Shigeto Kawahara

Testing Japanese loanword devoicing: Addressing task effects

Abstract: In the loanword phonology of Japanese, voiced obstruent geminates ([bb, dd, gg]) have been claimed to devoice when they co-occur with another voiced obstruent within the same morpheme (e.g., /beddo/ → [betto] 'bed'). This devoicing pattern has contributed much to address a number of theoretical issues in the recent phonological literature. However, the relevant data have been primarily based on intuition-based data provided by Nishimura (2003) and Kawahara (2006). Kawahara (2011a, 2011b) addressed this issue by conducting rating studies using naive native speakers of Japanese. The results generally supported the intuition-based data by Nishimura (2003) and Kawahara (2006). However, the rating studies also revealed several aspects of the devoicing pattern that go beyond the intuition-based data as well.

The current study further investigates the devoicing pattern by varying several task variables. In particular, this paper builds on Kawahara (2011a, 2011b) by adding (i) nonce word stimuli, (ii) a binary yes/no experiment, and (iii) auditory stimuli. The results show that (i) nonce words and real words behave similarly, but nonce words nevertheless show less variability across different grammatical conditions than real words; (ii) the binary yes/no experiment shows results similar to those of the scale-based experiment; and (iii) while auditory stimuli yield results comparable with those of orthographic stimuli, they also show an exaggerated effect of a phonetic implementation pattern. Overall, this paper uses Japanese as a case study, and finds some task effects in phonological judgment experiments. It is hoped that this paper stimulates further experimental research on phonological judgments of other phenomena in Japanese as well as in other languages.

Keywords: Japanese, geminates, devoicing, experimental phonology, task effects

Shigeto Kawahara: Institute of Cultural and Linguistic Studies, Keio University, 2–15–45 Mita, Minato-ku, Tokyo, JAPAN 108-8345. E-mail: kawahara@icl.keio.ac.jp

1 Introduction

1.1 The phenomenon

This paper is about devoicing of obstruents in the loanword phonology of Japanese. It has been known that voiced obstruent geminates ([bb, dd, gg]) in Japanese loanwords can be devoiced (Itô and Mester 1995, 1999; Quakenbusch 1989; Vance 1987), but exactly when such devoicing occurs remained unclear. For example, Itô and Mester (1999) argued that some items can undergo devoicing while other items cannot, and considered the first type of words as "assimilated foreign items" and the second, non-devoicing type as "unassimilated foreign items". Instead of relying on a (more or less) arbitrary etymological distinction, Nishimura (2003) proposes a *phonological* characterization of this distinction, claiming that voiced obstruent geminates optionally devoice when they co-occur with another voiced obstruent within the same stem, as exemplified by the data in (1). He further claims that this devoicing is due to a restriction against having two voiced obstruents within the same stem. In Japanese phonology, this restriction has long been known as Lyman's Law (Kawahara 2012b; Lyman 1894; Vance 2007), and has been formalized as OCP[voice] (Obligatory Contour Principle: Leben 1973; henceforth simply the OCP) (Itô and Mester 1986, 1998, 2003). In other words, devoicing is possible in (1) whereas it is impossible in non-OCP-violating voiced geminates, as shown in (2). Moreover, Nishimura (2003) argues that devoicing is also impossible in OCP-violating singletons, as in (3).

(1) Voiced obstruent geminates optionally devoice if they co-occur with another voiced obstruent; i.e., when they violate OCP[voice].

be**dd**o	→	be**tt**o	'bed'
ba**gg**u	→	ba**kk**u	'bag'
bi**gg**u	→	bi**kk**u	'big'

(2) Voiced obstruent geminates do not devoice if they do not violate OCP[voice].

suno**bb**u	→	suno**bb**u	*suno**pp**u	'snob'
he**dd**o	→	he**dd**o	*he**tt**o	'head'
re**gg**u	→	re**gg**u	*re**kk**u	'leg'

(3) Voiced singletons do not devoice even when they violate OCP[voice].

da**b**u	→	da**b**u	*da**p**u	'Dove'
do**g**uma	→	do**g**uma	*do**k**uma	'dogma'
da**g**u	→	da**g**u	*da**k**u	'Doug'

The patterns in (1)–(3) have attracted much attention in the recent phonological literature. It is beyond the scope of this paper to settle these debates; however, to briefly summarize, the devoicing pattern triggered three major theoretical debates: (i) how to explain the difference between singletons (= the data in (3)) and geminates (= the data in (1)) (Kawahara 2006, 2008; Rice 2006; Steriade 2004); (ii) how to capture the cumulative markedness requirement of devoicing in (1) (Farris-Trimble 2008; Nishimura 2003; Pater 2009, forthcoming; Tesar 2007); and (iii) how the spontaneous emergence of loanword devoicing in (1) bears on the theory of lexical stratification – a theory of how loanword phonology is related to native phonology (Crawford 2009; Itô and Mester 2003, 2008; Tateishi 2002). See Kawahara (2011a) and Kawahara (2012a) for recent summaries (the former in English and the latter in Japanese).

In short, the Japanese loanword devoicing pattern has contributed much to several theoretical debates in recent years. However, Kawahara (2011b) raises one issue: the Japanese loanword devoicing data are primarily based on the intuitions of two linguists, namely, Nishimura (2003) and Kawahara (2006); i.e., the grammaticality judgments in (1)–(3) primarily come from the authors themselves.[1] Many studies have raised concerns about research exclusively relying on authors' own introspections (e.g., Dąbrowska 2010; Gibson and Fedorenko 2010; Griner 2001; Labov 1996; Myers 2009; Ohala 1986; Schütze 1996). To address this problem, Kawahara (2011b) conducted a rating experiment with 38 native Japanese speakers who did not know about the devoicing pattern. The experiment indeed showed that Japanese speakers generally judge devoicing of OCP-violating geminates as more natural than devoicing of non-OCP-violating geminates or devoicing of OCP-violating singletons. In this regard, Kawahara (2011b) succeeded in supporting the empirical basis of the claims made about the patterns in (1)–(3). Kawahara (2011a) reports a follow-up experiment using a larger set of stimuli with 49 naive native speakers, which again supported the idea that devoicing of OCP-violating geminates is the most natural environment for native speakers of Japanese.

1.2 The current study

There are some remaining questions, however. First, both Kawahara (2011a) and Kawahara (2011b) used only real words. In the case of Japanese loanword

---

[1] See Nishimura (2003), Kawahara and Sano (2013) and Sano and Kawahara (forthcoming) for some evidence based on corpus data.

devoicing, it is of some interest to investigate whether the results obtained for real words generalize to nonce words. An often-used test on phonological productivity is a wug-test (Berko 1958), in which participants are asked to inflect nonce words. Some previous wug-tests have failed to replicate phonological patterns that apply to real words, in which case it is often concluded that the alleged phonological patterns are not productive; i.e., they are lexicalized (Griner 2001; Ohala 1974; Sanders 2003). (See also Shademan (2007) for some related discussion.) If the phonological pattern under discussion is not productive with nonce words, the pattern should probably not be used for phonological argumentation.

In fact, there is an example from Japanese phonology whose productivity has been questioned by way of experiments using nonce words. Several phonological changes occurring in Japanese verbal paradigms (Davis and Tsujimura 1991; Tsujimura 1996) were not replicated in nonce word experimentation (Batchelder 1999; Griner 2001; Vance 1987). In short, there is no guarantee that we can generalize the patterns of real words to nonce words, and it is vital to test the productivity of the phenomenon under question using nonce words. Experiment III in Kawahara (2012b) addressed this question, although in that paper, the comparison between real words and nonce words was not the main focus. The Section 2 of this paper therefore reports that experiment in more detail to address this issue of whether the previous results can be replicated with nonce words.

The first aim of this paper is to therefore re-examine Experiment III of Kawahara (2012) to mainly address the question of whether the results obtained in Kawahara (2011a) and Kawahara (2011b) can be replicated with nonce words. This report also allows us to compare the results of that experiment with two other experiments reported in the paper.

The second aim of this paper is to test the gradiency of judgment patterns found in the previous experiments. Kawahara (2011a, 2011b) found that Japanese speakers distinguish the naturalness of two processes that were both judged to be "ungrammatical" by Nishimura (2003) and Kawahara (2006), with the devoicing of non-OCP-violating geminates (= the examples in (2)) rated as more natural than the devoicing of OCP-violating singletons (= (3)). One may wonder whether this gradient effect was due to a task effect; the reason being that Kawahara (2011a, 2011b) uses a gradient scale. Testing this issue is in part motivated by the debate concerning the gradient nature of phonological judgments. It is known that grammatical judgments show distinctions beyond a simple, binary "grammatical" vs. "ungrammatical" dichotomy, especially in experimental settings (see e.g., Albright 2009; Coetzee 2008; Coleman and Pierrehumbert 1997; Daland et al. 2011; Dankovičová et al. 1998; Goldrick 2011; Greenberg and Jenkins 1964; Hayes 2000; Hayes and Wilson 2008; Pertz and Bever 1975; Pierrehumbert 2001; Shademan 2007 for phonological/phonotactic judgments; Chomsky 1965; Myers

2009; Schütze 1996; Sorace and Keller 2005 for syntactic judgments). However, one may contend the idea that we obtain gradient results in experimental settings because these experiments use scales. Therefore, the second aim of this paper is to test whether the gradient results that Kawahara (2011a, 2011b) obtained can be replicated using a binary yes/no task. Some previous studies (Bader and Mäussler 2010; Coleman and Pierrehumbert 1997; Dankovičová et al. 1998; Frisch et al. 2004) raised similar issues and found gradient results using a binary yes/no format. The current study thus builds on them and aims to address the gradient nature of phonological judgments in the case of Japanese loanword devoicing.

Finally, the third issue that this paper addresses is that Kawahara (2011a, 2011b) used visual, orthographic stimuli, although the instructions in these studies encouraged the participants to read the stimuli in their heads and use an auditory impression to make judgments. While many judgment experiments in linguistics are run with orthography, it is worth running the same experiment with auditory stimuli for a few reasons. First, one explanation for why voiced geminates, but not voiced singletons, can devoice is because a phonological voicing contrast is auditorily less perceptible in geminates than in singletons (Kawahara 2006, 2008). An auditory judgment experiment would help to address this specific hypothesis. Second, it would be interesting to investigate whether the results of Kawahara (2011a, 2011b) can be replicated with auditory stimuli, because phonology is concerned with sounds. Replicating the Japanese devoicing pattern with auditory stimuli is therefore the third aim of this paper.

To summarize, there are three issues that this paper aims to address: (i) the judgment patterns on devoicing as revealed by nonce words, (ii) the effect of using a binary yes/no format, and (iii) the effect of using auditory stimuli. This paper reexamines Kawahara (2012b) and reports two additional experiments in order to address these three issues. More generally, by varying experimental variables, the current project aims to further examine the empirical basis of the theoretical debates reviewed in Section 1.1, beyond Kawahara (2011a, 2011b).

Before reporting the actual experiments, a few remarks are in order. First, the experiments reported in this paper are judgment experiments for a phonological process, i.e., devoicing. The task is for native speakers to judge the naturalness or possibility of a phonological pattern, or in other words, a pairing between one form and another form (i.e., in this case, a phonological form and its optional variant). This task therefore differs from phonotactic wellformedness judgment tasks in which speakers judge the wellformedness of surface forms only (e.g., Bailey and Hahn 2001; Coetzee 2008; Coleman and Pierrehumbert 1997; Daland et al. 2011; Dankovičová et al. 1998; Greenberg and Jenkins 1964; Shademan 2007). Second, this paper offers a case study in Japanese of such a phonological judgment study. Although its scope is thus limited, it is hoped that this paper will stimulate

further studies on different phonological phenomena in different languages (including Japanese).

2 Kawahara (2012b) Experiment III and beyond

This experiment, briefly reported as Experiment III in Kawahara (2012b), is an orthography-based rating experiment. This section reexamines this experiment in detail, since some details and analyses of this experiment were omitted from Kawahara 2012b (Kawahara 2012b was written after the current paper),[2] and also since the other two experiments in the current paper crucially build on this experiment and I will make many cross-experimental comparisons in what follows. In particular, this section reexamines the experiment from the perspectives that are discussed in the introduction, mainly with the focus of comparing real words and nonce words.

This section thus addresses three issues: (i) to replicate Kawahara (2011a, 2011b); and, more importantly, (ii) to test whether the results obtained with real words in the previous studies generalize to nonce words; and finally, (iii) to compare the patterns of real words and nonce words.

2.1 Method

2.1.1 Stimuli

The stimuli consisted of four grammatical conditions: (i) OCP-violating geminates, (ii) non-OCP-violating geminates, (iii) OCP-violating singletons, and (iv) non-OCP-violating singletons, as summarized in (4), each with a representative example. In this design, two factors – OCP and GEM – were fully crossed. This paper uses CAPITAL LETTERS to represent variable names.

---

2 In Kawahara (2012b), together with two other experiments on rendaku, this experiment is also reported (in much less detail) to show the activity of the OCP – or Lyman's Law – in both loanwords and nonce words. That paper does not extensively discuss the effect of geminacy, its interaction with OCP, or on the comparison between real words and nonce words. Neither does that paper report the issue of gradiency at all; i.e., the analyses presented in Figures 2 and 3 below are new to this paper. Figure 1 is reproduced with permission from Elsevier.

(4) The four grammatical conditions
    a. OCP-violating geminates (e.g., [ba**gg**u])
    b. non-OCP-violating geminates (e.g., [e**gg**u])
    c. OCP-violating singletons (e.g., [da**g**u])
    d. non-OCP-violating singletons (e.g., [ma**g**u])

The experiment had 9 items per each condition. All the stimulus items were disyllabic, and all the target consonants were word-internal (since all lexical geminates in Japanese appear word-internally: Kawahara (forthcoming)). The stimulus set was constructed in the following way: first, real disyllabic words containing OCP-violating geminates were chosen; this case has the least number of existing items in the Japanese lexicon. This selection process resulted in 9 items. Among those 9 items, 6 items contained [dd] followed by epenthetic [o], and the remaining 3 items contained [gg] followed by epenthetic [u]. No stimuli with [bb] were found; in fact, no disyllabic words with OCP-violating [bb] exist which is not unexpected given that [bb] is very rare in Japanese loanwords (Katayama 1998). Then the items for the other three conditions were selected, consisting of 6 items for [d(d)] and 3 items for [g(g)], as listed in Table 1. Across all conditions, the number of items for each place of articulation was controlled for. Short vowels were used before geminates and singleton [g]. Long vowels and diphthongs had to be used before singleton [d], because disyllabic loanwords with an initial short vowel almost always have a geminate [dd], and not a singleton [d]. This pattern is due to a productive gemination process in loanword adaptation (e.g., [baddo] 'bad'; see e.g., Katayama 1998; Kubozono et al. 2009).[3] All of the stimuli have a pitch accent on the initial syllable, which is phonetically realized as a HL falling F0 contour.

Table 1: The list of the stimuli that are real words.

OCP-Gem	Gem	OCP-Sing	Sing
baddo 'bad'	heddo 'head'	bado 'badminton'	muudo 'mood'
beddo 'bed'	reddo 'red'	gaido 'guide'	waido 'wide'
daddo 'dad'	uddo 'wood'	zoido common name	haido 'hide'
deddo 'dead'	kiddo 'kid'	boodo 'board'	roodo 'road'
guddo 'good'	maddo 'mad'	gaado 'guard'	riido 'lead'
goddo 'god'	roddo 'rod'	baado 'bird'	huudo 'food'
baggu 'bag'	eggu 'egg'	dagu 'Doug'	hagu 'hug'
biggu 'big'	reggu 'leg'	bagu 'bug'	magu 'mag'
doggu 'dog'	taggu 'tag'	jogu 'jog'	ragu 'rag'

---

3 [bado] is a truncated form of [badominton].

Table 2: The list of the stimuli that are nonce words.

OCP-Gem	Gem	OCP-Sing	Sing
buddo	keddo	budo	hudo
boddo	koddo	dado	rado
doddo	ruddo	dodo	rudo
geddo	yuddo	dedo	rido
gaddo	taddo	gado	yudo
giddo	kuddo	gudo	wado
boggu	uggu	degu	hegu
gaggu	oggu	dogu	negu
goggu	naggu	gegu	mugu

The nonce word stimuli are listed in Table 2. These stimuli had the same phonological structures as the real word stimuli, except that all nonce word stimuli had short initial vowels, including those nonce words that contain a singleton [d], which can also have a short vowel.

2.1.2 Task

In this experiment Japanese speakers were asked to rate the naturalness of devoicing in the four grammatical conditions. The instructions stated that the questionnaire was about the naturalness of devoicing in Japanese loanwords. Using the same format as Kawahara (2011a, 2011b), for each question, the participants were presented with one stimulus item and asked to judge the naturalness of the form that undergoes devoicing of word-internal consonants (e.g., given [baddo], how natural would you find it to pronounce it as [batto]?). The instructions and stimuli were presented in Japanese orthography. The *katakana* orthography was used for the stimuli (for both [baddo] and [batto] in the example above), for both real words and nonce words, since *katakana* is conventionally used for loanwords and nonce words in standard Japanese orthography (Labrune 2012). Although the test was based on orthography, the participants were asked to read each stimulus in their heads, and make judgments based on their auditory impression rather than on the orthography.

Following Kawahara (2011a, 2011b), in this experiment, the speakers judged the naturalness of devoicing using a 5-point scale, as follows: A. "very natural", B. "somewhat natural", C. "neither natural nor unnatural", D. "somewhat unnatural", and E. "very unnatural". Since the software that ran the experiment (see Section 2.1.3) could not present the scale numerically, the responses were later converted to a numerical scale.

The main session was blocked into two parts. The first block presented all the real word stimuli, followed by a break sign. The second block presented all the nonce word stimuli. The entire experiment was structured in this way because it was assumed that making judgments about real words would be easier than making judgments about nonce words for the participants, allowing the participants to first gain familiarity with the task before giving judgments for nonce words.[4]

2.1.3 Procedure

Sakai, an online system which runs questionnaires, was used to run the current online experiment. An advantage of this internet-based methodology is the fact that it is easy to get a large number of participants. This advantage is particularly important when the researcher does not reside in an area where there are many local speakers of the target language. A potential disadvantage is that the researcher cannot control the environment in which the participants take the experiment, although Sprouse (2011) shows that linguistic judgment data gathered with this sort of method are comparable with the data gathered in the laboratory (see Reips (2002) and Sprouse (2011) for further, general discussion about online experimentation in psychology and linguistics).

The experimental website first presented a consent form and the instructions of the experiment. Then the main session started, with one trial presented per page. The order of the stimuli within each block was randomized.

2.1.4 Participants

Thirty-three native speakers of Japanese, who were mainly students at a Japanese university, participated in the experiment. One speaker reported that they are familiar with the devoicing pattern, and hence his/her data were excluded from the following analysis.

---

4 Kawahara (2010) reports an experiment that addresses the question of how this organization may have affected the results.

2.1.5 Statistics

The responses were first converted to numerical values as follows: A. "very natural" = 5; B. "somewhat natural" = 4; C. "neither natural nor unnatural" = 3; D. "somewhat unnatural" = 2; E. "very unnatural" = 1. For statistical analyses, a general linear mixed model was run (Baayen et al. 2008; Baayen 2008) using R (R Development Core Team, 1993–2013) with the lme4 package (Bates et al. 2011). The fixed factors were OCP and GEM.[5] The p-values were calculated by the Markov chain Monte Carlo method using the languageR package (Baayen 2009).

2.2 Results

Figure 1 illustrates average rating scores. In real words, the average naturalness ratings showed the following order: OCP-violating geminates (4.23) > non-OCP-violating geminates (3.29) > OCP-violating singletons (2.69) > non-OCP-violating singletons (2.21). Simply put, devoicing of OCP-violating geminates was rated as more natural than the devoicing of non-OCP-violating geminates which, in turn, was rated as more natural than the devoicing of OCP-violating singletons, which was rated as more natural than devoicing of non-OCP-violating singletons, replicating the results of previous studies (Kawahara, 2011a, 2011b). Statistically, for real words, all factors are significant: OCP ($t = 5.29$, $p < .001$), GEM ($t = 11.81$, $p < .001$), and the interaction between OCP and GEM ($t = 2.68$, $p < .01$). The significance of the main effects shows that OCP and GEM each affect naturalness ratings on devoicing, and the significant interaction term indicates that the effect of OCP is bigger on the geminate pair ($4.23 - 3.29 = 0.94$) than on the singleton pair ($2.69 - 2.21 = 0.48$).

For nonce words, the order of the naturalness ratings is the same as the real word condition: OCP-violating geminates (3.64) > non-OCP-violating geminates (3.41) > OCP-violating singletons (3.06) > non-OCP-violating singletons (2.81). The statistical analysis shows that both OCP ($t = 2.56$, $p < .05$) and GEM ($t = 6.44$, $p < .001$) are significant, but, unlike the result of real words, their interaction is

---

[5] To make the interpretation of the statistical analyses simpler, this model left out the effect of lexical usage frequencies on naturalness ratings. See Coetzee and Kawahara (2013), Kawahara (2011a) and Kawahara and Sano (2013) for discussion and also modeling of lexical frequency effects in the Japanese loanword devoicing pattern. Also, to avoid interpreting complex interaction terms, the difference between real words and nonce words was not coded in this model. The targeted comparison between real words and nonce words is provided in the discussion section (Section 2.3).
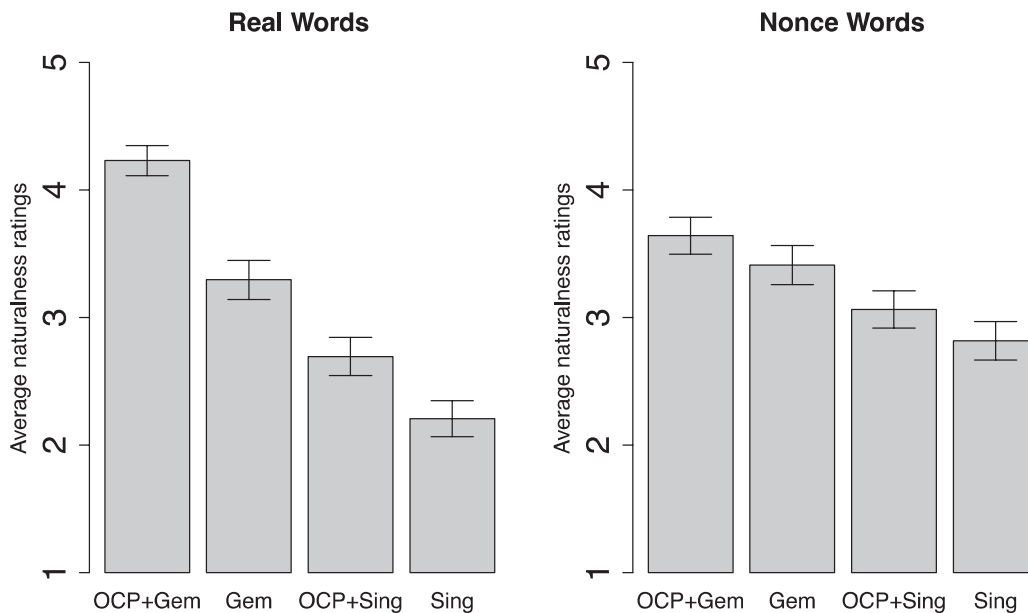
Fig. 1: The average naturalness ratings in the orthography-based rating experiment. Error bars represent 95% confidence intervals. (Reprinted with permission from Elsevier)

not ($t = 0.06$, *n.s.*). For nonce words, the effect of OCP on naturalness ratings is comparable between the singleton condition ($3.64 - 3.41 = 0.23$) and the geminate condition ($3.06 - 2.81 = 0.25$).

2.3 Discussion

2.3.1 Real words vs. nonce words

First, we observe the same order of the four grammatical conditions across real words and nonce words. This order also matches with the results of two previous studies using real words (Kawahara 2011a, 2011b). In this sense, the current experiment has shown that the results of the previous studies that use real words generalize to nonce words. Most importantly, even in nonce words, OCP-violating geminates received the highest naturalness ratings, supporting the original observation by Nishimura (2003). The current experiment thus contributes further empirical support for theoretical claims made about the Japanese loanword devoicing pattern (see Section 1.1).

    At the same time, we observe a difference between real words and nonce words: in nonce words, there is less variability in naturalness ratings across the four grammatical conditions than in real words. In other words, devoicing in

nonce words showed less variation in naturalness ratings across the four conditions than devoicing in real words did. The condition rated as having the most natural devoicing pattern is OCP-violating geminates; devoicing in OCP-violating geminates is judged to be less natural in nonce words than in real words. The least natural devoicing pattern is non-OCP-violating singletons; this condition is judged to be more natural in nonce words than in real words.

To statistically assess this difference between real words and nonce words, for each speaker, the standard deviations across all tokens were calculated separately for real words and nonce words. These standard deviations were then compared between the two conditions using a non-parametric within-subject Wilcoxon test. This analysis shows that the average standard deviations are 1.30 for the real words and 1.03 for the nonce words, and that the difference is significant ($p < .001$).

This reduction of variability across the four grammatical conditions in nonce words could be responsible for the absence of a significant interaction between OCP and GEM in nonce words; there may not be a space left for OCP-violating geminates to have naturalness ratings that are high enough to yield a significant interaction between OCP and GEM.

A question arises as to where the difference between real words and nonce words comes from. Presumably the participants have encountered real instances of devoicing in real words, which would make them more confident about what would happen to each target word. On the other hand, the participants have not seen nonce words before, and therefore they may feel less committed about making extreme judgments in general; i.e., they are reluctant to use endpoints of judgment scales. Despite this difference between real words and nonce words, as discussed, we observe the same ordering between the four grammatical conditions in real words and nonce words.

2.3.2 Gradiency

Second, the current study found gradient grammatical distinctions among the four grammatical conditions, just like the two previous studies (Kawahara, 2011a, 2011b). It does not seem possible to divide the judgment patterns simply into the "grammatical" category and the "ungrammatical" category. In this sense, the current results agree with the previous studies in finding distinctions that go beyond what Nishimura (2003) first proposed.[6]

---

6 Kawahara (2011a, 2011b) speculates about why Japanese speakers find the devoicing of non-OCP-violating geminates more natural than that of OCP-violating singletons. Beyond the specu-

One question that arises is whether this four-way distinction is due to a non-homogeneous speech community. That is, one could argue that the response from each speaker is always binary which follows a "grammatical" vs. "ungrammatical" dichotomy, but averaging over the responses from different speakers results in a gradient pattern. This hypothesis predicts that distributions of responses are at two extremes, because people should consistently rate each devoicing pattern either as completely natural (i.e., grammatical = 5 in rating) or completely unnatural (i.e., ungrammatical = 1 in rating). In this view, the differences between the four grammatical conditions arise from the difference in the number of speakers who assign grammatical status (= 5 in rating) to each condition. To examine this prediction, Figures 2 and 3 provide histograms that show the
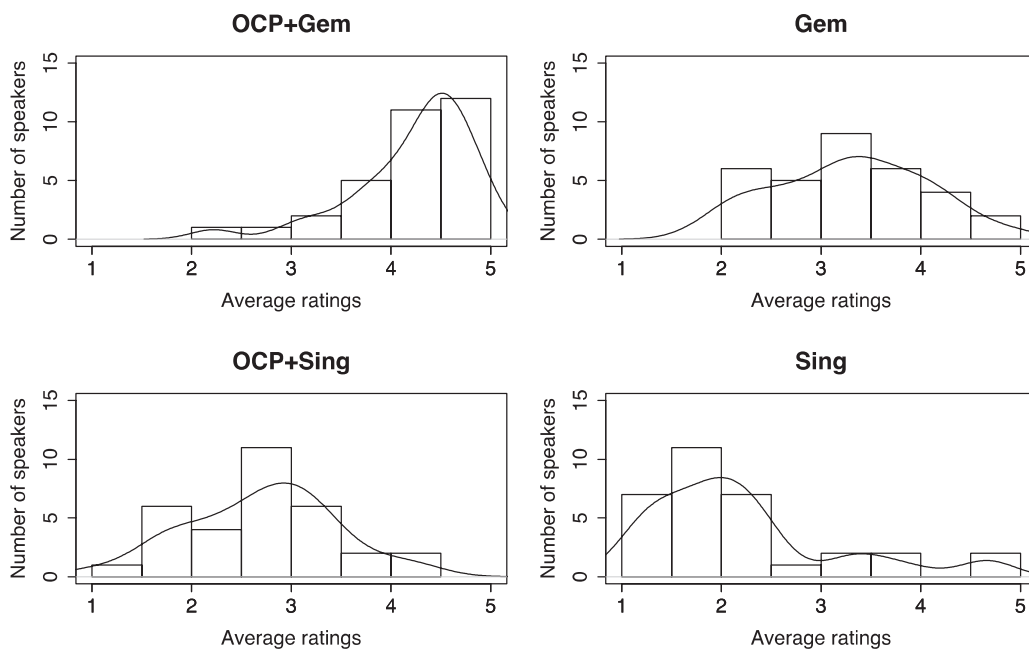

Fig. 2: A histogram of naturalness ratings (number of speakers) for real words.

lation presented there, yet another possibility is that a constraint against voiced geminates is a phonetically natural one (Ohala 1983), whereas OCP[voice] in Japanese is not (Kawahara 2008; Ohala 1981). In fact, there is evidence that children acquiring Japanese show a stage in which they apparently do not show the effect of OCP[voice] (Fukuda and Fukuda 1994), implying that this constraint may have to be learned rather than being innate (Kawahara 2008). Given the characteristics of OCP[voice] in Japanese, the speakers may have found the *phonetically* natural devoicing (= geminate devoicing) more *grammatically* natural. This possibility was brought to my attention by Armin Mester (p.c. August 2011).
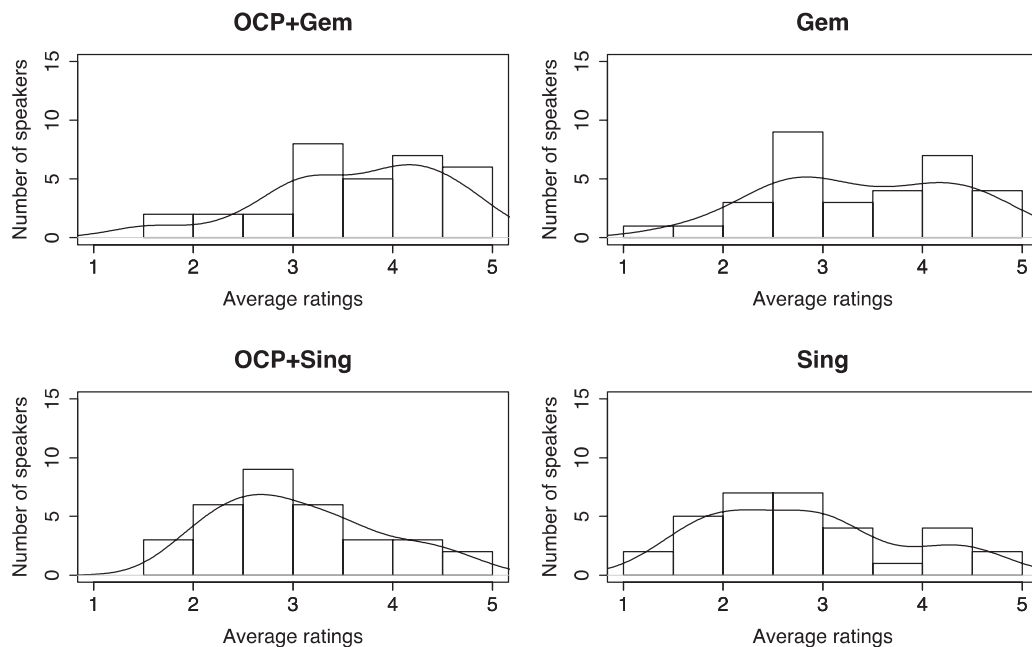
Fig. 3: A histogram of naturalness ratings (number of speakers) for nonce words.

distributions of average scores for each speaker in each grammatical condition. We observe that, contra the hypothesis, there are many speakers who show intermediate average scores in each grammatical condition.

An alternative to the hypothesis we examine in Figures 2 and 3 is to say that items within each grammatical condition showed a binary grammatical vs. ungrammatical pattern, but averaging over non-homogeneous set of items resulted in a gradient pattern. To check this possibility, Figures 4 and 5 illustrate the distributions of average naturalness ratings for each individual item. The hypothesis predicts that average scores for each item distribute at the two extreme ends, around grammatical (= 5 in rating) and ungrammatical (= 1 in rating). This prediction, however, is not supported by the actual data in Figures 4 and 5.

In summary, gradiency does not come from averaging over a non-homogeneous speech community or a non-homogeneous set of test items. It seems safe to conclude that the naturalness patterns in the Japanese devoicing case show a gradient distinction, which goes beyond the "grammatical" vs. "ungrammatical" dichotomy (Albright, 2009; Coetzee 2008; Coleman and Pierrehumbert 1997; Daland et al. 2011; Dankovičová et al. 1998; Goldrick 2011; Greenberg and Jenkins 1964; Hayes 2000; Hayes and Wilson 2008; Pertz and Bever 1975; Pierrehumbert, 2001; Shademan 2007).
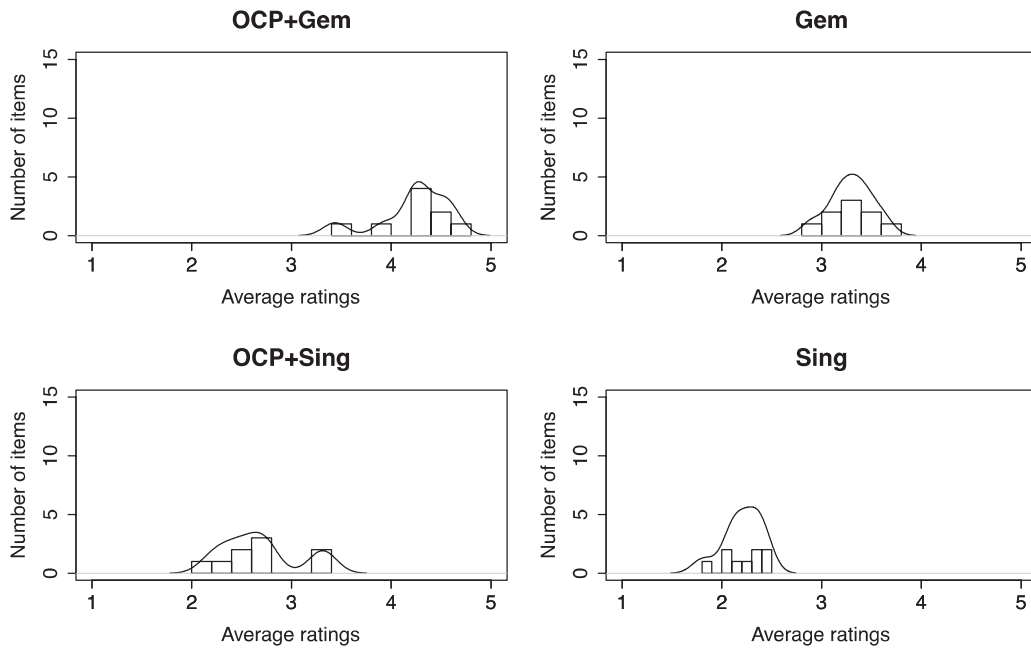
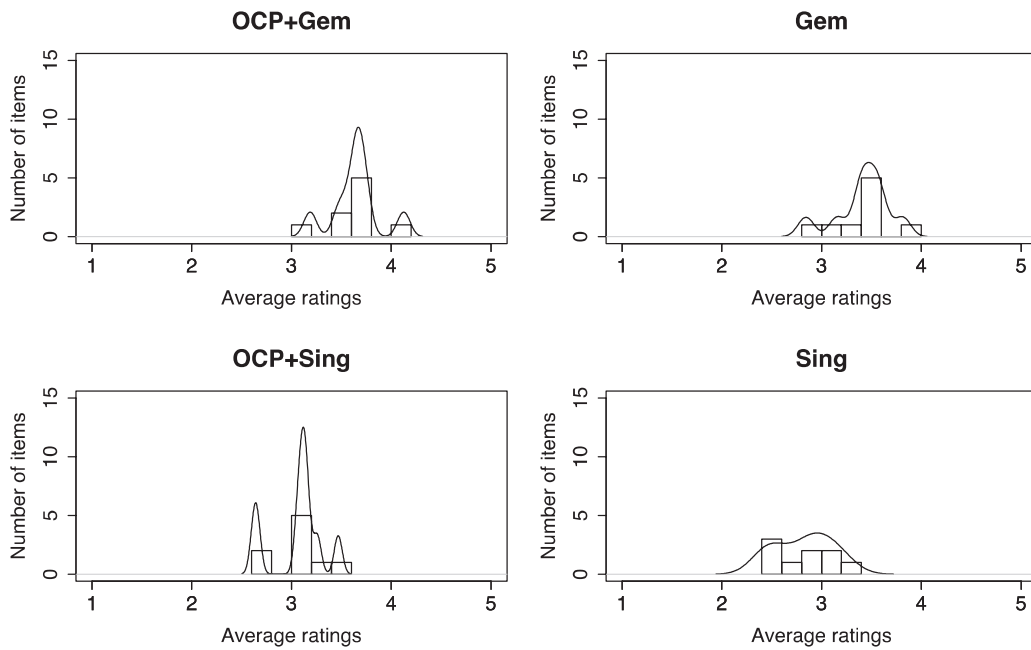**Fig. 4:** A histogram of naturalness ratings (number of items) for real words.

Fig. 5: A histogram of naturalness ratings (number of items) for nonce words.

3 Experiment I: Orthography-based yes/no experiment

Building on Kawahara (2012b) and to a lesser extent on Kawahara (2011a) and Kawahara (2011b), Experiment I is an orthography-based experiment that uses a binary yes/no format, rather than a scale-based rating one. The primary aim of this experiment is to address whether the gradient effect we observed in the previous three studies (Kawahara 2011a, 2011b, 2012b) can be replicated using a binary yes/no format. In these studies, given a 5-point scale, the participants may have felt obliged to use intermediate points (Schütze 2011). To avoid this task effect, the current experiment used a binary yes/no format.

3.1 Method

Experiment I is similar to Experiment III of Kawahara (2012b) reviewed and discussed in Section 2, but it instead asked native speakers whether devoicing in each of the four grammatical conditions is possible or not using a binary yes/no format. Experiment I used the same set of stimuli as Kawahara (2012b). Thirty-seven native speakers of Japanese, again mainly university students in Japan, participated in this experiment. There was no overlap between the participants of Kawahara (2012b) and those of Experiment I. No participants reported that they were familiar with the theoretical issues surrounding the devoicing phenomenon. Since the responses were binary, a logistic linear mixed model was used to analyze the results (Jaeger 2008; Quené and van den Berg 2008).

3.2 Results

Figure 6 illustrates the average ratios of DEVOICING POSSIBLE responses – the numbers of items participants chose DEVOICING POSSIBLE divided by the total number of items – of each condition, both for real words and nonce words. The ratio followed the same hierarchy as the rating experiment for both real words and nonce words: OCP-violating geminates (0.90) > non-OCP-violating geminates (0.62) > OCP-violating singletons (0.34) > non-OCP-violating singletons (0.22) for real words, and OCP-violating geminates (0.76) > non-OCP-violating geminates (0.62) > OCP-violating singletons (0.40) > non-OCP-violating singletons (0.33) for nonce words.
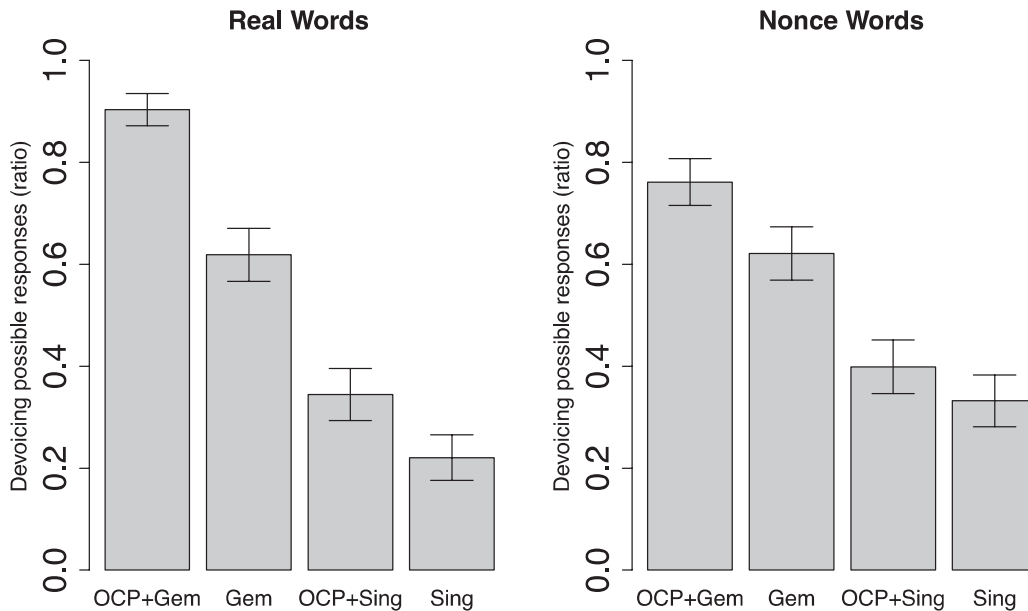
Fig. 6: Average DEVOICING POSSIBLE response ratios in the orthography-based yes/no test. Error bars represent 95% confidence intervals.

A logistic linear mixed model run on real words shows that OCP ($z = 4.17$, $p < .001$), GEM ($z = 11.09$, $p < .001$), and their interaction ($z = 3.67$, $p < .01$) are all significant. OCP and GEM each increase the possibility of devoicing. The significant interaction shows that the effect of OCP is bigger on the geminate pair (0.28 increase in ratio (0.90 – 0.62)) than on the singleton pair (0.12 increase in ratio (0.34 – 0.22)).

For nonce words, OCP ($z = 2.17$, $p < .05$) and GEM ($z = 8.56$, $p < .001$) are significant, but their interaction is not ($z = 1.65$, *n.s.*). There is some difference in the effect of OCP between the geminate pair (0.76 – 0.62 = 0.14) and the singleton pair (0.40 – 0.33 = 0.07), but this difference did not reach statistical significance.

3.3 Discussion

3.3.1 The rating experiment vs. the yes/no experiment

First of all, the rating experiment (Kawahara 2012b) and the binary yes/no experiment (the current experiment) yielded the same ordering between the four grammatical conditions. The results further support Nishimura's (2003) original observation, since naive Japanese speakers find devoicing of OCP-violating geminates possible more frequently than devoicing of non-OCP-violating geminates.

The results extend beyond Kawahara (2011a), Kawahara (2011b), and Kawahara (2012b) by showing this pattern with a yes/no format.

Second, even when the speakers made binary yes/no judgments, we observe a four-way grammatical distinction. This result shows that the gradient pattern obtained in Kawahara (2012b) was not due to the fact that the participants used a scale for their judgments (see Coleman and Pierrehumbert 1997; Dankovičová et al. 1998; Frisch et al. 2004 for similar results in wellformedness/word-likeliness judgment tasks). The phonological judgment pattern, at least in the case of Japanese devoicing, shows a gradient distinction that goes beyond a "grammatical" vs. "ungrammatical" dichotomy, regardless of whether we use a scale-based task or a binary yes/no task as the experimental format.

One may argue that this four-way grammatical distinction had arisen from averaging over a non-homogeneous speech community or a non-homogenous set of items. To address this possibility, analyses similar to those reported in Figures 2–5 were run for the current experiment, and these analyses showed that the four-way grammatical distinction did not arise from averaging over a non-homogeneous speech community or a non-homogeneous set of items.

3.3.2 Real words vs. nonce words

As with Kawahara (2012b), we again observe a reduction of variability across the four grammatical conditions in nonce words. As observed in Figure 6, OCP-violating geminates show fewer DEVOICING POSSIBLE responses in nonce words than in real words, and non-OCP violating singletons show more DEVOICING POSSIBLE responses in nonce words than in real words. To assess this decrease in variability in nonce words with respect to real words, standard deviations across the four grammatical conditions in the number of DEVOICING POSSIBLE responses for each condition were calculated. The average standard deviations in the numbers of DEVOICING POSSIBLE responses were 3.04 for the real word condition and 2.36 for the nonce word condition, and the difference is significant according to a within-subject Wilcoxon test ($p < .001$). Speakers make less consistent, less committed responses to each grammatical condition in nonce words than in real words, which results in less variability across the four grammatical conditions in nonce words.

4 Experiment II: Audio-based yes/no experiment

The second experiment is an audio-based experiment that used a yes/no format. The primary purpose of the experiment is to investigate whether the results of the

previous orthography-based experiments (the previous two experiments as well as those reported in Kawahara 2011a, 2011b) can be replicated with auditory stimuli.

4.1 Method

4.1.1 Stimuli

Experiment II used the same set of stimuli as the previous experiments (Kawahara 2012b and Experiment I above). To obtain the auditory stimuli, a female native speaker of Japanese, who was naive to the purpose of this paper, pronounced all the stimuli (both faithful renditions of the stimuli (e.g., [doggu]) and forms undergoing devoicing (e.g., [dokku])) seven times in a sound-attenuated booth. She was asked to read all the stimuli with a pitch accent on the initial syllable, i.e., with HL tonal contour.

Her speech was recorded with an AT4040 Cardioid Capacitor microphone with a pop filter and amplified through an ART TubeMP microphone pre-amplifier (JVC RX 554V), digitized at a 44K sampling rate. From the seven repetitions, tokens that have phonetic deviance – such as heavy creakiness or unusual F0 contours – were first excluded. Among those that did not have such problems, one token was chosen for each test item. To equalize the amplitudes of the stimuli, the peak amplitude of all stimuli was modified to 0.8 by Praat (Boersma and Weenink 1999–2013). Then the files were converted to mp3 files and embedded in a Sakai test. In her pronunciation, as expected, voiced geminates were semi-devoiced phonetically (Kawahara 2006; see also Hirose and Ashby 2007 and Matsuura 2012). As illustrated in the right panel of Figure 7, voicing during closure ceases at an early phase of the constriction interval. (However, see Kawahara 2006 for evidence that this phonetic semi-devoicing does not itself result in neutralization of a phonological voicing contrast in geminates.)

4.1.2 Participants and procedure

Experiment II was a judgment experiment using a yes/no format; the participants were presented with an original form and a form that undergoes the devoicing in audio formats, and were asked if the second form was a possible pronunciation of the original form. Twenty-five speakers participated in this experiment. The experiments were run in a quiet room at a Japanese university, using headphones. Other aspects of the experiment were identical to the previous two experiments,
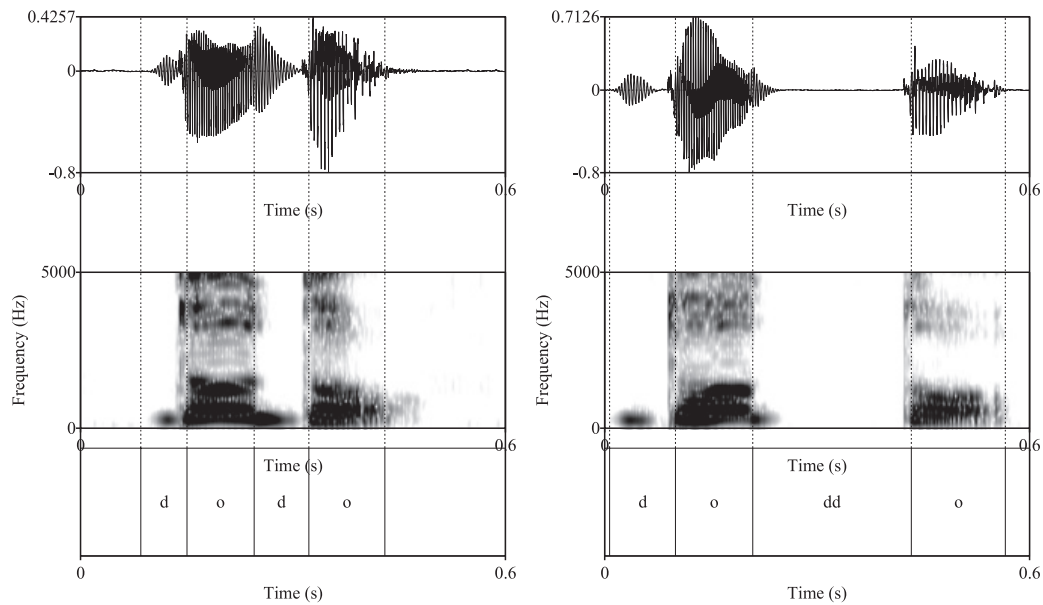
Fig. 7: A comparison of a singleton [d] and a geminate [dd] in the current stimuli.

except that the experimenter sat with the participants. As with Experiment I, within each trial, the participants were presented with an original form (e.g., [doggu] 'dog') and the form that undergoes devoicing (e.g., [dokku]). They were asked whether the second form is a possible pronunciation of the original form or not. No orthographic representations of the stimuli were given – the participants only saw play buttons. Since the two stimuli were presented as two separate play buttons, there was no fixed inter-stimulus interval. Participants were allowed to listen to the stimuli as many times as they liked.

4.2 Results

Figure 8 illustrates the results of Experiment II. The real words show the by-now familiar order: OCP-violating geminates (0.87) > non-OCP-violating geminates (0.68) > OCP-violating singletons (0.17) > non-OCP-violating singletons (0.12). For real words, GEM ($z = 11.12$, $p < .001$) is significant, and OCP is not ($z = 1.42$, $n.s.$). However, the interaction is significant ($z = 2.18$, $p < .05$), reflecting the fact that OCP has a more tangible effect on the geminate pair than on the singleton pair. Within the geminate pair, OCP is significant ($z = 4.94$, $p < .001$).

The nonce words show non-significant reversals within the geminate and the singleton pairs: non-OCP-violating geminates (0.87) > OCP-violating geminates (0.84) > non-OCP-violating singletons (0.36) > OCP-violating singletons (0.35).
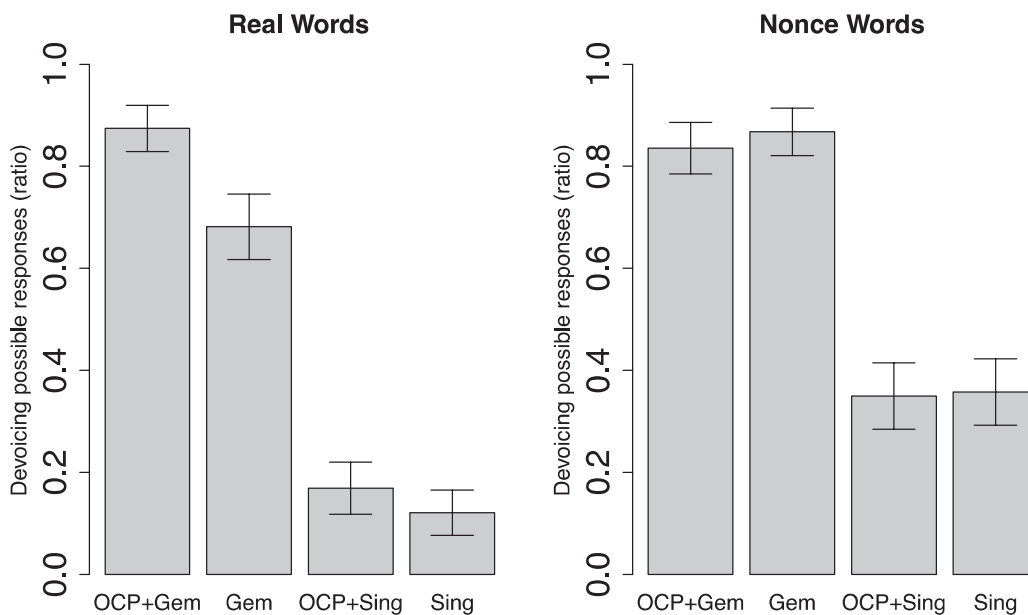
Fig. 8: Average DEVOICING POSSIBLE response ratios in the audio-based yes/no test.

The statistical test shows that only GEM ($z = 10.78$, $p < .001$) is significant, but not OCP ($z = -0.12$, *n.s.*) or the interaction ($z = -0.76$, *n.s.*). The reversal is not significant in the geminate pair ($z = -1.15$, *n.s.*) nor in the singleton pair ($z = -0.13$, *n.s.*).

4.3 Discussion

4.3.1 Orthography stimuli vs. auditory stimuli

The ordering between the four grammatical conditions in real words in Experiment II is identical to that observed in Experiments I and the previous studies (Kawahara 2011a, 2011b, 2012b). At least in the real word condition, the experiment with auditory stimuli yielded results similar to those in the orthography-based tests. In nonce words, the difference due to the OCP disappeared in both the singleton pair and the geminate pair.

One noticeable difference between auditory stimuli and orthographic stimuli is that the effect of GEM is larger in the current audio-based experiment than in the orthography-based experiment (Experiment I). The average difference between the geminate conditions and the singleton conditions in the number of DEVOICING POSSIBLE responses is 14.43 in Experiment I and 20.17 in Experiment II. To assess this difference statistically, a between-subject Wilcoxon test was run and it showed a significant effect ($p < .001$). The magnified effect of GEM in the

auditory condition may be responsible for the lack of effect of OCP in nonce words, since the participants' attention was directed to the difference due to GEM more in the audio-based experiment. The difference due to OCP was diminished in nonce words, since the variability between the four conditions was reduced in general in nonce words (see below in Section 4.3.2).

The reason for this magnified effect of GEM in Experiment II perhaps lies in the phonetic semi-devoicing in Japanese voiced geminates. As we observe in Figure 7, Japanese voiced geminates are phonetically semi-devoiced. Therefore, the participants heard renditions of voiced geminates that were already close to voiceless counterparts. On the other hand, voiced singleton stops were fully voiced, which sound more different from their voiceless counterparts. This difference in the perceptibility of the [voice] contrasts was demonstrated in the perception experiment reported in Kawahara (2006). Therefore, the effect of a particular phonetic implementation pattern – semi-devoicing in this case – is likely to have affected the possibility of devoicing in the current experiment. The current result thus accords well with Kawahara's (2006) hypothesis that the higher voicing neutralizability of geminates may have its roots in the phonetic semi-devoicing of voiced geminates in Japanese.

4.3.2 Reduction of variability in nonce words

Again, similar to the previous two experiments, differences in naturalness ratings across the four different conditions are reduced in nonce words. Average standard deviations in the numbers of DEVOICING POSSIBLE responses are 3.54 for the real words and 2.77 for the nonce words ($p < .001$).

5 General discussion

5.1 Summary

To summarize, we started with three questions regarding the judgment patterns of devoicing in Japanese: (i) the similarity and the difference between real words and nonce words, (ii) the difference between scale-based judgments and yes/no judgments, and (iii) the difference between orthographic stimuli and auditory stimuli. The findings are that, throughout all the experiments, nonce words and real words generally show similar patterns, but nonce words show less variability across the four grammatical conditions than real words. The comparison between

Kawahara (2012b) and Experiment I shows that experiments using a scale-based rating and those using a binary yes/no format show very similar results. The comparison between Experiment I and Kawahara (2012b) on the one hand and Experiment II on the other shows that auditory stimuli and orthographic stimuli yield comparable results, especially in real words. However, the effect of a particular phonetic implementation – semi-devoicing in Japanese voiced geminates – is exaggerated in the audio-based experiment.

5.2 Supporting the intuition-based data

Concerning the status of OCP-violating geminates, which were treated as special by Nishimura (2003) and Kawahara (2006), all the experiments, except for the nonce word condition in Experiment II, showed that OCP-violating geminates received highest naturalness scores, or were judged to be most likely to devoice. In the current experiments, this status of OCP-violating geminates is thus shown to hold even under different modes of phonological judgments, including nonce words. In this regard, the experiments further support the intuition-based data provided by Nishimura (2003) and Kawahara (2006). Therefore, expanding on Kawahara (2011a, 2011b) by testing various modes of phonological judgment, the current experiments contribute to further secure the empirical bases of the debates that were based on Japanese loanword devoicing phenomena, briefly reviewed in Section 1.1. In other words, we can perhaps conclude that the use of intuition-based data by Nishimura (2003) and Kawahara (2006) was reliable.

More generally, the current results are in line with the body of recent experimental work by Sprouse and his colleagues (Sprouse and Almeida 2010; Sprouse et al. 2011; Sprouse and Almeida 2011, 2012) showing that intuition-based data used in generative syntax are generally reliable given that they are replicated by experiments using naive native speakers. I do not wish to imply that experimental verification of linguistic data is hence not necessary. Given some cases that cannot be replicated by experiments (recall the discussion in Section 1.1), we should continue to experimentally verify the quality of the phonological data that we use in building phonological theories and, indeed as a result we may discover finer-grained distinctions, as was the case here.

5.3 Beyond the intuition-based data

While the experimental results generally agree with the introspection-based data by Nishimura (2003) and Kawahara (2006), the experiments have also

demonstrated that both the naturalness hierarchy (Kawahara 2012b) and de-voiceability hierarchy (Experiments I and II) show a distinction that goes beyond a binary "grammatical" vs "ungrammatical" distinction. This gradient pattern is observed even when the participants use a binary yes/no method (see also Bader and Mäussler 2010; Coleman and Pierrehumbert 1997; Dankovičová et al. 1998; Frisch et al. 2004 for similar results). The current experiments thus show that gradient judgment patterns do not necessarily arise because many experiments in the past have used a rating scale; i.e., that it is not a task effect (c.f. Gorman (forthcoming) and Schütze (2011)). In this sense, experimentation can reveal subtle aspects of our linguistic knowledge which can be missed by an approach that is exclusively based on intuition. Therefore, experimental approaches to phonological patterns can complement – but not replace – a more-traditional approach to phonology.

5.4 Where does gradience come from?

The current experiments show that Japanese speakers' judgments on devoicing is generally gradient, even when a yes/no format is used. One question that arises is where this gradience comes from. Even given this result, one could still hold that grammar is dichotomous, and that it is performance that is gradient (e.g., Sprouse 2007). However, recall that generally OCP and GEM both contribute to the naturalness/possibility of devoicing, and these two forces are most likely grammatical. A remaining question therefore is to identify where the gradience comes from – whether it be the grammar or performance – and if it is performance, how the two grammatical factors can derive such gradiency in performance. See Gorman (forthcoming) for recent related discussion.

5.5 Conclusion

To conclude, the three experiments generally replicated the results of the previous studies on Japanese loanword devoicing (Kawahara 2011a, 2011b) with different experimental settings. However, they also revealed interesting differences between certain conditions (for example, the difference between real words and nonce words). Although this paper used Japanese loanword devoicing as a case study, and thus its contribution is limited in its scope, it is hoped that further experimentation will reveal how systematic these differences are across different phonological phenomena and across different languages. To the extent that they

are different, further theoretical research should address how to model such differences.

Acknowledgments: The experiments reported in this paper were first reported in a longer manuscript circulated in 2010 (Kawahara 2010). For their insightful comments on this paper and Kawahara (2010), I am grateful to Osamu Fujimura, Kazu Kurisu, Julien Musolino, Jeremy Perkins, Jason Shaw, Mariko Sugahara, Kristen Syrett, Kyoko Yamaguchi, four anonymous reviewers, and the audience at the colloquium talks at the University of Pennsylvania and Johns Hopkins University, and the participants of Japan Phonology Forum 2011. Sophia Kao and Melanie Pangilinan helped me with the stimulus preparation for Experiment II. Finally, I would like to thank Hope McManus for her edits at the final stage of the manuscript preparation. This research is partly supported by Research Council grant to the author from Rutgers University. All remaining errors are mine.

References

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26. 9–41.

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

Baayen, R. Harald. 2009. LanguageR: R package. http://cran.r-project.org/web/packages/language (accessed 5 September 2013)

Baayen, R. Harald., 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.

Bader, Markus & Jana Mäussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46. 273–330.

Bailey, Todd & Ulrike Hahn. 2001. Determinants of wordlikeliness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44. 568–591.

Batchelder, Eleanor Olds. 1999. Rule or rote? Native-speaker knowledge of Japanese verb inflection. *Proceedings of the Second International Conference on Cognitive Science* and the 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference (ICCS/JCSS99). 141–146.

Bates, Douglas, Martin Maechler & Ben Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R package. http://cran.r-project.org/web/packages/lme4/index.html (accessed 5 September 2013)

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Boersma, Paul & David Weenink. 1999–2013. Praat: Doing phonetics by computer. http://www.praat.org/ (accessed 5 September 2013)

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Coetzee, Andries W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84. 218–257.

Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 30. 47–89.

Coleman, John & Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational phonology: Third meeting of the ACL special interest group in computational phonology*, 49–56. Somerset: Association for Computational Linguistics.

Crawford, J. Clifford. 2009. *Adaptation and transmission in Japanese loanword phonology*. Ithaca, NY: Cornell University dissertation.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28. 197–234.

Dankovičová, Jana, Paula West, John Coleman & Andrew Slater. 1998. Phonotatic grammaticality is gradient. Poster presented at LabPhon 6, University of York, 4 July.

Davis, Stuart & Natsuko Tsujimura. 1991. An autosegmental account of Japanese verbal conjugation. *Journal of Japanese Linguistics* 13. 117–144.

Dąbrowska, Ewa. 2010. Naive vs. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27. 1–23.

Farris-Trimble, Ashley. 2008. *Cumulative faithfulness effects in phonology*. Bloomington, IN: Indiana University dissertation.

Frisch, Stefan, Nathan Large & David Pisoni. 2004. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42. 481–496.

Fukuda, Suzy & Shinji Fukuda. 1994. To voice or not to voice: The operation of rendaku in the Japanese developmentally language-impaired. *McGill Working Papers in Linguistics* 10. 178–193.

Gibson, Edward & Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14. 233–234.

Goldrick, Matthew. 2011. Utilizing psychological realism to advance phonological theory. In John A. Goldsmith, Jason Riggle & Alan Yu (eds.) *The handbook of phonological theory,* 2nd edn., 631–660. Oxford: Blackwell-Wiley.

Gorman, Kyle. forthcoming. Gradient and categorical aspects of wordlikeness judgments. *Proceedings of the 43rd annual meeting of North East Linguistic Society*.

Greenberg, Joseph & James Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20. 157–177.

Griner, Barry. 2001. *Productivity of Japanese verb tense inflection: A case study*. Los Angeles, CA: University of California Los Angeles MA Thesis.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, syntax, and acquisition*, 88–120. Oxford: Oxford University Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry* 39. 379–440.

Hirose, Aki & Michael Ashby. 2007. An acoustic study of devoicing of the geminate obstruents in Japanese. *Proceedings of ICPhS* XVI. 909–912.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*, 817–838. Oxford: Blackwell.

Itô, Junko & Armin Mester. 1998. Markedness and word structure: OCP effects in Japanese. Ms. University of California Santa Cruz.

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

Itô, Junko & Armin Mester. 2003. *Japanese morphophonemics*. Cambridge, MA: MIT Press.

Itô, Junko & Armin Mester. 2008. Lexical classes in phonology. In Shigeru Miyagawa & Mamoru Saito (eds.), *The Oxford handbook of Japanese linguistics*, 84–106. Oxford: Oxford University Press.

Jaeger, Florian T. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59. 434–446.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology*. Santa Cruz, CA: University of California Santa Cruz dissertation.

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

Kawahara, Shigeto. 2008. Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* 17. 317–330.

Kawahara, Shigeto. 2010. Modes of phonological judgments. Ms. Rutgers University.

Kawahara, Shigeto. 2011a. Aspects of Japanese loanword devoicing. *Journal of East Asian Linguistics* 20. 169–194.

Kawahara, Shigeto. 2011b. Japanese loanword devoicing revisited: A rating study. *Natural language and Linguistic Theory* 29. 705–723.

Kawahara, Shigeto. 2012a. Gairaigo yuusei sokuon no museika: Rironteki kouken [Japanese loanword devoicing: A review]. In Phonological Society of Japan (ed.), *On'in kenkyu* [Phonological studies] 15, 93–104. Tokyo: Kaitakusha.

Kawahara, Shigeto. 2012b. Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment experiments. *Lingua* 122. 1193–1206.

Kawahara, Shigeto. forthcoming. The phonetics of obstruent geminates, *sokuon*. In Haruo Kubozono (ed.), *The handbooks of Japanese language and linguistics: Phonetics and phonology*. Berlin & Boston: De Gruyter Mouton.

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

Kubozono, Haruo, Junko Itô & Armin Mester. 2009. Consonant gemination in Japanese loanword phonology. In Linguistic Society of Korea (ed.), *Current issues in unity and diversity of languages: Collection of papers selected from the 18th International Congress of Linguists*, 953–973. Seoul: Dongam Publishing Co.

Labov, William. 1996. When intuitions fail. In Lisa McNair, Kora Singer, Lise Dolbrin, & Michelle Aucon (eds.), *Proceedings of Chicago Linguistic Society 32: Papers from the parasession on theory and data in linguistics*, 77–106. Chicago: Chicago Linguistics Society.

Labrune, Laurence. 2012. *The phonology of Japanese*. Oxford: Oxford University Press.

Leben, Will. 1973. *Suprasegmental phonology*. Cambridge, MA: MIT dissertation.

Lyman, Benjamin S. 1894. Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia*.

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]. Paper presented at the Phonetic Society of Japan, September 29th.

Myers, James. 2009. Syntactic judgment experiments. *Language and Linguistic Compass* 3. 406–423.

Nishimura, Kohei. 2003. *Lyman's Law in loanwords*. Nagoya: Nagoya University MA thesis.

Ohala, John J. 1974. Experimental historical phonology. In John M. Naderson & Charles Jones (eds.), *Historical linguistics II: Theory and description in phonology. Proceedings of the first international linguistic conference on historical linguistics*, 353–389. New York: Elsevier.

Ohala, John J. 1981. The listener as a source of sound change. In T. Myers, J. Laver, & Anderson J. (eds.), *Proceedings of Chicago Linguistic Society* 17, 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed), *The production of speech*, 189–216. New York: Springer.

Ohala, John J. 1986. Consumer's guide to evidence in phonology. *Phonology* 3. 3–26.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035.

Pater, Joe. forthcoming. Universal grammar with weighted constraints. In John J. McCarthy & Joe Pater (eds.), *Harmonic grammar and harmonic serialism*. London: Equinox.

Pertz, Doris L. & Thomas G. Bever. 1975. Sensitivity to phonological universals in children and adolescents. *Language* 51. 149–162.

Pierrehumbert, Janet B. 2001. Stochastic phonology. *GLOT* 5. 1–13.

Quakenbush, Hiroko. 1989. Gairaigo-no sokuonka-ni tsuite [On geminaton in loanwords]. *Ryuugakusei Nihongo Kyouiku-ni Kansuru Rironteki Jissenteki Kenkyuu* [Theoretical and practical research on Japanese language education for foreign students]. 1–7.

Quené, Hugo & Huub van den Berg. 2008. Examples of mixed effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59. 413–425.

R Development Core Team. 1993–2013. *R: A language and environment for statistical computing*. Vienna, Austria. R Foundation for Statistical Computing. Software, available at http://www.R-project.org.

Reips, Ulf-Dietrich. 2002. Standards for internet-based experimenting. *Experimental Psychology* 49. 243–256.

Rice, Keren. 2006. On the patterning of voiced stops in loanwords in Japanese. *Toronto Working Papers in Linguistics* 26. 11–22.

Sanders, Nathan. 2003. *Opacity and sound change in the Polish lexicon*. Santa Cruz, CA: University of California Santa Cruz dissertation.

Sano, Shinichiro & Shigeto Kawahara. forthcoming. A corpus-based study of geminate devoicing in Japanese: The role of the OCP and external factors. Gengo Kenkyuu [The Journal of the Linguistic Society of Japan].

Schütze, Carlson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Schütze, Carlson. 2011. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2. 206–211.

Shademan, Shabhame. 2007. *Grammar and analogy in phonotactic well-formedness judgments*. Los Angeles, CA: University of California Los Angeles dissertation.

Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115. 1497–1524.

Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1. 123–134.

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior and Research Methods* 43. 155–167.

Sprouse, Jon & Diogo Almeida. 2010. A quantitative defense of linguistic methodology. Ms. University of California, Irvine.

Sprouse, Jon & Diogo Almeida. 2011. Power in acceptability judgment experiments. Ms. University of California, Irvine.

Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics* 48. 609–652.

Sprouse, Jon, Carlson Schütze & Diogo Almeida. 2011. Assessing the reliability of journal data in syntax: Linguistic Inquiry 2001–2010. Ms. University of California, Irvine, University of California Los Angeles & Michigan State University.

Steriade, Donca. 2004. Projecting non-lexical phonology from phonetic knowledge. Paper presented at the Symposium on Phonological Theory: Representations and Architectures (City University of New York, February 20–21).

Tateishi, Koichi. 2002. Bunpou-no ichibutoshite-no goisou-no zehi [Lexical strata as a part of grammar]. *Onsei Kenkyuu* [Journal of the Phonetic Society of Japan] 6. 34–43.

Tesar, Bruce. 2007. A comparison of lexicographic and linear numeric optimization using violation difference ratios. Ms. Rutgers University.

Tsujimura, Natsuko. 1996. *An introduction to Japanese linguistics* (Blackwell Textbooks in Linguistics 10). Cambridge, MA & Oxford: Blackwell Publishers.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. New York: SUNY Press.

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