THE EFFECT OF SINGING ON IMPROVING SYLLABIC PRONUNCIATION – VOWEL EPENTHESIS IN JAPANESE

Hitomi Nakata[a] *& Linda Shockey*[b]

[a]Dokkyo University, Japan; [b]Reading University, UK
h_nakata@dokkyo.ac.jp; l.shockey@reading.ac.uk

ABSTRACT

One of the difficulties in pronouncing English for Japanese learners can be traced to unfamiliar English syllable structures (such as CCCV or CVC). It is commonly believed that native speakers of Japanese overcome this problem through the use of epenthetic vowels. This study examines to what extent these extra vowels are suppressed to approximate more closely the syllable structure of English after singing lessons specifically targeted to improve the pronunciation of consonant clusters.

Preliminary results suggest that university students of 18-20 yrs majoring in English performed quite well in pronouncing consonant clusters in connected speech even *before* having singing lessons: only 27% of tokens contained unnecessary vowels. After three months of weekly singing lessons, the rate of vowel insertion was further diminished to 13%. The effect was statistically supported in comparison to a control group. The results therefore suggest that at least young adult native speakers in Japanese do have sensitivity to the syllabic structure in English, and this can be reinforced by some specified tasks such as songs of focusing syllable timing.

Keywords: singing, syllabic pronunciation, Japanese, epenthesis, acoustic analysis

1. INTRODUCTION

It has been reported (e.g. [5]) that native speakers of Japanese tend to add unnecessary vowels when speaking English, as a result of misapplication of the phonological principles of Japanese. But some researchers (e.g. [6]) argue that younger Japanese demonstrate their sensitivity to syllable timing even in the mora-timed language.

The use of singing to enhance either listening or pronouncing skills in ESL/EFL contexts is recommended by a number of authors [2, 3] For example, BBC TeachingEnglish (British Council) says on their website [11], “Songs, and especially the chorus, provide real and 'catchy' examples of how whole phrases are pronounced often to the extent that students find it difficult to pick out individual words. The music further emphasises the 'flow' of the words”. However, this sort of claim seems to be vague, impressionistic and possibly even based on a wishful thinking. While the use of music in the classroom for acquisition of native-like prosody is not a new idea, few studies have tested the idea empirically.

The present study investigates the degree of vowel epenthesis in Japanese learners of English using an acoustic method of detecting/measuring vowel insertions in connected speech. Second, (if it is found that there is a margin to improve), we will further see if the frequency of insertion changes after having singing lessons which emphasise consonant clusters in English. Results will be compared before/after the lessons, and with results of a control group.

2. SUBJECTS AND PROCEDURES

16 students of English pronunciation in a class at a Japanese university constituted the experimental group (hereafter, EG). 11 students from an English listening class at another university in Japan formed the control group (hereafter, CG). Both groups consisted of roughly half men and half women, and were instructed by the same teacher (the first author) who is not a native speaker of English.

In the beginning of the term, both groups were asked to read aloud five short sentences of English which include consonant clusters. The sentences are from the textbook used by both groups [9]. Each utterance was recorded onto an IC recorder in an acoustically favourable environment and analyzed using AVS Audio Editor, in which both waveforms and spectrograms can be observed. For EG, singing sessions were given for about 20 min. a week during the term for 10 weeks. CG was not given these lessons. At the end of the term, both groups were asked to read aloud the same sentences, and the performances were recorded in the same process as the first time.

3. MATERIALS

3.1. Linguistic tokens (for measurement)

Five carrier sentences were given for subjects to read aloud: 'Spring is here', 'My sister is very strong', 'I have a friend in Australia', 'Our children always help me with the dishes after dinner', and 'The host introduced me to the other guests'. Table 1 shows a list of targeted words including consonant clusters which can be potentially broken up by a vowel insertion. The rightmost column 'Codes' represents abbreviations for hypothesised epenthetic positions and vowels. For example, if /sprɪŋ/ is pronounced as /sp**ʊ**rɪŋ/, it is labeled as 'mʊ1' meaning 'the 1st /ʊ/ appeared in the middle of the word'; and 'fɔ2' as 'the 2nd /ə/ appeared in the final', and so on. Altogether eight tokens are targeted in five sentences. The bolded/shaded segments are epenthetic vowels.

Table 1: Targeted eight words which potentially include epenthetic vowels.

Words	Potential epenthesis	Codes
Spring	sp**ʊ**rɪŋ …	mʊ1
is (here).	ɪz**ʊ**	fʊ1
My sister is	ɪz**ʊ**	fʊ2
very strong.	st**ə**rɔ (:)ŋ	mɔ1
I have a friend	frend**ə**	fɔ1
in Australia.	ɔ:st**ə**reɪlj**ə**	mɔ2
Our children	ʧɪld**ə**rən	mɔ3
… introduced	ɪnt**ə**rədju:st	mɔ4

Items above are re-organized in Table 2, categorising the ordering of the codes into the position of the tokens. The shaded parts are the 'middle' group (x 5) and non-shaded parts are the 'final' group (x 3).

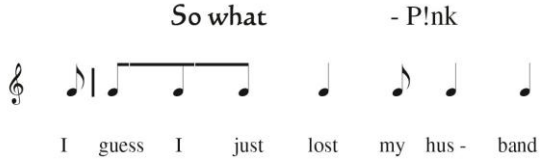
Table 2: Rearranged codes according to positions of potential epenthesis.

mʊ1	mɔ1	mɔ2	mɔ3
sp**ʊ**rɪŋ	st**ə**rɔ (:)ŋ	ɔ:st**ə**reɪlj**ə**	ʧɪld**ə**rən
mɔ4	fʊ1	fʊ2	fɔ1
ɪnt**ə**rədju:st	ɪz**ʊ**	ɪz**ʊ**	frend**ə**

3.2. Song material (for training)

English pop songs of relatively fast tempo and/or having clear vocal articulation were adopted and used for practice, e.g., '*We will rock you* (sung by *Queen*)' and '*A whole new world*' (sung by *Peabo Bryson & Regina Belle,* a.k.a. *Aladdin & Jasmine*)'. These songs are also relatively easy to sing, not having complicated melodies or extremely high tones. The general idea of 'phrasal practice' is employed from [1] here. The method goes like this. Students first listened to the song, and then sang it without a melody contour in order to focus on the rhythmic structure hence the syllabic structure two to three times. Then, they were asked to repeat the phrase-by-phrase singing performed by an instructor (the first author), which follows the whole part singing. At this stage, it is more like chanting rather than singing. Finally they sang the same song along with the original, karaoke versions, and/or movies such as YouTube. An example of the 'rhythmic' session is given below.

Figure 1: An example of phrasal practices in singing.



4. SEGMENTATION CRITERIA

Every utterance (altogether 256 tokens for the experimental group, and 176 for the control group) was analyzed. The criteria of the measurement are listed below.

- In each consonant cluster, vocalic parts were detected if they show periodic wave patterns and high amplitude between two consonants.
- Voiced segments of 50 ms or more were counted. (The value was set out from Patel's documentation '… a rapid acoustic transition (~ 50 ms)' [7] p. 74).
- Formant energy or intensity patterns were adopted for detecting the boundary of obscure parts such as a succession of a low vowel (e.g. [o]) and a sonorant consonant (e.g. [r]) [4] p. 149-151).
- If a targeted voicing part was too fused to be detected, it was compared to the same vowel of other parts of non-clustered parts.

5. RESULTS

5.1. Performances by the experiment group

Results of EG were shown in Table 3 below. The vocalic insertion was plotted as ● in the matrix disregarding the actual value of lengthening. (Except one case, G-mo1, all plots ranged from 50 – 100 ms). The alignment of the tokens accords with the one in Table 2; five items from the leftmost constitute a group of epenthetic vowels in the middle position, and three in the right make a group of those in the final position. 'Ss Cd'

indicates the subjects' code numbers: A – P; 'Ttl #' refers to the total number of plotted positions of vowel insertion. Individual totals are shown vertically, and the totals of each token are shown horizontally.

Table 3: Results of EG (Experimental Group) before/after the singing training.

Ss	Before lessons (Oct.2010)								Ttl	After lessons (Jan.2011)								Ttl
Co	mu	no	mo2	mo3	no4	fu1	fu2	fo1	#	mu	no	mo2	mo3	no4	fu1	fu2	fo1	#
A			●	●		●			3				●		●			2
B	●	●	●	●					4				●					1
C			●	●					2									0
D			●		●				2						●			1
E						●			1							●		1
F		●	●						2				●					1
G		●	●	●				●	4									0
H	●	●	●	●		●			5				●		●		●	3
I									0									0
J			●						1			●	●					2
K							●		1									0
L									0									0
M	●		●	●					3				●					1
N									0									0
O	●			●				●	3	●		●						2
P		●	●	●					3	●		●	●					3
#	4	5	#	8	1	3	1	2	34	2	0	3	7	0	3	1	1	17

As can be seen from the diagram above, it is immediately noticeable that the frequency of the vowel insertion was higher *before* having singing lessons. The total number of this pre-training period against the post-training was exactly 2.0 times higher: 34 vs. 17. But the general ratio of epenthesis was not so high in both periods; 27% in the pre-training and 13% in the past-training. A one tail test confirmed that the variances between two periods were statistically significant: ($F\ [0, 15] = 3.30$, $p = .002 < .05$). Therefore we can generalize two things: 1) Japanese learners of English do *not* insert vowels in consonant clusters very often when speaking English; 2) a certain period of singing lessons seems to facilitate the improvement of syllabic pronunciation.

The most frequent position of insertion was in the word-middle /ɔ/ in 'Australia' /ɔːstɔreɪljə/ (mo2) in the pre-training period: 10, but the number decreased to three after the training. The second highest position was another /ɔ/ in 'children' /ʧɪldɔrən/ (mo3): eight, and this number rarely changed even after the training: seven. The possible factor of the major improvement from /ɔːstɔreɪljə/ to /ɔːstreɪljə/ may be attributed to subjects' uncertainty of the pronunciation of the word. In measuring, a long pause between /ɔːst/ and /reɪljə/ was often observed, which consequently produced an unnecessary vowels /ɔ/ after /ɔːst/. Why /ɔ/? It may be due to the conversion to the moraic structure. In Japanese, 'Australia' is pronounced as /ɔːs(ʊ)tɔrarɪa/ which forms CVCV structure. If the second half of the word /reɪljə/ was not certain, it is speculated that the learners applied their L1 structure in order to fill the gap between 'Aust' and 'ralia'.

In measuring another word-final token /frend(ɔ)/, a consonant /d/ deletion was frequently observed. The phenomenon is not a major concern in this study but we would like to briefly mention; it can be attributed to a learners' strategy to approximate 'native-like speech' by reducing a final cluster. Tarone [10] p. 74 points out that learners produce some unexpected processes such as overgeneralization and approximation in dealing with their L2; as well as transferring their L1 to certain aspects of the interlanguage. Alternatively, it is possible that the subjects employed a native like pronunciation 'friend in' as /fren^{d}ɪn/. Shockey [8] p. 97 provides a similar example of pronouncing 'didn't' as /dɪ^{d}n̩/, where "… the (second) 'd' was a passing, short closure before a nasal release and the final 't' did not appear at all. Shockey observed this reduced speech from native speakers of English, but the same process can be applied to non-native speakers.

5.2. Comparison to the control group

Figure 2 and Figure 3 compare the results of EG and CG. For EG (Figure 2), the dotted/light marks indicate the number of epenthesis (max. 8: Y axis) for 16 subjects (coded A-P: X axis) in the pre-training period, while the square/dark marks refers to the resulted number for the same subject in the post-training period. For CG, the marking design signifies the same as EG, only is different in the subjects' number: 11 and the coding, a-k.

Figure 2: Number of epenthesis by EG before/after the training.

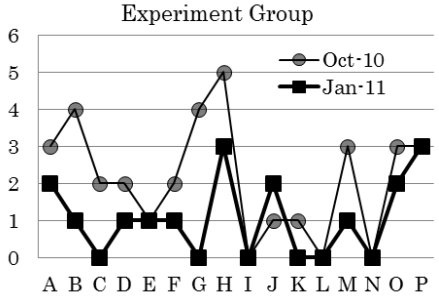
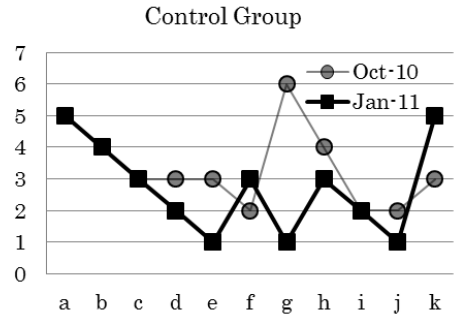


Figure 3: Number of epenthesis by CG before/after the training.



- Before singing practice

For both cases, the degree of inserting vowels into consonant clusters was relatively low. It ranges between 0 – 6 times with the average 2.13 for EG (27%) and 3.36 for CG (42%). Thus the notion of 'Japanese preference of epenthesis' was not supported.

- After singing practice

Almost all subjects showed less epenthesis than the pre-training period in both group. (The first three subjects (a – c) of CG are missing in the pre-training session due to the same number superimposed on the post-training). The variance of CG was *not* statistically significant: *(F [0, 9] = 1.17, p = .14 > .05*). The discrepancy between EG and CG hence supports the effect of singing lessons.

6. DISCUSSION AND SUMMARY

Results support our claim that for Japanese learners closed syllables are not a challenge, thus providing evidence that a consonant cluster in English is quite learnable for native speakers of a mora-timed language. The results also show that the degree of approximating English prosody can be further improved by having deliberate training such as singing. The training reduced the ratio of epenthesised tokens from 27% to only 13% in the short period of three months. We also obtained significant variances in comparison to CG.

Other factors may have contributed to the improvement. The subjects of EG had other (traditional) 'pronunciation-related' activities in classes such as repetitions after model speech, conversations in pairs, and even mini lectures on phonetics, which may enhance more native like speech. These factors should be examined more closely in future studies.

The good performance can also be attributed to the current Japanese pop songs or J-pop. We observed in karaoke bars that English lyrics are frequently used in modern Japanese songs. It is popular pastime for university students, so it is quite possible that they naturally "practice" their English pronunciation in leisure time. Subjects here are only young adults around 20-yrs old, so it would also be interesting if we compare performances both from older and younger generations, and from native speakers of English.

7. REFERENCES

[1] Geberina, P. 1990. Verbo-Tonal method. In Roberge, C., Kimura, M., Kawaguchi, Y. (eds.), *Nihongo no hatsuon shidoo – VT hoo no riron to jissai (Methodology of Japanese Pronunciation – Theory and Practice of VT Method)*. Tokyo. Bonjinsha.

[2] Graham, C. 1979. *Jazz Chants for Children*. New York: Oxford University Press.

[3] Kumai, N., Timson, S. 2010. *Smash Hit Listening – Second Edition*. Tokyo: MacMillan Language House.

[4] Ladefoged, P. 2003. *Phonetic Data Analysis – An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell.

[5] Mazuka, R., Cao, Y., Dupoux, E., Christophe A. 2010. The development of a phonological illusion: a cross-linguistic study with Japanese and French infants. *http://www.riken.jp/r-world/info/release/press/2010/101012/detail.html* (in Japanese)

[6] Nakata, H. 2010. *Timing Relationship Between Spoken and Sung Utterances in Japanese – Speech Rhythm and Musical Rhythm*. VDM Publishing House.

[7] Patel, A.D. 2008. *Music, Language, and the Brain*. Oxford: Oxford University Press.

[8] Shockey, L. 2003. *Sound Patterns of Spoken English*. Oxford: Blackwell.

[9] Sugimori, M., Yuminiwa, K., Ootsuka, A. 2005. *Sounds of English for Communication*. Kinseido.

[10] Tarone, E. 1987a. The phonology of Interlanguage. In Ioup, G., Weinberger, S. (eds.), *Interlanguage Phonology*. Rowley, MA: Newbury House, 70-85.

[11] TE Editor of British Council. 2006. *http://www.teachingenglish.org.uk/think/articles/developing-pronunciation-through-songs*