FACTORS AFFECTING UTTERANCE-FINAL VOWEL DEVOICING IN SPONTANEOUS JAPANESE

Yasuharu Den[1] and Hanae Koiso[2]

[1]Faculty of Letters, Chiba University, Japan

[2]The National Institute for Japanese Language and Linguistics
den@chiba-u.jp, koiso@ninjal.ac.jp

ABSTRACT

Investigation of spontaneous speech corpora has shown that vowel devoicing in Japanese is a statistical phenomena. However, factors behind vowel devoicing have not been fully studied. In addition, there have been no studies that specifically examined pre-pausal vowel devoicing. In this paper, we investigate vowel devoicing in the pre-pausal position, in particular, vowel devoicing occurring at the utterance-final position, using a corpus of spontaneous Japanese. We first show overall devoicing rates of morae appearing at the phrase-final, pre-pausal position, and illustrate that they are attributed to a small set of frequent lexical items. We then examine some factors that may affect the probability of vowel devoicing, focusing on some lexical items typically appearing at the utterance-final position. The results suggest that utterance-final vowel devoicing is influenced by final lengthening and the speaker's cognitive load.

Keywords: Vowel devoicing, pre-pausal position, utterance-final position, corpus-based study.

1. INTRODUCTION

In Japanese, high vowels (/i/ and /u/) are devoiced when preceded by a voiceless consonant and followed by another voiceless consonant or a silent pause, e.g., the second vowel in *asita* "tomorrow." Investigation of spontaneous speech corpora has shown that vowel devoicing is a statistical phenomena rather than a strict rule [7]. High vowels in these environments are not always devoiced, and they are also devoiced in other environments although at a lower rate. Furthermore, non-high vowels (/a/, /e/, and /o/) are sometimes devoiced.

The segmental types of the preceding and the following consonants have been found to influence the devoicing rate [7]. The devoicing rate is the highest when a high vowel is preceded by a fricative and followed by a stop, while the devoicing rate is the lowest when it is preceded by an affricate and followed by a fricative. These phonological factors, however, do not answer why vowels in the same phonological context are sometimes devoiced and sometimes not.

Very few studies have examined vowel devoicing in the pre-pausal position. Although a pause following a vowel is generally considered as constituting a devoicing environment in the same way as a voiceless consonant, empirical facts have shown that it is not the case. Based on read speech data produced by speakers of the Tottori dialect, Maekawa [5] showed that the devoicing rates of /ki/ and /si/ when followed by a pause within a sentence were only 6.1% and 10.3%, respectively; these numbers were far smaller than the devoicing rates of these morae when followed by a voiceless consonant, i.e., 61.0% and 97.6%. Investigation of a read speech corpus of Standard Japanese [8] also showed that the devoicing rate in the voiceless-consonant environment was 87.5%, whereas the devoicing rate in the pre-pausal environment was only 7.0%. There have been, however, no studies that specifically examined pre-pausal vowel devoicing in Japanese based on a statistical analysis of a spontaneous speech corpus.

In this paper, we investigate vowel devoicing in the pre-pausal position, in particular, vowel devoicing occurring at the utterance-final position, using a corpus of spontaneous Japanese. We first show overall devoicing rates of morae appearing at the phrase-final, pre-pausal position, and illustrate that they are attributed mainly to a small set of frequent lexical items. We then examine some factors that may affect the probability of vowel devoicing, focusing on some lexical items typically appearing at the utterance-final position.

2. DATA

In this study, a subset of the *Corpus of Spontaneous Japanese* (CSJ) [6] was used. The CSJ is a large-scale corpus of spontaneous Japanese, consisting mainly of monologs. Its *core* part includes hand-corrected annotations of various sorts, including clause units, *bunsetsu* phrases, words, phonetic

segments, and prosodic information. From among the CSJ-Core, the simulated public speech (SPS) subset was selected for the present analysis. The SPS subset consists of 107 casual 10- to 12-minutes narratives on everyday topics given by laypeople in front of small, friendly audiences.

In the annotations, the starting and the ending times of phonetic segments are precisely identified, which enables us to calculate durations of units at various levels such as phoneme, mora, word, phrase, and utterance. When a boundary is uncertain, it is marked as such. In addition, vowels are annotated as to whether or not they are devoiced. All the annotations are compiled into a relational database [4], and can be easily accessed via SQL queries.

3. DEVOICING RATE IN PHRASE-FINAL, PRE-PAUSAL POSITION

As a baseline, we first calculated overall devoicing rates of morae appearing at the phrase-final, pre-pausal position.

3.1. Method

In calculating devoicing rates in the corpus, it is important to consider the hierarchical nature of corpus data. That is, the entire sample is heterogeneously clustered by speakers. It is very crude to use a simple calculation based on the number of devoicing cases divided by the total number of cases, since a majority of devoicing cases may come from a few speakers. The current progress in the use of statistical models in psycholinguistics and corpus linguistics [1] can provide us a more sophisticated way.

The devoicing rate was calculated by using a logistic regression model with the devoicing status of the vowel as a dichotomous response variable and with speakers and mora types as a crossed random intercept. The estimated coefficient, β_m, for mora m was, first, given as $\beta_m = \beta_0 + \delta_m$, where β_0 and δ_m corresponded to the grand mean of the intercept and the deviation from it for mora m, respectively; and then the devoicing rate, r_m, of mora m was derived by transformation using a sigmoid function, $r_m = \frac{1}{1+\exp(-\beta_m)}$, which is an inverse-link function of the logistic regression model.

In the current analysis, only morae located at the end of a *bunsetsu* phrase and followed immediately by a silent pause were considered. When the mora involved an uncertain segment boundary or the *bunsetsu* phrase containing the mora involved a disfluency (filler, word fragment, or non-standard pronunciation), it was excluded from the analysis. Furthermore, to reduce the number of mora types and to stabilize the parameter estimation, morae whose occurrence frequencies were less than 100 were removed. Parameters of mixed-effects logistic regression models were estimated by using the `lme4` package of the R language.

Table 1: Top 5 morae with high devoicing rate (Freq. $>$ 100).

Mora	Freq.	Devoicing rate
/su/	1686	77.75%
/ta/	1255	8.00%
/to/	1268	3.34%
/si/	276	1.81%
/ku/	440	1.66%

Table 2: Frequent lexical items ending with /su/, /ta/, /to/, and /si/ (Freq. $>$ 100), and their devoicing rates.

Mora	Lexical item	Freq.	Devoicing rate
/su/	*masu* (Aux. verb)	1056	81.33%
	desu (Aux. verb)	596	79.54%
/ta/	*ta* (Aux. verb)	1167	7.66%
/to/	*to* (Case part.)	688	5.38%
	to (Conj. part.)	400	0.02%
/si/	*si* (Conj. part.)	174	1.34%

3.2. Results

Table 1 shows the top 5 morae with high devoicing rate. As pointed out in previous studies [5, 8], the devoicing rates of vowels in the pre-pausal position were, in general, not high, i.e., less than 10%; one notable exception was the devoicing rate of /su/, which was as much as 77.75%.

A closer look at the data showed that the devoicing rates of pre-pausal vowels were likely to be attributed to a small set of lexical items. Table 2 showed frequent lexical items (Freq. $>$ 100) ending with /su/, /ta/, /to/, and /si/, which were the four most frequent morae in Table 1. (There was no lexical items ending with /ku/ whose frequency was greater than 100.)

The polite marker *masu* and the polite copula *desu* occupied the vast majority of the /su/ cases (98.0%), and the devoicing rates of /su/ in these lexical items were both higher (81.33% and 79.54%, respectively) than the overall devoicing rate of /su/ (77.75%). The past tense maker *ta* constituted 93.0% of the /ta/ cases, and its devoicing rate (7.66%) was as much as that for the entire /ta/ cases (8.00%). By contrast, there were two frequent lexical items ending with

/to/, the case and quotation particle *to* and the conjunctive particle *to*, and their devoicing rates were quite different, 5.38% for the former vs. 0.02% for the latter; the overall devoicing rate of /to/ (3.34%) was between these numbers. As for /si/, the conjunctive particle *si* occupied a majority of cases (63.0%), although there were many other lexical items (such as *sukosi* "little," *sikasi* "but," and *watasi* "I") whose frequencies did not reach 100. The devoicing rate of the conjunctive particle *si* (1.34%) was comparable with that for the entire /si/ cases (1.81%), although a bit smaller.

3.3. Discussion

The devoicing rates of pre-pausal vowels were, in general, not high, i.e., less than 10%, with the exception of devoicing in /su/, which was as much as 78% (Table 1). Even in a typical devoicing environment—high vowels preceded by a voiceless consonant and followed by a pause—, the devoicing rate was usually very low (1.81% for /si/, 1.66% for /ku/, and less than 1% for the remaining cases other than /su/). By contrast, the devoicing rate was sometimes higher than these values in an atypical devoicing environment—non-high vowels preceded by a voiceless consonant and followed by a pause (8.00% for /ta/ and 3.34% for /to/). These values are, in fact, greater than the overall devoicing rates of these non-high vowels in the CSJ, reported in [7], 1.09% for /a/ and 1.28% for /o/.

These results are difficult to explain only from the phonological point of view, but it is possible that some syntactic and/or lexical factors are involved. As shown in Table 2, the high devoicing rate of /su/ and the higher-than-expected devoicing rate of /ta/ are tightly connected to the high devoicing rates of particular lexical items, i.e., *masu*, *desu*, and *ta*. These lexical items typically appear at the utterance-final position. Interestingly, this does not apply to the remaining three lexical items in Table 2, whose devoicing rates were lower than those of the above three; the case and quotation particle *to* is usually followed by a verb like *iu* "say" and *omou* "think" within an utterance, and the conjunctive particles *to* and *si* are often used to connect clauses within an utterance. Thus, the high devoicing rate observed in the pre-pausal position may be related to the finality of the utterance.

To see vowel devoicing in the utterance-final position more precisely, we next examine some factors that may affect the probability of vowel devoicing in the utterance-final position, focusing on the three lexical items, *masu*, *desu*, and *ta*.

4. FACTORS AFFECTING DEVOICING RATE IN UTTERANCE-FINAL POSITION

4.1. Method

To elucidate possible factors influencing devoicing rate in the utterance-final position, a logistic regression model with the following fixed effects was applied separately to each of the data subsets for *masu*, *desu*, and *ta*.

1. The duration of the mora (/su/ or /ta/)
2. The duration of the following pause
3. The duration of the following utterance

Firstly, it is known that devoicing rates of long vowels are far lower than those of the corresponding short vowels [7]. If this inhibition effect is also imposed by non-lexical lengthening of vowels, an increase in the duration of the mora in question would reduce the probability of devoicing the vowel in that mora. Secondly, the *duration*, not the *presence*, of a following pause has not been studied as a possible factor affecting vowel devoicing. In the current analysis, the duration of the following pause was included in our statistical model. Finally, the duration of the following utterance (clause unit in the terminology of the CSJ [6]) is an interesting factor to be investigated. Several studies have suggested that the duration of the utterance reflects the speaker's cognitive load at the time of planning that utterance and that it is exposed by such phenomena as insertion of a filler [9], lengthening of a speech segment [2], and addition of a boundary pitch movement [3]. Vowel devoicing, or non-devoicing, can be another candidate for such manifestation. All these duration variables were centered and scaled, after log-transformation, so that they could be compared in the same magnitude.

In addition to these fixed effects, a random intercept for speakers was used. Since we have already obtained, in the previous analysis, the random effects of speakers (the per-speaker deviations from the mean) that are relevant to the devoicing rates of *morae in the phrase-final, pre-pausal position*, these values were also introduced in the model. If the devoicing rate in the utterance-final position is determined merely by the overall devoicing rate of the mora and the speaker variance, only the per-speaker deviation variable would be significant, the above three fixed effects being non-significant.

4.2. Results

Vowel devoicing in *masu* Table 3 shows the estimated coefficients of the model for the *masu* data. In addition to the highly significant speaker deviation

variable, the effect of the duration of the mora was also significant. When /su/ had a longer duration, /u/ was less likely to be devoiced. The duration of the following utterance also had a nearly significant effect on the devoicing rate of *masu*. When *masu* was followed by a longer utterance, its last vowel /u/ was less likely to be devoiced. By contrast, the duration of the following pause had no significant effect.

Table 3: Estimated coefficients of the *masu* model. σ_S indicates the standard deviation of the random intercept for speakers.

	Coef.	SE	z	p
Intercept	1.67	.17	10.08	$< .001$
Speaker dev.	1.41	.12	12.03	$< .001$
Mora dur.	$-.62$	.15	-4.07	$< .001$
Fol. pause dur.	.16	.13	1.23	.219
Fol. utt. dur.	$-.20$	.10	-1.89	$< .06$
				$\sigma_S = .56$

Vowel devoicing in *desu* Table 4 shows the estimated coefficients of the model for the *desu* data. The duration of the mora had a nearly significant effect on the devoicing rate of *desu*. When /su/ had a longer duration, /u/ was less likely to be devoiced. The other two factors, the durations of the following pause and utterance, had no significant effects.

Table 4: Estimated coefficients for the *desu* model.

	Coef.	SE	z	p
Intercept	2.15	.41	5.22	$< .001$
Speaker Dev.	1.93	.31	6.32	$< .001$
Mora dur.	$-.50$	.26	-1.91	$< .06$
Fol. pause dur.	.24	.16	1.44	.149
Fol. utt. dur.	.04	.20	.18	.856
				$\sigma_S = 1.95$

Vowel devoicing in *ta* Table 5 shows the estimated coefficients of the model for the *ta* data. Both the effects of the duration of the mora and the duration of the following utterance were significant. When *ta* had a longer duration or was followed by a longer utterance, /a/ was less likely to be devoiced. The duration of the following pause, again, had no significant effect.

Table 5: Estimated coefficients for the *ta* model.

	Coef.	SE	z	p
Intercept	-3.90	.41	-9.48	$< .001$
Speaker Dev.	1.70	.20	8.44	$< .001$
Mora dur.	$-.62$	.21	-3.00	$< .003$
Fol. pause dur.	.01	.16	.03	.973
Fol. utt. dur.	$-.31$	.13	-2.45	$< .02$
				$\sigma_S = 1.42$

4.3. Discussion

In addition to the speaker deviation variable, the effect of the duration of the mora was always significant, or nearly significant, across all of the *masu*, *desu*, and *ta* models. When the last mora of the utterance had a longer duration, its last vowel was less likely to be devoiced. This is consistent with the previous observation that devoicing rates of long vowels are lower than those of the corresponding short vowels [7]. A notable difference of the current finding and the previous one, however, is that the previous finding was concerned with lexically-specified long vowels, such as the last vowel in /hoNshuH/ "the main island (of Japan)," while the current finding is concerned with non-lexical lengthening of vowels. It is widely known that utterance-final vowels are usually lengthened. Our results suggest that this final lengthening phenomenon may interact with vowel devoicing; when the degree of final lengthening is large, the final vowel of the utterance is not devoiced, even when devoicable. A typical context in which the final vowel is considerably lengthened is when it bears a boundary pitch movement such as a rising and a rising-falling intonation. In such situation, it would be likely that a devoicable final vowel is in fact not devoiced to bear a pitch movement. This also explains why the devoicing rates of some conjunctive particles were very low (*to* and *si* in Table 2); they are often accompanied by a continuous rising-falling intonation.

Another interesting finding is that the duration of the following utterance also influenced the devoicing rate of the uttterance-final vowels in *masu* and *ta*. When the following utterance became longer, the utterance-final vowel was less likely to be devoiced. As discussed in previous studies [9, 3, **?**], the duration of an utterance can be a measure of the speaker's cognitive load in planning the utterance. Our results suggest that vowel devoicing may be related to this cognitive factor; when speakers experience a heavy cognitive load, they may inhibit devoicing of utterance-final vowels. The cognitive load in planning the following utterance may be related to other variables such as the duration of the utterance-final mora and the duration of the following pause. Thus, some of the fixed effects used in our analysis may be correlated with each other. Further analyses taking this correlation into account are necessary to draw a solid conclusion.

5. REFERENCES

[1] Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

[2] Den, Y. in press. Some phonological, syntactic, and cognitive factors behind phrase-final lengthening in spontaneous Japanese: A corpus-based study. *Laboratory Phonology* 14.

[3] Koiso, H., Den, Y. 2013. Acoustic and linguistic features related to speech planning appearing at weak clause boundaries in Japanese monologs. *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* Stockholm, Sweden. 37–40.

[4] Koiso, H., Den, Y., Nishikawa, K., Maekawa, K. 2014. Design and development of an RDB version of the Corpus of Spontaneous Japanese. *Proceedings of the 9th International Conference on Language Resources and Evaluation* Reykjavik, Iceland. 1471–1476.

[5] Maekawa, K. 1989. Boin no musei-ka (in Japanese). In: Sugito, M., (ed), *Nihon-go no Onsei-On'in (1)*. Tokyo: Meiji Shoin 135–153.

[6] Maekawa, K. 2003. Corpus of Spontaneous Japanese: Its design and evaluation. *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition* Tokyo. 7–12.

[7] Maekawa, K., Kikuchi, H. 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report. In: van de Weijer, J. M., Nanjo, K., Nishihara, T., (eds), *Voicing in Japanese*. Berlin: Mouton de Gruyter 205–228.

[8] Nagano-Madsen, Y. 1994. Vowel devoicing rates in Japanese from a sentence corpus. *Lund Working Papers in Linguistics* 42, 117–127.

[9] Watanabe, M., Hirose, K., Den, Y., Miwa, S., Minematsu, N. 2006. Factors influencing ratios of filled pauses at clause boundaries in Japanese. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* Athens, Greece. 253–256.