 5 duration patterns of transition). The edited tokens were put back in the original frame sentence and presented to the participants.

2.2. Participants

Twenty-four Japanese first-year and second-year students of Chuo University participated in the experiment. None of them had lived or studied abroad, and had a hearing impairment. They learnt English at school for 6 years (first-years) or 7 years (second-years), and their mean TOEIC L&R score was 615. This, and informal evaluation by the author, established their English proficiency level around CEFR A2-B1. They received a small monetary compensation for their participation.

2.3. Procedure

The participants were tested in a quiet Language Laboratory room at Chuo University. They were seated individually in front of laptop PCs, and their task was to listen to the stimulus token through covered-ear headphones, and using a mouse, click the target SC word that appeared on the screen.

The locations of the words on the screen were randomised, and each stimulus token was played in a random order. The stimuli were played through covered-ear headphones at a sound level adjusted by them, and none of them reported that their attention had been diverted by extraneous noises or by the presence of other participants. The whole experiment process was controlled by Praat, utilising Experiment MFC objects.

Since the experiment had to be conducted during one 100-minute class, the twenty-four participants were randomly assigned to two groups of 12 participants, and one group listened to /bVb/ DP + DN, while the other to /dVd/ DP + DN. This arrangement reduced the number of stimulus presentation to 90 (15 token types x 6 target word position on the screen) + 40 (20 token types x 2 target position on the screen) =130 and minimalised the effect of reduced concentration. All the members of each group were tested at the same time.

The experiment was preceded by a task demonstration by the author designed to make the participants familiar with the experimental setting and the nature of the stimuli. Care was taken that the participants knew that they must not click a word on their screen before it was played. After that, the participants joined the trial session of three test stimuli, and the questions raised were answered at this stage. In the main session, short break was inserted after every 10 stimulus presentations. At the end of the experiment, the general purpose of experiment was explained to the participants but no individual feedback was given.

2.4. Results

Figures 1 and 2 are for DP results. The vertical axis represents the identification rate, while the horizontal axis is for the transition duration ('full' means there is no SC). In these Figures, /iː/ tokens show a steady decline in % correct of identification as the duration of transitions become shorter, and also there is a big drop in 10ms for /e/. These observations are in concordance with [5]. However, in these Figures, /ɪ/ shows an overall rise in % correct, and this rise was not observed in the previous studies. This pattern is replicated for DN results in Figures 3 and 4, where /iː/ tokens shows a steady decline in contrast with the rise of /ɪ/ tokens. This rise was also unreported in [5], where the performance of the participants shows no significant difference or fall.

Furthermore, there is a difference in the response patterns of /ɪ/ between DP and DN tokens. For DP, the rise occurs later in Figures 1 and 2: when the duration of the /CV/ and /VC/ transition is shorter, around 10ms or 20ms, while for DN, in Figures 3 and 4, the rise occurs earlier: when the duration of the transition is still 40ms. Furthermore, overall identification rate for 'full' (i.e. no SC) /ɪ/ is lower in DN than DP. (N.B. The full token has no SC but the duration has been manipulated.)

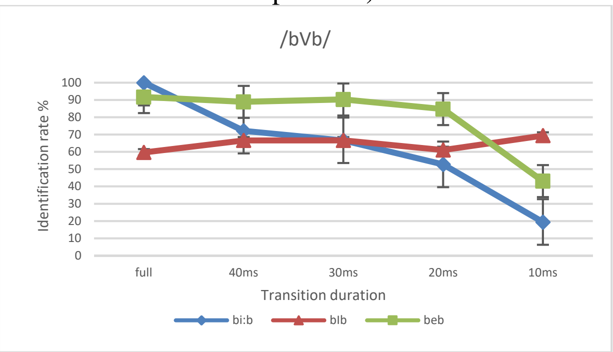

Figure 1: Results of DP /bVb/

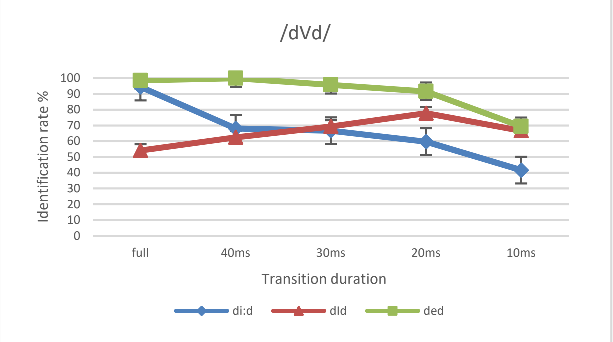

Figure 2: Results of DP /dVd/

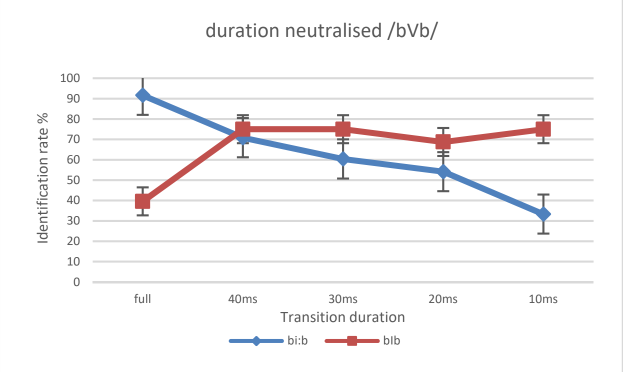

Figure 3: Results of DN /bVb/

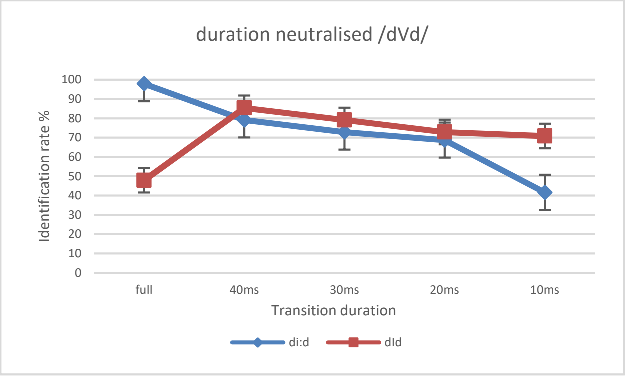

Figure 4: Results of DN /dVd/

3. DISCUSSION

Overall, as mentioned in Section 2.4, there are some concordances in the results between [5] and this study: the falling trend of /iː/ and /e/ in DP and DN conditions, and the sudden dip in 10ms for /e/ (See Figure 4 in [5]). However, the rise pattern of /ɪ/, which was not reported in their studies or others, as well as the difference of the point where /ɪ/ rises between DP and DN (or the higher % correct in full DN tokens), was observed in this study. This issue needs to be discussed in detail.

As mentioned before, Japanese listeners are heavily dependent on durational cues when they listen to /iː/-/ɪ/ difference, and they hardly pay any attention to the spectral cues. When the DN stimuli were created, to neutralise the durational difference, the nucleus duration of /ɪ/ was elongated (See Table 1), and consequently, this longer nucleus duration was likely to evoke a perceptual preference for longer /iː/, hence the lower % correct for /ɪ/.

Another explanation comes from the influence of Japanese sound system. In Standard Japanese, there is a phonological contrast based on consonantal gemination: *saka* 'slope' and *sakka* 'writer'. Many loan words from English /CVC/ containing /iː/ are written in Japanese with non-gemination and a long /iː/, hence 'sheep' having become *shiipu*, while those containing /ɪ/ with gemination and a short /i/, thus 'ship' having become *shippu*. It is known that the Japanese consonantal geminate gives an auditory impression of vowel clipping. Vance [12] illustrates how a Japanese vowel in an emphatic word, followed by a glottal stop (with auditory clipping impression) is spelt with a letter representing geminates.

Although the extra care was taken to minimise the clipping impression, the finally produced SC tokens were not free from it, and therefore they tend to be perceived as having consonantal geminates, providing an auditory bias towards /ɪ/. This influence of clipping impression also explains why there is a rise in /ɪ/ as the duration of the /CV/ and /VC/ transitions becomes shorter: the shorter transitions give a stronger impression of clipping because of the shorter vowel duration.

Moreover, the fact that the % correct is higher for DN tokens than DP ones is also explained by this process. If the duration of the English vowel /ɪ/ is longer and closer to that of /iː/, Japanese listeners, according to Morrison [11], will more likely to perceive it as /iː/, but the results were that they preferred /ɪ/ in DN, which has a longer duration.

The length of the silence is crucial for the perception of Japanese plosive geminates (e.g. [13]). When creating the DN stimuli, the duration of /ɪ/ was increased. This means that the SC interval between the transitions is longer in DN than in DP. The longer silence duration certainly evokes the perception of geminates, and this contributes to the perceptual preference of /ɪ/. Interestingly, the peak of % /ɪ/ in /dVd/, in Figure 3, falls in 20ms stimuli, while in Figure 5, it is in 40ms stimuli. The 20ms shift in peak could be explained by the durational difference between DN and DP: 197.1(DN)-168.0(DP)=20.1ms. However, this hypothesis requires further verification.

In the perception test, the data for the goodness-of-fit for each token was also taken but not yet analysed. The scope of the future research includes its analysis and the investigation of individual difference in performance in correlation of their English proficiency.

4. CONCLUSION

This study examined the perception of English SC vowels /iː/, /ɪ/, /e/ by Japanese listeners. The results demonstrate that: (1) the Japanese sound system, in particular the importance of duration on vowels and consonantal geminates as a perceptual cue, plays a significant role in English SC perception of /iː/ and /ɪ/; and (2) certainly for some vowels, DN tokens creates some perceptual confusions, but the influence of Japanese sound system outweighs the perceptual effect of neutralised duration.

5. ACKNOWLEDGEMENT

This research is partially funded by Chuo University Personal Research Grant.

6. REFERENCES

[1] Lindblom, B. 1963. Spectrographic study of vowel reduction. *J. Acoust Soc. Am.* 35, 1773–1781.

[2] Lindblom, B., Studdert-Kennedy, M. 1967. On the role of formant transitions in vowel recognition. *J. Acoust Soc. Am.* 42, 830–843.

[3] Strange, W. 1989. Dynamic specification of coarticulated vowels spoken in sentence context. *J. Acoust Soc. Am.* 85, 2135-2153.

[4] Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57.

[5] Rogers, C., Lopez, A. 2008. Perception of silent-center syllables by native and non-native English speakers. *J. Acoust Soc. Am.* 124, 1278-1293.

[6] Schwartz, G., Aperliński, G., Jekiel, M., Malarski, K. 2016. Spectral Dynamics in L1 and L2 Vowel Perception. *Research in Language*. 14, 61-77.

[7] Schwartz, G., Dzierla, J. 2018. Polish listeners' perception of vowel inherent spectral change in L2 English. *Poznan Studies in Contemporary Linguistics.* 54. 307-332.

[8] Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K. 2001. Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *J. Acoust Soc. Am.* 109. 1691-1704.

[9] Nozawa, T., Wayland, R. 2012. Effects of Consonantal Contexts on the Discrimination and Identification of American English Vowels by Native Speakers of Japanese. *Kotoba-no-Kagaku-Kenkyu.* *（ことばの科学研究）*13. 19-40.

[10] Boersma, P., Weenink, D. Praat: doing phonetics by computer [computer programme]. Version 6.2.10, retrieved on 20 March 2022 from http://www.praat.org/

[11] Morrison, G. 2002. *Effects of L1 duration experience on Japanese and Spanish Listeners' perception of English high front vowels.* Unpublished MA dissertation. Dept. of Linguistics, Simon Fraser University.

[12] Vance, T. 2008. *The Sounds of Japanese.* Cambridge University Press.

[13] Kawahara, S. 2015. The phonetics of sokuon, or geminate obstruents. In: Kubozono, H. (ed). *Handbook of Japanese Phonetics and Phonology*. De Gruyter Mouton, 43-78.