

CORPUS-BASED ANALYSIS OF VOWEL DEVOICING IN SPONTANEOUS JAPANESE

AN INTERIM REPORT

Kikuo Maekawa[†] and Hideaki Kikuchi^{‡†}

[†]Department of Language Research, National Institute for Japanese Language

[‡]School of Human Science, Waseda University

1. INTRODUCTION

Introductory textbooks of phonetics or pronunciation dictionary of Japanese often state that close vowel (/i/ and /u/) are devoiced when they are preceded and followed both by voiceless consonants. This description turns out quickly to be incorrect when we look at real data. Close vowels are not always devoiced even in the above-mentioned environment, for one thing, and on the other hand, close vowels followed by voiced consonants can be devoiced to some extent when they are preceded by voiceless consonants. Moreover, non-close vowels like /a/ are devoiced occasionally.

These facts, which we will examine more closely in this paper, indicate that vowel devoicing is a probabilistic event: an event whose occurrence cannot be predicted with 100% accuracy. Vowel devoicing, accordingly, should be analyzed in a statistical perspective. In this perspective, phoneticians, including the first of the current authors, conducted statistical analyses of vowel devoicing in order to find out factors that determine the probability of vowel devoicing in a given phonological context.

The reported results, however, did not coincide always. For example, there is a disagreement regarding the influence of the manner of the following consonants. Han (1962) claimed that close vowel followed by an affricate or fricative was more likely to be devoiced than those followed by a plosive, but Takeda and Kuwabara (1987) obtained exactly the opposite result. The latter study also reported that one of the devoicing rules proposed in NHK (1985), namely “low-pitched mora in pre-pause position is likely to be devoiced” was almost useless in interpreting the devoicing pattern observed in a read-speech corpus.

There are several supposed-to-be reasons of disagreements. First, some descriptions of devoicing were based upon introspection. Generally speaking, introspection alone is not an appropriate analysis method for a probabilistic event like devoicing.

Second, the experimental data examined in at least some of the past studies were too small to be able to arrive at stable conclusion. This problem is likely to happen when the occurrence probability of an event is inherently very low, and/or, multiple factors and their complex interactions are involved.

Third, the data analyzed in different studies were not homogeneous with respect to the data collection method. At least three different methods were used in the previous studies: reading isolated words, reading words in a carrier sentence, and reading prose.

It is important to note, at this point, that no previous study examined devoicing in spontaneous speech. Observation of spontaneous speech is necessary because vowel devoicing may be influenced by the difference in speaking style like many other linguistic variations.

Theoretically, it is not impossible to conceive an experiment designed to solve all three problems mentioned

above, but from a practical point of view, it is virtually impossible to conduct such experiment. The cost of the experiment should be too high to be supported if the aim of the experiment is nothing but the analysis of devoicing.

Recent development of speech corpora, however, opened up a new vista for the study of vowel devoicing and other phonetic variations. Since the size and coverage of speech corpora are growing rapidly, we can use them for the study of phonetic variation. As a matter of fact, Takeda and Kuwabara (1978) and Yoshida and Sagisaka (1990) analyzed the ATR speech database developed for speech synthesis and recognition. As exemplified by these studies, use of large-scale corpora gives solution to the first two problems mentioned above.

The problem of speaking style, however, remained unsolved if the corpora contained only read speech, which is the case for most of the existing speech corpora. This last problem might be solved if there is a large corpus of spontaneous speech. In the rest of this paper, we will examine how devoiced vowels distribute in a corpus of spontaneous Japanese.

2. THE DATA

2.1 THE CSJ

The data we analyzed was an excerpt from the CSJ, or the *Corpus of Spontaneous Japanese*, that we have been developing since 1999 aiming at the public release in the spring of 2004. CSJ is a large-scale speech database designed mainly for the study of speech recognition and phonetics-linguistics (See Maekawa, Koiso, Furui and Isahara 2000 for the blue print of the CSJ).

The whole body of the CSJ contains about 7,000,000 words spoken by native speakers of so-called Standard, or Common, Japanese. This corresponds roughly to about 650 hours of speech. The main body of the corpus is monologue taken from two sources: academic presentation speech (APS) and simulated public speech (SPS).

The APS is the live recording of academic presentations done in meetings of nine different academic societies covering both humanities and natural or engineering fields. The SPS, on the other hand, is the public speech of every-day topic done by recruited layman subjects in front of a small number of audiences. The sex and age of the SPS speakers are balanced.

The speech data was recorded using head-worn directional microphone and a DAT with the sampling frequency of 48 kHz and 16-bit precision. The speech data was then down-sampled to 16,000 Hz and stored in computer.

All recorded speech were transcribed and morphologically analyzed in terms of word boundary and part-of-speech information. In addition to these research resources, there is a true subset of the corpus to which we concentrate the cost of annotation; this subset is called the Core.

The Core contains about 500,000 words or about 45 hours of speech, to all of which we provide segment and intonation label information¹. The tag set used in the segmental labeling of the Core is shown in Table 1. When this segment label information is coupled with the X-JToBI intonation labels that we developed for the CSJ (Maekawa, Kikuchi, Igarashi, and Venditti 2002), the Core can be an excellent resource for the phonetic study of spontaneous speech.

¹ We provide more research information like discourse segment label and dependency-structure label for the Core. But these information are not relevant to the current paper.

By the way, the tag set is a mixture of phonemic and sub-phonemic labels. This inconsistency was a deliberate choice of ours to enrich the value of the Core as resource of the study of phonetic variation.

Table 1 here

The segment labeling of the Core was performed in three steps. First, the initial labels were generated from the transcription text and aligned automatically to speech signal using a Hidden Markov Model based speech recognition toolkit (Young et al., 1999). The accuracy of automatic alignment in terms of phoneme boundary location, averaged over all phonemes, is currently -3.84 ms average and 21 ms standard deviation (Kikuchi and Maekawa, 2002).

Then, human labelers checked the appropriateness of the generated labels and their location on time axis. Finally, trained phoneticians checked inter-labeler Inconsistency before fixing the final labels.

During the course of manual corrections, the voicing of vowel segments was judged to be either voiced or voiceless. Informations like wide-band spectrogram, speech waveform, extracted speech fundamental frequency, peak value of the autocorrelation function, in addition to audio playback were available in this work, but the most important criteria of the judgment consisted in audio-playback and presence versus absence of speech fundamental frequency. In our speech-analysis environment, fundamental frequency was judged to be present if the probability of voicing of an analysis frame was higher than 0.5, and this probability was determined according to two-dimensional normal distribution of speech intensity and periodicity.

2.2 THE CURRENT DATA SET

Because compilation of the CSJ is on the way, we cannot use the whole body of the Core. The current data set used for the analyses reported below involved about 23 hours of segment-labeled speech containing 427,973 vowel segments.

This data set involved 29 female and 56 male speakers whose averaged age and standard deviation were 32.2 ± 5.5 and 32.3 ± 6.6 years old respectively. 65 subjects were born in Tokyo and all others were born in three surrounding prefectures of Tokyo, namely, Saitama, Kanagawa, and Chiba. From a dialectological point of view all subjects spoke so-called Standard Japanese. As for the type of monologue, 41 APS and 44 SPS monologue were involved. 6 APS and 23 SPS were by female speakers and 35 APS and 21 SPS were by male speakers. Most of these monologues lasted from 10 to 15 minutes.

During the course of transcription work, speech signal is divided into chunks by longer than 200 ms pause. This chunk we will call an ‘utterance’, but utterance in this sense may or may not correspond to syntactically meaningful construction.

Lastly, the following notation is adopted in the rest of this paper. Symbols ‘C’ and ‘V’ stand for consonant and vowels, and the latter stands usually for short vowels. ‘Co’ and ‘Cv’ stand respectively for voiceless and voiced consonants. ‘Vn’ and ‘Vnc’ stand respectively for close and non-close vowels. The combination of these symbols embraced with slashes represent phonological environment; for example, /CoVnCo/ stands for phonological environment where close vowels are preceded and followed both by voiceless consonants, while /CoVnCv/ stands

for the environment where close vowels are preceded by voiceless consonant and followed by voiced consonant. When it is necessary to make distinction of the preceding and following consonants, integers 1 and 2 are used as index. ‘C1’ and ‘C2’ stand respectively for preceding and following consonant.

3. OVERVIEW OF VOWEL VOICING

We start our analysis by giving an overview of the vowel voicing in the current data set. Table 2 tabulates the number of vowel samples and the average devoicing rate represented in percentage. Devoicing rates of long vowels (/aH/, /eH/, /iH/, /oH/, and /uH/) remained consistently the lowest. Among short vowels, close vowels showed distinctively higher devoicing rate than non-close vowels, as expected.

Table 2 here

Table 3 shows the distribution of devoicing rate as the function of the voicing of the C1 and C2 in the /C1VC2/ environment where 300,018 vowels were involved. In addition to the expected fact that devoicing rate is by far the highest in the /CoVnCo/ environment, this table reveals interesting findings about the nature of vowel devoicing.

First, devoicing rate of close vowels in the ‘typical’ /CoVnCo/ environment was not 100%. Second, close vowels were devoiced with modest probability in the /CoVnCv/ environment (17.37 and 20.91% for /i/ and /u/ respectively), too. Third, non-close vowels could be devoiced in the /CoVncCo/ environment (2.10, 3.31, and 3.45 % for /a/, /e/, and /o/ respectively). Moreover, there was no environment where devoicing was completely blocked. Vowel could be devoiced even in the /CvVnCv/ environment (i.e., non-close vowels preceded and followed by voiced consonants), which is regarded to be the most atypical environment of vowel devoicing.

To examine whether the devoicing occurring in environments other than /CoVnCo/ is phonetically the same as the devoicing in /CoVnCo/ is an interesting research question. In the next section, we will examine devoicing in three different environments, i.e., /CoVnCo/, /CoVnCv/, and /CoVncCo/.

Table 3 here

4. ANALYSIS OF VOWEL DEVOICING

4.1 THE /CoVnCo/ ENVIRONMENT

To begin with, we will analyze devoicing in the /CoVnCo/ environment. As we saw already in table 3, the devoicing rates in this ‘typical’ environment were less than 90%. So, the essential task here is to clarify the conditions that lowered the probability of vowel devoicing.

Tables 4 and 5 summarize the voicing status of /i/ and /u/ according to the phonemic classification of C1 and C2. These tables, as well as all the following tables, need some mentions. First, because C1 and C2 were phonemically classified, allophones shown in table 1 were merged into phonemes. Also, we presuppose a voiceless affricate phoneme /c/ adopting the phonemic analysis of Hattori (1950).

Second, the combinations of C1 and C2 where the total number of sample occurrence was less than 10 were

omitted from the tables. Third, all phonemically palatalized consonants were omitted altogether, because in most of the C1-C2 combinations involving the palatalized consonants, the occurrences of sample were less than 10.

Table 4 here

Table 5 here

4.1.1 Interaction of consonant manners

Tables 4 and 5 show the importance of the manners of C1 and C2 as the factors of vowel devoicing, as suggested by many previous studies (See introduction and discussion for reference). Tables 6 and 7 are summaries of tables 4 and 5 from this point of view.

Table 6 here

Table 7 here

These tables show several interesting tendencies. First, the devoicing rate was the highest when fricative C1 was followed by stop C2 in both tables, and the second highest devoicing rate was observed when fricative C1 was followed by affricate C2 in both tables. On the contrary, the devoicing rate was the lowest when affricate C1 is followed by fricative C2, and the second lowest rate was observed when fricative C1 is followed by fricative C2 in both tables. Also, it is worth noting that, in terms of the peripheral distribution, the highest devoicing rate was observed when C2 was stop, and the lowest devoicing rate was observed when C2 was fricative.

These facts show clearly that there is an interaction between the manners of C1 and C2. Two-way ANOVA between the manners of C1 and C2 applied for data pooled over /i/ and /u/ showed that main effects of C1 and C2 and their interaction were all significant (For C1, DF=2, F=44.38, P<0.0001; For C2 DF=2, F=1959.43, P<0.0001; For C1*C2, DF=4, F=263.24, P<0.0001). Phonetic interpretation of the manner interaction will be discussed in section 5.1 below.

In the calculation of tables 6 and 7, samples where C2 was a geminate /Q/ were omitted, because the manner of /Q/ per se is not specified from a phonological point of view, and, it seemed that following geminate constituted a special environment of devoicing as shown below.

Table 8 compares devoicing rates of close vowels (pooled over /i/ and /u/) in cases where C2 was and was not a geminate. This table shows that devoicing rate was lower when C2 was a geminate regardless of the C1 manner (DF=758, t=24.84, P<0.0001, unequal variance). Further analysis revealed that devoicing rate was the highest for the combination of fricative C1 and stops geminate (namely the geminate followed by stop), and was the lowest for the combination of fricative C1 and fricative geminate (namely the geminate followed by fricative). These were the same tendency as observed in tables 6 and 7.

Table 8 here

4.1.2 Consecutive devoicing

Because the initial and final consonants of the /CoVnCo/ environment are both voiceless, it happens that more than two consecutive vowels belong to this environment; when it happens it is called the environment of consecutive, or sequential, devoicing. Experimental studies showed that more than two consecutive close vowels could be devoiced in this environment (Maekawa 1990a,b).

At the same time, however, it is widely believed that there is a tendency to avoid consecutive devoicing (See Sakuma 1929 and Maekawa 1989 among many others). If this tendency does exist in spontaneous speech, it may help us to understand why devoicing rate in the /CoVnCo/ environment could not be 100%.

Although the environment of consecutive devoicing can be formed both word-internally and across word boundary, we examine only the word-internal environment in order to exclude potential influence of word boundary (Kondo, 1997).

The current data set involved 318 samples where consecutive devoicing could happen word internally. Table 9 shows the distribution of voicing status with respect to the first two vowels in the environment of consecutive devoicing. For example, if /niNsiki/ ('recognition') is followed by verb-forming suffix (i.e., *sahen* verb) of /suru/, the last two vowels of /niNsiki/ are in the environment of consecutive devoicing.

According to this table, 84 samples out of the total of 318 showed consecutive devoicing (26.4%), while in all other cases consecutive devoicing was avoided. The table also showed that the most frequent pattern of vowel voicing in this environment was the devoiced first vowel followed by voiced second vowel.

Table 9 here

Figure 1 here

Figure 1 compares devoicing rates of the first and second vowels involved in the environment of consecutive devoicing. Its abscissa represents combination of the manner of C1 and C2, and is sorted in the descending order of the observed devoicing rate of the first vowel. Letters, 'A', 'F', and 'S' stand respectively for affricate, fricative, and stop; and are combined in the order of C1/C2. This figure shows two devoicing rates were, by and large, inversely proportional, reflecting the 'one or the other' relationship between the two vowels.² The graph also shows that when fricative was combined with affricate or stop, it was always the vowel associated with the fricative that showed higher devoicing rate, and, when both consonants were fricatives, it was the second vowel that showed high devoicing rate.

4.2 THE /CoVnCv/ ENVIRONMENT

From now on, we will examine vowel devoicing in 'atypical' environments. This section deals with the /CoVnCv/ environment. Tables 10 and 11 show the devoicing rate of /i/ and /u/ as a function of the manner of C1 and C2. Two-way ANOVA between the manners of C1 and C2 applied for data pooled over /i/ and /u/ showed that main effects of C1 and C2 and their interaction were all significant (For C1, DF=2, F=440.24, P<.0001; For C2 DF=4, F=344.15, P<.0001; For C1*C2, DF=8, F=155.35, P<.0001).

² It seems that 'S/S' is aberrant from the general tendency of inverse proportion. See section 5.2 for discussion.

Table 10 here

Table 11 here

As long as C1 is concerned, the effect of consonant manner was similar to that observed in /CoVnCo/ environment in that fricatives and stops showed the highest and lowest devoicing rate respectively. As for C2, the effect of consonant manner was drastically different from that observed in /CoVnCo/ samples. The manner that showed the highest devoicing rate here was nasal. This is congruent with the results of Maekawa (1989 and 1990a).

Also, approximants, i.e., /w/ and /y/, enhanced devoicing more than stops. When C2 was approximant, the highest devoicing rate was observed for vowel /u/ preceded by fricatives.

Closer look at the data, however, revealed that this enhancing effect of approximant was the result of high devoicing rate in a few lexical items, namely, /desu/ (polite form of copula /da/) and /masu/ (an auxiliary verb of politeness). In the /CoVnCv/ samples, /desu/ was followed by sentence-ending particle /yo/ 138 times and devoiced 107 times (devoicing rate was 77.54%). Also, /masu/ was followed by particles /yo/ or /wa/ 28 times and devoiced 14 times (50% devoicing). If we remove these two lexical items from the data set, the resulting devoicing rate was 18% and is lower than that observed in the corresponding cell of table 10.

Figure 2 shows the relation between word-frequency and devoicing rate of words in the /CoVnCv/ environment. Note that individual symbol in the figure represents the averaged devoicing rate of a word. Note also that both axes are in logarithm, and, words whose frequency was lower than 10 or whose devoicing rate was 0 were excluded from analysis.

The locations of /desu/ and /masu/ in this figure were likely to be outliers of the overall trend of slight negative correlation (N=293, $r=-0.146$)³. The effect of the following approximant should be regarded, at least partly, as the consequence of word idiosyncrasy of high frequency function words.

Figure 2 here

4.3 THE /CoVncCo/ ENVIRONMENT

The last environment is /CoVncCo/, namely, non-close vowels preceded and followed both by voiceless consonants. Tables 12-14 show the devoicing rate of three non-close vowels as a function of the manner of C1 and C2.

It is difficult to extract phonetically meaningful generalization from these tables. Fricative C1 and stop C2 seem to enhance devoicing more than other manners, but the difference was not salient. As a matter of fact, three-way ANOVA of vowels (/a/, /e/, /o/), C1 manner, and C2 manner revealed that none of the main effects was significant (For vowels, DF=2, F=2.57, P>0.0766; For C1 manner, DF=2, F=1.82, P>0.1616; For C2, DF=3, F=0.64,

³ The sample locating in between /desu/ and /masu/ in figure 1 was /si/, a suffix that turn a noun or adjectival into a verb (so called *sahen* verb).

$P > 0.5890$). The C1-C2 manner interaction was not significant either ($DF=6$, $F=0.98$, $P > 0.4354$).

Table 12 here

Table 13 here

Table 14 here

In tables 12-14, devoicing rate stayed nearly the same regardless of the combination of consonant manners in these tables, and it is this very fact that characterizes the devoicing of non-close vowels. Devoicing in the /CoVncCo/ environment is special in that the manners of adjacent consonants do not play crucial role in the prediction of devoicing rates. But it does not mean that devoicing of non-close vowels was completely free from phonological conditioning. There is at least one phonological factor that influences devoicing rate of /CoVncCo/ vowels: consecutive identical morae, or, the repetition of the same mora.

Sakuma (1929) noted that in words like /kokoro/ ('mind') and /haha/ ('mother'), the vowel in the first mora could be devoiced. Table 15 summarizes the devoicing rate of the first vowels of 1260 samples that involves consecutive identical morae in the /CoVncCo/ environment. Devoicing rates of /a/ and /o/ shown in the table were higher than the overall devoicing rate shown in tables 12 and 14.

Table 15 here

In addition to this phonological conditioning, extra-linguistic factors played important role in the devoicing of /CoVncCo/ samples. First, figure 3 shows the effect of speaking rate on the devoicing of non-close vowels. In this figure, speaking rate was measured based upon the histogram of speaking rate (number of mora per second) averaged over an utterance for each speaker. Speaking rate 1 means that the average speaking rate of the utterance involving the vowel in question is within the lowest 25 % of the speaker's histogram, and, speaking rate 4 means the top 25%. With the exception of /o/, devoicing rate of non-close vowels increased monotonically as a function of the speaking rate.

Figure 3 here

Lastly, table 16 shows the effect of 'laughter' on non-close vowel devoicing. In the transcription of CSJ, a tag was given if the speaker was speaking while laughing. Although this difference was not statistically significant ($DF=27$, $t=-0.86$, $P < 0.3967$, unequal variance), devoicing rate of the in utterances involving the laughter-tag was consistently higher than in utterances without the tag.

Table 16 here

5. DISCUSSIONS

5.1 INTERPRETATION OF MANNER INTERACTION

The results of our analysis about the manner of C1 and C2 are congruent with most of the past studies. For

example, Takeda and Kuwabara (1987) reported that devoicing rate of vowels in general were higher when C1 was fricative, and the devoicing rate of the vowel in /si/ mora was highest when the mora was followed by stops. Similarly, Yoshida and Sagisaka (1990) reported that devoicing rate of close vowels preceded by voiceless consonants became the highest when they were followed by stops. However, these studies examined the effects of C1 and C2 independently, and did not pay attention for their interaction.

Recently, Yoshida (2002) and Fujimoto (submitted) examined the interaction of adjacent consonants and arrived at conclusions similar to ours. But their experiments examined only a subset of all possible manner combinations. Yoshida's experiment examined /k/ and /s/ only, and Fujimoto's examined /k, t, s/ and /h/.

Our results revealed the validity of the manner interaction in much wider phonetic context, and in more naturalistic setting. This is probably the most valuable finding of the current study.

In the analysis of /CoVnCo/ environment, we found that the interaction between the manners of C1 and C2 was statistically significant. The fact that the combinations of fricative-fricative and affricate-fricative resulted in low devoicing rate is interpreted naturally if we pay attention for the easiness of mora boundary perception.

In a CV mora whose consonant is fricative or affricate, the devoiced vowel is phonetically realized as the extension of the frication noise. So, devoicing of vowels in the above-mentioned phonetic context results in succession of frication noises, of which the first and last halves belong to different morae. Devoicing of this sort is likely to be avoided because it is difficult to perceive the boundary of two successive frication noises that corresponds to mora boundary.

Similar perceptual difficulty is likely to arise in the combinations of stop followed by fricative, also. In this combination, mora boundary is formed by the aspiration noise of the stop and the frication noise of the fricative. Perception of mora boundary in this context, however, is not as difficult as the combinations of fricative or affricate followed by fricative, because the presence of a stop can easily be perceived by the presence of its burst, and, the aspiration noise of a stop is phonetically different from frication noise with respect to its quality and quantity.

On the other hand, in the manner combinations having a stop as C2, it is relatively easy to perceive mora boundary, because the boundary is formed by acoustically salient feature, i.e., the burst of the stop. This salience is preserved when C2 is an affricate, too, because the first half of an affricate is phonetically nothing but a stop.

Lastly, the negative effect of the following geminate can be interpreted from a perceptual point of view, too. Devoicing of a vowel before a geminate requires, on the part of the listener, perception of two mora boundaries embedded within a stretch of voiceless sounds. For example, if the first vowels of /hiQsori/ ('quietly') is devoiced, listener is required to perceive the first mora boundary at the point where palatal fricative changed its color into alveolar fricative, and, the second mora boundary somewhere in the long stretch of the alveolar fricative. It is not surprising that the language has tendency to avoid such difficult perception.

5.2 CONSECUTIVE DEVOICING

The second valuable finding of the current study is the quantitative confirmation of the tendency to avoid consecutive devoicing and the role played by the combination of consecutive consonants. In 4.1.2, we noted that it was the vowels associated with fricative that showed higher devoicing rate. It is interesting, in this respect, to see that observed devoicing rates of the first vowel were, by and large, close to those observed in the /CoVnCo/

environment as in Table 17. This similarity suggests that consecutive devoicing is basically a simple process. No special forward-looking processing is needed to determine the devoicing rate of the first vowel. The devoicing rate of the second vowel, on the other hand, should involve backward reference to the voicing status of the preceding (i.e. the first) vowel.

At this point, it is important to note that the combination 'S/S' makes aberrant case both in figure 1 and table 17. Currently, we are unable to explain this exception, but it is noteworthy that the number of samples used in the analyses of consecutive devoicing is small depending on the manner combination (see figure 1). An increase in data will make it possible to decide if this is really an exception.

Lastly, the finding that sequential devoicing does play an important role in the devoicing of close vowels requires revision of the past analysis presented by the first author. Maekawa (1989 and 1990a) reported that devoicing rate of close vowels could be higher when the vowel of the following mora had non-close vowel. Although we eschew from presenting data, this tendency was observed clearly in the current data set. The tendency, however, should be interpreted, at least partly, as a by-product of the avoidance of consecutive devoicing. That a close vowel has non-close following vowel means automatically that the vowel in question is not in the environment of sequential devoicing, hence the devoicing rate of the vowel is expected to be higher than elsewhere.

Table 17 here

5.3 ATYPICAL ENVIRONMENTS

The third contribution of this study is the observation of devoicing in atypical environments, namely, the /CoVnCv/ and /CoVnCo/ samples. Our analyses suggested that devoicing of /CoVnCv/ close vowels were similar to that of /CoVnCo/ close vowels in that they were deeply conditioned by the manner of adjacent consonants. Although the influence of C2 was quite different depending on the voicing of C2, it seemed that they constituted one large class of vowel devoicing. Devoicing of non-close vowels, on the other hand, was a radically different phenomenon from close vowel devoicing in that manners of adjacent consonants had almost no influence on devoicing rate.

With respect to the influence of extra-linguistic factors that we presented in the analysis of non-close vowels, it is worth noting that both speaking rate and laughter showed exactly the same influence upon the devoicing of close vowels. Devoicing rate of close vowels increased monotonically as a function of speaking rate without exception, and, vowels uttered with laughter showed higher devoicing rate than those uttered without laughter.

Effect of speaking rate upon devoicing rate was repeatedly confirmed in the studies like Maekawa (1990a) and Kondo (1997), and confirmed here for spontaneous speech data.

Recent studies of linguistic variations recorded in CSJ revealed that the presence of laughter was an excellent indicator of speaker's relaxedness and the resulting low speaking style. Probably, vowels are more likely to be devoiced in low speaking style than high speaking style where speakers pay more attention for their speech. This view is congruent with the finding of Imaizumi, Hayashi and Deguchi (1995) that close vowel devoicing is less prominent when schoolteachers spoke to hearing-impaired pupils than they spoke to normal hearings.

In the current data, as a matter of fact, the averaged devoicing rates of SPS type samples were significantly higher than that of APS type samples as shown in table 18. According to the two-way ANOVA between environment and speech type, both main effects were significant and the interaction was not significant (Environment: DF=2, F=536000.9, P<0.0001; Speech type: DF=1, F=39.32, P<.0001; Environment*Speech type: DF=2, F=2.95, P<0.0524).

Table 18 here

7. CONCLUDING REMARKS

Use of spontaneous speech corpus revealed its effectiveness in the analysis of vowel devoicing. The data presented here is one of the most reliable resources for the study of vowel voicing both in quality and quantity. The full coverage of the C1-C2 manner combination would have been impossible if the amount of data was substantially smaller than the current data set. Needless to say, however, the current data set is not ample enough for statistically complex phenomenon like consecutive devoicing analyzed in section 4.1.2. More reliable conclusion will be achieved once we have access to the whole of the CSJ-Core whose data size is more than twice of the current data.

Most of the analyses done in this paper are linguistic analysis in the sense that phonological conditions were used as the factors of vowel devoicing. Yet, as suggested in the analysis of non-close vowel devoicing, it is obvious that extra-linguistic factors played certain role. Extensive analyses of extra-linguistic factors and the integration of linguistic and extra-linguistic factors is an important step towards full understanding of vowel devoicing phenomenon. Lastly, intonation label information of the CSJ-Core will make it possible to examine the effect of prosodic conditionings like pitch accent. All these analyses should be the theme of future study.

ACKNOWLEDGMENT

The authors are grateful for all speakers of the CSJ. Our gratitude also goes to Professor Hisao Kuwabara of Tekyo Science University who sent us his paper upon our request.

REFERENCES

- Fujimoto, Masako. and Shigeru. Kiritani (Submitted). Comparison of vowel devoicing for speakers of Tokyo- and Kinki dialects. (Submitted to the *Journal of Phonetic Society of Japan*).
- Han, Mieko, S. (1962). Unvoicing of vowels in Japanese. *Study of Sounds*, 10, 81-100.
- Hattori, Shiro (1950). Phoneme, phone, and compound phone. *Gengo Kenkyu*, 16, 92-109 (Revised version is in S. Hattori. *Gengogaku no houhou*. Tokyo; Iwanami, 1960).
- Imaizumi, Satoshi, A. Hayashi, and T. Deguchi (1995). Listener adaptive characteristics of vowel devoicing in Japanese Dialogue. *J. Acoust. Soc. Amer.* 98(2), 768-778.
- Kikuchi, Hideaki and Kikuo Maekawa (2002). Accuracy of automatic phoneme labeling on spontaneous speech. *Proceedings of the 2002 Spring meeting of the Acoustical Society of Japan*, 97-98.

- Kondo, Mariko (1997). *Mechanism of vowel devoicing in Japanese*. Ph.D. Thesis. Faculty of Arts, the University of Edinburgh.
- Maekawa, Kikuo (1989). Boin no musei-ka. In Sugito, M. (Ed.) *Nihon-go no Onsei-On'in* (1), 135-153, Meiji Shoin.
- (1990a). Effects of speaking rate on the voicing variation in Japanese. *Technical Report of the Institute of Electronics, Information and Communication Engineers* (SP89-148), 47-53.
- (1990b). Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese. *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, 2, 517-520, Kobe.
- (2002a). Hanashikotobani okeru chouboinno tanko. *Kokugogakkai 2002 nendo syunkitakai youshisyuu*, 43-50.
- (2002b). Study of language variation using Corpus of Spontaneous Japanese. *Journal of Phonetic Society of Japan*, 6(3),???
- Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara (2000). Spontaneous speech corpus of Japanese. *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC)*, 2, 947-952.
- Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer Venditti (2002). X-JToBI: An extended J_ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*, 3, 1545-1548, Denver.
- NHK (1985). *Nihongo hatsuon akusento jiten*. Tokyo: Nihon Housou Kyoukai.
- Sakuma, Kanae (1929). *Nihon Onseigaku*. Tokyo: Kyobunsha.
- Takeda, Kazuya and Hisao Kuwabara (1987). Boin museika no youin bunseki to yosoku syuhou no kentou. *Proceedings of 1987 Autumn Meetings of the Acoustical Society of Japan*, 1, 105-106.
- Yoshida, Natsuya (2002). The effect of phonetic environment on vowel devoicing in Japanese. *Kokugogaku*, 53 (3), 34-47.
- Yoshida, Natsuya and Yoshinori Sagisaka (1990). Boin museika no youin bunseki. *Technical Report of ATR Interpreting Telephony Research Laboratories* (TR-I-0159).
- Young, Steve, J. Jansen, J. Odell, D. Ollason and P. Woodland (1999). *The HTK Handbook*. Entropic Research Laboratories.