The phonetics of [voice] in singletons and geminates in Japanese: An acoustic and electroglottography study*

Shigeto Kawahara
Keio University
The Institute of Cultural and Linguistic Studies

Short title: Japanese [voice] in singletons and geminates

Postal address:
2-15-45 Mita, Minato-ku
Tokyo, JAPAN 108-8345

Tel:
(+81)-(0)3-5427-1595

Fax:
(+81)-(0)3-5427-1594

Email:
kawahara@icl.keio.ac.jp

*Thanks to Atsuo Suemitsu at JAIST for his help with EGG recording and analysis, and Mayuko Kawano and Mio Tsuchiya for their help with the acoustic analysis. I received useful feedback from Hirokai Kato on the EGG analysis. Thanks also to Donna Erickson and Michinao Matsui for comments on a previous version of this paper.

Abstract

Maintaining glottal vibration during stop closure presents an aerodynamic challenge to speakers. The intraoral air pressure rises as air goes into the oral cavity, which makes it difficult to sustain transglottal air pressure drop to maintain glottal vibration. This problem is particularly challenging for geminates, because they have long closure. Japanese phonology thus has lacked voiced stop geminates from its inventory until recently. However, recent loanwords do have some words with voiced geminates. This paper explores the phonetic nature of this newly-created contrast with an acoustic study as well as with electroglottography (EGG), the primary question being how Japanese speakers implement the aerodynamically challenging contrast, when the contrast is new and its role is limited in loanword phonology. The EGG results show that Japanese speakers give up maintaining glottal vibration during geminate closure; i.e. Japanese voiced geminates are semi-devoiced. However, the acoustic results show that Japanese speakers nevertheless maintain the phonological [voice] contrast in geminates in terms of other acoustic dimensions, such as closure duration and F1 in the following vowel. The current study thus shows that Japanese speakers have now established a [voice] contrast in geminates, but not to the extent that they overcome the aerodynamic problem to sustain full vocal fold vibration during geminate closure. Furthermore, a cross-linguistic comparison with Arabic supports the view that speakers implement a contrast with more information (i.e. higher entropy) more robustly (Aylett & Turk, 2004, 2006; Cohen-Priva, 2012, 2015; Hume, 2016; Shaw et al., 2014).

1 Introduction

1.1 Motivations of the study

Maintaining glottal vibration during stop closure presents an aerodynamic challenge to speakers. Intraoral air pressure rises as voicing continues, because air from the lungs flows into the oral cavity and is trapped inside the intraoral cavity with stop closure. The rise in intraoral air pressure makes it difficult to maintain transglottal air pressure drop that is necessary to sustain glottal vibration (Hayes, 1999; Jaeger, 1978; Kingston, 2007; Ohala, 1983; Ohala & Riordan, 1979; Westbury, 1979; Westbury & Keating, 1986). Speakers therefore need to expand their oral cavity to counteract the rise in the intraoral air pressure (Ohala, 1983; Ohala & Riordan, 1979). This aerodynamic challenge is particularly difficult to deal with in long consonants—also known as geminates—because speakers face this aerodynamic challenge for a long period of time. For this reason, there are a number of languages in which a [voice] feature is not contrastive in geminates, while the feature is contrastive in singletons (Hayes & Steriade, 2004; Jaeger, 1978; Ohala, 1983; Podesva, 2000, 2002).[1]

Such was the case in Japanese, until recent loanwords from English and other languages resulted in voiced geminate consonants (Kawahara, 2006). In native words, there are no voiced obstruent geminates, and when a phonological gemination process targets a voiced obstruent, it creates a nasal-obstruent cluster (Ito & Mester, 1995, 1999; Kuroda, 1965). For example, the suffix /-ri/ causes gemination of the last consonant of a root (e.g. /uka-ri/ → [u**kk**ari] 'absent-mindedly'), but it creates a nasal-stop cluster in a word like /zabu+ri/ → [za**mb**uri] 'splashing'. Syncope which occurs in Sino-Japanese words (e.g. /hat**u**+tatu/ → [ha**tt**atu] 'development') is blocked when it would result in a voiced geminate (e.g. /ha**tu+b**ai/ → [ha**tub**ai], *[ha**bb**ai]) 'start being sold' (Ito & Mester, 1996; Kurisu, 2000).

In recent loanwords from English and other languages, however, Japanese speakers often borrow word-final consonants as geminates (e.g. /bakku/ for *back*). This gemination process has resulted in voiced geminates; e.g. /be**dd**o/ 'bed' and /e**gg**u/ 'egg' (Katayama, 1998; Kubozono et al., 2008; Shirai, 2002). Although this gemination process is not exception-less, Shirai (2002) found that word-final [d] is borrowed as geminates more than 50% the time, and word-final [g] is geminated about 50% of the time. We thus now witness a minimal pair like [re**dd**o] 'red' vs. [re**tt**o] 'let' in Japanese loanwords, although, to reiterate, such a lexical contrast is limited to recent loanwords.[2]

[1] Some languages, including Japanese as we will see below, have a [voice] contrast in geminates, but some of them do not maintain full glottal vibration during geminate closures (e.g. Tashlhiyt Berber: Ridouane 2010).

[2] Emphatic gemination, which is generally not structure-preserving, can also create voiced geminates, even in native words; e.g. /hi**dd**oi/ 'very awful' from /hi**d**oi/ 'awful' (Kawahara & Braver, 2014). This process however does not create a lexical contrast.

One question that is addressed in this paper is how Japanese speakers implement this newly-created [voice] contrast. This general question—how a new phonological contrast in loanwords is implemented phonetically—is in and of itself interesting, but we can view this question from another perspective. Since the [voice] contrast in geminates is contrastive only in loanwords (Ito & Mester, 1995, 1999), when we consider Japanese phonology as a whole, its functional load is very low; in other words, there are not very many minimal pairs that are distinguished in terms of a geminate [voice] contrast in Japanese. To illustrate, the frequency counts of /t/, /d/, /tt/, /dd/ from a Japanese corpus (Amano & Kondo, 2000) are provided in Table 1.

Table 1: Token and type frequencies of /t/-/d/ and /tt/-/dd/ in Japanese. Based on Amano & Kondo (2000).

	Sounds	Counts	Sounds	Counts	Shannon's Entropy
Token	/t/	6,166,896	/d/	1,986,985	0.8 bits
	/tt/	478,525	/dd/	7,727	0.12 bits
Type	/t/	4,615	/d/	2,366	0.92 bits
	/tt/	469	/dd/	42	0.41 bits

The frequency of /dd/ with respect to /tt/ is much smaller than the frequency of /d/ with respect to /t/. Therefore, there is a sense in which the [voice] contrast in geminate is still "not so important", compared to the [voice] contrast in singletons, as the former contrast does not serve to make many lexical contrasts in the whole phonology of Japanese.[3] We can capture this intuition that the [voice] contrast in geminates is not very informative by resorting to Shannon's entropy (Shannon, 1948). Shannon's entropy is calculated as $-\sum P(A) log_2 P(A)$ where A is a set of possible events (here "voiced" or "voiceless"), and can be taken as a measure of informativity. For a binary contrast, Shannon's entropy varies from 0 (no information) to 1, and the higher values indicate more informativity.

Based on token frequency counts in Table 1, the [voice] contrast in the /t/-/d/ pair is 0.8 bits, whereas the [voice] contrast in the /tt/-/dd/ pair is only 0.12 bits (Kawahara, 2016). This means that a [voice] contrast is more unpredictable and more informative in singleton pairs than in geminate pairs. Recent studies show that speakers implement a phonological contrast with higher entropy (i.e. more information) more robustly (Aylett & Turk, 2004, 2006; Cohen-Priva, 2012, 2015; Hall et al., 2016; Hume, 2016; Shaw et al., 2014); for example, Aylett & Turk (2004) show that in English, more predictable vowels are shorter

[3]This idea is expressed in more formal terms by Rice (2006), although she treats this as a matter of a categorical "yes-contrastive" vs. "not-contrastive" distinction.

and more centralized. Given this observation, the prediction is that Japanese speakers may not implement the [voice] contrast in geminates so robustly.

There is yet another sense in which exploring the phonetics of the geminate [voice] contrast is interesting. Nishimura (2006) points out that voiced geminates can optionally devoice when there is another voiced obstruent within the same stem. For example, /be**dd**o/ 'bed' can be pronounced as [be**tt**o], although this neutralization is optional (Kawahara & Sano, 2013; Sano & Kawahara, 2013) (see also Kawahara 2015a for a recent overview). Previous studies argue that this devoicing is triggered by a phonotactic restriction, called Lyman's Law (henceforth LL), which prohibits two voiced obstruents within the same morpheme (*{[+voice]...[+voice]}$_{morpheme}$: Ito & Mester 1995; Lyman 1894; Vance 2007). Kawahara (2006) argues that this neutralization is a case of "perceptually tolerated articulatory simplification" (Hura et al., 1992; Johnson, 2003; Kohler, 1990; Steriade, 2008): neutralizing the [voice] contrast in geminates is perceptually not conspicuous, and hence tolerated by the phonology of Japanese.

In summary, the current study investigates the phonetics of voiced geminates, which have three properties: (i) maintaining glottal vibration during geminate stops is aerodynamically challenging; (ii) its entropy—or informativity—is still low in the whole Japanese lexicon; (iii) it can be neutralized by an optional phonological process in Japanese phonology.

1.2 Previous studies

There have been a few studies of the phonetics of geminate voicing in Japanese. Kawahara (2006) recorded three female speakers, producing [kVC(C)V] where V ∈ [a, e, o], and C(C) ∈ [p, t, k, b, d, g, pp, tt, kk, bb, dd, gg]. Kawahara (2006) found, among others, that voiced geminates are semi-devoiced, showing on average about 40% of glottal vibration during closure (see also Hirose & Ashby 2007 for the same finding), and that perceptually, the [voice] contrast is less perceptible in geminates than in singletons.

In that study, some stimuli were real words and some were nonce words, and the ratio between real words and nonce words was not controlled. Neither does Kawahara (2006) report individual speakers' behavior.[4] The current study thus updates Kawahara (2006) by using real words only, and, more importantly, by measuring vocal fold vibration during closure using electroglottography (EGG), which is demonstrably a better device to use than an acoustic analysis to measure glottal vibration (for discussion, see Baer et al. 1983; Childers et al. 1984; Rothenberg 1981; Roubeau et al. 1987—see also Figure 5 below, which illustrates that EGG may provide a more conservative estimate of actual glottal vibration.). This study

[4]The individual differences in Kawahara (2006) were reported in Kawahara (2005), a much less widely circulated paper which was published in a working paper volume.

also balances gender by including male speakers of Japanese. Finally, this study examines whether there are any tangible phonetic differences in terms of Lyman's Law's violations. Including this factor was motivated by the phonological observation that voiced geminates can be devoiced by Lyman's Law (Kawahara, 2015a; Nishimura, 2006)—would we observe any phonetic differences among those consonants that can potentially devoice and those that do not?

There has also been some work on voiced geminates in dialects where voiced geminates do occur in their native vocabulary (Matsuura, 2012; Takada, 2011, 2013). In addition to the fact that these dialects are very different from the Standard Tokyo Japanese, these studies focused only on voicing duration closure, and do not consider other phonetic correlates of [voice] contrast. This study combines acoustic analyses with EGG analysis to explore the nature of the geminate [voice] contrast in Standard Tokyo Japanese.

2 Method

The aim of the current experiment is to explore how the [voice] contrast is implemented in singletons and geminates in Japanese. In addition, this experiment explores whether the presence of another voiced obstruent, which can trigger optional phonological devoicing of geminates, affects the implementation of the phonetics of geminates.

2.1 Stimuli

The experiment had six conditions: (i) voiced geminates that occur with another voiced obstruent (i.e. those that violate Lyman's Law, henceforth LL); (ii) voiced geminates that do not violate LL; (iii) voiced singletons that violate LL; (iv) voiced singletons that do not violate LL; (v) voiceless geminates; and (vi) voiceless singletons. Table 2 provides the list of the stimuli. All the stimulus words are existing loanwords, and all of them were disyllabic. Each condition had nine items: six of them had a coronal target stop (/t, d, tt, dd/); three of them had a dorsal target stop (/k, g, kk, gg/). There were no stimuli with a labial target consonant, because [bb] rarely occurs even in loanwords (Katayama, 1998; Kawahara, 2006; Shirai, 2002). All the stimuli had accent on the initial syllable, with H tone on the first syllable and L tone on the second syllable (Kawahara, 2015c). In the singleton conditions ((iii), (iv) and (vi)), the vowels preceding /d/ or /t/ were long or diphthongs. This choice was inevitable, since coronal gemination is very common when the preceding vowel is short (Katayama, 1998; Shirai, 2002).

Table 2: The list of stimuli.

(i) LL-violating geminates		(ii) Non-LL-violating geminates		(iii) LL-violating singletons	
baddo	'bad'	heddo	'head'	gaido	'guide'
beddo	'bed'	reddo	'red'	zoido	'(toy name)'
deddo	'dead'	uddo	'wood'	baado	'bird'
guddo	'good'	kiddo	'kid'	boodo	'board'
goddo	'God'	maddo	'mad'	gaado	'guard'
baggu	'bag'	eggu	'egg'	dagu	'Doug'
biggu	'big'	reggu	'leg'	bagu	'bug'
doggu	'dog'	taggu	'tag'	ʤogu	'jog'
(iv) non-LL-violating singletons		(v) Voiceless geminates		(vi) Voiceless singletons	
muudo	'mood'	katto	'cut'	aato	'art'
waido	'wide'	kitto	'kit'	ooto	'auto'
roodo	'road'	nitto	'knit'	kaato	'cart'
riido	'lead'	matto	'mattress'	kooto	'coart'
huudo	'food'	metto	'helmet'	ʃiito	'seat'
hagu	'hug'	makku	'mac'	maiku	'mic'
magu	'mug'	sakku	'sack'	reiku	'lake'
ragu	'rag'	tʃekku	'check'	piiku	'peak'

2.2 Recording

The experiment took place in a sound-attenuated recording room in Japan Advanced Institute of Science and Technology (JAIST), Kanazawa, Japan. EGG (Portable Electro-Laryngograph: Laryngograph Ltd) was used to detect glottal vibration during closure of target stops. Each sensor was placed near the vocal folds, as illustrated in Figure 1. EGG detects vocal fold contact by recording changes in the electrical impedance of the larynx (Baer et al., 1983; Childers et al., 1984; Rothenberg, 1981; Roubeau et al., 1987). The signal was boosted with its built-in amplifier and recorded to an audio recorder PMD671 (Marantz).

Figure 1: Placement of EGG sensors.

The utterances were simultaneously recorded with a shot-gun microphone (NTG-3: RODE), amplified via AR501 (FOSTEX), with 48k sampling rate and 16 bit quantization level. The recording was conducted in a stereo format, where one channel recorded acoustic signals and the other EGG signals. Input gains of each channel were adjusted to an appropriate level with the Marantz recorder. The recording took place after an EMA experiment, which was for a different project. EMA sensors were removed before the recording of the current experiment.

2.3 Speakers

Four speakers participated in this experiment. Speakers 1 and 4 were male, and Speakers 2 and 3 were female. They were all in their 30's at the time of the experiment. All speakers spoke Standard Japanese in daily life. None of the participants reported a history of speech or hearing disorder.

2.4 Procedure

The stimuli were presented in a randomized order using Superlab (Cedrus Corporation, 2010) on a mac laptop machine. The stimuli were presented in isolation in the *katakana* orthography, which is the standard orthography for loanwords. Within each block, all the stimuli were presented once. The speakers went through five blocks of trials. The order was re-randomized for each repetition.

2.5 Analysis

This study analyzed the following measures, based on previous studies of acoustic manifestations of a [voice] contrast across different languages (Kawahara 2006; Kingston & Diehl 1994; Lisker 1986 and see below for more).

(1) Measures analyzed in this study
   a. Vocal fold vibration duration based on the EGG signal.
   b. Closure duration based on the acoustic signal based on spectrograms.
   c. Percentage of (a) with respect to (b).
   d. F0 of the following vowel after coronal consonants.
   e. F1 of the following vowel after coronal consonants.

These properties are known to be acoustic correlates of a [voice] contrast in several languages. Voiced stops involve longer glottal vibration during closure than voiceless stops (Stevens &

Blumstein, 1981). Voiced stops are known to be shorter than voiceless stops (Kluender et al., 1988; Pickett, 1980; Port & Dalby, 1982). We can also calculate glottal vibration duration percentage with respect to closure duration, which represent how much of the closure involves vocal fold vibration. More of the closure should be voiced for voiced stops than for voiceless stops. F0 and F1 are often lower next to voiced stops than voiceless stops (Caisse, 1982; Diehl & Molis, 1995; Kingston & Diehl, 1994, 1995).

There are other properties that are known to co-vary with a [voice] contrast, such as preceding vowel duration (Chen, 1970; Port & Dalby, 1982; Raphael, 1972). Preceding vowel duration was not measured in this study, because neither phonological length nor quality is constant across the stimuli. (Recall that this was inevitable because the current experiment used real words.) F0 and F1 of the preceding vowels were not measured for the same reason. F0 and F1 of the following vowels were measured only after coronal consonants, because all the following vowels were [o] (this vowel is epenthetic—see Table 2). F0 and F1 after dorsal consonants were not measured, because the vowels were [u], which was devoiced after [k] (see e.g. Fujimoto 2015; Tsuchida 1997). Recall that there were six items for each condition containing coronal target stops.

For the acoustic analysis, Praat was used to place segmental boundaries between the target consonants and surrounding vowels (Boersma, 2001). The onset of the target consonants were placed based on the weakening of energies in the high formants and the energies in the waveform, as illustrated in Figure 2. The onset of the following vowel was determined as the point when the periodic energy starts after the target consonants.

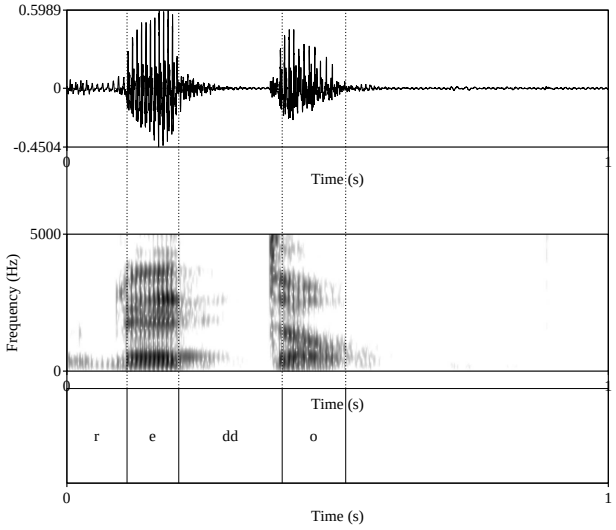

Figure 2: Annotation with Praat based on the acoustic signal. The token is /reddo/ by Speaker 4. The time scale is 1 sec.

Based on this annotation, closure duration was extracted for all target consonants. A 20 ms analysis window was created 10 ms after the offset of the target consonants. Average F0 and F1 values within this analysis window were calculated. F0 was measured using the autocorrelation method, with the "very accurate option". F1 was based on the burg method.

Using the word segmentation based on acoustic analysis, vocal fold vibration duration closure was measured using the EGG signal, as shown in Figure 4. The left edge of the voicing interval in the EGG signal is aligned with the left edge of the target consonants in the acoustic signal. The right edge of the voicing in the EGG signal was determined based on the presence of energy in the waveform and spectrogram of the EGG signal. The left figure shows a sample of a geminate [dd], and the right figure shows a sample of a singleton [d]. Voiced geminates are typically semi-devoiced, as a number of previous acoustic-based studies have found (Hirose & Ashby, 2007; Kawahara, 2006). In such cases, the right edge of the vocal fold vibration interval was determined as the point when the EGG energy dies out. For voiced singleton stops, they were almost always fully voiced, so that their right edge was aligned with their right edge of the closure, determined based on the acoustic signal.[5]

[5]There were two tokens of voiced geminates in which glottal vibration stopped during the closure, but revived before the release (Figure 3). In this case, the durations of these two intervals were summed. Takada (2011, 2013) reports that this sort of "revival" of vocal fold vibration can occur in Japanese.

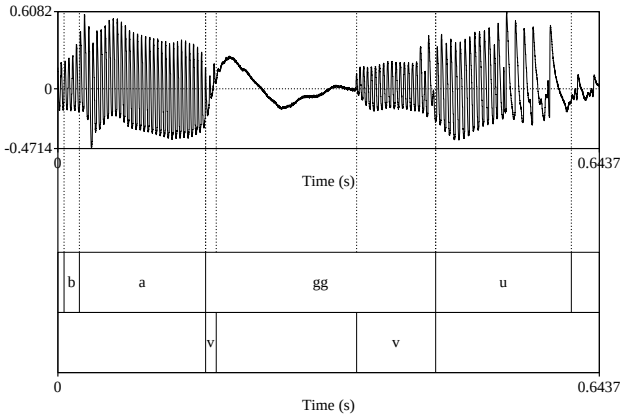

Figure 3: A case in which glottal vibration dies out once and revives during closure. The interval annotated as "v" (bottom) represents the vocal fold vibration interval in the EGG signal.

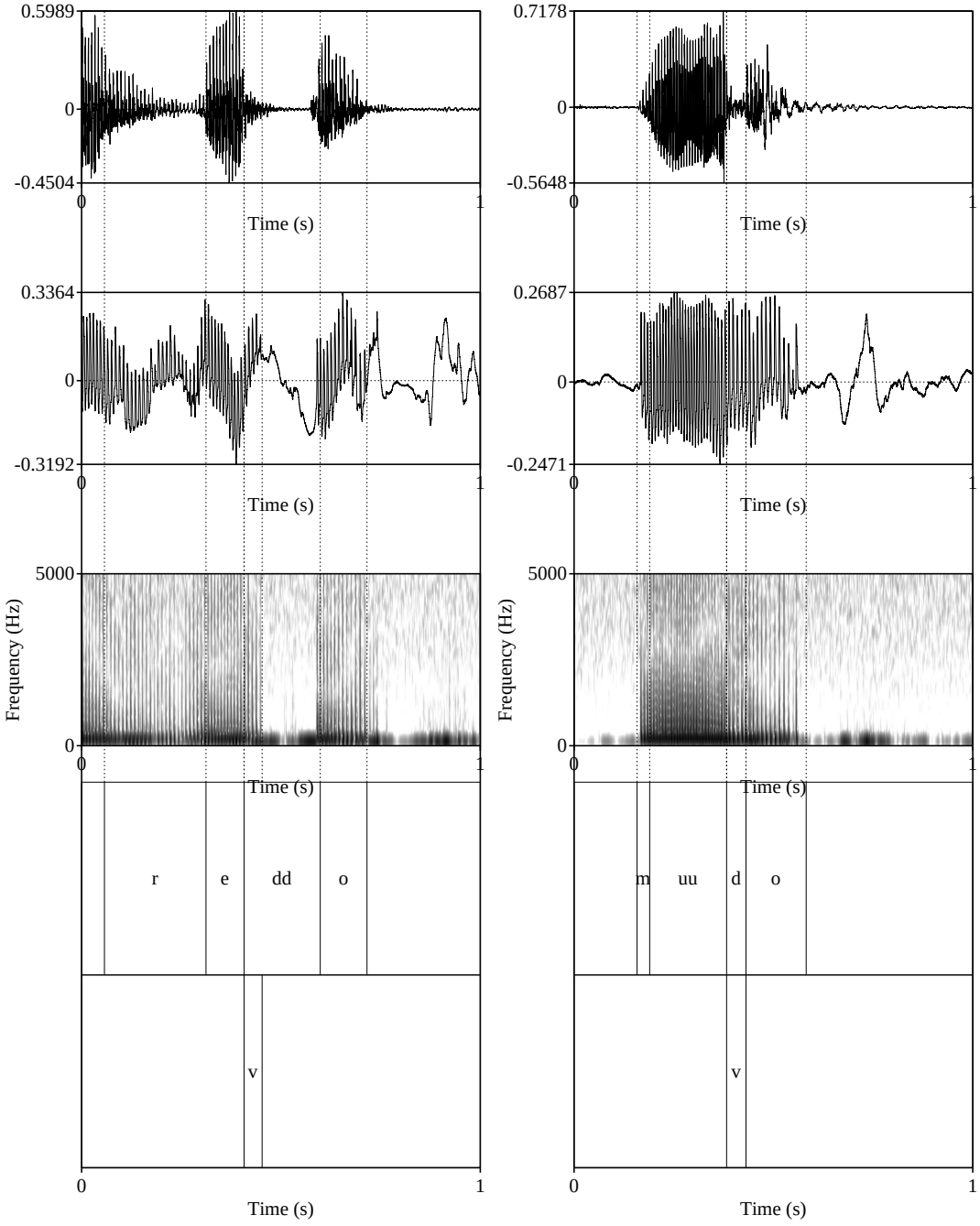

Figure 4: Illustration of estimation of glottal vibration for [dd] and [d] using EGG signal, indicated as "v". The top panel shows the acoustic waveforms, and the second panel shows the EGG signals, the third panel shows the spectrograms of the EGG signals. The left figure=geminate [dd] in [reddo]; the right figure=singleton [d] in [muudo]. The left edge of the voicing interval was always aligned with the left edge of the target consonant determined by the acoustic signal. The time scale is 1 sec. See Figure 5 for more on the comparison between the acoustic signal and the EGG signal.

Figure 5 compares the acoustic and EGG signals of /reddo/. Voicing, as indicated in the acoustic spectrogram (top panel) by the "voice bar" in the lower frequencies, as well as the continuation of formants in the upper frequencies, is longer than the glottal vibration detected in the EGG signal (the middle panel). This lingering of the voicing information in the acoustic signal may be due to a type of reverberation. Studying voicing during closure using acoustic information may overestimate the actual vocal fold vibration duration. This comparison highlights the importance of studying vocal fold vibration using EGG.[6]

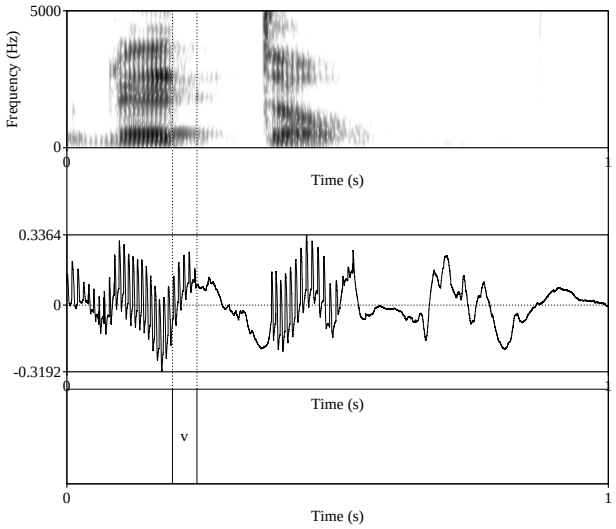

Figure 5: Comparison of the acoustics signal (top) and EGG signal (middle) of the same token of /reddo/. The interval annotated as "v" (bottom) represents the vocal fold vibration interval in the EGG signal. The time scale is 1 sec.

2.6 LL-driven phonological devoicing?

Overall, based on the auditory impression, none of the speakers apparently devoiced LL-violating voiced singletons or voiced geminates. This is not too surprising, because the LL-driven devoicing is an optional process (Kawahara & Sano, 2013; Sano & Kawahara, 2013), and speakers in general do not often apply such an optional process in lab speech. See the result section for more on this point.

[6]It should be noted, however, that it is possible to achieve glottal vibration without completely closing vocal folds (Kawahara et al., 1999), which can take place for some women's speech, especially when they are speaking in a soft voice (Hiroaki Kato, p.c., 2016). In such cases, EGG would underestimate actual vocal fold vibration. EGG-based estimates would nevertheless be more conservative.

2.7 Statistics

Since the factors are not fully crossed (i.e. LL-violation is irrelevant for voiceless stops), two linear-mixed models were fit to each dependent measure, with speaker, item, and repetition as random variables (Baayen et al., 2008; Baayen, 2008; Bates, 2005). The first model excluded LL-violating words and tested the effect of the singleton/geminate difference and the voiced/voiceless difference on each type of measurement. The second model excluded voiceless consonants and tested the effects of LL-violation and the singleton/geminate difference. All statistical computation was implemented in R (R Development Core Team, 1993–2016) with the `lme4` package (Bates et al., 2011), which was also used to create the result figures. The significance of the fixed effects was calculated by the Markov chain Monte Carlo method using the `pval.fnc()` function in the `languageR` package (Baayen, 2009).

3 Results

We first begin with the vocal fold vibration duration, which is arguably the most important cue for a phonological [voice] contrast (Lisker, 1978; Parker et al., 1986; Raphael, 1981; Stevens & Blumstein, 1981). Figure 6 shows vocal fold vibration duration, measured based on the EGG signals. In this and following figures, the following legends are used to represent the six conditions: (i) /D..DD/=LL-violating geminates, (ii) /...DD/=non-LL-violating geminates, (iii) /D..D/=LL-violating singletons, (iv) /..D/=non-LL-violating singletons, (v) /TT/=voiceless geminates, and (vi) /T/=voiceless singletons (where D represents a singleton voiced stop; DD represents a geminate voiced stop; T(T) represents voiceless stops). The error bars indicate 95% confidence intervals here and throughout.

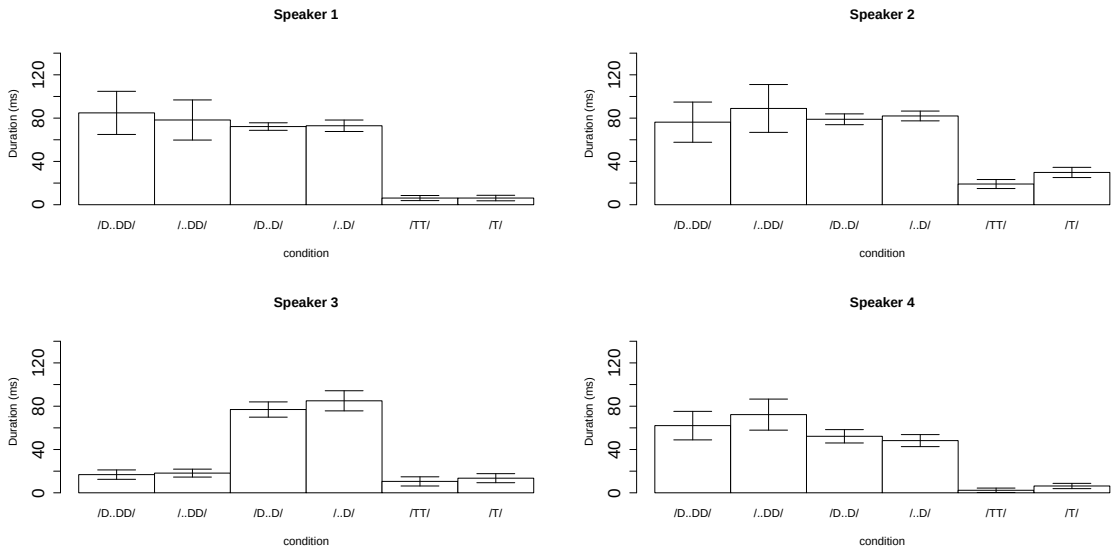

Figure 6: Vocal fold vibration duration (ms) measured based on the EGG signals. The error bars represent 95% confidence intervals.

Voiced stops generally show longer vocal fold vibration duration than voiceless stops. The first linear-mixed model, which excludes the first and the third bars, examines the effect of [voice] and geminacy, and shows that [voice] has a significant impact on vocal fold vibration duration ($t = 9.4, p < .001$), but geminacy does not ($t = -0.7, n.s.$); no significant interaction was found ($t = -0.3, n.s.$). The second linear mixed model comparing the first four bars shows that none of the factors are significant (geminacy: $t = -0.34, n.s.$; LL: $t = -0.1, n.s.$; interaction: $t = -0.4, n.s.$). The results thus show that voiced stops generally show longer vocal fold vibration than voiceless stops, and in terms of absolute vocal fold vibration duration, voiced geminates and voiced singletons are comparable. The results imply that voiced geminates are semi-devoiced, having vocal fold vibration duration that is only as long as voiced singletons.

A closer inspection of the patterns in Figure 6 shows that Speaker 3 gives up vocal fold vibration in geminates so that vocal fold vibration in "voiced geminates" is of comparable duration to vocal fold vibration in voiceless stops. A post-hoc test examining the first four bars of Speaker 3 shows that indeed the effect of geminacy is significant ($t = -13.8, p < .001$), suggesting that this speaker has shorter vocal fold vibration in voiced geminates than in voiced singletons. Nevertheless, comparing voiced geminates and voiceless stops reveals that the former has slightly longer vocal fold vibration than the latter ($t = 2.4, p < .05$)

For Speakers 1 and 2, voiced geminates and voiced singletons seem to show comparable duration of vocal fold vibration, despite the former having long closure duration. Speaker 4 shows slightly longer vocal fold vibration during closure for geminates than for singletons,

which a post-hoc test reveals to be significant ($t = 2.0, p < .05$).

Overall, the sizes of error bars show that voiced geminates show more variation than voiced singletons. The effect of LL seems negligible both for geminates and singletons, confirming the auditory impression that the speakers did not devoice LL-violating geminates.

Next, Figure 7 shows closure duration of each type of stop. Expectedly, geminates are longer than singletons (Homma, 1981; Hirata & Whiton, 2005; Idemaru & Guion, 2008; Idemaru & Guion-Anderson, 2010; Kawahara, 2015b). Voiceless stops are longer than voiced stops (Homma, 1981; Kawahara, 2006).

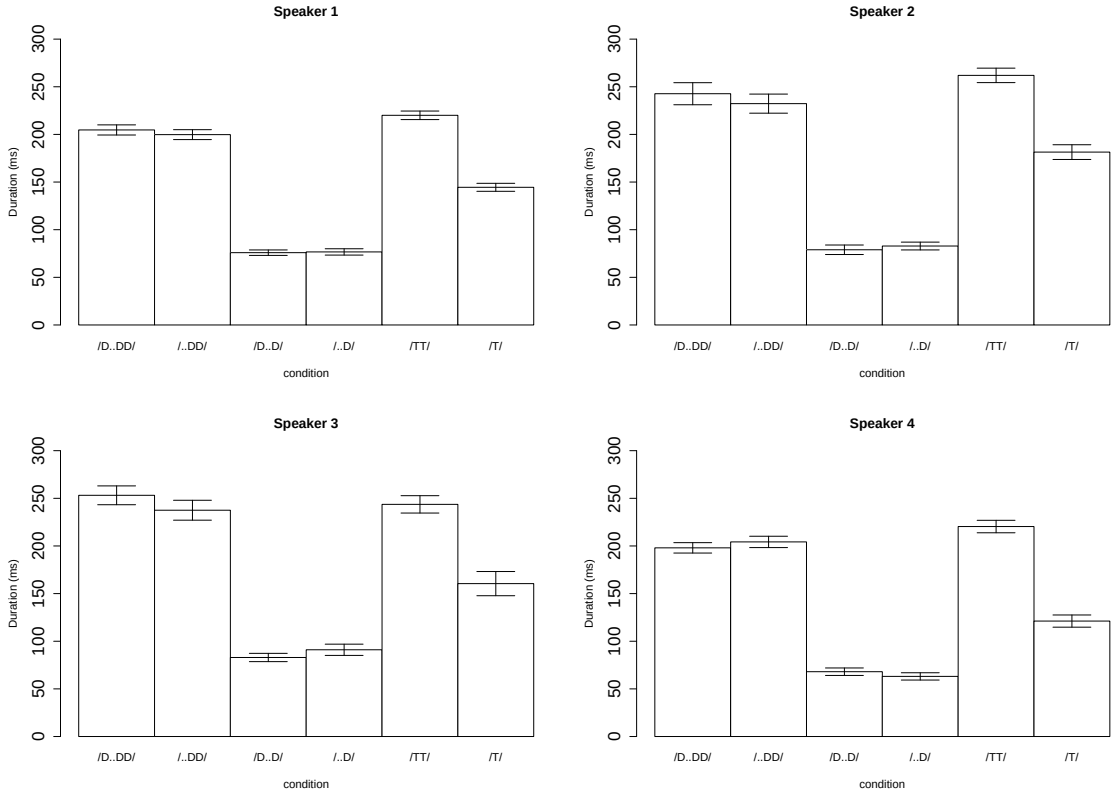

Figure 7: Closure duration (ms).

The first linear-mixed model, which excludes the first and third bars, shows that geminates are longer than singletons ($t = 17.1, p < .001$), and voiceless stops are longer than voiced stops ($t = -15.1, p < .001$). The interaction was significant ($t = 8.04, p < .001$), because the difference between voiced stops and voiceless stops is more pronounced in singleton pairs than in geminate pairs. The second linear-mixed model, which compares the first four bars, confirms that geminates are longer than singletons ($t = 30.9, p < .001$), but LL has no effects ($t = -0.1, n.s.$); the interaction was not significant ($t = 1.0, n.s.$).

Figure 8 shows the percentage of vocal fold vibration with respect to closure duration.

Voiced geminates are voiced about 40% of the closure, except for Speaker 3, who shows only less than 10% of vocal fold vibration during closure. Voiced singletons seem to be fully voiced, except for Speaker 4 who shows only about 80% of vocal fold vibration. Voiceless stops show a little bit of vocal fold vibration, i.e., "a leakage of voicing" from the preceding vowels.

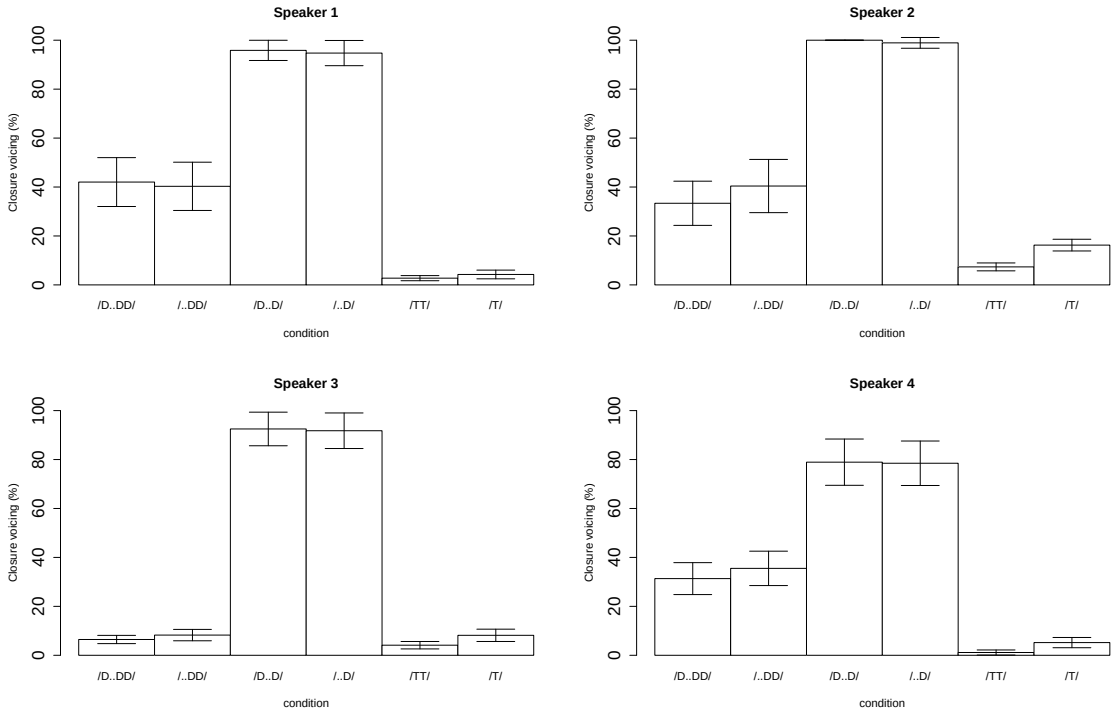

Figure 8: vocal fold vibration duration percentage (%).

The first linear mixed model, excluding the first and third bars, shows that there is no effect of geminacy ($t = -1.3, n.s.$), but it significantly interacted with [voice] ($t = -11.1, p < .001$), which itself is significant ($t = 23.6, p < .001$). These results show that voiced stops have larger portions of glottal vibration during closure than voiceless stops, and singletons more so than geminates. The second linear mixed model confirms that singleton voiced stops have more of their closures voiced than geminate voiced stops ($t = -13.9, p < .001.$), but LL had no impact ($t = 0.1, n.s.$). The interaction was not significant either ($t = -0.5, n.s.$).

Recall that Speaker 3 had only slightly longer glottal vibration duration for voiced geminates than for voiceless stops (Figure 6). When relativized with respect to closure duration, there was no significant difference between voiced geminates and voiceless stops ($t = 0.9, n.s.$).

Figure 9 shows the results of F0. There does not seem to be a clear effect of [voice] on the F0 of the preceding vowels in any of the speakers. The first linear mixed model shows no

effects of geminacy ($t = 1.3, n.s.$), [voice] ($t = 0.4, n.s.$), or their interaction ($t = 0.1, n.s.$). The second linear mixed model shows no significant effects either (geminacy: $t = 1.2, n.s.$; LL: $t = -0.5, n.s.$; interaction: $t = 1.3, n.s.$).

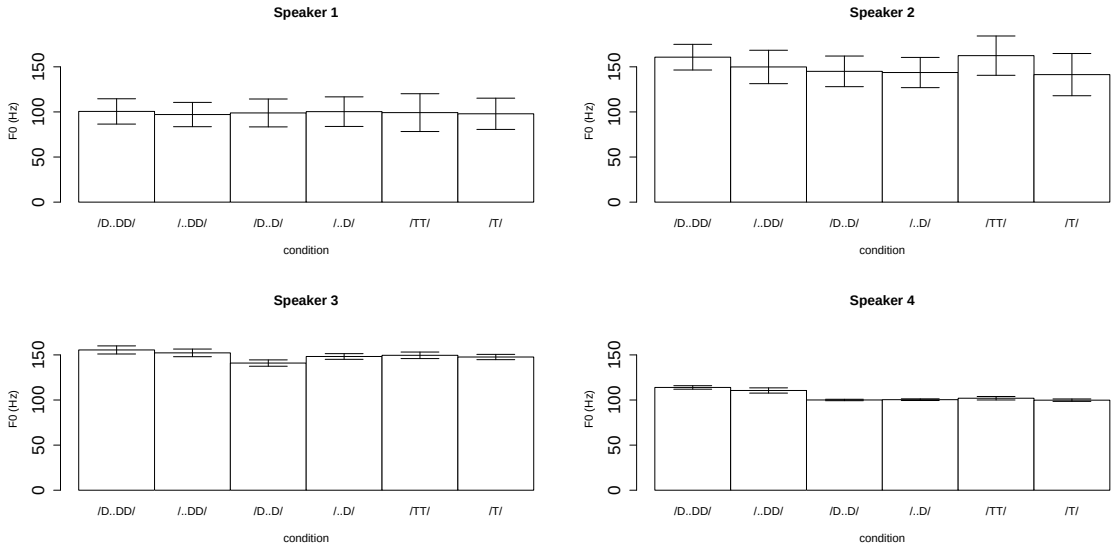

Figure 9: F0 (Hz).

This result was unexpected, but there is a language like Tamil, which does not show a difference in F0 next to voiced/voiceless consonants (Kingston & Diehl, 1994). However, Kawahara 2006 did find depression of F0 after voiced stops in Japanese. Even more unexpectedly, it seems that Speaker 4 has higher F0 after voiced geminates than the other consonants. A post-hoc comparison of voiced geminates and voiceless geminates turned out to be significant ($t = 5.9, p < .001$). It is not clear why voiced geminates raise F0 in the following vowel, and only for this speaker.

The F0 values in the current experiment are all relatively low for all speakers (below 150 Hz), and this is because the measurements were done on L-toned syllables. Setting aside the behavior of Speaker 4, it could be that in the current experiment, since the target syllables are L-toned utterance-final syllables, F0 values may have been at the floor (see Kawahara & Shinya 2008 for final-lowering in Japanese); i.e. the F0 values were so low that they did not allow the effect of F0 depression due to voiced stops.

Figure 10 shows the results of F1. Generally, F1 is markedly higher after voiceless stops than after voiced stops. The first linear mixed model shows that [voice] has a significant impact on F1 ($t = -6.04, p < .001$). There was a significant effect of geminacy as well ($t = 5.1, p < .001$), which is probably due to the fact that /tt/ shows higher F1 than /t/, a trend that is clearly observed in Speakers 1, 2 and 4. The interaction was significant as

well ($t = -3.99, p < .001$), because there do not seem to be any clear differences between /dd/ and /d/. Within voiced stops, there were no effects of geminacy ($t = -0.5, n.s.$), LL ($t = -1.1, n.s.$), or its interaction ($t = 1.5, n.s.$).

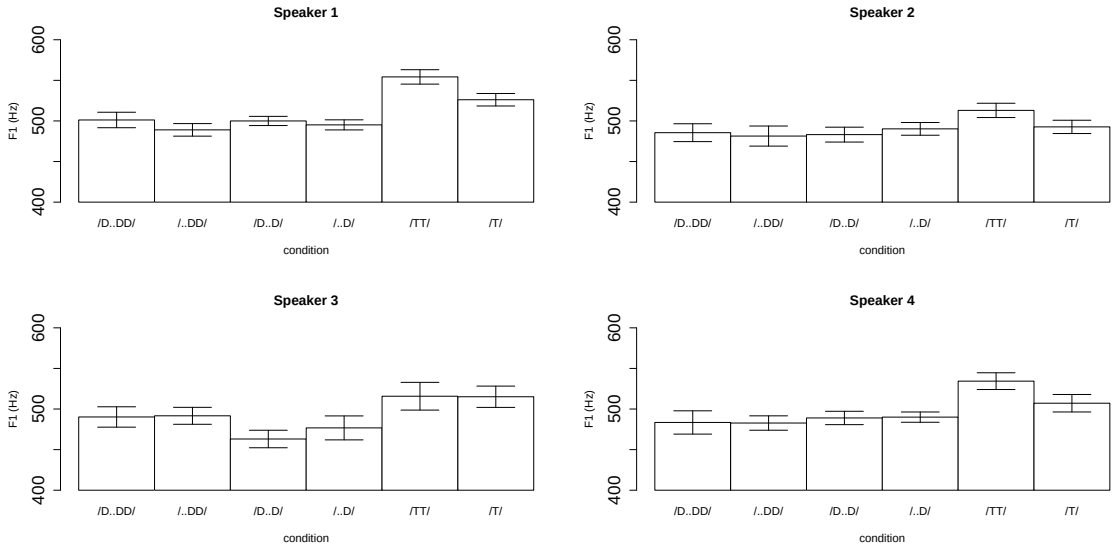

Figure 10: F1 (Hz).

4 Discussion

4.1 Summary

We have observed that in terms of absolute duration, voiced geminates and voiced singletons have comparable durations of glottal vibration during closure. Because voiced geminates are longer than voiced singletons, when viewed in terms of vocal fold vibration duration percentage with respect to whole closure duration, Japanese voiced geminates are semi-devoiced. Nevertheless, vocal fold vibration is longer for voiced geminates than for voiceless geminates. Moreover, all the speakers maintain the phonological [voice] contrast in geminates in other acoustic dimensions, such as closure duration and F1 of the following vowels. There were no clear effects on F0, except for the unexpected pattern observed in Speaker 4.

Throughout this experiment, no clear effects of Lyman's Law were found. The lack of Lyman's Law implies two things: (i) the participants of this experiment did not apply LL-driven geminate devoicing, which is not too surprising given that this devoicing is an optional process, which can be blocked in lab speech; (ii) Lyman's Law does not affect the phonetic implementation of [voice] itself. The second point further suggests that Lyman's Law is a *phonological* constraint, not a phonetic constraint.

4.2 A brief cross-linguistic comparison

As discussed in the introduction, maintaining glottal vibration during geminate closure is aerodynamically challenging. Furthermore, voiced geminates are contrastive only in a subset of the Japanese lexicon (i.e. loanwords), and hence they do not carry high functional loads. It is therefore not too surprising that Japanese speakers do not attempt to maintain full vocal fold vibration during geminate closure. This does not mean, however, that maintaining vocal fold vibration during geminate closure is physically or physiologically impossible, because speakers can expand their oral cavity to keep the intraoral air pressure low, thereby maintaining vocal fold vibration during geminates (Ohala, 1983; Ohala & Riordan, 1979). Figure 11 shows a sample spectrogram of four languages (Arabic, Hindi, Norwegian, and Swedish), all of which show full vocal fold vibration during geminates.

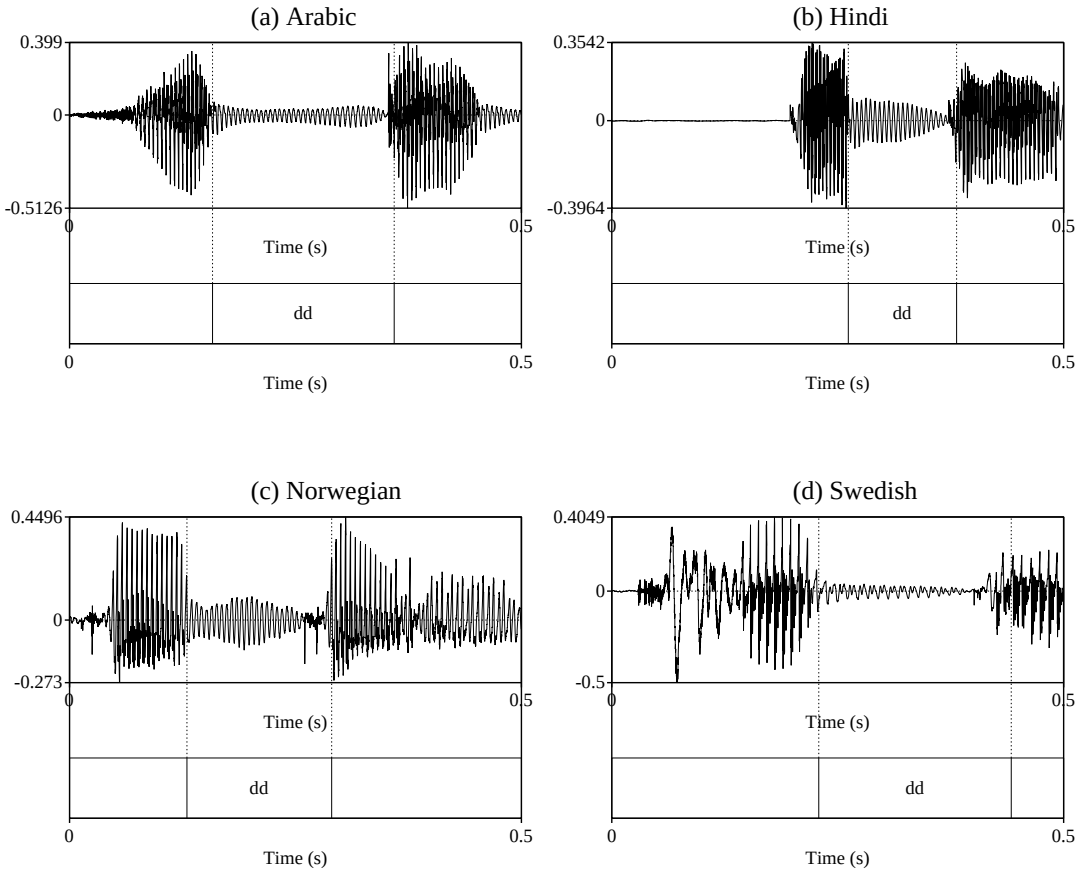

Figure 11: A voiced geminate in four languages. Top left=Arabic; top right=Hindi; Bottom left=Norwegian; Bottom right=Swedish. The time scale is 0.5 sec.

The difference between Japanese and these languages may be attributed to the fact that the [voice] contrast is fully contrastive in geminates in these languages. For example, a search

in the Arabic corpus (Kilany et al., 1997) shows that voiced geminates are as common as voiced singletons. In terms of entropy, a [voice] contrast in singletons is 0.97 bits in singletons, and 0.93 bits in geminates—a [voice] contrast is almost equally informative in singleton pairs and geminate pairs (Kawahara, 2016).

The current results are thus compatible with the view that speakers implement a contrast with more information (i.e. higher entropy) more robustly (Aylett & Turk, 2004, 2006; Cohen-Priva, 2012, 2015; Hall et al., 2016; Hume, 2016; Shaw et al., 2014). This theory makes an explicit prediction about the phonetic implementation of the geminate [voice] contrast in different languages; the higher the entropy of a [voice] contrast in geminate in a particular language, the more robustly that [voice] contrast is implemented. This initial comparison between Japanese and Arabic shows that this prediction is on the right track, but it should be examined more extensively in future studies.

References

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

Aylett, Matthew & Alice Turk (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* **47**(1): 31–56.

Aylett, Matthew & Alice Turk (2006) Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* **119**(5): 3048–3059.

Baayen, Harald R. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, Harald R. (2009) LanguageR. R package.

Baayen, Harald R., Doug.J. Davidson, & Douglas. M. Bates (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**: 390–412.

Baer, Thomas, Anders Löfqvist, & Nancy S. McGarr (1983) Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. *Journal of the Acoustical Society of America* **73**: 1304–1308.

Bates, Douglas (2005) Fitting linear mixed models in R. *R News* **5**: 27–30.

Bates, Douglas, Martin Maechler, & Ben Bolker (2011) lme4: Linear mixed-effects models using S4 classes. R package.

Boersma, Paul (2001) Praat, a system for doing phonetics by computer. *Glot International* **5**(9/10): 341–345.

Caisse, Michelle (1982) *Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops*. MA thesis, University of California, Berkeley.

Cedrus Corporation (2010) Superlab v. 4.0. Software.

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

Childers, D.G, A. M. Smith, & G.P. Moore (1984) Relationships between electroglottograph speech and vocal cord contact. *Folia Phoniatr* **36**: 105–118.

Cohen-Priva, Uriel (2012) *Deriving linguistic generalizations from information utility.* Doctoral dissertation, Stanford.

Cohen-Priva, Uriel (2015) Informativity affects consonant duration and deletion rates. *Journal of Laboratory Phonology* **6**(2): 243–278.

Diehl, Randy & Michelle Molis (1995) Effects of fundamental frequency on medial [voice] judgments. *Phonetica* **52**: 188–195.

Fujimoto, Masako (2015) Vowel devoicing. In *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, Haruo Kubozono, ed., 167-214, Mouton.

Hall, Kathleen Currie, Elizabeth Hume, Florian T Jaeger, & Andrew Wedel (2016) The message shapes phonology. Ms.

Hayes, Bruce (1999) Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In *Functionalism and Formalism in Linguistics, vol. 1: General Papers*, Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer, & Kathleen Wheatly, eds., Amsterdam: John Benjamins, 243–285.

Hayes, Bruce & Donca Steriade (2004) Introduction: The phonetic bases of phonological markedness. In *Phonetically Based Phonology.*, Bruce Hayes, Robert Kirchner, & Donca Steriade, eds., Cambridge: Cambridge University Press, 1–33.

Hirata, Yukari & Jacob Whiton (2005) Effects of speaking rate on the singleton/geminate distinction in Japanese. *Journal of the Acoustical Society of America* **118**: 1647–1660.

Hirose, Aki & Michael Ashby (2007) An acoustic study of devoicing of the geminate obstruents in Japanese. In *Proceedings of ICPhS*, vol. XVI, Trouvain Jürgen & William J. Barry, eds., Saarbrücken, 909–912.

Homma, Yayoi (1981) Durational relationship between Japanese stops and vowels. *Journal of Phonetics* **9**: 273–281.

Hume, Elizabeth (2016) Phonological markedness and its relation to the uncertainty of words. *On-in Kenkyu [Phonological Studies]* **19**.

Hura, Susan, Björn Lindblom, & Randy Diehl (1992) On the role of perception in shaping phonological assimilation rules. *Language and Speech* **35**: 59–72.

Idemaru, Kaori & Susan Guion (2008) Acoustic covariants of length contrast in Japanese stops. *Journal of International Phonetic Association* **38**(2): 167–186.

Idemaru, Kaori & Susan Guion-Anderson (2010) Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica* **67**: 25–46.

Ito, Junko & Armin Mester (1995) Japanese phonology. In *The Handbook of Phonological Theory*, John Goldsmith, ed., Oxford: Blackwell, 817–838.

Ito, Junko & Armin Mester (1996) Stem and word in Sino-Japanese. In *Phonological Structure and Language Processing: Cross-Linguistic Studies*, T. Otake & A. Cutler, eds., Berlin: Mouton de Gruyter, 13–44.

Ito, Junko & Armin Mester (1999) The phonological lexicon. In *The Handbook of Japanese Linguistics*, Natsuko Tsujimura, ed., Oxford: Blackwell, 62–100.

Jaeger, Jeri J. (1978) Speech aerodynamics and phonological universals. In *Proceedings of Berkeley Linguistic Society 4*, Jeri Jaeger, Anthony Woodbury, Farrel Ackerman, Christine Chiavello, Orin Gensler, John Kingston, Eve Sweetner, Henry Chompson, & Kenneth Whistler, eds., Berkeley: Berkeley Linguistics Society, 311–325.

Johnson, Keith (2003) *Acoustic and Auditory Phonetics: 2nd Edition*. Malden and Oxford: Blackwell.

Katayama, Motoko (1998) *Optimality Theory and Japanese loanword phonology*. Doctoral dissertation, University of California, Santa Cruz.

Kawahara, Hideki, Parham Zolfaghari, Alain de Cheveigne, & Roy Patterson (1999) Source information extraction using fixed points of a frequency to instantaneous frequency map. *IEICE technical report. Speech* **99**: 1–8.

Kawahara, Shigeto (2005) Voicing and geminacy in Japanese: An acoustic and perceptual study. In *Univeristy of Massachusetts Occasional Papers in Linguistics 31*, Kathryn Flack & Shigeto Kawahara, eds., Amherst: GLSA, 87–120.

Kawahara, Shigeto (2006) A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* **82**(3): 536–574.

Kawahara, Shigeto (2015a) Geminate devoicing in Japanese loanwords: Theoretical and experimental investigations. *Language and Linguistic Compass* **9**(4): 168–182.

Kawahara, Shigeto (2015b) The phonetics of *sokuon*, obstruent geminates. In *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, Haruo Kubozono, ed., Mouton, 43–73.

Kawahara, Shigeto (2015c) The phonology of Japanese accent. In *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, Haruo Kubozono, ed., Mouton: Mouton de Gruyter, 445–492.

Kawahara, Shigeto (2016) Japanese geminate devoicing once again: Insights from information theory. A plenary talk at FAJL 8.

Kawahara, Shigeto & Aaron Braver (2014) Durational properties of emphatically-lengthened consonants in Japanese. *Journal of International Phonetic Association* **44**(3): 237–260.

Kawahara, Shigeto & Shin-ichiro Sano (2013) A corpus-based study of geminate devoicing in Japanese: Linguistic factors. *Language Sciences* **40**: 300–307.

Kawahara, Shigeto & Takahito Shinya (2008) The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase and Utterance. *Phonetica* **65**(1-2): 62–105.

Kilany, Hanaa, H. Gadalla, H. Arram, A. Yacoub, A. El-Habashi, A. Shalaby, K. Karins, E. Rowson, R. MacIntyre, P. Kingsbury, & C. McLemore (1997) LDC Egyptian Colloquial Arabic Lexicon. Linguistic Data Consortium, University of Pennsylvania.

Kingston, John (2007) The phonetics-phonology interface. In *The Cambridge Handbook of Phonology*, Paul de Lacy, ed., Cambridge: Cambridge University Press, 401–434.

Kingston, John & Randy Diehl (1994) Phonetic knowledge. *Language* **70**: 419–454.

Kingston, John & Randy Diehl (1995) Intermediate properties in the perception of distinctive feature values. In *Papers in laboratory phonology IV: Phonology and phonetic evidence*, B. Connell & A. Arvaniti, eds., Cambridge: Cambridge University Press, 7–27.

Kluender, Keith, Randy Diehl, & Beverly Wright (1988) Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* **16**: 153–169.

Kohler, Klaus (1990) Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In *Speech Production and Speech Modeling*, William J. Hardcastle & Alain Marchal, eds., Dordrecht: Kluwer, 69–92.

Kubozono, Haruo, Junko Ito, & Armin Mester (2008) Consonant gemination in Japanese loanword phonology. In *Current Issues in Unity and Diversity of Languages. Collection of*

Papers Selected from the 18th International Congress of Linguists, The Linguistic Society of Korea, ed., Republic of Korea: Dongam Publishing Co, 953–973.
Kurisu, Kazutaka (2000) Richness of the base and root-fusion in Sino-Japanese. *Journal of East Asian Linguistics* **9**: 147–185.
Kuroda, S.-Y. (1965) *Generative Grammatical Studies in the Japanese Language*. Doctoral dissertation, MIT.
Lisker, Leigh (1978) On buzzing the English /b/. *Haskins Laboratories Status Report on Speech Research* **SR-55/56**: 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.
Lyman, Benjamin S. (1894) Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia* : 1–17.
Matsuura, Toshio (2012) Yuusei sogai jyuushion-no onsei jitsugen-niokeru chiikisa-ni kansuru yobitekibunseki [A preliminary analysis on regional variation in phonetic realization of voiced obstruent geminates]. Talk presented at the Phonetic Society of Japan, Sept 29th.
Nishimura, Kohei (2006) Lyman's Law in loanwords. *On'in Kenkyu [Phonological Studies]* **9**: 83–90.
Ohala, John J. (1983) The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, Peter MacNeilage, ed., New York: Springer-Verlag, 189–216.
Ohala, John J. & Carol J. Riordan (1979) Passive vocal tract enlargement during voiced stops. In *Speech Communication Papers*, Jared. J. Wolf & Dennis H. Klatt, eds., New York: Acoustical Society of America, 89–92.
Parker, Ellen, Randy Diehl, & Keith Kluender (1986) Trading relations in speech and non-speech. *Perception & Psychophysics* **39**: 129–142.
Pickett, J. M. (1980) *The Sounds of Speech Communication*. Baltimore: University Park Press.
Podesva, Robert (2000) Constraints on geminates in Burmese and Selayarese. In *Proceedings of West Coast Conference on Formal Linguistics 19*, Roger Bilerey-Mosier & Brook Danielle Lillehaugen, eds., Somerville: Cascadilla Press, 343–356.
Podesva, Robert (2002) Segmental constraints on geminates and their implications for typology. Talk given at the LSA Annual Meeting.
Port, Robert & Johnathan Dalby (1982) Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* **32**: 141–152.
R Development Core Team (1993–2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
Raphael, Lawrence (1972) Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical 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.
Rice, Keren (2006) On the patterning of voiced stops in loanwords in Japanese. *Toronto Working Papers in Linguistics* **26**: 11–22.
Ridouane, Rachid (2010) Geminate at the junction of phonetics and phonology. In *Papers in Laboratory Phonology X*, Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio, & Nathalie Valleé, eds., Berlin: Mouton de Gruyter, 61–90.

Rothenberg, M. (1981) Some relations between glottal air flow and vocal fold contact area. *Proceedings of the conference on the assessment of vocal pathology* : 88–96.

Roubeau, B., C. Chevrie-Muller, & C. Arabia-Guidet (1987) Electroglottographic study of changes of voice registers. *Folia Phoniatr* **39**: 280–289.

Sano, Shinichiro & Shigeto Kawahara (2013) A corpus-based study of geminate devoicing in Japanese: The role of the OCP and external factors. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]* **144**: 103–118.

Shannon, Claude (1948) *A mathematical theory of communication*. Ma thesis, MIT.

Shaw, Jason, Chong Han, & Yuan Ma (2014) Surviving truncation: informativity at the interface of morphology and phonology. *Morphology* **24**: 407–432.

Shirai, Setsuko (2002) *Gemination in Loans from English to Japanese*. MA thesis, University of Washington.

Steriade, Donca (2008) The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In *The nature of the word*, Kristin Hanson & Sharon Inkelas, eds., Cambridge: MIT Press, 151–179.

Stevens, Kenneth & Sheila Blumstein (1981) The search for invariant acoustic correlates of phonetic features. In *Perspectives on the study of speech*, Peter Eimas & Joanne D. Miller, eds., New Jersey: Earlbaum, 1–38.

Takada, Mieko (2011) *Nihongo-no gotou heisaon-no kenkyuu [A study of word-initial stops in Japanese]*. Tokyo: Kuroshio Publications.

Takada, Mieko (2013) Regional differences in sound patterns during the closure of Japanese voiced geminates. Paper presented at 3rd ICPP at NINJAL, Japan.

Tsuchida, Ayako (1997) *Phonetics and Phonology of Japanese Vowel Devoicing*. Doctoral dissertation, Cornell University.

Vance, Timothy (2007) Have we learned anything about *rendaku* that Lyman didn't already know? In *Current issues in the history and structure of Japanese*, Bjarke Frellesvig, Masayoshi Shibatani, & John Carles Smith, eds., Tokyo: Kurosio, 153–170.

Westbury, John R. (1979) *Aspects of the Temporal Control of Voicing in Consonant Clusters in English*. Doctoral dissertation, University of Texas, Austin.

Westbury, John R. & Patricia Keating (1986) On the naturalness of stop consonant voicing. *Journal of Linguistics* **22**: 145–166.