Voicing and geminacy in Japanese: An acoustic and perceptual study*

Shigeto Kawahara

University of Massachusetts, Amherst

1. Introduction

Maintaining voicing in obstruents is articulatorily challenging. During obstruent closure, intraoral air pressure goes up quickly, and as a consequence it becomes difficult to maintain a sufficient transglottal air pressure drop to produce voicing. This difficulty becomes more problematic in geminates, which have long closures (Hayes and Steriade 2004; Jaeger 1978; Ohala 1983; Westbury 1979). Reflecting this articulatory difficulty, historically, Japanese allowed no voiced geminates. Various alternations support this distributional restriction. Coda nasalization in mimetic gemination in (1b) is induced to avoid voiced geminates (Kuroda 1965; Itô and Mester 1999). A root-final vowel in Sino-Japanese is syncopated in compounds, with subsequent place assimilation of the root-final consonant to the following consonant, as shown in (2a) (Itô and Mester 1996); however, such syncope is blocked when it would result in a voiced geminate, as in (2b):

(1)	a. /tapu+μ+ri/ →	[tappuri]	*[tampuri]	'a lot of'
	b. /zabu+μ+ri/ →	[zamburi]	*[zabburi]	'splashing sound'
(2)	a. /hatu+kaku/ →	[hakkatsu]	*[hatsutatsu]	'revelation'
	b. /hatu+gen/ →	[hatsugen]	*[haggen]	'remarks'

Although a restriction against voiced geminates in Japanese is clearly motivated, in recent loanwords, we do find voiced geminates (McCawley 1968; Itô and Mester 1999 among others). A word-final sound in a donor language is geminated, and the following vowel is epenthesized, even when this results in a voiced geminate, as in 'dog' borrowed as [doggu] (Lovins 1973; Katayama 1998; Shirai 1999; Takagi and Mann 1994). Such a

* I am thankful to José Benki, Kathryn Flack, Ben Gelbart, Joe Pater, Chris Potts, John McCarthy, Caren Rotello and Taka Shinya for their comments and suggestions on this project. I would like to thank especially John Kingston for his support in virtually every aspect of the studies reported here. Finally, I wish to express my gratitude for all the Japanese speakers who participated in the experiments reported below. All remaining errors are mine.

voiced geminate minimally contrasts with a voiceless geminate; there are minimal pairs like [kiddo] 'kid' and [kitto] 'kit' that show that voicing is indeed phonemic for geminates in the loanword phonology. Yet, as mentioned above, voicing in geminates is aerodynamically challenging, and there is evidence that voicing in singletons and voicing in geminates behave differently in Japanese phonology; Nishimura (2003) and Kawahara (2005) show that only voiced geminates, but not voiced singletons, devoice when they cooccur with another voiced obstruent. In other words, only voicing in geminates can be lost in response to the OCP(voi), which prohibits more than one voiced obstruent within a stem (e.g. Itô and Mester 2003).

(3) Voicing in geminates can optionally be lost in response to the OCP(voi)

gebberusu	~	gepperusu	'Göbbels (proper name)'
guddo	~	gutto	'good'
beddo	~	betto	'bed'
doggu	~	dokku	'dog'
baggu	~	bakku	'bag'

(4) Voicing in singletons is not lost

bagii	'buggy'	bogii	'bogey'
bobu	'Bob'	doguma	'dogma'
dagu	'Doug'	daibu	'dive'
giga	'giga- (10^9)'	gaburieru	'Gabriel'

To account for this asymmetry, following the P-Map hypothesis (Steriade 2001), Kawahara (2005) hypothesizes that voicing in geminates is more easily lost because voicing in geminates is harder to hear. Cross-linguistically, contrasts that are signaled by weaker cues are more prone to phonological neutralization (Hura et al. 1992; Jun 2004; Kohler 1990). For example, preconsonantal consonants have many disadvantages in signaling their place: they suffer from the lack of CV transitions which provide primary cues for place distinction, and they are often unreleased, which again weakens place cues (see Jun 2004 and references cited therein). As is well-known, preconsonantal consonants undergo place neutralizations much more often than prevocalic consonants.

Kawahara (2005) applies the same logic to explain the contrast between (3) and (4). He hypothesized that voicing is harder to detect in geminates, and therefore it is more prone to phonological neutralization. More concretely, for example, the [atta]~[adda] contrast is less reliably perceived than the [ata]~[ada] contrast; so it would not have a large perceptual impact if [adda] became [atta], while if [ada] became [ata], it would be more perceptually conspicuous. In other words, neutralizing voicing in geminates is regarded as "perceptually tolerated articulatory simplification" (Kohler 1990): since voicing in geminates is hard to perceive, its loss does not have a large perceptual consequence, hence tolerated. Just as preconsonantal consonants are more likely to undergo place neutralization than prevocalic consonants, perceptually weak voicing in geminates is more easily lost than more robustly cued voicing in singletons. See Kawahara (2005) on why the loss of voicing in (4) cannot be purely due to the articulatory difficulty of

voicing in geminates.

This paper reports phonetic studies that aim to verify the hypothesis that voicing in geminates is less reliably perceived than voicing in singletons. As little is known about voicing cues in Japanese voiced geminates, I began with an acoustic experiment that identified a set of acoustic cues that distinguishes voiceless and voiced consonants. The primary aim of this experiment was to see whether such cues manifest themselves differently in singletons and geminates, and if so, in what ways. In other words, the experiment looked for evidence from acoustics bearing on whether voicing is harder to detect in geminates. The result of this experiment shows that some cues are indeed weakened in geminates, which might lead to higher confusability of voicing in geminates.

With these observations in mind, the second experiment more directly tested the core hypothesis of this paper, which is that voicing is harder to detect in geminates than in singletons. In order to most closely replicate the natural environments in which Japanese listeners hear voiced geminates, the natural tokens recorded in the first experiment were used. In the experiment, Japanese speakers identified the presence (or the absence) of voicing in a noisy environment. The result clearly shows that voicing is hard to perceive in geminates, while voicing in singletons is accurately perceived. In summary, the two experiments reported in this paper show the following points:

(5) a) Some phonetic correlates of voicing are weakened in geminates.
b) Some phonetic differences are enhanced in vowels next to geminates.
c) Voicing in geminates is not well perceived.

2. Experiment I: Acoustics of voicing and geminacy

The first experiment was designed to investigate the following three questions:

(6) 1. What are the phonetic cues that signal voicing in Japanese?
2. How are such cues different in singleton and geminates?
3. Do geminates have a disadvantage in signaling voicing?

2.1. Methods

2.1.1. The speakers and recording

Three native speakers of Japanese were recruited at the University of Massachusetts, Amherst. They were all female and in their mid twenties. An informed consent form was obtained from each speaker in accordance with the University of Massachusetts human research subjects guidelines. The dialects the subjects spoke were Shizuoka Japanese (Speaker E), Tokyo Japanese (Speaker T) and Hiroshima Japanese (Speaker W). The frame sentence used in the experiment was Standard (Tokyo) Japanese, and the subjects were asked to read the sentences in Standard Japanese as well. They were all paid for their time. The speech was recorded through a microphone (MicroMic II C420 by AKG) by a CD-recorder (TAS-CAM CD RW-700) in a sound attenuated booth at the University of Massachusetts. The recorded tokens were then digitized with a 22.050 KHz sampling

rate and 16 bit quantization level. Including short breaks between each repetition, the recording session lasted about 45 minutes.

2.1.2. The stimuli

The stimuli consisted of 36 words, which were mostly nonce words.[1] In addition, 36 nonce words were added as fillers. The target words were all disyllabic: the first consonant was [k], the second consonant was the target ([p], [t], [k], [pp], [tt], [kk], [b], [d], [g], [bb], [dd], [gg]) and three different vowels were used ([a], [e], [o]) for both the first syllable and the second syllable (henceforth V1 and V2, respectively); some examples are *kappa, kaba, kege, kokko, kodo*. The speakers were asked to pronounce these tokens with a HL tonal contour, which is a default pattern in loanword and nonce word pronunciation.

Each word was written on a card in *katakana* orthography, which is conventionally used for loanwords. This was because voiced geminates are found only in loanwords. Six repetitions of each set were recorded, with a short break between each repetition. The order of the stimuli was randomized after each repetition In order to solicit natural utterances and avoid domain-edge strengthening effects on target words (e.g. Fourgeron and Keating 1997), the stimuli were embedded in the following frame sentence:

(7) jyaa ____ de onegai
then ____ with please
'Please, (do something) with ___. (casual register)'

In order to avoid the hyper-articulation of the materials in an experimental environment, the speakers were encouraged to produce sentences in a natural speech style. Specifically, they were asked to imagine a situation where they were preparing a party and they wanted their friend to fetch the things whose names were the target words.

2.1.3. Measurement and analysis

All measurements were done using Boersma and Weenink's (1992) Praat. Following the past literature on acoustic and perceptual correlates of voicing (Lisker 1987; Kingston and Diehl 1994; Raphael 1981; Stevens and Blumstein 1981), the following values were measured, which is visually illustrated in Figure 1:

(8) a) closure voicing
b) duration of the preceding vowel
c) closure duration
d) F0 of the surrounding vowels
e) F1 of the surrounding vowels

[1] It was impossible to completely exclude real words in this set; [kaka], [kaba], [kakka] are real words. Yet as they were all written in *katakana* orthography, at least [kaka] and [kakka], which are usually written in *hiragana*, should have been hard to recognize as real words.

Closure voicing is the glottal vibration during obstruent closure; this acoustically appears as a voice bar, energy observed during closure near the baseline of the spectrogram. This should appear in only voiced consonants. The second cue lies in the immediately preceding vowel, which is known to be longer before voiced consonants. The third correlate of voicing is closure duration: cross-linguistically, voiceless consonants are longer than voiced consonants. Finally, F0 and F1 are generally higher next to voiceless consonants in both the preceding and following vowels (V1 and V2). These measurement points (except for F0) are illustrated in Figure 1:

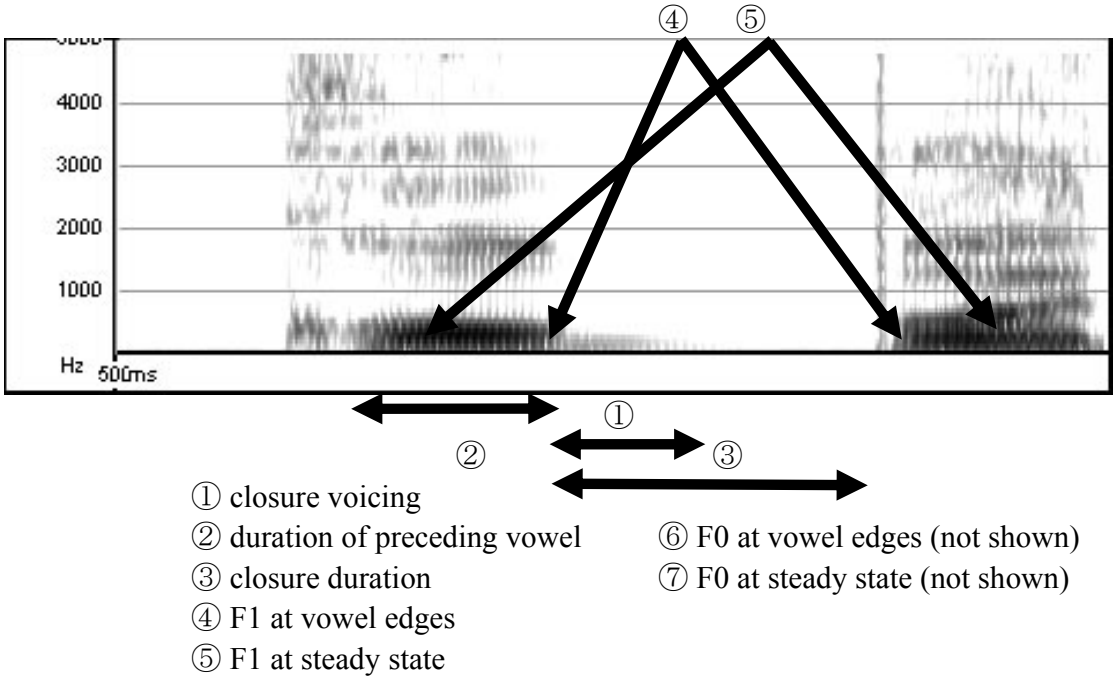


Figure 1: Illustration of measurement points. The spectrogram is that of [kobbo] uttered by Speaker E.

More detailed explanations of how these values were measured are provided below.

To analyze these acoustic measures, an ANOVA was run with CONSONANTAL LENGTH[2] (2-level), VOICING (2-level) and SUBJECT (3-level) as between-subject independent variables. This is because what is of interest is how a voicing difference manifests itself in these acoustic values, and how they vary in singleton and geminate environments. I treated SUBJECT as an independent variable as well to test for any inter-speaker variability.

2.2. Results

The overall results show that the phonemic difference in voicing is cued in both singletons and geminates by all of the measurements taken here. However, some of these

[2] In this paper, "duration" refers to a phonetic temporal property while "length" refers to a phonological geminacy contrast.

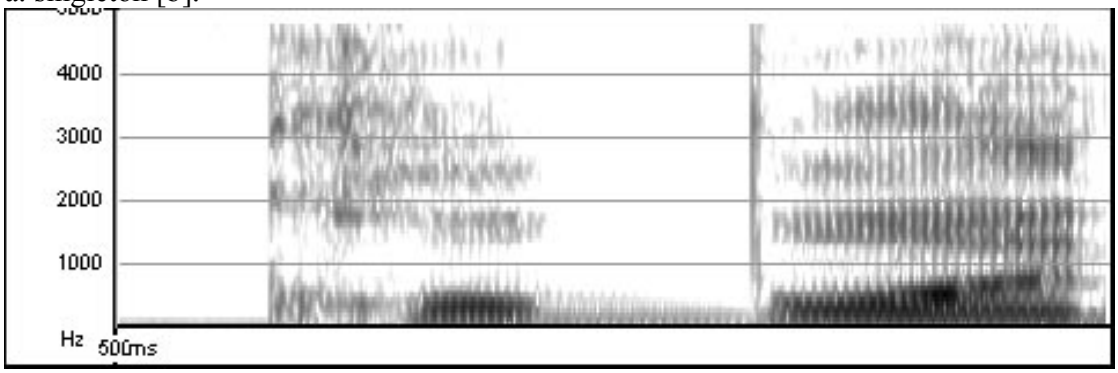
cues, most notably closure voicing, are weakened in geminates. Yet all speakers attempt to compensate for this weakening in some way or another, by enhancing some phonetic difference next to voiced geminates or by producing a phonetic difference not observed in voiced singletons. Each of the acoustic measures will be discussed in more detail below.

2.2.1. Closure voicing

One of the most important voicing cues is the extent to which voicing continues during closure, acoustically realized as a voice bar (Lisker 1986; Raphael 1981; Stevens and Blumenstein 1981). The duration of the voice bar was measured for each token, and the ratio of closure voicing with respect to duration of closure was calculated. The duration of a voice bar was measured based on the presence of energy in low frequency ranges. The onset of closure was in most cases acoustically unambiguous, signaled by abrupt disappearance of formants. In case of gradual closure, which was sometimes observed for dorsals, the disappearance of F2 and F3 was used as a criterion. The offset was set at the release of the closure, which was cued by the appearance of the burst noise. The values reported here do not include the burst noise in the closure duration.

One of the most noticeable differences between voiced singletons and geminates is that while voiced singletons maintain voicing throughout the closure, there are very few tokens of geminates in which such full voicing is observed. Figure 2 illustrates representative tokens uttered by Speaker W:

a. singleton [b].



k o b o

b. geminate [b]

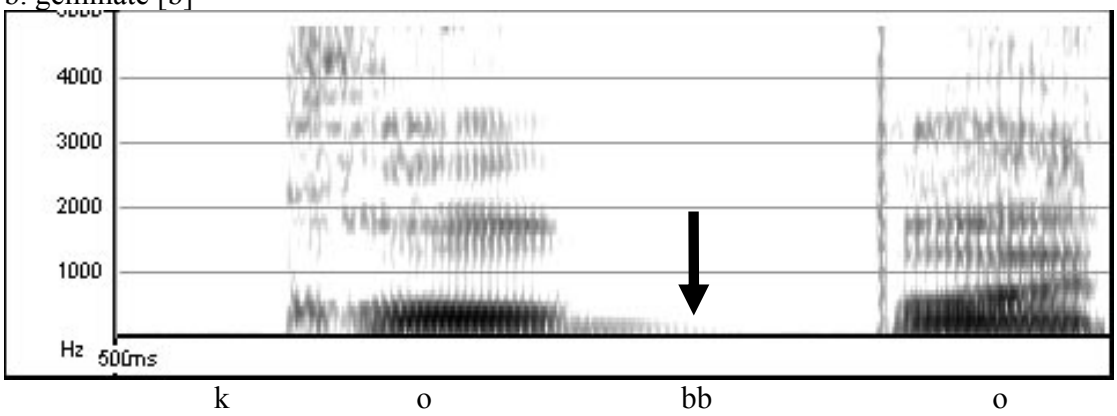


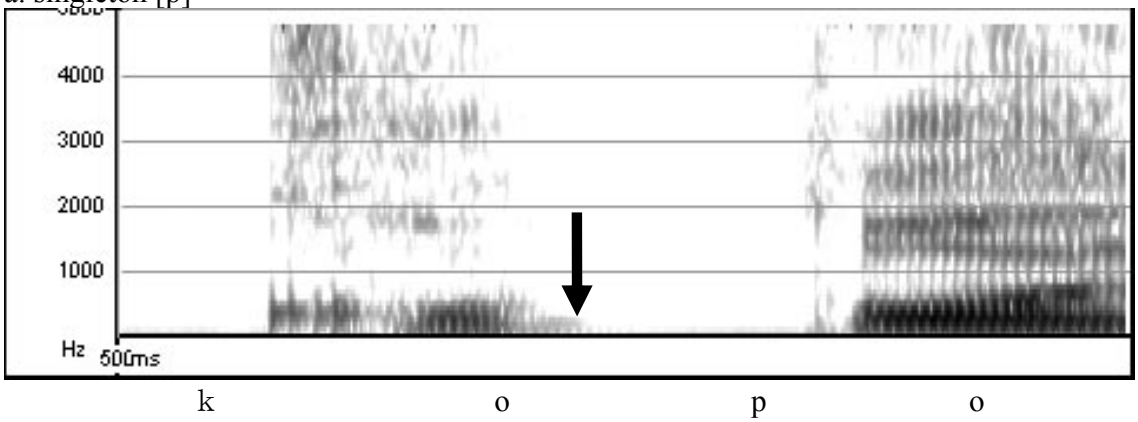
Figure 2: Spectrograms of singleton and geminate [b] pronounced by Speaker W. While voicing is fully maintained in the singleton [b] (a), partial devoicing is observed after the arrow in the geminate [bb] (b).

The first spectrogram is that of a singleton [b], and as seen, closure voicing continues throughout the closure. In contrast, in the second spectrogram of a geminate [bb], voicing stops in an early phase of the closure, at the point indicated by an arrow.

This contrast between singleton and geminate voiced consonants is a very general pattern observed for all speakers; for example, Speaker T shows full voicing for all through 54 singletons, while she exhibits no tokens of geminates in which voicing is maintained more than 80 percent of the closure. Speaker E and Speaker W show two instances of exhaustively voiced geminates, [dd] and [bb], respectively, but all other tokens are partially devoiced.

Consider next Figure 3 which shows the spectrograms of [p] and [pp] from Speaker W:

a. singleton [p]



b. geminate [p]

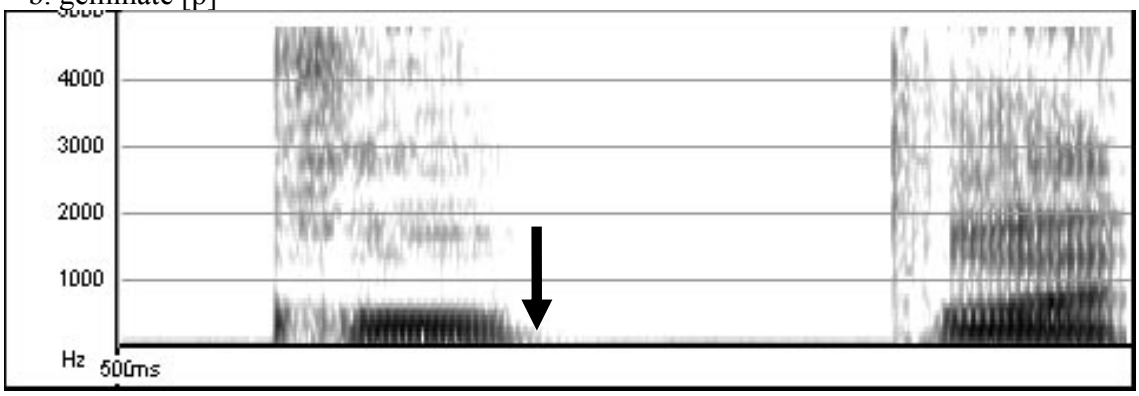
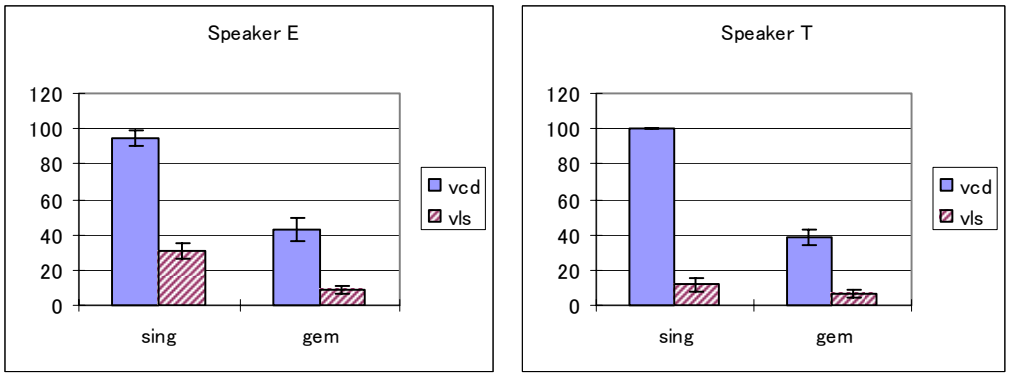


Figure 3: Spectrograms of [p] and [pp] pronounced by Speaker W. Voicing leakage is observed, indicating that voicing cessation and closure do not completely coincide.

Even voiceless consonants have a small amount of voicing leakage; there is short closure voicing after the closure of a voiceless [p]. For singleton pairs like [p]~[b], even with such voicing leakage in [p], a voicing contrast is still clear, since the closure voicing is exhaustive for [b]. However, in geminate pairs like [pp]~[bb], given that even [pp] has some closure voicing and [bb] is partially devoiced, the acoustic difference between voiced and voiceless consonants is very small.

To numerically analyze these observations, the proportion of closure voicing with respect to closure duration was calculated. The results are summarized in Figure 4. Here and throughout, in summary figures, the first pairs of bars in each graph represents singleton values while the second pair shows geminate values. Within each block, the first (solid) bar represents voiced consonants, and the second (striped) bar represents voiceless consonants. Error bars represent 95% confidence intervals, calculated as $t_{0.05} \times$ standard error of the mean (s.e.).



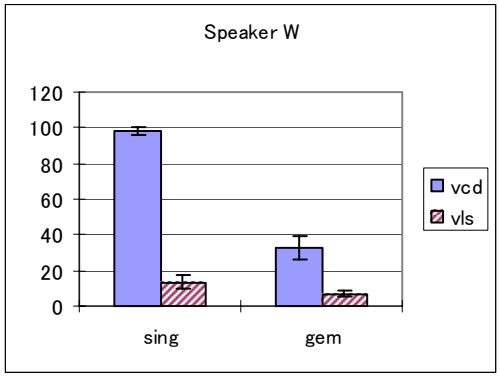


Figure 4: The ratio of closure voicing with respect to closure duration (%) for each speaker. Error bars represent 95 confidence intervals.

For all speakers, singleton voiced consonants are voiced through almost 100 percent of their closure; on the other hand, voiced geminates are voiced through only around 30 to 40 percent of their closure, indicating that partial devoicing is prevalent. The acoustic voicing difference between voiced and voiceless singletons is thus drastically reduced in geminates. Such weakening of closure voicing in geminates should have a strong impact on the perceptibility of voicing in geminates, as closure voicing is presumably an important cue to phonemic voicing (Lisker 1986; Raphael 1981; Stevens and Blumstein 1981), and about 60 percent of the closure, the geminates are "voiceless." Further, since such phonemically "voiced" consonants are acoustically "voiceless" at the time of release, this should again attenuate overall voicing perception in geminates, because it is known that the onset cues have primacy over offset cues (e.g. Rapahel 1981; Slis 1986).

The result of ANOVA suggests that VOICING and CONSONANTAL LENGTH both significantly affect the ratio of closure voicing: $F(1, 608)=2073.928$, $p<.0001$ and $F(1, 608)=877.896$, $p<.0001$, respectively. It does not come as a surprise that a phonemic voicing distinction affects the proportion of closure voicing. More interesting is the fact that the length difference has an effect on closure voicing as well. This is because voiced geminates are frequently partially devoiced, as seen above. This is also indicated by the fact that the LENGTH-VOICE interaction is significant: $F(1, 608)=414.107$, $p<.0001$: only voiced geminates, not voiced singletons, undergo partial devoicing.

Though partial devoicing is prevalent, the voicing contrast seems always maintained: an independent sample t-test shows a significant difference between voiced and voiceless geminates ($t(332)=16.450$, $p<.0001$). Compared to voiced geminates, which have around 30~40 percent closure voicing, voiceless geminates have on average less than 10 percent closure voicing. The average absolute duration of voicing in geminates is around 40 ms across all the speakers (Speaker E=42.2 ms, Speaker T=42.4 ms, Speaker W=38.3 ms), which is small, but not negligible. These values are different from those for voiceless geminates, which are about 10 ms (Speaker E=10.5 ms, Speaker T=9.0 ms, Speaker W=9.7 ms).

2.2.2. Duration of preceding vowels

The second phonetic difference that correlates with the voicing distinction is the duration of the immediately preceding vowel (V1). To measure this, the onset of the V1 is set at the first periodic wave after the aspiration of the preceding [k], judged based on the beginning of a periodic wave in the waveform. The offset is set at the onset of consonantal closure, signaled by the disappearance of F2 and F3.

An ANOVA shows that VOICING and LENGTH both have a statistically significant impact on the duration of V1 ($F(1, 603)=166.344$, $p<.0001$ and $F(1, 603)=453.184$, $p<.0001$). This reflects the tendency for V1 to be longer before voiced consonants as well as before geminates, as illustrated in Figure 5:

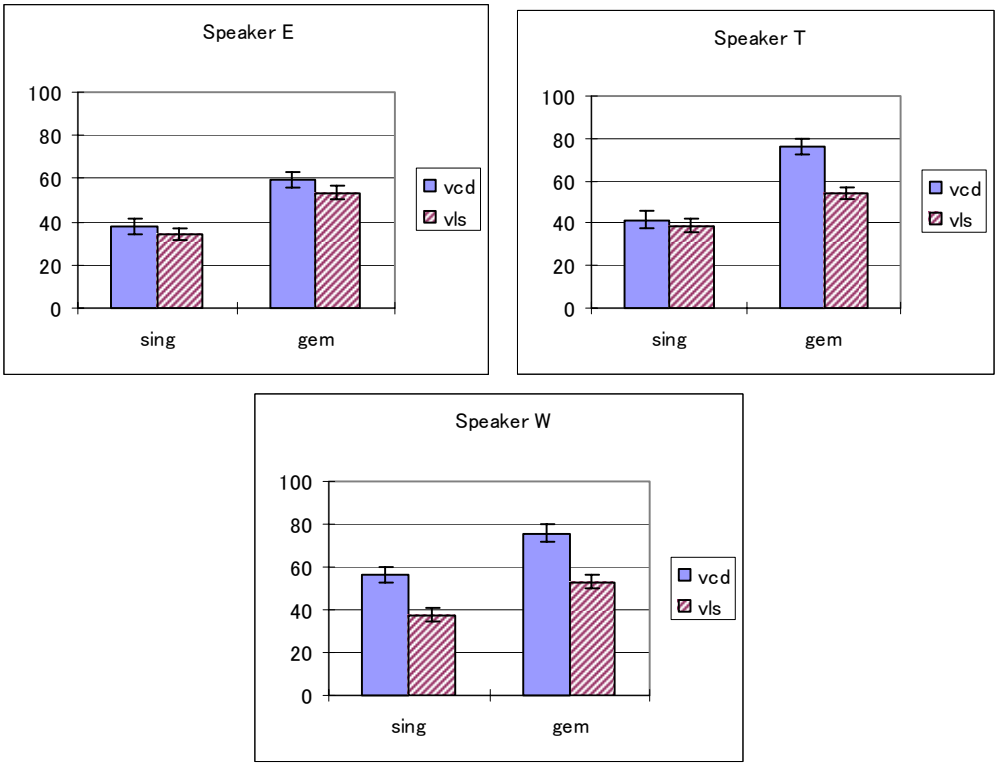


Figure 5: The duration of the preceding vowel (milliseconds).

Vowels are cross-linguistically longer before voiced obstruents than before voiceless ones (Chen 1970; Raphael 1972, 1981). This is true in Japanese for all the speakers both before singletons and geminates, as confirmed by the ANOVA result.

The fact that LENGTH has a main effect on V1 duration captures the tendency for preceding vowels to be longer before geminates. This is contrary to the cross-linguistic tendency that vowels are shorter in closed syllables than in open syllables (see Maddieson 1985; though see Smith 1995 who argues that this tendency is not universal, using data from Japanese). One might suspect that a Japanese geminate does not close a preceding syllable, but this postulation is not tenable because geminates count as moraic,

and thus appear to be coda consonants (see e.g. McCawley 1968; Poser 1990 for evidence). Even with such lengthening of V1, however, a voicing contrast is still maintained before geminates: $t(322)=11.116$, $p<.0001$.

According to the ANOVA, the interaction of VOICING and LENGTH is significant ($F(1, 603)=19.487$, $p<.0001$); this shows that the extent to which voicing affects V1 duration before voiced versus voiceless consonants. This is most clearly observed in Speaker T; the V1 difference due to voicing is larger before geminates. A related observation is the interactions of SPEAKER with VOICING and LENGTH are both statistically significant ($F(2, 603)=22.648$, $p=.004$ and $F(2, 603)=4.584$, $p=.011$, respectively). The significance of the SPEAKER-VOICE interaction shows that there is inter-speaker variation for the extent of V1 difference before voiceless versus voiced consonants. The significance of the SPEAKER-LENGTH interaction indicates that the degree to which geminacy affects V1 duration also differs among the three speakers. In Figure 5, we can see that Speaker E has relatively small differences between voiceless and voiced environments before both singletons and geminates. On the other hand, Speaker W shows relatively large differences in both environments. Finally, Speaker T makes the V1 difference greater before geminates than before singletons. Reflecting this, the interaction of the all variables is significant: $F(2, 603)=7.895$, $p<.0001$).

The fact that Speaker T has a larger difference before geminates than before singletons might be captured as a compensation effect: as geminates suffer from partial devoicing, the speaker might be attempting to enhance the contrast in V1 as an alternative means signaling voicing. In other words, to make up for the weakening of closure voicing, she enhances another cue. We observe below that a similar effect is exhibited by the other two speakers in other acoustic dimensions.

2.2.3. Closure duration

The third difference between voiced and voiceless consonants is closure duration. How closure duration was measured is stated in §2.2.1. An ANOVA suggests that, as cross-linguistically often observed (Westbury 1979; Ohala 1983: 195), voiced consonants are shorter in duration than voiceless consonants ($F(1, 602)=182.938$, $p<.0001$), which presumably contributes to perception of voicing (Lisker 1957, 1981, 1986; Kingston and Diehl 1994). LENGTH, quite naturally, exhibits a large significance; by definition, geminates have longer closure duration ($F(1, 603)=3220.478$, $p<.0001$). The interaction of VOICING and LENGTH is not significant ($F(1, 603)<1$). This means that the difference in closure duration due to a voicing contrast is preserved both in singletons and geminates.

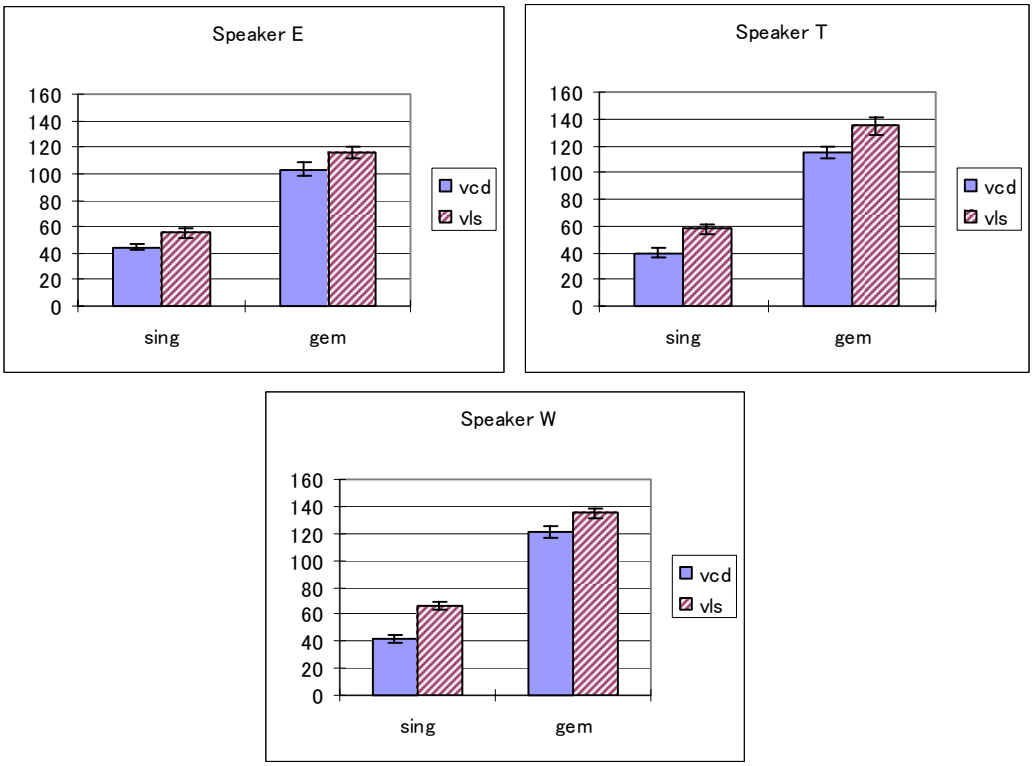


Figure 6: Closure duration of each consonant type (milliseconds).

As revealed by ANOVA, the closure duration difference is consistently present in singletons and geminates; the absolute magnitude of the differences between voiced and voiceless consonants is about the same in singletons and geminates, as indicated by the fact that the VOICE-LENGTH interaction was non-significant. However, given the consistent difference, geminate pairs are more similar to each other than singleton pairs because geminates have inherently longer duration. To numerically show this, the proportion of voiced consonants with respect to voiceless consonants was calculated. The following chart summarizes the ratio of mean closure duration for voiced consonants with respect to voiceless ones at each place of articulation.

(a) Speaker E

	Singletons			Geminates		
	vls (ms)	vcd (ms)	vcd/vls (%)	vls (ms)	vcd (ms)	vcd/vls (%)
lab	63	49	78 (13)	115	107	93 (8)
cor	59	43	72 (15)	120	107	90 (10)
dor	43	39	92 (9)	113	97	86 (12)

(b) Speaker T

	Singletons			Geminates		
	vls (ms)	vcd (ms)	vcd/vls (%)	vls (ms)	vcd (ms)	vcd/vls (%)
lab	66	48	72 (15)	140	118	84 (12)
cor	52	30	58 (17)	137	107	78 (13)
dor	55	52	95 (13)	129	119	92 (9)

(c) Speaker W

	Singletons			Geminates		
	vls (ms)	vcd (ms)	vcd/vls (%)	vls (ms)	vcd (ms)	vcd/vls (%)
lab	77	50	64 (16)	145	125	86 (16)
cor	63	35	56 (16)	130	123	94 (15)
dor	60	39	64 (17)	123	115	94 (14)

Table 1: The ratio of voiced consonant with respect to voiceless consonants in terms of closure duration. The numbers in parentheses represent margins of error, calculated as $t_{0.05}(n-1)\times ((p(1-p)/n)^{0.5})$ where p is the proportion of vcd/vls, and n is the number of data points.

What is evident is that the ratio of voiced/voiceless is higher in geminates. This means that geminate pairs are more similar to each other than singleton pairs in terms of closure duration. In some cases (Speaker W's coronal and dorsal and Speaker E's labial), the ratio is above 90 percent, which means that voiceless and voiced consonants are almost identical in their duration. This further implies that a closure duration difference, which is presumably one of the perceptual cues for voicing, is harder to detect in geminates. This is yet another factor that might make a voicing distinction in geminates harder to hear.

2.2.4. F0 at V2 onset

The fourth voicing cue is F0 frequency at the onset of the following vowel (V2). F0 was measured at the first periodic wave right after the consonantal burst, using autocorrelation function of Praat. Cross-linguistically, it is observed that F0 is higher in vowels next to voiceless consonants (see Kingston and Diehl 1994 among others), and this is in general true for the Japanese speakers as well. Figure 7 illustrates:

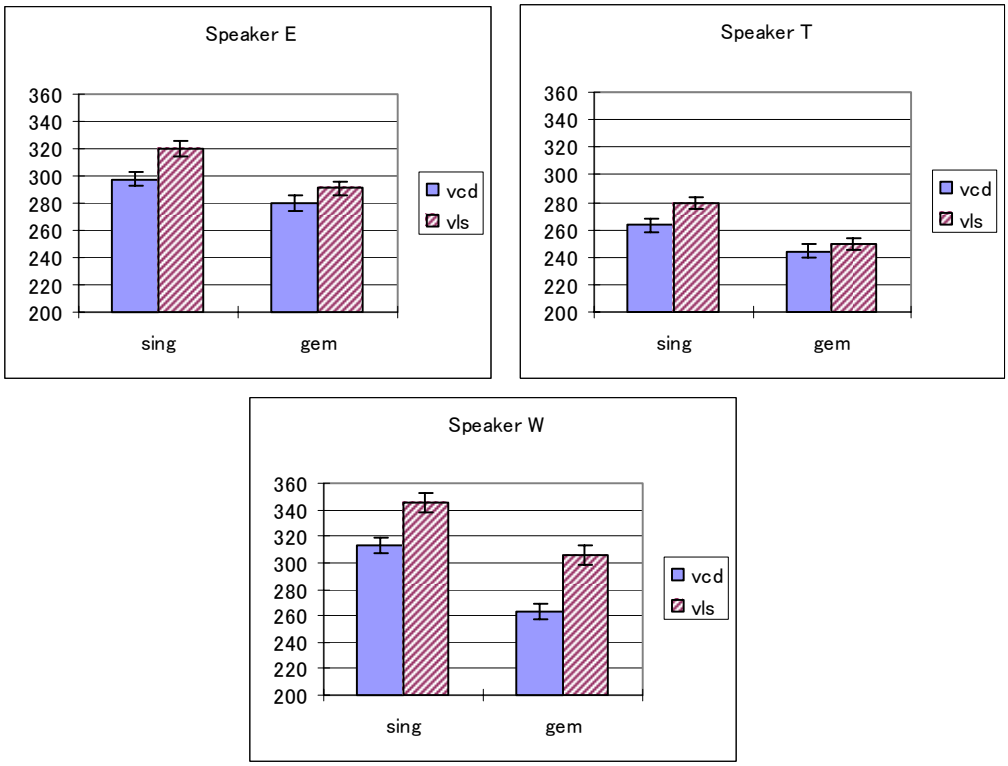


Figure 7: F0 at V2 onset (Hz).

An ANOVA shows that VOICING and LENGTH both have a statistically significant influence on F0 at V2 onset ($F(1, 604)=175.945$ $p<.0001$ and $F(1, 604)=365.276$, $p<.0001$). F0 is in general higher after voiceless consonants, although Speaker T does not show a difference after geminates ($t(106)=1.341$, $p=.182$). The interaction of VOICING and LENGTH is not significant ($F(1, 604)=1.803$, $p=.180$); however, if we look at each speaker separately, the interaction of these factors is clearly observed. Speaker E and T have a smaller difference after geminates (and in fact Speaker T's difference is lost after geminates). On the other hand, Speaker W has a larger difference after geminates (around 32Hz for singletons and 40Hz for geminates). This observation is statistically supported by the fact that the interaction of VOICING, LENGTH and SPEAKER is significant: $F(2, 604)=5.387$, $p=.005$).

Another observation is that F0 is lower after geminates (recall that LENGTH has a statistically significant impact on F0). Perhaps this is because the tonal contour of the recorded tokens is HL; given longer closure, the F0 fall is more drastic after geminates because there is more time to implement the HL fall (in a heavy syllable the fall starts at the first mora of the syllable (e.g. McCawley 1968:133-134)).

Anther point that merits discussion here is the fact that Speaker W has a larger F0 difference after geminates. This can be captured as a compensation effect in which the speaker attempts to enhance the voicing cue by F0 manipulation after geminates, whose closure voicing is weakened. Another related point is that, for Speakers E and W, the F0 difference is maintained after geminates, despite the fact that glottal vibration usually

stops before release.[3] These two points suggest that manipulation of F0 is not automatic but intentional (Kingston and Diehl 1994). If it were automatic, we could not explain the fact that semi-devoiced voiced geminates have a lower F0 in the following vowel. Also, the fact that a speaker can enhance an F0 difference after geminates suggests that it is possible to intentionally control F0. Finally, to the extent that this manipulation is to enhance the voicing contrast, this is in line with Kingston and Diehl (1994)'s view that such manipulation is essentially to enhance phonological contrasts.

2.2.5. F0 at V2 steady state

As seen above, F0 at V2 onset is higher after voiceless consonants. F0 at V2 steady state was also measured at the sixth glottal pulse after the onset of V2 (about 10 to 20 ms away from the onset). In this position also, F0 is lower after geminates ($F(1, 604)=683.250$, $p<.0001$), and after voiced consonants ($F(1, 604)=44.470$, $p<.0001$), although the second generalization is only true for Speaker E and W, as is discussed more fully below. The interaction of LENGTH and VOICING is also significant ($F(1, 604)=4.960$, $p=.026$). This is because Speaker E and W have larger differences after geminates:

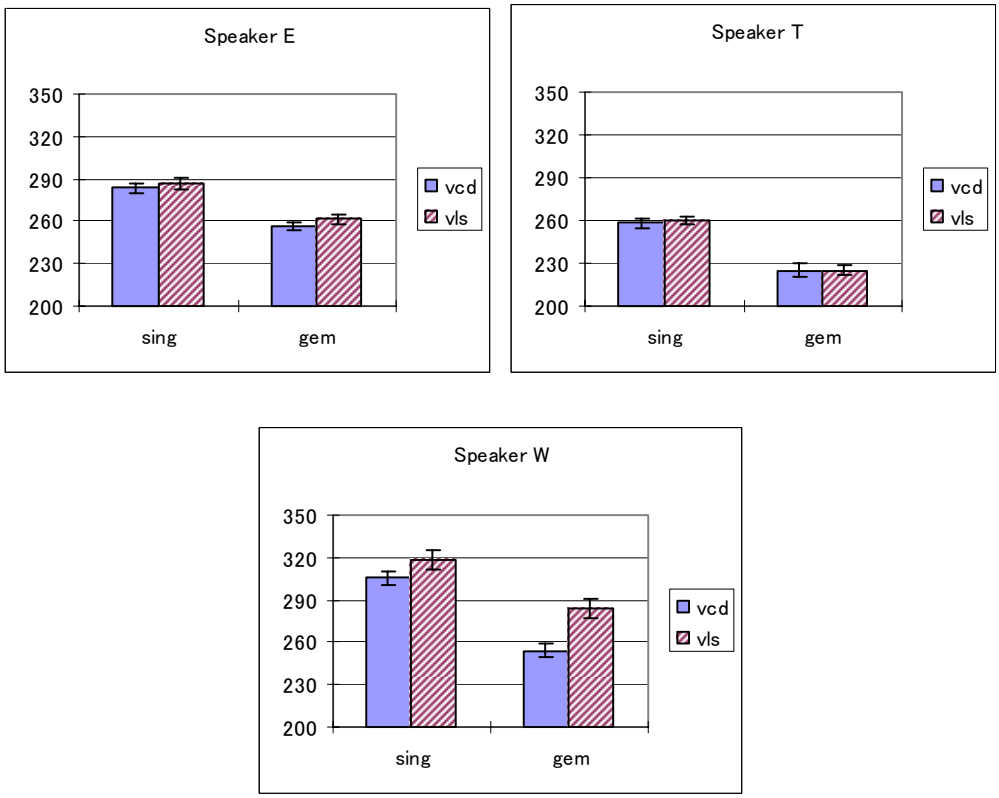


Figure 8: F0 at V2 steady state.

Looking at each speaker's behavior, Speaker T does not show any difference in terms of

[3] A similar fact is reported in English [+voi] consonants where F0 depression is observed next to [+voi] regardless of the presence of the actual voicing during the closure (Kingston and Diehl 1994 citing an unpublished work by Caisse (1982)).

VOICING (t(194)=1.371, p=.172). What is more interesting is Speaker E, for whom the difference after singletons is not statistically significant, t(102)=1.166, p=.267, but the difference after geminates is, t(106)=2.529, p=.013. This pattern observed in Speaker E - that an F0 contrast emerges only after geminates - can again be captured as a compensation effect. Voicing is weakened in geminates, so that the speaker attempts to signal a voicing contrast in a way that is specific to geminates. Similarly, Speaker W has a larger difference after geminates, which can also be captured as a compensation effect. Reflecting such inter-speaker variability, the interaction of all the variables is highly significant (F(2, 604)=4.889, p=.008).

2.2.6. F0 at V1 offset

A voicing contrast is also cued by the F0 of the preceding vowel (V1), which is higher before voiceless consonants. Figure 9 illustrates the general pattern of the three speakers:

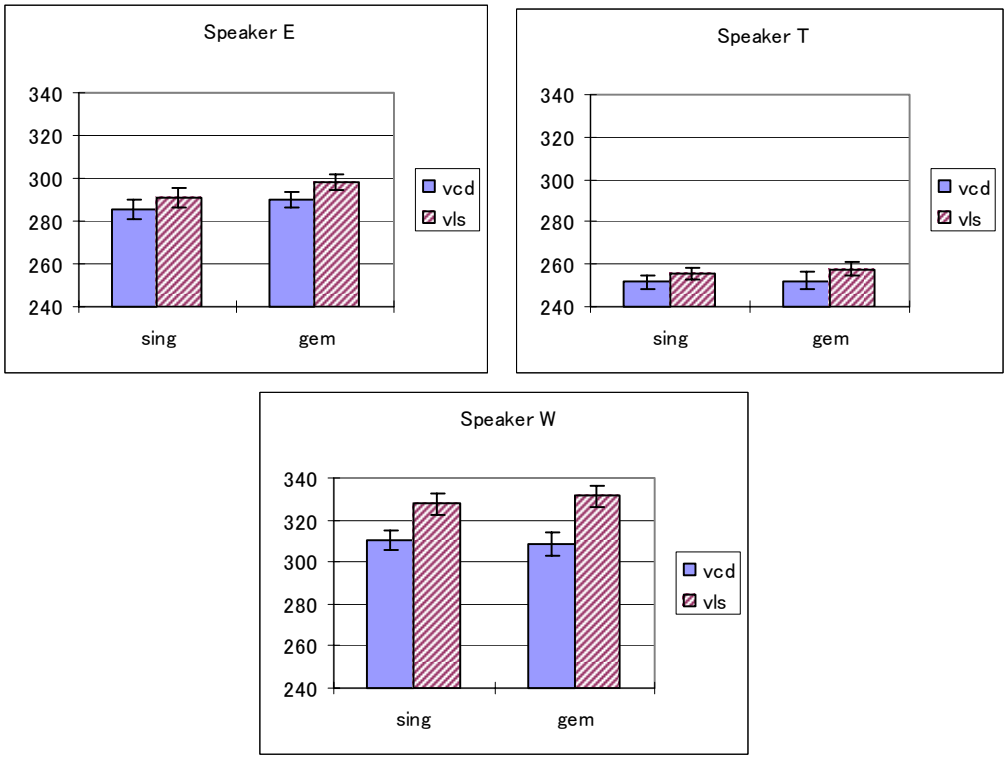


Figure 9: F0 at V1 offset.

An ANOVA shows that the influence of VOICING on F0 at V1 offset is significant: F(1, 601)= 71.288, p<.0001. A smaller main effect was observed for LENGTH: F(1,604)=4.191, p=.041. As seen in Figure 9, F0 is higher after geminates. Finally, Speaker W has a larger F0 difference than Speaker E and T, and thus the interaction of VOICING and SPEAKER is significant (F(2, 601)=14.764, p<.0001).

2.2.7. F0 at V1 steady state

The F0 values during the steady state of V1 were also measured. The measurement point was set at the sixth glottal pulse away from the offset of the vowel. There are some cases before voiceless consonants in which V1 is so short that the sixth pulse is located very close to the transitional state from the first consonant [k]. In such cases, the midpoint of the vowel was calculated, and F0 was measured at that point.

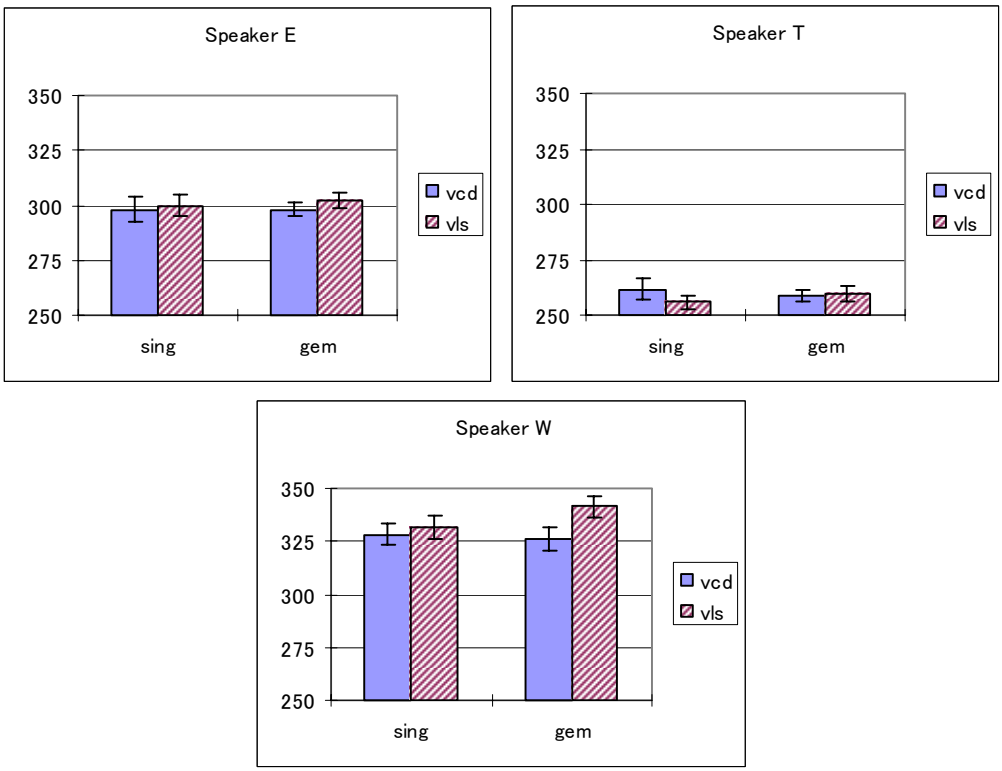


Figure 10: F0 at V1 steady state.

Overall, the difference in F0 after voiceless and voiced consonants is statistically reliable, $F(1, 599)=6.339$, $p=.012$, though not all speakers show this pattern. Speaker E has no difference before singletons ($t(100)=.375$, $p=.709$), but shows a difference before geminates (marginally significant, $t(105)=1.912$, $p=.059$). Speaker W exhibits a larger difference before geminates, while Speaker T does not show any difference either before singletons or geminates. LENGTH has no effect on F0 at the steady state of V1, $F(1, 599)<1.972$, $p=.161$. This is the tendency observed throughout the speakers; hence there is no interaction between SPEAKER and LENGTH ($F(2, 599)<1$). The interaction of VOICE and LENGTH is significant, $F(2, 599)=7.586$, $p=.006$, reflecting the fact that Speaker E and Speaker W make larger differences after geminates.

2.2.8. F1 at V2 onset

As is the case with F0, F1 is cross-linguistically known to be higher next to voiceless consonants (e.g. Kingston and Diehl 1994). To check for such a tendency in Japanese,

the F1 frequency at both V2 onset and V2 steady state was measured, calculated by Praat's LPC analysis, setting the LPC coefficient to 10. The onset measurement point was set at the first periodic wave after the burst, and the steady state measurement point was at the sixth glottal pulse after burst.

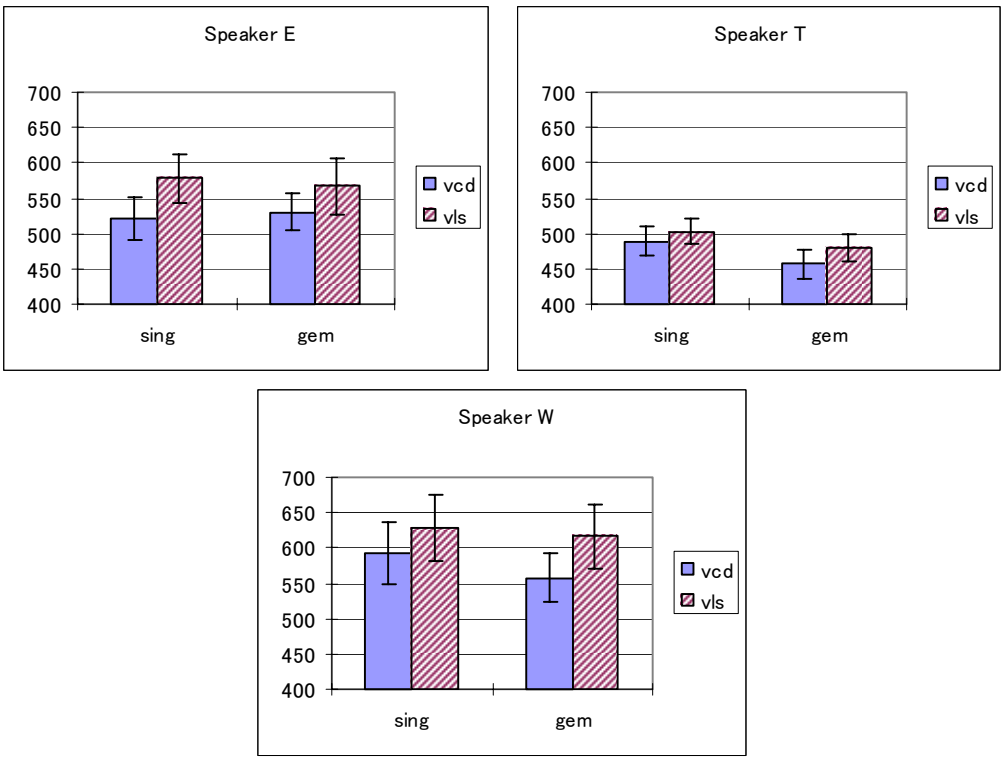


Figure 11: F1 at V2 onset. Vowel qualities are averaged over.

An ANOVA shows that VOICING affects F1 at V2 onset, $F(1, 600)=14.564$, $p<.0001$. As expected, for all the speakers, F1 is higher after voiceless consonants. LENGTH has a marginally significant effect, $F(1, 600)=3.177$, $p=.075$): F1 is lower after geminates. The size of F1 differences after voiced and voiceless consonants is similar in post-singleton and post-geminate positions, hence, no interaction of LENGTH and VOICING ($F(1, 600) <1$).

Here again, as was the case in F0 at V2 onset, a phonological distinction between voiceless and voiced consonants has a significant effect on F1 value, despite the fact that glottal vibration itself stops before release. This suggests that the F1 difference appearing next to voiceless/voiced consonants is not an automatic effect due to glottal vibration, but instead speakers can intentionally manipulate its values.

2.2.9. F1 at V2 steady state

F1 values during V2 steady state exhibit very consistent patterns across the speakers. Overall, both VOICING and LENGTH have a significant effect ($F(1, 601)=4.378$, $p=.037$ and $F(1, 601)=12.462$, $p<.0001$). More interestingly, no F1 differences are observed before

singletons (t(287)=.118, p=.862) but a difference emerges after geminates (t(321)=2.213, p=.028). These generalizations are illustrated in Figure 12:

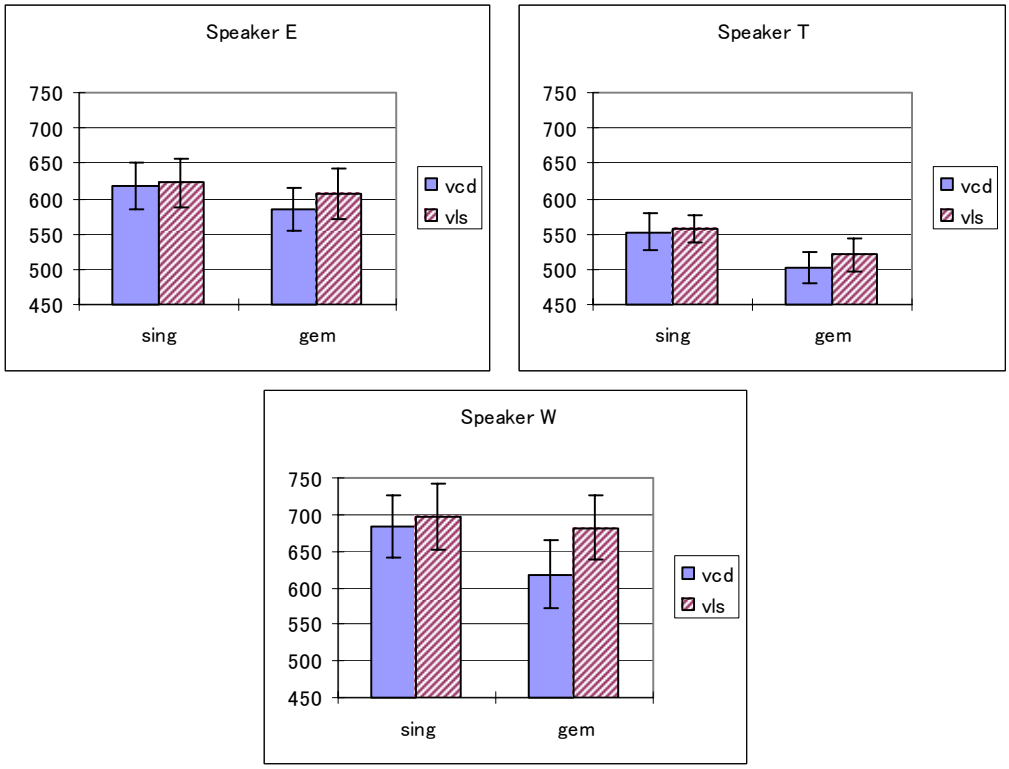


Figure 12: F1 at V2 steady state.

2.2.10. F1 at V1

The tendency for F1 to be higher next to voiceless consonants is not observed at V1, either at the offset ($F(1, 600)<1$) or at the steady state ($F(1, 600)<1$). LENGTH has an effect only at the steady state $F(1, 600)=7.192$, $p=.008$), but not at the offset ($F(1, 600)<1$. I do not have a good explanation on why an effect can emerge only at the steady state.

2.3. Discussion

The purpose of the acoustic experiment described above was to see what kinds of the acoustic cues are used to signal voicing in Japanese, and how differently these cues are realized in singletons and geminates. The experiment revealed that Japanese utilizes many of the cues that are known to signal voicing cross-linguistically. It also showed that some phonetic differences that signal a phonemic voicing difference are attenuated in geminates (most notably closure voicing and closure duration).[4] On the other hand, the

[4] Another factor that might weaken a voicing distinction in geminates is the lack of spirantization. Voiced singletons, especially [g], spirantize whereas voiceless singletons do not. As a result, singleton pairs like /g/~/k/ are phonetically distinguished in terms of continuancy as well. However, voiced geminates do not spirantize, and as a result, for geminate pairs, a continuancy difference is not a cue to signal phonemic voicing.

speakers attempt to compensate for the weakened cues by showing some phonetic differences between voiceless and voiced consonants only in the environment of geminates, or by making general phonetic correlates of voicing more prominent surrounding geminates. The overall results are summarized in Table 2:

Phonetic cues	Change in geminates
closure voicing	Weakened in geminates.
V1 duration	A difference is larger before geminates for Speaker T.
closure duration	Geminate pairs are more similar to each other than singleton pairs; the vcd/vls ratio is closer to 1 in geminates.
F0 at V2 onset	A larger difference after geminates for Speaker W A smaller difference after geminates for Speakers E and T.
F0 at V2 steady state	A difference appears only after geminates for Speaker E. A larger difference after geminates for Speaker W.
F0 at V1 offset	None.
F0 at V1 steady state	A difference appears only after geminates for Speaker E. A larger difference before geminates for Speaker W
F1 at V2 onset	Speakers T and W have a larger difference before geminates.
F1 at V2 steady state	A difference emerges after geminates.

Table 2: Summary of the acoustic cues of Japanese voicing, and how they are affected by a singleton/geminate difference.

One generalization that holds for all the speakers is that closure voicing and closure duration cues to voicing are attenuated in geminates, but F0 and F1 differences are enhanced in one way or another surrounding geminates (modulo F0 at V2 onset for Speaker E and T).

Despite the speakers' attempt for compensation, however, it seems reasonable to speculate that overall, voicing cues are weakened in geminates. First, closure voicing, which arguably constitutes an important cue for voicing perception (Lisker 1986, Raphael 1981), is weakened in geminates. Second, the compensation effects observed above are subject to inter-speaker variation. In fact, none of the strategies is taken by all three speakers, except for the F1 difference enhancement at V2 steady state. For example, although an F0 difference at V2 onset is enhanced after geminates for Speaker W, the opposite pattern holds for Speaker E and T. Thus, unless such cues are integrated in some way (e.g. Kingston and Diehl 1995) so that such integrated cues are consistently enhanced in the context of geminates, it is doubtful that such enhancements provide reliable perceptual cues. Even if such enhancements indeed partially compensate for the weakening of other cues, it is also doubtful that the amount of compensation is enough. For example, Speaker E's F0 at V2 steady state exhibits a difference only after geminates, but the difference that emerges is around 6 Hz. Speaker W's enhancement of F0 differences after geminates at V2 onset is only 8-10 Hz. It seems unlikely that such small differences have a large perceptual effect. In sum, compared to the systematic weakening of closure voicing in geminates, the attempts for compensation are subject to inter-

speaker variability, and the effects seem very small. Thus from the acoustic point of view, it seems likely that voicing cues are overall weakened in geminates. This conclusion is supported by the result of the perceptual experiment reported below in §3.

3. Experiment II: Perceptual experiment

In order to more directly test the hypothesis that voicing is harder to hear in geminates than in singletons, a perceptual experiment was conducted. The primary aim of this experiment was to see how well Japanese speakers perceive voicing in singletons and geminates in natural environments. In order to most accurately replicate the situation in which Japanese speakers hear voicing in geminates and singletons, the natural tokens recorded in the first experiment were used. However, if I had used natural tokens and nothing else, Japanese speakers might have performed at ceiling. To overcome this problem, the stimuli were covered by cocktail party noise so as to confuse the listeners. Following the observation from the first experiment that the acoustic cues for voicing in geminates are overall attenuated, the prediction is that voicing in geminates is perceived relatively poorly compared to voicing in singletons.

3.1. Methods

3.1.1. Stimuli

From the pool of tokens obtained in Experiment I, for each speaker, one representative example of each type of stimulus was chosen. The total number of the stimuli was therefore 108 (3 speakers × 3 vowels × 3 places of articulation × 2 consonantal lengths × 2 voicing types). Tokens that contained phonetic irregularity (such as transient sounds or devoiced V1) or spirantization were not used; for the case of singleton [g]s, which very frequently undergo spirantization, tokens with least spirantization were chosen. Among the tokens of voiced geminates at each place of articulation with no phonetic distortions, those used were the ones with closure voicing duration closest to that place of articulation's average. This was in order to use representative tokens of natural voiced geminates. See the Appendix for acoustic values of the tokens used.

Cocktail party noise was used to cover the tokens. This particular kind of noise was used because to cover voicing, it was necessary to use speech-like noise that has energy in low spectra range; voicing would not be covered well by white masking noise (Miller and Nicely 1955). To obtain cocktail party noise, a party was recorded at the linguistics department of the University of Massachusetts using a SONY TCD-D8 portable DAT recorder. The recorded sound was divided into 3-second noise stretches. Six files of such stretches were randomly chosen and superimposed on top of one another. Twelve such noise files were created. To equalize the amplitudes of all the stimuli, the peak amplitudes were adjusted to 0.50 Pascal by Praat; the peak amplitudes of the noise files were modified to 0.45 Pascal. Since $dB=10\times\log_{10}(Pascal^2/0.00002^2)$, the peak amplitudes of the stimuli and the noise are 87.95dB and 87.04dB, respectively. Thus the signal-to-noise ratio (S/N ratio) is 87.95dB-87.04=0.91dB (since dB is a logarithmic function, the ratio is calculated as the numerator minus the denominator). Then, one noise file was randomly chosen and was superimposed on each stimulus. After the stimuli and the noise

were combined, the edges of the combined files where only noise was present were trimmed off. After this process, all stimuli were approximately 1.5 second long, including the frame sentence.

3.1.2. Subjects

In the main experiment, 15 female and 2 male native speakers of Japanese were recruited from the University of Massachusetts community. They were all in their twenties or early thirties. The speakers that participated in the first experiment were excluded since they might have some advantage hearing their own voice. All the subjects had normal hearing and were free of any speech disorders. Some had a basic knowledge of linguistics, but none had had extensive phonetic training. The range of dialects that the speakers spoke was diverse, including Chiba Japanese, Tokyo Japanese, Shizuoka Japanese, Ibaragi Japanese and Osaka Japanese. No report has been made of a difference in the behavior of voiced geminates among these dialects, so this dialectal variation was not expected to impact the results. Two listeners were complete bilingual speakers of Japanese and English, but their results were very similar to the results of the other subjects; hence they are included in the results reported below. All the subjects were paid or given extra credit for linguistics classes. An Informed consent form was obtained from each subject in accordance with the University of Massachusetts human research subjects guidelines.

3.1.3. Task

The experiment was conducted in a sound-attenuated booth at University of Massachusetts, Amherst. Superlab pro software (by Cedrus) was used for audio and visual presentation of each stimulus. This automatically randomizes the order of presentation. The subjects listened to stimuli over headphones (DT 250 by Beyerdynamic). They heard one stimulus at a time; as soon as a listener heard a stimulus, two choices showed up on a computer screen. The choices were minimally different in terms of voicing e.g. for [kappa], the two choices were ‘kappa’ and ‘kabba’. The task was to make a judgment about the voicing quality based on what they heard. *Katakana* orthography was used for the visual stimuli so that people would perceive the stimuli as foreign words, in which voiced geminates are allowed. In order to make sure that speakers respond to all stimuli, there were no time limits. The listeners were not given feedback about the correctness of their response.

Before the testing sessions, they had a practice session where they did the same task for each kind of 36 tokens pronounced by one speaker. In the practice session, however, stimuli were not covered by noise, and they were given feedback about the correctness of their answers. They were also instructed to adjust the volume to a comfortable listening level during the practice session.

One testing session consisted of three blocks; each block contained all the types of stimuli pronounced by one speaker. One block thus contained 36 types of tokens (3 vowels × 3 places of articulation × 2 consonantal lengths × 2 voicing types)), and therefore one session contained 108 stimuli as a total. One session usually ended in a few minutes. The entire experiment consisted of eight such sessions. The subjects were

encouraged to take short breaks once or twice during the whole experiment. Including the instructions at the beginning and the post-experiment debriefing explanation, the entire experiment lasted about one hour.

3.2. Results

The results of this experiment clearly show voiced geminates are misperceived as voiceless much more frequently than voiced singletons. This supports the general hypothesis of this paper that voicing is indeed harder to detect in geminates than in singletons. First, the listeners' accuracy (i.e. the proportion of correct answers across all eight trials) for each item was calculated. Averaging over the results of 17 listeners, Figure 13 summarizes the general results in terms of voicing and geminacy. This shows that voiced geminates suffer from misperception:

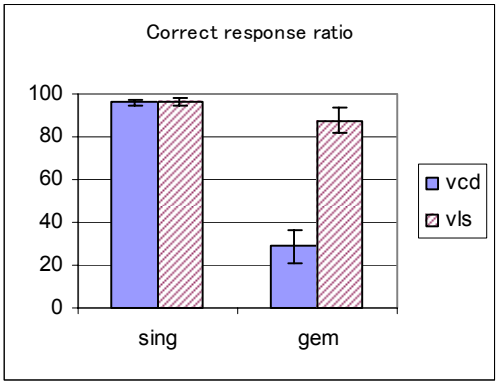


Figure 13**:** The average of correct response percentage out of eight trials averaged over 17 listeners.

As illustrated in the first two bars, when the target is a singleton consonant, both voiceless and voiced consonants are judged correctly more than 95 percent of the time (vls=96.4%; vcd=96.0%). Voiceless geminates are judged nearly as well (87.6%). On the other hand, voiced geminates are often misjudged: the accuracy goes down to 28.7%. This shows that voicing is indeed hard to detect in geminates, while voiced singleton consonants do not suffer from such a problem.

To statistically verify these observations, a repeated-measures ANOVA was run with VOICING (2-levels), LENGTH (2-levels), and PLACE (3-levels) as independent variables. To simplify the analysis, the two other factors (SPEAKER and VOWEL QUALITY) were averaged over. A vowel quality difference is not quite significant ($F(2,32)=3.028$, $p=.062$). Although the speaker variable exhibits a statistical difference ($F(2,32)=4.754$, $p=.016$), the mean values are not so different (Speaker E=78.5%, Speaker T=76.2, Speaker W=76.7%).

The results of the ANOVA are as follows. First, there is a large, statistically significant difference in speakers' performance between singletons and geminates consonants: $F(1, 16)=980.955$, $p <.0001$. VOICING has a main effect as well, $F(1, 16)=35.941$, $p<.0001$. These are likely to be due the fact that voiced geminates are frequently misjudged as

voiceless. This conclusion is supported by the fact that the interaction of VOICING and GEMINACY is also highly significant: $F(1,16)=45.437$, $p<.0001$. Its significance shows that voiced and voiceless consonants are judged differently in singleton and geminate context: only voiced geminates were poorly identified. No main effect is observed for PLACE ($F(2, 32)<1$, $p=.916$). Overall, the claim that voiced geminates suffer from misperception is supported.

Next, Figure 14 shows the listeners' performance on the judgment of voiced consonants at each place of articulation. As seen, the tendency of voiced geminates to be poorly judged holds across the three places (see below for more on differences due to PLACE):

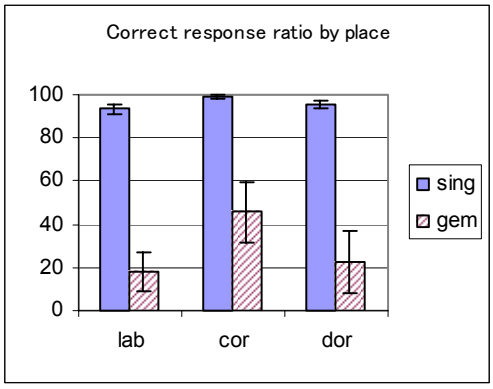
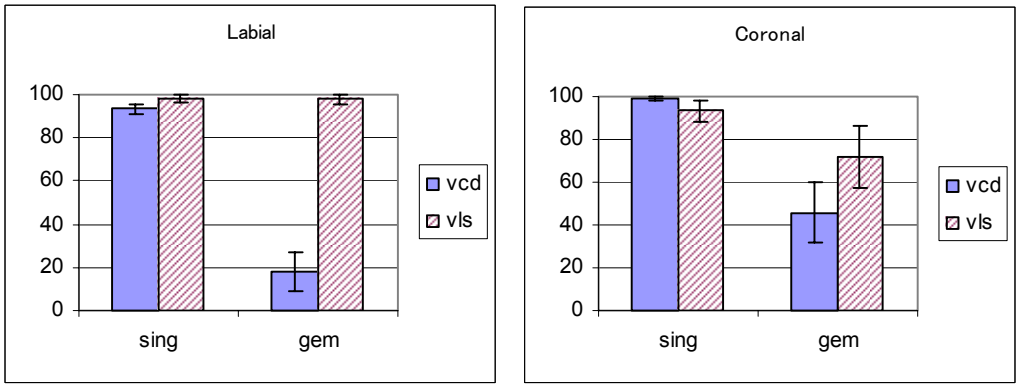


Figure 14**:** Correct identification rate of voiced consonants at each place of articulation.

Finally, consider Figure 15 which illustrates the correct identification proportion for each segment type, classified according to the place of articulation. This shows that the performance of Japanese speakers to identify voicing in singletons is consistently high across all places of articulations, whereas voicing in geminates is very frequently misperceived:



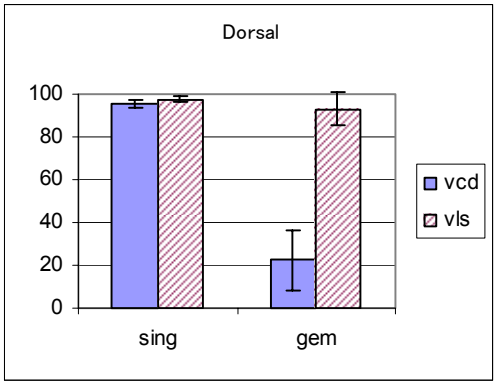


Figure 15**:** Correct identification proportion of each segment type at each place of articulation.

Interesting differences are observed between coronals on the one hand and labials and dorsals on the other. In labials and dorsals, voiceless geminates (as well as singletons) are judged correctly almost 100 percent of the time ([pp]=98.0%, [kk]=93.0%) while for coronals, some [tt] tokens are misheard as voiced (72%). In addition, the correct response proportion for [dd] is much higher than that for [bb] and [gg] ([bb]=17.9%, [dd]=46.6% and [gg]=22.3%). To see if these differences among the three places were statistically significant, I performed a post-hoc contrast analysis. It reveals that there is a significant difference between coronals on the one hand and labials and dorsals on the other: coronal vs. labial: $F(1, 16)= 37.067$, $p <.0001$, coronal vs. dorsal: $F(1, 16)= 74.897$, $p <.0001$.

3.3. Discussion

3.3.1. Bias against perceiving voicing in geminates

The perceptual test supports the hypothesis that voicing is harder to perceive in geminates. This suggests that the weakened cues - closure voicing and closure duration - might not be compensated for by speakers' manipulation of F0 and F1 surrounding geminates (see §2.4). We can further conclude from this that closure voicing and/or closure duration are important cues to voicing in Japanese.

One interesting aspect of the results is that, listeners rarely misperceived voiceless geminates as voiced. On the other hand, voiced geminates are often misperceived as voiceless. To see how sensitive Japanese speakers are to voicing in singletons and geminates and to see if there is any perceptual bias against hearing voicing, sensitivity (d') (MacMillan and Creelman 1991) was computed for each subject: d' is a measurement of sensitivity based on z-scores of hit and false alarm rates (where 'hit' is the probability of the listeners' correctly identifying voiced consonants as voiced, and 'false alarm' is the probability of the listeners' falsely identifying voiceless consonants as voiced).[5] A d' of

[5] Since z-scores are not defined for 0 and 1, I followed MacMillan and Creelman (1991) to add or subtract the equivalent of half of one response (i.e. 1/2×n) from each perfect score. For example, if a listener identified voiceless geminates as voiceless 100 percent of the time, the proportion is 1-(1/2*216)=0.998 where 216 is the number of voiceless geminate tokens they heard.

zero indicates that hit and false alarm rates are the same, and that subjects have no sensitivity to voicing. The results are that the average of d' for singletons across all of the speakers is 3.794, which is significantly different from zero, $t(16)=34.15$, $p<.0001$; the average d' for geminates is 0.705, which again is significantly different from zero, $t(16)=11.47$, $p<.0001$. This shows that the Japanese listeners are sensitive to a voicing distinction both in singletons and geminates. However, they show a much higher sensitivity for singletons; a paired t-test comparing d' for singletons and geminates reveals a significant difference: $t(16)=27.27$, $p<.0001$.

Also, following MacMillan and Creelman (1991), the bias function c was calculated, which is the sum of z-scores of hit and false alarm rates multiplied by -0.5. The mean c for singletons is 0.08, which does not significantly deviate from zero ($t(16)=1.007$, $p=.33$). On the other hand, the mean c for geminates is 1.079, which is significantly different from zero ($t(16)=5.569$, $p<.0001$). This again shows that there is a perceptual bias against responding "voiced" when geminate stimuli are indeed voiced, but this is not the case for singleton stimuli. These results essentially indicate that when Japanese speakers are confused as to the voicing quality in a geminate in a noisy environment, they perceive it as voiceless by default.

To explain this observation, in addition to weakening of acoustic cues, it might be the case that lexical frequency and/or phonological constraints antagonistic against voiced geminates are also at work here. Since voicing is phonemic on geminates only in loanwords, voiced geminates are overall much less frequent than voiced singletons are in the entire Japanese lexicon. As a consequence, less frequently heard voicing in geminates might have a disadvantage in being perceived, as frequency can cause such a perceptual bias (e.g. Hey et al. in press for a recent overview). Also, grammatical constraints antagonistic to voiced geminates might be also at work: phonologically illegal sounds or sound sequences cause a perceptual bias (e.g. Moreton 2002). From the result of the experiment *per se*, it is not clear which factor is responsible for the poor performance in judging voiced geminates (though see below for evidence in favor of a grammatical account); however, what is important is the fact that voicing in geminates is less well perceived compared to voicing in singletons, and this is perhaps related to its tendency to be easily neutralized, as argued by Kawahara (2005).

3.3.2. Place differences

The difference between coronals on the one hand and labials and dorsals on the other is noteworthy. For labials and dorsals, the Japanese speakers exhibit a bias against hearing voiced geminates; they rarely hear voiceless consonants as voiced. On the other hand, for coronals, the Japanese speakers seem to be more confused when faced with coronal geminates: as a result, the voicing judgment of coronal geminates is much closer to what is expected if people are responding by chance (in which case both [tt] and [dd] would be judged correctly 50% of the time). The calculation of the bias function c indicates this. On average, $c(cor)=0.459$, $c(lab)=1.575$, and $c(dors)=1.62$; labials and dorsals show more bias against voiced geminates. A within-subject contrast analysis comparing coronals and the average of labials and dorsals shows a significant difference, $t(16)=6.11$, $p<.0001$.

It can be speculated that [dd] is more acceptable for Japanese speakers so when they listen to a coronal geminate in a noisy environment, they are confused about the voicing quality, and respond more or less at chance, rather than merely rejecting the possibility of a [+voice] perception.[6] On the other hand, when they are confused with the presence of voicing in labial and dorsal geminates, listeners tend to reject the possibility of [+voice] perception.

This finding about a difference between labials and dorsals versus coronals is partially replicated by Gelbart's (2005) finding that Japanese listeners are very reluctant to hear a geminate [bb], compared to [dd] (he does not test [gg]). He reports that given a continuum of a full voiced obstruent, along which closure duration is varied, a labial is not heard as a geminate unless it has very long closure. In fact, even with the longest closure (about 200ms), a labial voiced consonant is not judged as geminate 100% of the time (a bit above 80%). In addition, the reaction time for length judgment for labial voiced consonants is longer than that for coronal consonants. This is in line with my result in that there is an extra bias against [bb] (compared to [dd]) in Japanese people's perception.

That [bb] and [gg] are particularly disfavored compared to [dd] is also reflected in the lexical frequency of [dd] compared to [bb] and [gg]. According to Amano and Kondo (2000), a database based on Asahi Newspaper issues from 1985 to 1998, [dd] is much more frequent than [bb] or [gg] (based on tokens; the numbers in parentheses represent the number of types), as shown in Table 3:

Following Vowel	bb	dd	gg
a	91 (4)	493(14)	142(2)
i	0	13(3)	41(6)
u	398 (6)	0	996(147)
e	0	15(4)	0
o	1 (1)	22375(316)	22(2)
sum	490 (11)	22896 (337)	1201 (157)

Table 3: Frequency of geminate [b], [d], and [g]. The numbers in parentheses stand for type frequency. The data is from Amano and Kondo (2000).

The small number of [bb] tokens reflects the fact that a labial often fails to geminate in the environment where gemination is otherwise expected (Katayama 1998; Shirai 1999): compare *knob*, borrowed as [nobu], with *dog*, borrowed as [doggu], and *God*, borrowed as [goddo].[7] From this, it seems reasonable to posit a grammatical constraint against [bb] in Japanese, which blocks gemination of [b] and hinders the perception of [bb]. Note that this blockage of gemination of [b] (and [g]) cannot be explained in terms of lexical

[6] After the experiment, some subjects did report that coronal geminates are harder to distinguish their voicing quality compared to others.

[7] [g] is less likely to undergo gemination than [d], although it does not resist gemination as much as [b].

frequency: prior to borrowing, no voiced geminates were present in the Japanese lexicon: all of [bb], [dd], and [gg] had zero frequency.[8] Therefore, some kind of grammatical constraint must have been at work. In sum, the fact that Japanese speakers do not hear [bb] and [gg] as often as they hear [dd] might suggest that there might be an additional grammatical constraint against [bb] and [gg] - [bb] and [gg] are more marked, and hence people are more reluctant to hear them. A constraint against [dd] does exist, but the requirement is not very strong; this means that people get confused between [tt] and [dd], rather than rejecting [dd].

3.3.3. Japanese-specific markedness hierarchy?

The results discussed above is inconsistent with a purely aerodynamic view of markedness, which predicts that [bb] is least marked voiced geminate (Hayes and Steraide 2004; Ohala 1983). The size of the oral cavity is biggest for labials, and cheek muscles are susceptible to passive expansion, so maintaining voicing should be easiest during [bb]. Cross-linguistically, it is indeed observed that [bb] is more frequent than [dd] (see the works cited above). So there must be a Japanese-specific constraint against [bb]. The reason for this language-specific behavior of Japanese is yet to be explained, however.

The higher markedness of [gg] over [dd] might be attributable to the general marked scale *[gg] » *[dd], which derives from aerodynamics for the same reasons described above. The size of oral cavity is smaller for [gg] than for [dd], and its capacity to actively and passively expand is much smaller (see the references above). Therefore, it is easier to maintain voicing in coronals than dorsals, hence the extra bias against [gg].

4. Conclusion

This paper has investigated acoustic and perceptual aspects of voicing in Japanese geminates. As Japanese only recently phonemicized voicing in geminates, little work has been done on the interaction of voicing and geminacy. I have identified various phonetic cues that signal voicing in Japanese: (i) closure voicing, (ii) closure duration, (iii) V1 duration (iv) F0 and F1 of surrounding vowels. Some of these cues are weakened in the context of geminates. Most notably, closure voicing is weakened in geminates because maintaining glottal vibration during a long obstruent closure is aerodynamically hard. On the other hand, we have also seen that some other cues are enhanced in the context of geminates, perhaps to compensate for the weakening of other cues. Compared to the systematic weakening of closure voicing, however, such attempts for compensation are subject to inter-speaker variability, and the effects seem small. So from the acoustic point of view, it seems likely that voicing cues are overall weakened in geminates.

This conclusion has been supported by the results of the perceptual experiment. The experiment revealed that voicing is much harder to discriminate in geminates in general, supporting the claim in Kawahara (2005). The experiment shows that weakened cues are

[8] One exception is the resulting forms of emphatic gemination, e.g. *suggoi* from *sugoi* 'formidable'. See Kawahara (2001) for a discussion of the non-structure preserving nature of this process.

important for voicing perception in Japanese, and the attempt for compensation is not sufficient. What has not been investigated in this research, however, is which acoustic cues contribute to voicing perception to what extent. This is a topic for future research.

Furthermore, I pointed out that there is a possibility that, in addition to weakening of acoustic cues, Japanese grammar might have constraints that yield a perceptual bias against labial and dorsal voiced geminates, which also manifest themselves through the patterns of gemination. This explains the particularly poor performance in identifying voicing in [bb] and [gg], as well as the fact that Japanese listeners almost never misidentified voiceless geminates as voiced. This provides another case in which grammatical constraints may affect people's perception (e.g. Moreton 2002). Lexical frequency might also be at work, though it fails to explain the blockage of gemination in loanwords. One remaining question is why [dd] is more favored than [bb], despite that aerodynamically [bb] is predicted to be more unmarked than [dd].

One general conclusion that can be drawn from the experiments reported in this paper is that phonology is at least partially driven by phonetics. Phonologically, voicing is more easily lost in geminates than in singletons in Japanese. In light of the result of this experiment - that voicing is harder to hear in geminates - we can regard this as another case in which contrasts signaled by phonetically weak cues are phonologically more prone to neutralization, just as preconsonantal place cues are much more easily lost than prevocalic place cues. This finding adds to the growing body of literature that shows phonological neutralization is closely tied to phonetic perceptibility (Hura et al. 1992; Jun 2004; Kohler 1990; Steriade 2001). This study thus provides an additional endorsement of the claim that phonology is, at least in part, affected by phonetic factors.

Appendix: Acoustic values of the tokens used in Experiment II.

Speaker E							
	closure duration (ms)	voicing duration (ms)	V1 duration (ms)	F0 at V1 (Hz)	F1 at V1 (Hz)	F0 at V2 (Hz)	F1 at V2 (Hz)
kapa	15	54	29	286	825	311	768
kepe	15	72	35	309	533	308	546
kopo	19	63	25	290	594	306	574
kaba	57	57	39	297	714	284	723
kebe	45	45	38	301	495	298	530
kobo	45	45	31	292	420	296	502
kappa	16	132	50	295	809	259	796
keppe	12	127	52	289	542	273	539
koppo	17	133	52	304	590	278	519
kabba	31	93	50	271	811	266	742
kebbe	36	113	72	308	571	267	512
kobbo	44	123	55	285	535	255	513
kata	17	62	30	300	660	320	621
kete	17	63	42	311	516	317	515
koto	24	63	38	295	443	309	489
kada	61	61	43	273	511	288	575
kede	46	46	51	283	354	284	454
kodo	43	43	38	283	421	288	462
katta	2	134	62	299	595	261	555
kette	29	159	62	290	431	312	480
kotto	24	146	62	288	471	267	475
kadda	41	105	74	295	601	284	588
kedde	37	121	76	285	464	266	448
koddo	45	149	79	169	428	274	457
kaka	8	39	33	290	613	321	781
keke	8	42	43	298	338	346	418
koko	2	55	41	295	458	331	538
kaga	36	36	42	286	483	279	630
kege	43	43	73	285	296	287	321
kogo	51	51	58	308	421	288	481
kakka	9	129	63	308	699	263	790
kekke	5	115	69	334	374	284	399
kokko	15	120	51	327	466	273	531
kagga	34	89	61	304	609	302	592
kegge	52	130	83	287	333	285	392
koggo	40	92	52	304	471	265	527

Speaker T							
	closure duration (ms)	voicing duration (ms)	V1 duration (ms)	F0 at V1 (Hz)	F1 at V1 (Hz)	F0 at V2 (Hz)	F1 at V2 (Hz)
kapa	6	74	33	270	439	283	569
kepe	17	81	52	248	449	282	498
kopo	21	66	27	250	480	266	482
kaba	43	43	31	257	561	207	547
kebe	46	46	21	244	386	259	429
kobo	46	46	34	255	504	258	502
kappa	17	115	53	259	532	249	542
keppe	16	136	59	251	494	243	505
koppo	14	131	56	264	484	254	478
kabba	30	139	83	354	505	219	574
kebbe	38	141	84	263	410	247	471
kobbo	42	120	51	249	488	237	454
kata	0	62	30	259	595	276	582
kete	11	88	50	255	459	254	442
koto	8	55	41	258	512	284	511
kada	26	26	40	258	600	281	538
kede	38	38	63	252	450	252	450
kodo	32	32	50	278	502	273	460
katta	21	107	58	259	600	284	564
kette	14	115	68	268	536	252	472
kotto	0	144	42	257	494	277	469
kadda	40	103	74	266	613	260	528
kedde	35	93	98	262	515	249	457
koddo	24	102	90	270	523	238	465
kaka	10	70	42	255	543	278	565
keke	9	52	49	251	485	297	472
koko	0	51	58	249	431	276	512
kaga	41	41	60	253	503	268	526
kege	39	39	85	254	339	257	370
kogo	52	52	55	244	360	249	319
kakka	7	126	54	260	541	252	548
kekke	0	130	59	241	462	240	456
kokko	9	140	52	260	467	246	457
kagga	41	139	98	242	418	212	504
kegge	46	122	73	243	424	225	427
koggo	44	122	88	261	354	239	403
Speaker W							
	closure duration (ms)	voicing duration (ms)	V1 duration (ms)	F0 at V1 (Hz)	F1 at V1 (Hz)	F0 at V2 (Hz)	F1 at V2 (Hz)
kapa	2	95	33	295	828	275	820

kepe	13	77	36	344	641	335	610
kopo	32	76	26	333	642	339	488
kaba	53	53	47	313	933	320	906
kebe	50	50	64	337	623	339	593
kobo	46	46	39	321	466	338	440
kappa	19	147	47	287	855	268	776
keppe	11	149	56	328	573	286	536
koppo	10	136	61	330	388	310	553
kabba	23	121	63	300	879	259	769
kebbe	37	87	104	234	661	294	409
kobbo	36	115	85	328	670	252	474
kata	0	81	25	288	848	270	810
kete	13	61	52	318	501	335	523
koto	11	62	42	336	557	346	597
kada	33	33	42	305	857	299	730
kede	38	38	62	304	443	314	400
kodo	31	31	71	275	491	266	495
katta	7	142	34	314	893	275	795
kette	8	130	66	328	461	307	482
kotto	0	138	66	317	470	295	566
kadda	35	121	39	307	917	265	820
kedde	34	116	79	318	377	264	477
koddo	35	116	74	349	647	281	525
kaka	0	56	34	343	666	359	951
keke	0	56	47	335	442	343	406
koko	11	66	46	340	566	356	618
kaga	40	40	59	307	594	324	761
kege	29	29	88	299	411	286	410
kogo	44	44	58	280	403	272	501
kakka	0	120	38	320	777	288	907
kekke	11	124	61	353	402	336	419
kokko	12	123	59	322	443	306	533
kagga	37	116	76	323	681	262	665
kegge	27	77	89	335	771	317	428
koggo	43	132	84	328	409	277	511

References

Amano, Shigeaki and Tadahisa Kondo (2000) *NTT database series: Lexical properties of Japanese, 2nd release*. Tokyo: Sanseido.

Boersma, Paul, and David Weenink (1992) *Praat: Doing Phonetics by Computer*.

Chen, Matthew (1970) Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129-159.

Fougeron, Cecile, and Patricia Keating (1997) Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 106: 3728-3740.

Gelbart, Ben (2005) *Perception of Foreignness*. Doctoral dissertation, University of Massachusetts, Amherst.
Hay, Jennifer, Janet Pierrehumbert and Mary Beckman (in press) Speech perception, well-formedness, and the statistics of the Lexicon. To appear in *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
Hayes, Bruce and Donca Steriade (2004) Introduction: The phonetic bases of phonological markedness. In B. Hayes, R. Kirchner, and D. Steriade (eds). *Phonetically-based Phonology*. Cambridge: Cambridge University Press. pp. 1-33.
Hura, Susan, Björn Lindblom, and Randy Diehl (1992) On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35: 59-72.
Itô, Junko and Armin Mester (1996) Stem and word in Sino-Japanese. In T. Otake and A. Cutler (eds.) *Phonological Structure and Language Processing: Cross-Linguistic Studies*. Berlin: Mouton de Gruyter. pp. 13-44.
Itô, Junko and Armin Mester (1999) The phonological lexicon. In N. Tsujimura. (ed.) *The Handbook of Japanese Linguistics*. Oxford: Blackwell. pp. 62-100.
Itô, Junko and Armin Mester (2003) *Japanese Morphophonemics*. Cambridge: MIT Press.
Jaeger, Jeri (1978) Speech aerodynamics and phonological universals. *Proceedings of Chicago Linguistics Society* 311-329.
Jun, Jongho. (2004) Place assimilation. In B. Hayes, R. Kirchner, and D. Steriade (eds). *Phonetically-based Phonology*. Cambridge: Cambridge University Press. pp. 35-57.
Katayama, Motoko (1998) *Loanword Phonology in Japanese and Optimality Theory*. Doctoral dissertation, University of California, Santa Cruz.
Kawahara, Shigeto (2001) *Similarity among variants: Output-variant faithfulness*. BA Thesis, International Christian University.
Kawahara, Shigeto (2005) A faithfulness scale projected from phonetic perceptibility: The case of voicing in Japanese. Ms., University of Massachusetts, Amherst.
Kingston, John and Randy Diehl (1994) Phonetic knowledge. *Language* 70: 419-455.
Kingston, John and Randy Diehl (1995) Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (eds.) *Phonology and phonetics: Papers in laboratory phonology IV*. Cambridge: Cambridge University Press. pp. 7-27.
Kohler, K. J. (1990) Segmental reduction in connected speech: Phonological facts and phonetic explanation. In W.J. Hardcastle and A. Marchals (eds.) *Speech Production and Speech Modeling*. Dordrecht: Kluwer Publishers. pp. 69-92.
Kuroda, Shige-Yuki (1965) *Generative Grammatical Studies in the Japanese Language*. Doctoral Dissertation, MIT.
Lisker, Leigh (1957) Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33: 42-49.
Lisker, Leigh (1981) On generalizing the rapid-rabid distinction based on silent gap duration. *Haskins Laboratories Status Report on Speech Research SR-65*: 251-259
Lisker, Leigh (1986) "Voicing" in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29: 3-11.
Lovins, Julie (1973) *Loanwords and the Phonological Structure*. Doctoral dissertation, University of Chicago.
Macmillian, Neil and Douglas Creelman (1991) *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.
Maddieson, Ian (1985) Phonetic cues to syllabification. In V. Fromkin (ed.) *Phonetic*

Linguistics. London: Academic Press. pp. 203-221
McCawley, James (1968) *The Phonological Component of a Grammar of Japanese*. Hague: Mouton.
Miller, George, and Patricia Nicely (1955) An analysis of perceptual confusions among some English consonants. *Journal of Acoustical Society of America* 27: 338-352.
Moreton, Elliot (2002) Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84: 55-71.
Nishimura, Kohei (2001) Lyman's Law in loanwords. Ms., University of Tokyo.
Ohala, John (1983) The origin of sound patterns in vocal tract constraints. In P. MacNeilage (ed.) *The Production of Speech*. New York: Springer Verlag. pp 189-216.
Poser, William (1990) Evidence for foot structure in Japanese. *Language* 66: 78-105.
Raphael, Lawrence (1972) Preceding vowel duration as a cue to perception of the voicing characteristic of word-final consonants in American English. *Journal of Acoustic Society of America* 51: 1296-1303.
Raphael, Lawrence (1981) Duration and contexts as cues to word-final cognate opposition in English. *Phonetica* 38: 126-47.
Shirai, Setsuko (1999) *Gemination in Japanese loan words*. MA Thesis, University of Washington.
Slis, Imas (1986) Assimilation of voice in Dutch as a function of stress, word boundaries and sex of speaker and listener. *Journal of Phonetics* 14: 311-326.
Smith, Caroline (1995) Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (eds.) *Phonology and phonetics: Papers in laboratory phonology IV*. Cambridge: Cambridge University Press. pp. 205-222.
Steriade, Donca (2001) The phonology of perceptibility effect: The P-map and its consequences for constraint organization. Ms., University of California, Los Angeles.
Stevens, Kenneth and Sheila Blumstein (1981) The search for invariant acoustic correlates of phonetic features. In P.D. Eimas and J.L. Millers (eds.) *Perspectives on Study of Speech*. Hillsdale: Erlbaum. pp. 1-38.
Takagi, Naoyuki and Virginia Mann (1994) A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22:343-356.
Westbury, John (1979) *Aspects of the Temporal Control of Voicing in Consonant Clusters in English*. Doctoral Dissertation, University of Texas, Austin.

Department of Linguistics
South College
University of Massachusetts, Amherst
Amherst, MA 01003

kawahara@linguist.umass.edu
http://www.people.umass.edu/kawahara/