VOWEL DEVOICING IN TOKYO JAPANESE: ITS LEXICAL STATUS

Natsuya YOSHIDA[1], Mafuyu KITAHARA[2], Ayako S. SHIROSE[3]

[1]National Institute for Japanese Language and Linguistics(NINJAL),Japan
[2]Sophia University, Japan, [3]Tokyo Gakugei University, Japan
[1]natsuya@ninjal.ac.jp

ABSTRACT

In Tokyo Japanese, high vowels can be devoiced typically between voiceless elements. According to the classic view, devoiced vowels are not phonemes but allophones of corresponding vowels. Consequently, the devoicing process belongs to the phonetic domain. However, many scholars have claimed that devoicing actually belongs to the phonological process, and some of them have even proposed that devoicing may partly be specified in the lexicon. To investigate the lexical status of vowel devoicing, we conducted a lexical decision task of natural and edited mismatched stimuli where high vowels were devoiced before voiced consonants or voiced before voiceless consonants in real words and nonce words. Reaction time (RT) for the stimulus was measured. The results suggest that the mismatch condition affected RT selectively according to real/nonce status of words and voicing. Though the idea of lexically specified devoicing is not directly supported, we can provide new insights into the issue of phonetics-phonology-lexicon interface.

Keywords: Japanese, vowel devoicing, lexicon, mismatch sound

1. INTRODUCTION AND BACKGROUND

In the present report, we investigate the lexical status of vowel devoicing in Tokyo Japanese. As numerous previous studies have pointed out, high vowels [i,u] that are surrounded by voiceless elements tend to be devoiced in many Japanese dialects (including Tokyo dialect). Broadly, two types of explanations have been proposed for the mechanism of vowel devoicing.

The first type is a phonetic account where a voiced segment surrounded by voiceless elements loses vocal vibration because the surrounding voiceless elements require the glottal adduction gesture [3]. In other words, vowel devoicing (hereafter devoicing) is the result of phonetic gestural overlapping, which belongs to a phonetic process, not a phonological process. This account leads to the claim that devoicing specification may not be a part of an entry in the mental lexicon.

The second type asserts that devoicing occurs at an abstract, phonological domain [6]. In this account, the status of the vowel had changed before its articulation began, and speakers decided in advance to produce a given vowel as devoiced. Some scholars have reported that devoiceable vowels before the boundary may not be devoiced. For example, in a compound /oshi#tsukeru/ (‘push' + 'against’), /i/ before the boundary and the /u/ in /shi/ can be devoiceable because both vowels are surrounded by the voiceless consonants. However, Vance [7] reported that in the two devoiceable vowels cited above, only /u/ could be devoiced. He suggested that the word boundary affects the devoicing of the preceding vowel. It is clear that word compounding belongs to a lexical process. How can the postlexical process affect the lexical process? This boundary affection and some other findings suggest that devoicing items are an entry in the lexicon.

To explore the lexical status of the devoicing vowel, a lexical decision task of natural and edited mismatched stimuli was conducted. The mismatched stimuli had *devoiced* high vowels before *voiced* consonant or *voiced* high vowels before *voiceless* consonant. The aim of this experiment is to test the following predictions: (1) if devoiced vowels are the entries of the lexicon, reaction time (RT) for mismatched stimuli may be longer than that for natural stimuli in the real word condition, and (2) because nonce words are considered not to be in the lexicon, the RT for mismatched stimuli and natural stimuli might show approximately the same value.

2. PRE-EXPERIMENT

2.1. Pre-experiment

To check the naturalness of the stimulus, a pre-experiment was carried out.

Participants: Five university students participated in the pre-experiment. All the participants were born and had grown up in Tokyo or its suburbs and studied in a university in the Tokyo area at the time

of the experiment. No hearing disorder was reported. The students participated only in this experiment.

Materials: Five pairs of words listed below.

/hutoo/ (‘wharf’) - /hudoo/ (‘stability’)
/suki/ (‘gap’) - /sugi/ (‘cedar’)
/kikoo/ (‘climate’) - /kigoo/ (‘sign’)
/supaisu/ (‘spice’) - /suraisu/ (‘slice’)
/huku/ (‘fortune’) - /hugu/ (‘puffer fish’).

The first word of each pair contains a devoicing context at the first mora (underlined vowel is devoiceable). The second word only differs in a voiced consonant at the beginning of the second mora. As a result, devoicing of the first vowel does not naturally occur in the second words. All other features, including word length and pitch accent, are the same between pairs. To control word familiarity, each pair has almost the same frequency in the Japanese corpus (Balanced Corpus of Contemporary Written Japanese [BCCWJ]) [1]. We gave priority to the familiarity of the word over other features, so the pairs have unbalanced environments such as vowel type (four out of five are /u/) and different phonemic levels (only one pair is not homorganic). To avoid the influence of these unbalanced environments, we selected the stimuli as the random factor in our statistic model. See details in 3.2.3.

Utterances of these words were produced by a Japanese female who was 21 years old at the time of recording and who was born and had grown up in Tokyo. She did not know the intent of this study and did not join the following experiments. All devoiceable vowels in her utterances were devoiced.

Using the splicing method, the first mora of the first word in each pair was exchanged with the first mora of the second word to make the mismatched stimuli. That is, the stimuli originally recorded with a fully voiced vowel was changed to a voiceless vowel and vice versa.

Procedure: A pair of words written in Japanese orthography on a computer screen presented by Praat (ver.6.0.40) [2]. Natural stimuli and mismatched stimuli were presented three times in a random order. Participants were asked to indicate which word they had heard using the mouse. If they wanted, they could hear the stimulus again (three times maximum).

2.2. Results

Based on the judgement for the second mora, we computed the accuracy of the response. The mean accuracy was as follows.

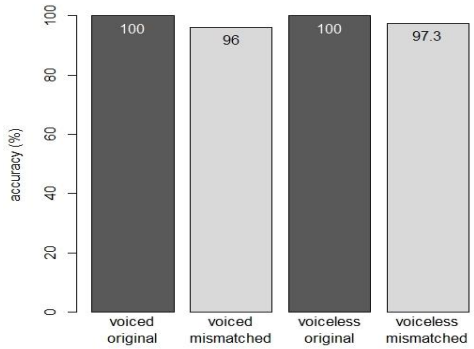


Figure 1: Accuracy of pre-experiment

No significant difference between original and mismatched stimuli was observed in Fisher’s test (voiced p=0.245, *n.s.*, voiceless p=0.497, *n.s.*).

2.3. Discussion

All the accuracy values were near 100%, and no significant difference was found in any stimulus type. Therefore, the intelligibility of the mismatched stimuli was considered as high as that of the natural stimuli.

3. LEXICAL DECISION TASK

3.1. Lexical decision task

If a devoiced vowel is a part of an entry in the mental lexicon, it is predicted that the RT of the mismatched devoiced stimuli will be longer than that of natural stimuli. If devoicing occurs in the lexical process, it is also predicted that the real/nonce status of words will affect the RT of mismatched stimuli.

Participants: Seven university students were recruited. All participants were born or had grown up in Tokyo or its suburbs. No hearing disorder was reported.

Materials: Ten nonce words were added that had a similar devoicing/voiced environment in the words checked in the pre-experiment.

real word	nonce word
/hutoo/ - /hudoo/	/hutora/ - /hudora/
/suki/ - /sugi/	/suko/ - /sugo/
/kikoo/ - /kigoo/	/kikome/ - /kigome/
/supaisu/ - /suraisu/	/supakuru/ - /surakuru/
/huku/ - /hugu/	/huko/ - /hugo/

Underlines denote devoiceable vowels. Ten devoicing/voiced environments, two lexical statuses (real or nonce word), two voicing conditions (natural or mismatch), and two positions on the display (left or right) yielded 80 stimuli for a trial. To avoid the ceiling effect, each stimulus was overlaid with white noise that had the same amplitude[1] of the stimulus.

Procedure: Each trial began with symbols for fixation (+++++) for 0.5 to 5 seconds[2]. Then, a pair of words written in Japanese orthography were displayed on a computer screen. Two seconds later, participants heard the stimulus. Participants were asked to quickly indicate which word they had heard using the left (←) or right (→) arrow keys. To motivate the participants, their RT was displayed on the screen right after the response for 1.5 seconds. PsychoPy software (ver. 1.84.2) was used to run the experiment and to record the data [4]. The trial was repeated twice[3].

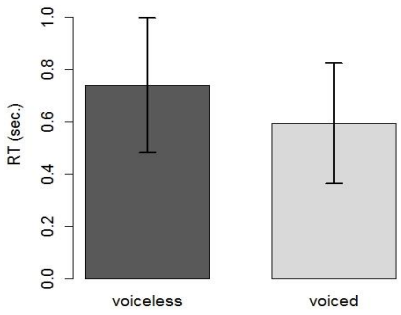
Measurement: Each stimulus varied in the duration of the first mora naturally. For example, the duration of a mora containing a devoiced vowel tended to be shorter than that of a normal mora. To avoid any confounding effect of this, the RTs were measured from the end of the first mora.

3.2. Results

3.2.1. RT (voicing)

Figure 2 shows the mean RT value sorted by the voicing of the second consonant.

Figure 2:Mean RT (voicing)

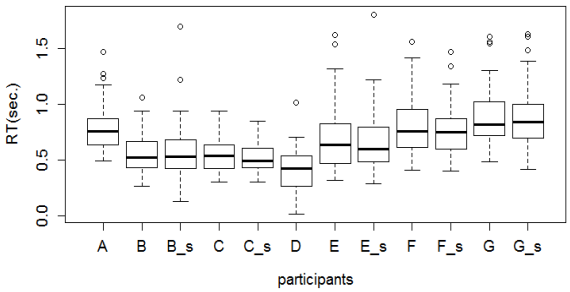


"Voiceless" and "voiced" in the figure above denote the voicing of the second consonant, respectively. The mean RT of the stimuli with voiced consonant (0.593sec.) was seemingly shorter than that of the stimuli with voiceless consonants (0.739sec.). This difference is significant according to Welch t-test (p<.001). We divided our data according to the voicing status of the second consonant thereafter.

3.2.2. Individual difference

Figure 3 shows the distribution of RTs. The alphabets under x axis denote each participant.

Figure 3: RT of each participants

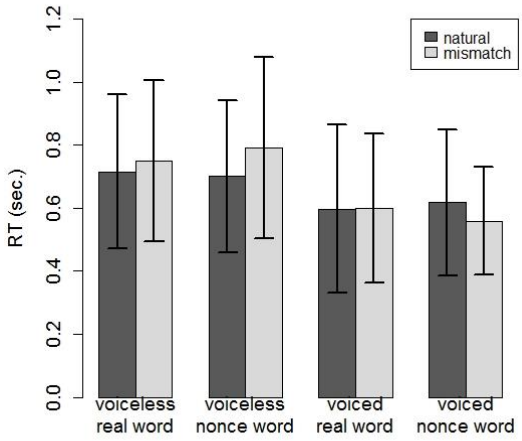


The mean RTs varied from 0.399sec. ('D') to 0.876sec. ('G') by participants. One-way ANOVA was conducted to examine the effect on participants. The results showed that the difference in participants was significant (p<.001). We treated this variable as random effect in the statistical analysis.

3.2.3. RT (summary)

Figure 4 shows the mean RT of the group. The group consisted of two criteria, voicing and matching.

Figure 4: Summary of the results



A generalized linear mixed model was used for further statistical tests. The model used RT as the dependent variable and contained the random factors PARTICIPANT and TARGET (stimulus). We also included the random slopes for REAL (real word or nonce word) by PARTICIPANT. REAL and MATCH (natural or mismatch) were included as fixed factors. Parameters for the model were estimated using Laplace approximation, implemented in the *glmer* function included in the *lme4* package for the R statistical program (version 3.4.3) [5].

Table 1: Resluts of the statistical tests
Second consonant: voiceless

Fixed effects	Estimate	SD	P

(Intercept)	1.130	0.047	<.001
REAL	0.093	0.068	0.169
MATCH	0.152	0.045	<.001
REAL:MATCH	-0.091	0.064	0.152

Second consonant: voiced

Fixed effects	Estimate	SD	P
(Intercept)	1.470	0.072	<.001
REAL	-0.043	0.100	0.661
MATCH	-0.160	0.063	<.03
REAL:MATCH	0.167	0.087	<.1

The MATCH factor had a significant effect on the RT in both voiceless and voiced conditions. In the former condition, the RT of mismatched stimuli was longer than that of the natural stimuli. On the other hand, in the latter, the effect was found to be in the reverse direction, as was evident from the negative estimate (-0.160). This was because the nonce words' RT of mismatched stimuli was shorter than that of natural stimuli, as shown in the rightmost pair of bars in Figure 4. There was no significant effect of REAL factor in both conditions. The interaction of REAL and MATCH was only weakly present in the voiced condition.

4. GENERAL DISCUSSION

The results of the lexical decision task suggest that Japanese listeners predict a voiced consonant after a fully voiced vowel. If a voiceless consonant comes after a voiced vowel, the discrepancy between the sound and their prediction may cause a slower RT. This indicates that Japanese speakers have some knowledge of devoicing. However, there is no significant effect of the REAL factor, suggesting that devoicing is not a part of the lexical information. The listeners' prediction might come from the phonetic domain.

On the other hand, when a voiced consonant comes after a devoiced vowel, we see a faster RT in mismatched stimuli of nonce words. This may suggest that the prediction by the listener is only one way from the full vowel to the following consonant. The information from a devoiced (and thus reduced) vowel is of less importance.

As for the faster RT in mismatch nonce words, further analysis of errors reveals another viewpoint. Participants sometimes hear a voiced segment as a voiceless segment or a voiceless segment as voiced segment. This type of misperception occurred in 136 responses as a whole. The error ratio of natural stimuli was 9.0%, while the error ratio of mismatched stimuli was 19.4% (see Figure 5).

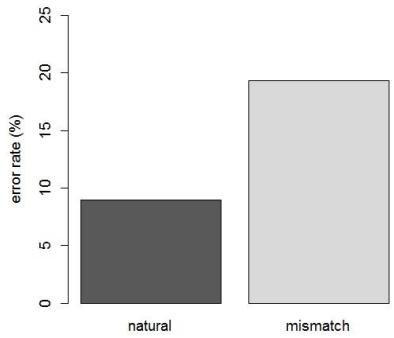


Figure 5: Error ratio

The misperception ratio for the natural stimuli was lower than that for the mismatched stimuli. We found significant differences between natural and mismatched stimuli using Fisher's test [<.0001]. There was no significant difference in the misperception ratio by the voicing condition of the second consonant [Fisher's test: p=0.397 *n.s.*]. A possible interpretation of this is that, depending on the phonetic nature of the first mora, participants anticipated the voicing feature of the second consonant. In the mismatched stimuli, however, they achieved faster RT under a time pressure by sacrificing response accuracy. In fact, two mismatched stimuli /supakuru/ and /hutora/ had shorter RTs than the natural stimuli had. However, others did not have the same tendency. Further research may be needed on this topic.

Another take on this is that voiced consonants may reset the prediction process. Although a voiceless consonant after a devoiced vowel is predicted, the voicing information in the closure region of the following consonant may be forced to reset the prediction, which would not hinder the lexical decision process.

5. SUMMARY AND CONCLUSION

In this report, we investigated the lexical status of vowel devoicing in Tokyo Japanese. The results of the lexical decision task suggest that the devoicing process belongs to the phonetic domain, and the lexicon does not hold information about devoicing.

ACKNOWLEDGEMENTS

An earlier version of this report was presented at the PSJ meeting (December 2018). We are grateful for the comments from the audience at the meeting. We also thank Michinao F. Matsui for his valuable comments. This report was supported by JSPS Grants-in-Aid (#18K00601) for the authors.

REFERENCES

[1] The Balanced Corpus of Contemporary Written Japanese. (BCCWJ) Center for corpus development, NINJAL.

[2] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer. University of Amsterdam.

[3] Fujimoto, M., 2012 Effects of consonantal environment and speech rate on vowel devoicing: An analysis of glottal opening pattern. (in Japanese) Journal of the Phonetic Society of Japan, 16 (3), 1-13

[4] Peirce, JW. 2007. PsychoPy - Psychophysics software in Python. Jounal of Neurosci Methods, 162(1-2), 8-13.

[5] R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

[6] Tsuchida, A. 1997. *Phonetics and phonology of Japanese vowel devoicing.* Doctoral dissertation. Cornell University.

[7] Vance, T. 1992. Lexical phonology and Japanese vowel devoicing. In Larson, G. et. al. (eds.) *The joy of grammar: A Festschirift in honor of James D. McCawley.*

[1] We measured the mean level of the amplitude of the whole length of the word (including pause) in root mean square.

[2] We selected a random duration from one of the set 0.5, 2, 3.5 and 5 seconds to avoid participants' anticipation.

[3] Two participants completed only one trial.