Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report

Kikuo Maekawa and Hideaki Kikuchi

1. Introduction

Introductory textbooks of phonetics or pronunciation dictionaries of Japanese often state that close vowels (/i/ and /u/) are devoiced when they are both preceded and followed by voiceless consonants. This description turns out quickly to be incorrect when we look at real data. For one thing, close vowels are not always devoiced, even in the above-mentioned environment, and in addition, 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 also 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 from a statistical perspective. In this perspective, phoneticians, including the first author of this paper, have in the past conducted statistical analyses of vowel devoicing in order to find out which factors determine the probability of vowel devoicing in a given phonological context.

The reported results, however, have not always coincided. For example, there is disagreement regarding the influence of the manner of articulation of the following consonant. Han (1962) claimed that close vowels followed by an affricate or fricative were 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 a "low-pitched mora in pre-pause position is likely to be devoiced", was almost useless in interpreting the devoicing patterns observed in a read-speech corpus.

There may be several possible reasons for such 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 previous 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 of isolated words, reading of words in a carrier sentence, and reading of 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 differences in speaking style, as is the case with 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 an experiment. The cost of the experiment would 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, has 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. In fact, Takeda and Kuwabara (1978) and Yoshida and Sagisaka (1990) have analyzed the ATR speech database developed for speech synthesis and recognition, and have shown that the use of large-scale corpora provide a solution to the first of the two problems mentioned above.

The problem of speaking style, however, has so far remained unsolved since most existing corpora contain only read speech. This last problem might be solved by a large corpus of spontaneous speech. In the rest of this paper, we will examine the distribution of devoiced vowels in a corpus of spontaneous Japanese.

2. The data

2.1. The Corpus of Spontaneous Japanese (CSJ)

The data we analyzed is an excerpt from the *Corpus of Spontaneous Japanese* (henceforth 'CSJ'), which we have been developing since 1999, aiming for 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 blueprint of the CSJ).

The whole body of the CSJ contains about 7.5 million words spoken by native speakers of so-called Standard, or Common, Japanese. This corresponds roughly to about 660 hours of speech. The main body of the corpus is monologue taken from two sources: academic presentation speech (APS) and simulated public speaking (SPS).

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

The speech data was recorded using a 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 kHz and stored in computer.

All recorded speech was transcribed and morphologically analyzed in terms of word boundary and part-of-speech information. In addition to this tagging of the entire corpus, we have done extensive annotation of a number of linguistic features to a subset of the corpus; we call this subset 'the Core'.

The Core contains about 500,000 words or about 45 hours of speech, all of which have been (sub-)phonemically segmented and labeled for intonation.[1] The tag set used in the segmental labeling of the Core is shown in Table 1. 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 for the study of phonetic variation. 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.

The segment labeling of the Core was preformed in three steps. First, the initial labels were generated from the transcription text and aligned automatically to the 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).

Table 1. Label set used for the segmental labeling of the CSJ

Vowels:	a, i, u, e, o (voiced)
	A, I, U, E, O (devoiced)
Plain Consonants:	
	k, g, G[ɤ], @[ŋ], s, z, t, c[ts], d, n, h, F, b, p, m, r[ɾ], w, y
Phonetically palatalized consonants:	
	kj, gj, Gj, @j, sj[ʃ], zj[ʒ], cj[tʃ], nj[ɲ], hj[ç]
Phonologically palatalized consonants (*'youon'*):	
	ky, gy, Gy, @y, sy, zy, cy, ny, hy, by, py, my, ry
Moraic phonemes:	
Long vowel:	H
Geminate (*'sokuon'*):	Q
Moraic nasal (*'hatsuon*):	N

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

During the course of manual corrections, the voicing of vowel segments was judged to be either voiced or voiceless. Information from the wide-band spectrogram, speech waveform, extracted speech fundamental frequency, peak value of the autocorrelation function, in addition to audio playback were all available for these judgments, but the most important criteria was the audio playback and presence versus absence of the 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 a two-dimensional normal distribution of speech intensity and periodicity.

2.2. The current data set

Because compilation of the CSJ is currently underway (as of february 2003), we are not able to use the whole body of the Core. The data set used for the

analyses reported below consists of about 23 hours of segment-labeled speech containing 427,973 vowel segments.

This data set contains 29 female and 56 male speakers whose average age and standard deviation were 32.2±5.5 and 32.3±6.6 years old, respectively. Sixty five 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 monologues are present in our data set. Six APS and 23 SPS monologues are by female speakers and 35 APS and 21 SPS monologues are by male speakers. Most of these monologues lasted from 10 to 15 minutes.

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

Lastly, the following notation is adopted in the rest of this paper. Symbols 'C' and 'V' stand for consonants and (short) vowels. 'Co' and 'Cv' stand respectively for voiceless and voiced consonants. 'Vc' and 'Vnc' stand respectively for close and non-close vowels. The combination of these symbols placed within forward slashes represents the phonological environment; for example, /CoVcCo/ stands for the phonological environment in which close vowels are both preceded and followed by voiceless consonants, while /CoVcCv/ stands for the environment in which close vowels are preceded by a voiceless consonant and followed by a voiced consonant. When it is necessary to make a distinction between the preceding and following consonant, integers 1 and 2 are used as an index: 'C1' and 'C2' stand for preceding and following consonant, respectively.

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 as a 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. Number of samples and averaged devoicing rate of all vowel segments

VOWEL	N	VOICED	DEVOICED	% DEVOICED
a	109,624	108,432	1,192	1.09
aH	3,956	3,954	2	0.05
e	58,154	57,401	753	1.29
eH	12,363	12,361	2	0.02
i	75,581	60,675	14,906	19.72
iH	2,650	2,646	4	0.15
o	88,412	87,282	1,130	1.28
oH	19,445	19,437	8	0.04
u	49,448	33,917	15,531	31.41
uH	8,340	8,307	33	0.40

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

First, the devoicing rate of close vowels in the 'typical' /CoVcCo/ environment was not 100%. Second, close vowels were also devoiced with modest probability in the /CoVcCv/ environment (17.37% and 20.91% for /i/ and /u/, respectively). Third, non-close vowels also 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 in which devoicing was completely blocked. Vowels could be devoiced even in the /CvVncCv/ environment (i.e., non-close vowels preceded and followed by voiced consonants), which is regarded to be the most atypical environment for vowel devoicing. Similar findings were reported earlier in Venditti and van Santen (1998).

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

Table 3. Devoicing in the /C1VC2/ environment as a function of the voicing of C1 and C2

VOWEL	C1	C2	VOICED	DEVOICED	% DEVOICED
a	Co	Co	12,214	262	2.10
	Co	Cv	18,570	92	0.49
	Cv	Co	24,943	481	1.89
	Cv	Cv	19,867	29	0.15
e	Co	Co	5,550	190	3.31
	Co	Cv	10,890	116	1.05
	Cv	Co	11,552	323	2.72
	Cv	Cv	11,388	29	0.25
i	Co	Co	1,475	12,124	89.15
	Co	Cv	10,556	2,219	17.37
	Cv	Co	9,200	126	1.35
	Cv	Cv	12,072	133	1.09
o	Co	Co	12,247	437	3.45
	Co	Cv	19,752	365	1.81
	Cv	Co	14,650	13	0.09
	Cv	Cv	16,802	14	0.08
u	Co	Co	1,732	9,267	84.25
	Co	Cv	11,851	3,133	20.91
	Cv	Co	5,562	127	2.23
	Cv	Cv	7,748	61	0.78

4. Analysis of vowel devoicing

4.1. The /CoVcCo/ environment

We will first analyze devoicing in the /CoVcCo/ 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 identify the conditions that decrease the probability of vowel devoicing in this context.

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 introduction. First, because C1 and C2 were phonemically classified, allophones shown in Table 1 were merged into phonemes. Also, we presuppose a voiceless (dental) affricate phoneme /c/ adopting the phonemic analysis of Hattori (1950).

Second, the combinations of C1 and C2 where the total number of samples 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 number of samples was less than 10.

Table 4. Cross-tabulation of the voicing of /i/ in the /CoVcCo/ environment by C1 and C2

VOWEL	C1	C2	VOICED	DEVOICED	% DEVOICED
i	c	c	16	73	82.02
		h	35	7	16.67
		k	31	358	92.03
		p	7	44	86.27
		Q	16	16	50.00
		s	64	41	39.05
		t	32	181	84.98
	h	c	5	80	94.12
		h	22	9	29.03
		k	15	342	95.80
		Q	21	39	65.00
		s	11	3	21.43
		t	21	883	97.68
	k	c	19	62	76.54
		h	167	65	28.02
		k	73	476	86.70
		Q	32	51	61.45
		s	144	262	64.53
		t	53	791	93.72
	p	Q	118	9	7.09
	s	c	7	259	97.37
		h	47	14	22.95
		k	50	1,102	95.66
		Q	25	92	78.63
		s	259	178	40.73
		t	49	6,507	99.25
	t	k	11	0	0.00
		Q	13	0	0.00

Table 5. Cross-tabulation of the voicing of /u/ in the /CoVcCo/ environment by C1 and C2

VOWEL	C1	C2	VOICED	DEVOICED	% DEVOICED
u	c	c	16	57	78.08
		h	24	10	29.41
		k	44	872	95.20
		Q	13	32	71.11
		s	137	140	50.54
		t	19	207	91.59
	h	c	4	86	95.56
		h	17	16	48.48
		k	15	227	93.80
		Q	25	7	21.88
		s	6	46	88.46
		t	10	248	96.12
	k	c	48	123	71.93
		h	132	56	29.79
		k	151	246	61.96
		p	3	21	87.50
		Q	114	26	18.57
		s	380	1,202	75.98
		t	148	1,021	87.34
	p	k	8	7	46.67
		s	12	18	60.00
		t	6	12	66.67
	s	c	3	8	72.73
		h	4	8	66.67
		k	31	2,207	98.61
		p	2	154	98.72
		Q	23	31	57.41
		s	60	195	76.47
		t	37	1,210	97.03

4.1.1. Interaction of consonant manners

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

Table 6. Devoicing rate [%] of /i/ in the /CoVcCo/ environment classified by the manner of C1 and C2

		C2			
		Affricate	Fricative	Stop	
C1	Affricate	81.1	33.3	89.4	78.3
	Fricative	96.3	38.1	98.4	94.6
	Stop	80.2	51.5	89.3	77.3
		91.0	43.8	47.7	

Table 7. Devoicing rate [%] of /u/ in the /CoVcCo/ environment classified by the manner of C1 and C2

		C2			
		Affricate	Fricative	Stop	
C1	Affricate	77.2	48.1	94.5	83.6
	Fricative	95.1	61.2	97.5	93.5
	Stop	80.8	74.0	80.1	77.1
		84.4	68.8	35.9	

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. In contrast, 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 articulation of C1 and C2. A two-way ANOVA between the manners of

C1 and C2 applied to 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 in which 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 a 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 the devoicing rate was lower when C2 was a geminate, regardless of the manner of C1 (DF=758, t=24.84, P<0.0001, unequal variance). Further analysis revealed that the devoicing rate was the highest for the combination of fricative C1 and a stop geminate (namely a geminate followed by a stop), and was the lowest for the combination of fricative C1 and a fricative geminate (namely a geminate followed by a fricative). These show the same tendency as observed in Tables 6 and 7.

Table 8. Effect of the following geminate on devoicing rate: Pooled data of /i/ and /u/

	C2 non /Q/			C2 /Q/		
C1	Voiced	Devoiced	% Devoiced	Voiced	Devoiced	% Devoiced
Affricate	454	2,021	81.7	29	49	62.8
Fricative	860	14,099	94.3	112	181	61.8
Stop	1,464	4,954	77.2	282	87	23.6

4.1.2. Consecutive devoicing

Because the initial and final consonants in the /CoVcCo/ environment are both voiceless (and due to the common CV syllable structure of Japanese), it happens that more than two consecutive vowels can belong to this environment (e.g. ...(CoVc)CoVcCo(VcCo)...) When this happens it is called consecutive, or sequential, devoicing. Experimental studies have shown that more than two consecutive close vowels can 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 the devoicing rate in the canonical /CoVcCo/ environment was not 100% in our data.

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

The current data set contains 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) /suru/, the last two vowels of /niNsiki/ are in the consecutive devoicing environment.

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

Table 9. Voicing of the first two vowels in the environment of consecutive devoicing

		SECOND VOWEL	
		VOICED	DEVOICED
FIRST VOWEL	VOICED	17	44
	DEVOICED	171	84

Figure 1 compares devoicing rates of the first and second vowels in the consecutive devoicing environment. Its abscissa represents the 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 that the two devoicing rates were, by and large, inversely proportional, reflecting a 'one or the other' relationship between the two vowels.[2] The graph also shows that when a fricative was combined with an affricate or stop, it was always the vowel associated with (i.e., in the same mora as) the fricative that showed the higher devoicing rate, and, when both consonants were fricatives, it was the second vowel that showed a high devoicing rate.

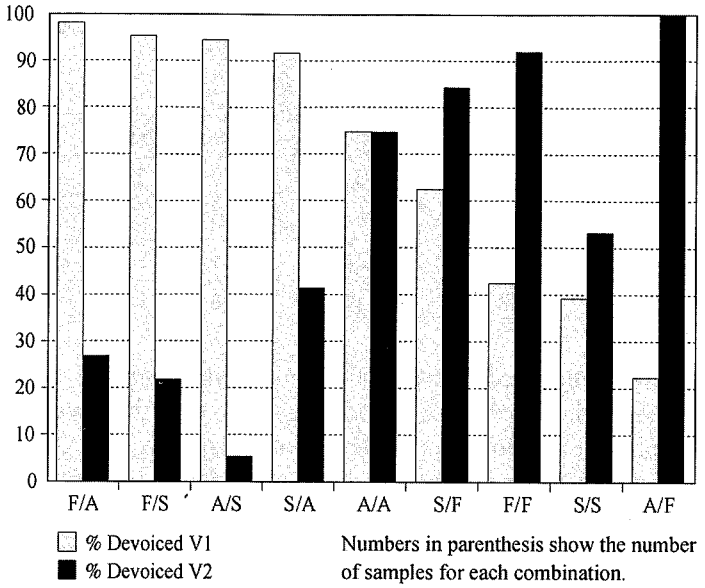

Figure 1. Devoicing rate of two vowels in the environment of consecutive devoicing

4.2. The /CoVcCv/ environment

From this point on, we will examine vowel devoicing in 'atypical' environments. This section deals with the /CoVcCv/ environment. Tables 10 and 11 show the devoicing rate of /i/ and /u/ as a function of the manner of C1 and C2. A two-way ANOVA between the manners of C1 and C2 applied to 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).

Table 10. Devoicing rate [%] of /i/ in the /CoVcCv/ environment classified by the manner of C1 and C2

		C2					
		Approximant	Fricative	Liquid	Nasal	Stop	
C1	Affricate	12.8	9.7	8.4	18.2	9.4	12.5
	Fricative	20.2	8.7	10.4	38.4	10.2	28.3
	Stop	5.8	5.7	2.0	9.6	5.9	7.8
		14.0	8.2	7.1	24.4	8.6	

Table 11. Devoicing rate [%] of /u/ in the /CoVcCv/ environment classified by the manner of C1 and C2

		C2					
		Approximant	Fricative	Liquid	Nasal	Stop	
C1	Affricate	13.8	20.6	6.9	28.3	12.9	19.8
	Fricative	46.5	16.7	5.5	65.6	22.1	36.8
	Stop	4.7	2.6	3.1	3.8	5.0	3.9
		18.4	12.6	4.6	35.8	15.9	

As far as C1 is concerned, the effect of consonant manner was similar to that observed in the /CoVcCo/ 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 the /CoVcCo/ samples. The manner of articulation 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 did. The highest devoicing rate of all was observed for vowel /u/ preceded by a fricative and followed by an approximant.A closer look at the data, however, revealed that this enhancing effect of an approximant was the result of a high devoicing rate in only a few lexical items, namely, /desu/ (polite form of copula /da/) and /masu/ (an auxiliary verb of politeness). In the /CoVcCv/ samples, /desu/ was followed by sentence-ending particle /yo/ 138 times and devoiced 107 times (the 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 only 18%, and is lower than the 46.5% reported in the C1 fricative/C2 approximant cell of Table 10.

Figure 2 shows the relation between word-frequency and devoicing rate of words in the /CoVcCv/ environment. Note that individual symbols in the figure represent the averaged devoicing rate of a given word. Note also that both axes are plotted on a logarithmic scale, and, words whose frequency was lower than 10 or whose devoicing rate was 0 were excluded from the analysis.

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

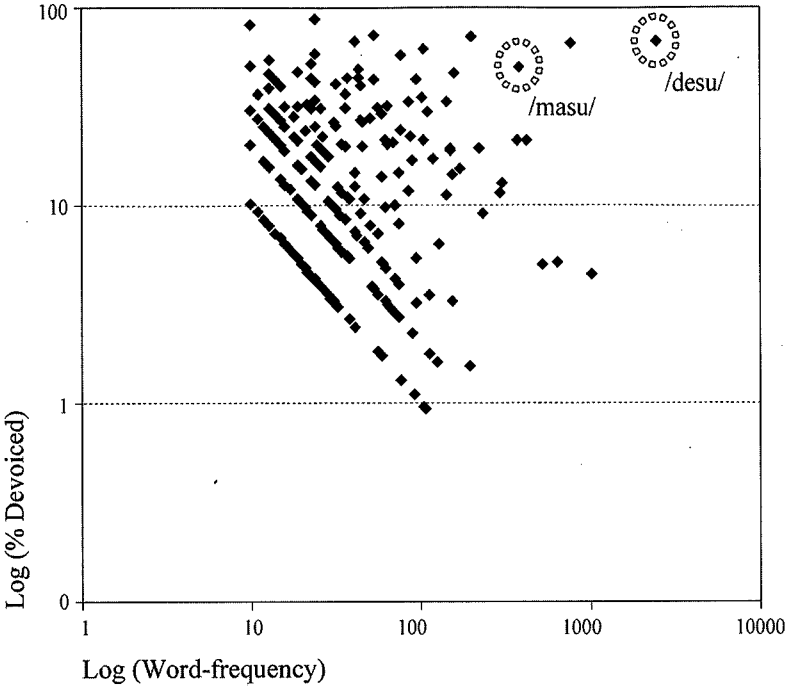

Figure 2. Word-frequency and devoicing rate in the /CoVcCv/ environment

4.3. The /CoVncCo/ environment

The last environment we will examine is /CoVncCo/, namely, non-close vowels both preceded and followed 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 any phonetically meaningful generalizations from these tables. Fricative C1 and stop C2 seem to enhance devoicing more than other manners, but the difference was not salient. Indeed, a three-way ANOVA of vowels (/a/, /e/, /o/), C1 manner, and C2 manner revealed that none of the main effects were 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, P>0.5890). The C1-C2 manner interaction was not significant either (DF=6, F=0.98, P>0.4354).

Table 12. Devoicing rate [%] of /a/ in the /CoVncCo/ environment by the manner of C1 and C2

		C2								
		Affricate		Fricative		Stop		Geminate		
C1	Affricate	–	(1)	–	(7)	0.0	(47)	1.3	(79)	0.7
	Fricative	3.1	(389)	5.0	(714)	1.5	(1206)	1.3	(316)	2.7
	Stop	0.8	(880)	1.1	(2537)	2.8	(5162)	1.1	(1134)	2.0
		1.5		1.9		2.5		1.1		

Numbers in parenthesis show the number of samples for each combination.

Table 13. Devoicing rate [%] of /e/ in the /CoVncCo/ environment by the manner of C1 and C2

		C2								
		Affricate		Fricative		Stop		Geminate		
C1	Affricate	–	(1)	–	(0)	–	(0)	–	(7)	0.0
	Fricative	0.9	(333)	2.2	(45)	7.5	(388)	0.0	(132)	3.7
	Stop	0.6	(176)	4.4	(1,083)	3.4	(2,925)	1.2	(650)	3.2
		0.8		4.3		3.9		1.0		

Numbers in parenthesis show the number of samples for each combination.

Table 14. Devoicing rate [%] of /o/ in the /CoVncCo/ environment by the manner of C1 and C2

		C2								
		Affricate		Fricative		Stop		Geminate		
C1	Affricate	–	(3)	–	(3)	1.7	(59)	4.0	(455)	3.8
	Fricative	2.6	(76)	4.1	(410)	4.3	(1,205)	1.7	(119)	4.0
	Stop	2.6	(721)	2.1	(2,070)	3.8	(7,208)	1.1	(355)	3.3
		2.8		2.5		3.9		2.6		

Numbers in parentheses show the number of samples for each combination.

In Tables 12–14, the devoicing rate stayed nearly the same regardless of the combination of consonant manners, 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 a crucial role in the prediction of devoicing rates. But this does not mean that devoicing of non-close vowels was completely free from phonological con-

ditioning. There is at least one phonological factor that influences the 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 contain 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. Devoicing of the first vowel of two identical morae in the /CoVncCo/ environment

VOWEL	VOICED	DEVOICED	% DEVOICED
a	458	54	10.5
e	112	5	4.3
o	490	141	22.3

In addition to this phonological conditioning, extra-linguistic factors played an important role in the devoicing of /CoVncCo/ samples. First, Figure 3 shows the effect of speaking rate on the devoicing of non-close vowels. The speaking rate of a given speaker's utterance was taken to be the number of mora per second, averaged over the entire utterance. A histogram of speaking rates was plotted for each speaker, and was divided into 4 intervals for purposes of the current analysis. In the figure, speaking rate 1 means that the average speaking rate of the utterance containing 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/, the devoicing rate of non-close vowels increased monotonically as a function of the speaking rate.

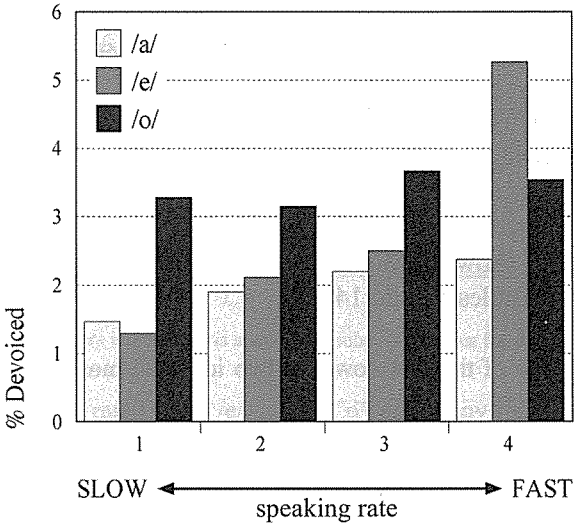

Figure 3. Effect of speaking rate on devoicing rate in the /CoVncCo/ environment

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), the devoicing rate of non-close vowels in utterances containing the laughter-tag was consistently higher than in utterances without the tag.

Table 16. Devoicing rate in the /CoVncCo/ environment as a function of laughter

VOWEL	LAUGHTER	N	VOICED	DEVOICED	% DEVOICED
a	0	12,184	11,940	244	2.00
	1	292	274	18	6.16
e	0	5,616	5,434	182	3.24
	1	124	116	8	6.45
o	0	12,396	11,977	419	3.38
	1	288	270	18	6.25

5. Discussion

5.1. Interpretation of manner interaction

The results of our analysis about the manner of C1 and C2 are congruent with most past studies. For example, Takeda and Kuwabara (1987) reported that the devoicing rate of vowels in general was higher when C1 was a fricative, and the devoicing rate of the vowel in the /si/ mora was highest when the mora was followed by a stop. Similarly, Yoshida and Sagisaka (1990) reported that the 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 to their interaction.

Recently, N. Yoshida (2002) and Fujimoto (2003) examined the interaction of adjacent consonants and arrived at conclusions similar to ours. However, 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 reveal the validity of the manner interaction in much wider phonetic context, and in a more naturalistic setting, namely, in spontaneous speech. This is probably the most valuable finding of the current study.

In our analysis of the /CoVcCo/ 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 a low devoicing rate is interpreted naturally if we think about the ease of mora boundary perception. In a CV mora whose consonant is a 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 (that is, Co[fric/affric]-Vc-Co[fric]) results in the succession of frication noise, 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 mora boundary within this extended frication.

Similar perceptual difficulty is also likely to arise when a devoiced vowel is preceded by a stop and followed by a fricative. In this combination, the mora boundary occurs between the aspiration noise of the stop and the frication noise of the fricative. Perception of a mora boundary in this context, however, is not as difficult as the combination of a fricative/affricate followed by a 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 a mora boundary, because the boundary is formed by an acoustically salient feature, i.e., the burst of the stop. This salience is also preserved when C2 is an affricate, since the first half of an affricate is phonetically nothing but a stop.

Lastly, the negative effect on devoicing of a following geminate can also be interpreted from a perceptual point of view. 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 vowel of /hiQsori/ ('quietly') is devoiced, the listener is required to perceive the first mora boundary at the point where palatal fricative (the conditional variant of /h/ before /i/) changes its color into a alveolar fricative, and, the second mora boundary somewhere within the long stretch of the alveolar fricative. It is not surprising that the language has a tendency to avoid such a difficult perceptual combination.

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 Section 4.1.2, we noted that it was vowels associated with (i.e., in the same mora as) fricatives that showed higher devoicing rates. It is interesting, in this respect, to see that the observed devoicing rates of the first vowel in a consecutive devoicing environment were, by and large, close to those observed in the /CoVcCo/ environment, as summarized 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, involves 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' in an exception in both Figure 1 and Table 17. The devoicing rate for this combination in the consecutive devoicing environment is low, yet the rate in the canonical /CoVcCo/ environment is high. 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 for many of the manner combinations (see Figure 1). An increase in data will make it possible to decide if this case is really an exception.

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

Table 17. Comparison of the devoicing rate of the first vowel in a consecutive devoicing environment with that of the vowel in the /CoVcCo/ environment, pooled over /i/ and /u/

MANNER	V1	/CoVcCo/
F/A	98.3	95.9
F/S	95.3	98.1
A/S	94.4	92.6
S/A	91.6	80.7
A/A	75.0	79.3
S/F	62.5	68.6
F/F	42.3	48.9
S/S	39.2	84.4
A/F	22.2	43.1

5.3. Atypical environments

The third contribution of this study is the observation of devoicing in atypical environments, namely, in /CoVcCv/ and /CoVncCo/ environments. Our analyses suggest that the devoicing of /CoVcCv/ close vowels were similar to that of /CoVcCo/ close vowels in that they were deeply conditioned by

the manner of articulation of adjacent consonants. Although the influence of C2 was quite different depending on the voicing of C2, it seems that these environments constitute 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 the 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. The devoicing rate of close vowels increased monotonically as a function of speaking rate without exception, and, vowels uttered with laughter showed higher a devoicing rate than those uttered without laughter.

The effect of speaking rate on the devoicing rate has been repeatedly confirmed in previous studies such as Maekawa (1990a) and Kondo (1997), and has been confirmed here for spontaneous speech data.

Recent studies of linguistic variations recorded in CSJ have revealed that the presence of laughter was an excellent indicator of the speaker's relaxation, resulting in a casual speaking style. Perhaps vowels are more likely to be devoiced in a casual speaking style than in a more formal speaking style in which speakers pay more attention to their speech. This view is consistent with the finding of Imaizumi, Hayashi and Deguchi (1995) that close vowel devoicing is less prominent when school teachers spoke to hearing-impaired pupils than when they spoke to normal hearing pupils.

In the current data, as a matter of fact, the average devoicing rates in SPS (simulated public speaking) samples were significantly higher than that in APS (academic presentation speech) samples, as shown in Table 18. According to a two-way ANOVA between phonetic 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. Difference of devoicing rate due to speech type

	APS		SPS	
Environment	N	% DEV	N	% DEV
CoVcCo	11,028	85.9	13,570	87.8
CoVcCv	10,943	18.4	16,816	19.9
CoVncCo	12,215	2.5	18,685	3.1

6. Concluding remarks

The use of a spontaneous speech corpus has 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 its quality and in its quantity. Full coverage of the many C1–C2 manner combinations 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 still not large enough for a complete analysis of the statistically complex phenomena like consecutive devoicing discussed in Section 4.1.2. More reliable conclusions will be achieved once we have access to the entire CSJ-Core whose data size is more than twice the current data.

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

Acknowledgments

The authors are grateful to all speakers in the Corpus Spoken Japanese. Our gratitude also goes to Professor Hisao Kuwabara of Teikyo Science University who sent us his paper upon our request, and Dr. Jennifer Venditti whose comments on an earlier version of this paper helped us greatly.

Notes

1. The Core is also labeled for other research information such as clause boundary, discourse segmentation and dependency structure, but this information is not relevant to the current paper. Visit the following URL for more information about CSJ; http://www2.kokken.go.jp/~csj/public/index.html

2. It seems that 'S/S' is an exception to the general tendency of inverse proportion. See section 5.2 for discussion.
3. The sample located in between /desu/ and /masu/ in Figure 1 is /si/, a suffix that turns a noun or adjectival into a verb (i.e. a *sahen* verb).