# 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. /be**dd**o/ → [be**tt**o] 'bed'). This devoicing pattern has contributed much to address a number of theoretical issues in the recent phonological literature. However, the relevant data has been primarily based on intuition-based data provided by Nishimura (2003) and Kawahara (2006). Kawahara (2011a,b) 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,b) by adding (i) nonce word stimuli, (ii) a binary yes/no experiment, and (iii) auditory stimuli. The results of three experiments 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) a binary yes/no experiment shows results similar to those of a 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 and other languages.

# 1 Introduction

## 1.1 The phenomenon

This paper is about devoicing of obstruents in the loanword phonology of Japanese. It has been known that some 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 latter type as "unassimilated foreign items". Instead of relying on a (more or less) arbitrary etymological distinction, Nishimura (2003) proposed 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. This restriction restriction which has long been known as Lyman's Law in Japanese phonology (Lyman, 1894; Kawahara, 2012b; Vance, 2007), and has been formalized as the 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, as an interesting twist, Nishimura (2003) argues that devoicing is also impossible in OCP-violating singletons, as in (3).

(1) Voiced obstruent geminates can optionally devoice if they co-occur with another voiced obstruent; i.e. when they violate the OCP[voice]
   a. be**dd**o → be**tt**o 'bed'
   b. ba**gg**u → ba**kk**u 'bag'
   c. bi**gg**u → bi**kk**u 'big'

(2) Voiced obstruent geminates do not devoice if they do not violate the OCP[voice]
   a. suno**bb**u → *suno**pp**u 'snob'
   b. he**dd**o → *he**tt**o 'head'
   c. re**gg**u → *re**kk**u 'leg'

(3) Voiced singletons do not devoice even when they violate the OCP[voice]
   a. da**b**u → *da**p**u 'Dove'
   b. do**g**uma → *do**k**uma 'dogma'
   c. 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, but 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, to appear; Tesar, 2007); (iii) how the spontaneous emergence of the loanword devoicing in (1) bears on a 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) raised one issue: the Japanese loanword devoicing data is 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 against 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 do not know about the devoicing pattern. The experiment indeed showed that Japanese speakers generally judge devoicing of OCP-violating geminates more natural than devoicing in other environments. In this regard, Kawahara (2011b) succeeded in supporting the empirical basis of the claims made based on 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 most natural for native speakers of Japanese.

## 1.2 The current study

However, there are some remaining questions, which the current paper aims to address. First, both Kawahara (2011a) and Kawahara (2011b) used only real words, but it is of some interest to investigate whether the results obtained with real words generalize to nonce words in the case of Japanese loanword devoicing. An often-used test on phonological productivity is a wug-test (Berko, 1958), in which the 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 alleged phonological patterns are not productive i.e. 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 not probably 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 observed in Japanese verbal paradigms (Davis and Tsujimura, 1991; Tsujimura, 1996) have not been replicated by 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 true

[1] See Nishimura (2003), Kawahara and Sano (2012) and Sano and Kawahara (2012) for some evidence based on corpus data.

productivity of the phenomenon under question using nonce words. One of the questions that this paper addresses is thus whether the results of the previous rating studies can be replicated using nonce words as stimuli, and if so to what extent.

Second, Kawahara (2011a,b) found that Japanese speakers distinguish the naturalness of two processes that were both judged to be "ungrammatical" by Nishimura (2003) and Kawahara (2006); e.g. devoicing of non-OCP violating geminates (=(2)) and devoicing of OCP-violating singletons (=(3)). One may wonder whether this gradient effect was due to a task effect; the reason for the gradient result may be because Kawahara (2011a,b) used a gradient scale. Testing this issue is in part motivated by the debate concerning the gradient nature of phonological judgments. It has been known that grammatical judgments show distinctions beyond a simple "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 that we obtain gradient results in experimental settings because the these experiments often use scales. Therefore, the second aim of this paper is to test whether the gradient results that Kawahara (2011a,b) 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, Kawahara (2011a,b) used visual, orthographic stimuli, although the instructions in these studies encouraged the participants to read the stimuli in their heads and use the 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 only 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,b) can be replicated with auditory stimuli, because phonology is concerned with sounds. Testing the Japanese devoicing pattern with auditory stimuli is therefore the third aim of this paper.

To summarize, the three issues that this paper aims to address are: (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 reports three experiments that 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 what was found in Kawahara (2011a,b).

Before reporting the actual experiments, a few remarks are in order. First, the experiments reported in this paper are judgment experiments on 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. a phonological form and its optional variant in this case). This task therefore differs from phonotactic wellformendess 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 from Japanese of such phonological judgment studies. 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 Experiment I: Orthography-based rating experiment

The first experiment is an orthography-based rating experiment.[2] The main purposes of this experiment are (i) to replicate Kawahara (2011a,b) and, more importantly, (ii) to test whether the results obtained with real words in the previous studies generalize to nonce words, and (iii) to compare the patterns of real words and nonce words.

## 2.1 Method

### 2.1.1 Stimuli

All three experiments reported in this paper used the same set of stimuli, which consisted of four grammatical conditions: OCP-violating geminates, non-OCP-violating geminates, OCP-violating singletons, and non-OCP-violating singletons with a representative example, as summarized in (4). In this design, two factors—OCP and GEM—were fully crossed. This paper uses CAPITAL LETTERS to represent variable names.

(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])

[2]This experiment is also reported (in less detail) in Kawahara (2012b) to show the activity of the OCP—or Lyman's Law—in loanwords and nonce words, together with two other experiments on Rendaku.

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 (all lexical geminates in Japanese appear word-internally: Kawahara to appear). The stimulus set was constructed in the following way. First, real disyllabic words containing OCP-violating geminates—the case for which we have the least number of existing items in the Japanese lexicon—were chosen. This selection process resulted in 9 items. Among the 9 items, 6 items contained [dd] followed by epenthetic [o], and 3 items contained [gg] followed by epenthetic [u]. No stimuli with [bb] were found, because [bb] is very rare in Japanese loanwords (Katayama, 1998) and therefore no disyllabic words with OCP-violating [bb] exist.[3] Then the words for the other three conditions were selected with six items with [d(d)] and three items with [g(g)], controlling place of articulation, as listed in Table 1. 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], not a singleton [d], due to a productive gemination process in loanword adaptation.(Kubozono et al., 2008)[4] All the stimuli have a pitch accent on the initial syllable, which is phonetically realized as a HL falling pitch.

Table 1: The list of the stimuli, 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 | 'mug' |
| doggu | 'dog' | taggu | 'tag' | jogu | 'jog' | ragu | 'rag' |

The nonce word stimuli are listed in Table 2. The nonce word stimuli had the same phonological structures as the real word stimuli, except that all the nonce word stimuli had short initial vowels. (Nonce words can have a short vowel before a singleton [d]).

[3]The only lexical item that has an OCP-violating [bb] is *gebberusu* 'Göbbels', which is not disyllabic.
[4][bado] is a truncated form of [badominton].

Table 2: The list of the stimuli, 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 |

### 2.1.2 Task

In this experiment Japanese speakers rated the naturalness of devoicing in the four grammatical conditions. The instructions explained that the questionnaire was about the naturalness of devoicing in Japanese loanwords. Using the same format as Kawahara (2011a,b), for each question, the participants were presented with one stimulus 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 the 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, because *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 orthography.

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

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.[5]

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

### 2.1.3 Procedure

Sakai (https://sakai.rutgers.edu/portal), a java-based online system which runs online 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 we cannot control the environments 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 at a laboratory (see Reips 2002 and Sprouse 2011 for further, general discussion on online experimentation in psychology and linguistics).

The first page of the experimental website presented a consent form, which was followed by the instructions of the experiment. After the instructions, the main session started where one trial was presented on one page. The order of the stimuli within each block was randomized by Sakai. At the end of the experiment, as a pre-screening procedure before the data analysis, the participants were asked if they were familiar with theoretical issues surrounding the devoicing phenomenon.

### 2.1.4 Participants

Thirty-three native speakers of Japanese, mainly students at a Japanese university, participated in this online experiment. One speaker was familiar with the devoicing phenomenon, and therefore his/her data did not enter the following analysis.

### 2.1.5 Statistics

The responses were first converted to numerical values as follows: "very natural"=5; "somewhat natural"=4; "neither natural nor unnatural"=3; "somewhat unnatural"=2; "very unnatural"=1. For statistical analyses, a general linear mixed model was run (Baayen et al., 2008; Baayen, 2008) in which OCP and GEM were fixed factors, using R (R Development Core Team, 1993–2013) with the `lme4` package (Bates et al., 2011).[6] The p-values were calculated by the Markov chain Monte Carlo method using the `languageR` package (Baayen, 2009).

[6]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 (2012) for discussion and 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 either. The targeted comparison between real words and nonce words is provided in the discussion section (section 2.3).

## 2.2 Results

Figure 1 illustrates average rating scores in Experiment I. 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), replicating the previous studies (Kawahara, 2011a,b). Statistically, for real words, all factors are significant: OCP ($t = 5.29, p < .001$), GEM ($t = 11.81, p < .001$), and the interaction ($t = 2.68, p < .01$). 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).

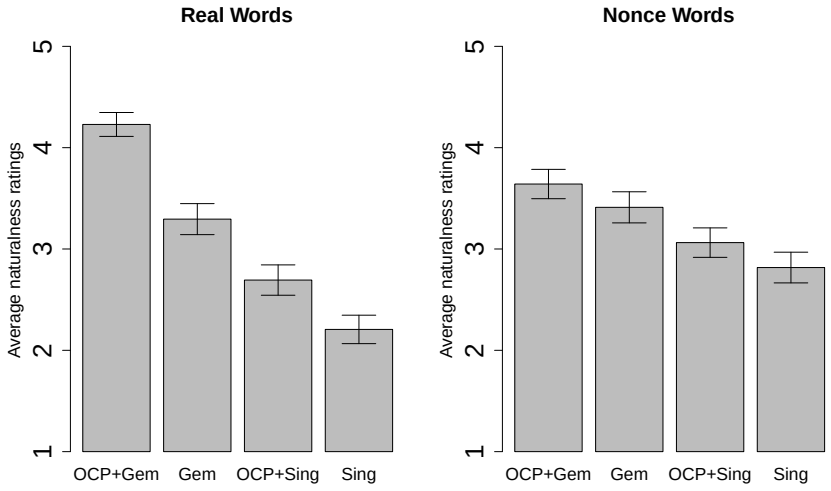


Figure 1: The average naturalness ratings in the orthography-based rating experiment (Experiment I). The error bars represent 95% confidence intervals.

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 OCP ($t = 2.56, p < .05$) and GEM ($t = 6.44, p < .001$) are both significant, but their interaction is 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 between real words and nonce words. The order also matches with what is found in the two previous studies using real words (Kawahara, 2011a,b). In this sense, the current experiment has shown that the results of previous studies using 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 further contributes to support the empirical foundation of the theoretical claims made based on the Japanese loanword devoicing pattern (see section 1.1).

At the same time, we observe a difference between real words and nonce words: there is less variability in naturalness ratings across the four grammatical conditions in nonce words than in real words. In other words, ratings differ less between the four grammatical conditions in nonce words than in real words. The most natural devoicing (OCP-violating geminates) is judged to be less natural in nonce words than in real words, and the least natural devoicing (non-OCP-violating geminates) 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 be responsible for the lack 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 clear-cut grammatical judgments in general; i.e. they are reluctant to use endpoints of judgment scales. Despite this difference between real words and nonce words, 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,b). It does not seem possible to divide the grammatical 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.[7]

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 resulted in gradient patterns. This hypothesis predicts distributions of responses at two extremes, because people should consistently rate each devoicing pattern either as completely grammatical (=5 in rating) or completely 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 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,

[7] Kawahara (2011a,b) presents some speculations about why Japanese speakers find the devoicing of non-OCP-violating geminates more natural than that of OCP-violating singletons. Beyond the speculations presented there, a 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. Given these 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).

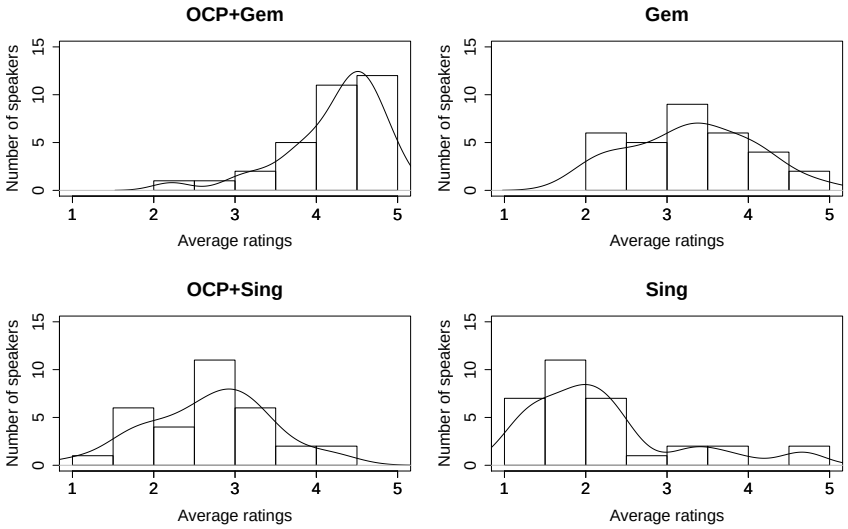


Figure 2: A histogram of naturalness ratings (number of speakers), real words.

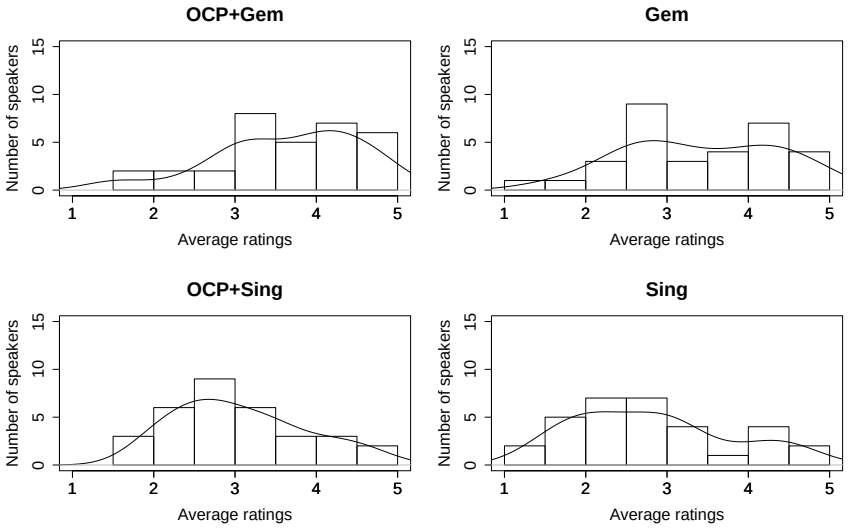


Figure 3: A histogram of naturalness ratings (number of speakers), nonce words.

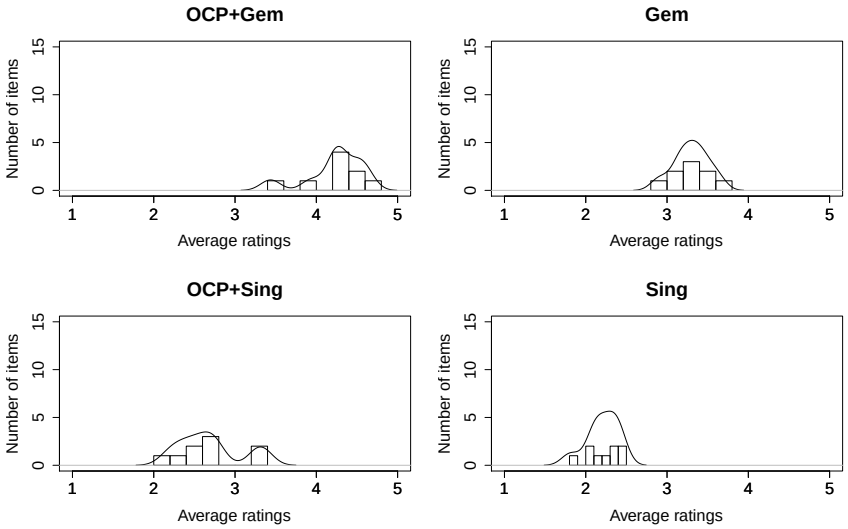


Figure 4: A histogram of naturalness ratings (number of items), real words.

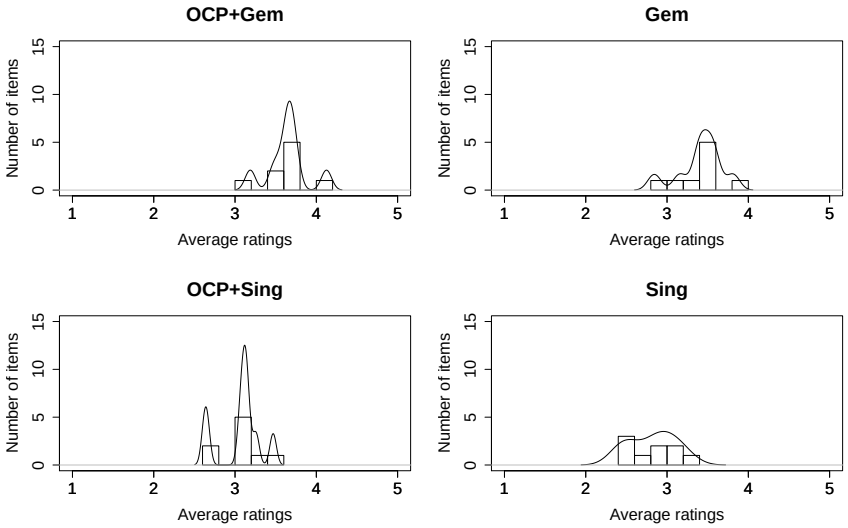


Figure 5: A histogram of naturalness ratings (number of items), nonce words.

2007).

# 3 Experiment II: Orthography-based yes/no experiment

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

## 3.1 Method

Experiment II is similar to Experiment I, but it instead asked native speakers whether devoicing in each of the four grammatical conditions is possible or not in a binary yes/no format. Thirty-seven native speakers of Japanese, again mainly university students in Japan, participated in this experiment. There is no overlap between the participants of Experiment I and those of Experiment II. 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.

A logistic linear mixed model 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

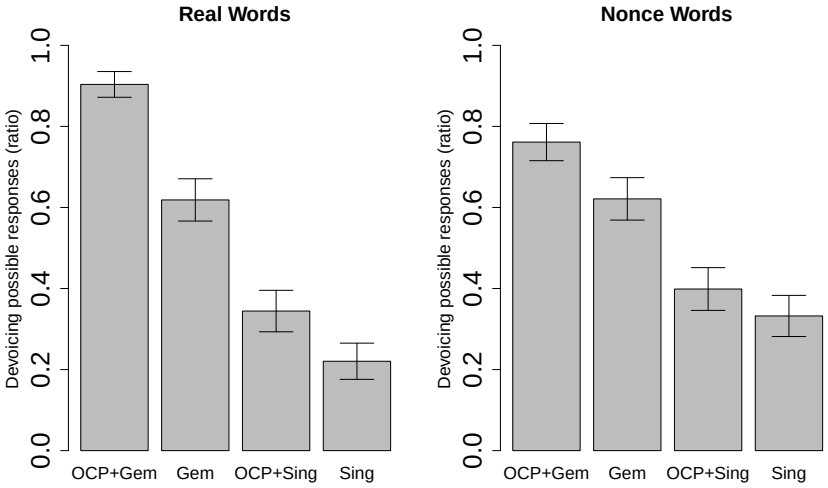


Figure 6: Average DEVOICING POSSIBLE response ratios in a orthography-based yes/no test (Experiment II). The error bars represent 95% confidence intervals.

geminate pair (0.76-0.62=0.14) and the singleton pair (0.40-0.33=0.07), but the 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 (Experiment I) and the binary yes/no experiment (Experiment II) yielded the same ordering between the four grammatical conditions. The results thus further support Nishimura's (2003) original observation in that naive Japanese speakers most frequently find the devoicing of OCP-violating geminates possible. The results extend beyond Kawahara (2011a) and Kawahara (2011b) by showing this 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 Experiment I 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 an 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 pos-

sibility, analyses similar to those reported in Figures 2-5 were run for Experiment II, 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 Experiment I, we again observe 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 III: Audio-based yes/no experiment

The final experiment is an audio-based experiment which 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,b) can be replicated with auditory stimuli.

## 4.1 Method

### 4.1.1 Stimuli

Experiment III used the same set of stimuli as the previous two experiments. 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 at 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 through 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 do not have such

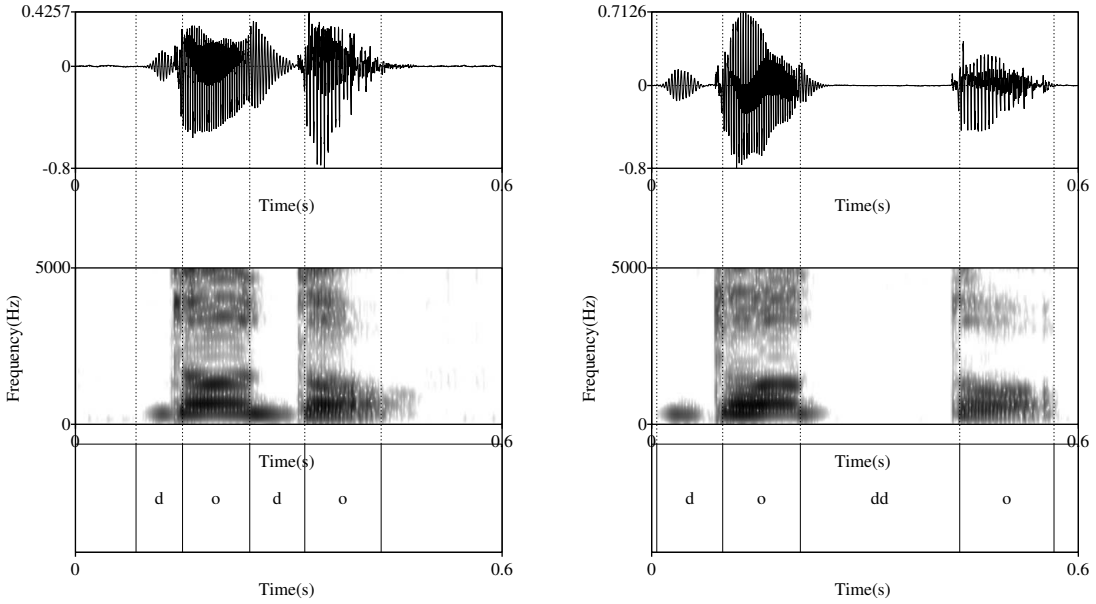


Figure 7: A comparison of a singleton [d] and a geminate [dd] in the current stimuli.

problems, one token was chosen for each test item. To equalize the amplitudes of the stimuli, peak amplitude of all the 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 III 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, except that the experimenter sat with the participants. As with Experiment II, 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 the second form is a possible pronunciation of the original form or not. No orthographic representations of the stimuli were given—the partic-

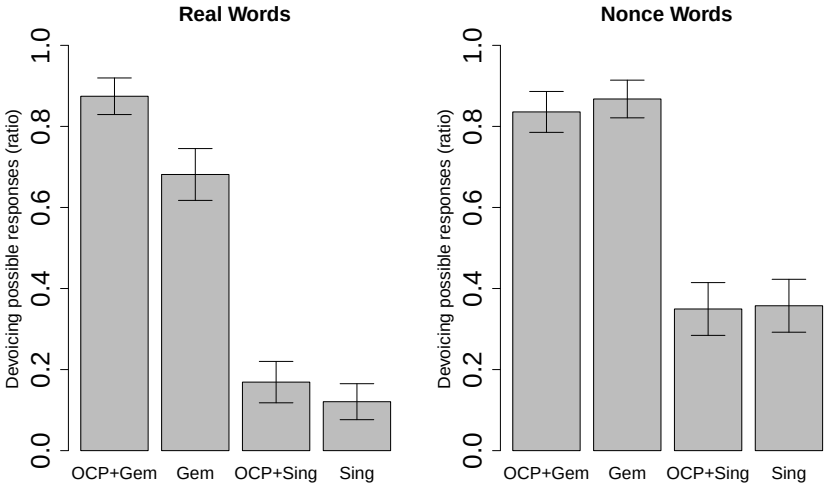


Figure 8: Average DEVOICING POSSIBLE response ratios in a audio-based yes/no test in Experiment III.

ipants 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 like.

## 4.2 Results

Figure 8 illustrates the results of Experiment III. The real words show the by-now familiar hierarchy: 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). 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 ($z = -1.15, n.s.$) or 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 III is identical to that observed in Experiments I and II. 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 II). The average difference between the geminate conditions and the singleton conditions in the number of DEVOICING POSSIBLE responses is 14.43 in Experiment II and 20.17 in Experiment III. To assess this difference statistically, a between-subject Wilcoxon test was run and it showed a significant effect ($p < .001$). This magnified effect of GEM 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, and since the variability between the four conditions was reduced in general in nonce words (see below in section 4.3.2), the difference due to OCP was diminished in nonce words.

The reason for this magnified effect of GEM in Experiment III 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 very 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 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 two previous 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 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 Experiment I and Experiment II shows that experiments using a scale-based rating and those using a binary yes/no format show very similar results. The comparison between Experiments I-II and Experiment III 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 but the nonce word condition in Experiment III showed that OCP-violating geminates received highest naturalness scores, or were judged to be most likely to undergo devoicing. 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,b) 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, 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 not inappropriate.

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 can be replicated by experiments using naive native speakers, and are hence generally reliable. I do not wish to imply that experimental verification of linguistic data is hence not necessary; given some cases which 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.

## 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 (Experiment I) and devoiceability hierarchy (Experiments II and III) show a distinction that goes beyond a dichotomous "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 judgement patterns do not necessarily arise because many experiments in the past has used a rating scale; i.e. that it is not a task effect (cf. Gorman to appear and Schütze 2011 who suggest that gradient effects are partly due to a task effect). 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' judgment on devoicing is generally gradient, even when a yes/no format was 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—grammar or performance—and if the latter, how the two grammatical factors can derive such gradiency in performance (see Gorman to appear 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,b) with different experimental settings. However, they revealed interesting differences between certain conditions (for example, the difference between real words and nonce words) as well. 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, further theoretical research should address how to model such differences.

## Acknowledgements

The experiments reported in this paper were reported in a longer manuscript circulated in 2010 (Kawahara, 2010), and Experiment I was reported in less detail in Kawahara (2012b) together with two other experiments on Rendaku. 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 III. 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 (1): 9–41.

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, and 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, and Jana Mäussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46: 273–330.

Bailey, Todd, and 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*.

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

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

Boersma, Paul, and David Weenink. 1999–2013. Praat: Doing phonetics by computer. Software.

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

Coetzee, Andries W. 2008. Grammaticality and ungrammaticality in phonology. *Language* 84 (2): 218–257.

Coetzee, Andries W., and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 30 (1).

Coleman, John, and 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. Doctoral dissertation, Cornell University.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28 (2): 197–234.

Dankovičová, Jana, Paula West, John Coleman, and Andrew Slater. 1998. Phonotatic grammaticality is gradient. Poster presented at LabPhon 6.

Davis, Stuart, and Natsuko Tsujimura. 1991. An autosegmental account of Japanese verbal conjugation. *Journal of Japanese Linguistics* 13: 117–44.

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

Farris-Trimble, Ashley. 2008. Cumulative faithfulness effects in phonology. Doctoral dissertation, Indiana University.

Frisch, Stefan, Nathan Large, and 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, and 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, and Evelina Fedorenko. 2010. Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences* 14 (6): 233–234.

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

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

Greenberg, Joseph, and 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. MA thesis, University of California Los Angeles.

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

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotac-

tic learning. *Linguistic inquiry* 39: 379–440.

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

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

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

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

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

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

Itô, Junko, and Armin Mester. 2008. Lexical classes in phonology. In *The Oxford handbook of Japanese linguistics*, eds. Shigeru Miyagawa and Mamoru Saito, 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. Doctoral dissertation, University of California, Santa Cruz.

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

Kawahara, Shigeto. 2008. Phonetic naturalness and unnaturalness in Japanese loanword phonology. *Journal of East Asian Linguistics* 17 (4): 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 (2): 169–194.

Kawahara, Shigeto. 2011b. Japanese loanword devoicing revisited: A rating study. *Natural language and Linguistic Theory* 29 (3): 705–723.

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

Kawahara, Shigeto. 2012b. Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment experiments. *Lingua* 122 (11): 1193–1206.

Kawahara, Shigeto. to appear. The phonetics of obstruent geminates, *sokuon*. In *The handbook of Japanese language and linguistics: Phonetics and phonology*, ed. Haruo Kubozono. Mouton.

Kawahara, Shigeto, and Shinichiro Sano. 2012. A corpus-based study of geminate devoicing in

Japanese: Internal factors. Ms. Rutgers University and International Christian University.

Kubozono, Haruo, Junko Itô, and 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*, ed. The Linguistic Society of Korea, 953–973. Republic of Korea: Dongam Publishing Co.

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

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

Leben, Will. 1973. Suprasegmental phonology. Doctoral dissertation, MIT.

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]. Talk presented at the Phonetic Society of Japan, Sept 29th.

Myers, James. 2009. Syntactic judgment experiments. *Language and Linguistic Compass* 3 (1): 406–423.

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

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

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

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. Peter MacNeilage, 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. to appear. Universal grammar with weighted constraints. In *Harmonic Grammar and Harmonic Serialism*, eds. John J. McCarthy and Joe Pater. London: Equinox.

Pertz, D. L., and T. 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.

Quakenbusch, Hiroko. 1989. Gairaigo-no sokuonka-ni tsuite [On geminaton in loanwords]. *Ryuugakusei Nihongo Kyouiku-ni Kansuru Rironteki Jissenteki Kenkyuu*.

Quené, Hugo, and 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 (4): 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. Doctoral dissertation, University of California, Santa Cruz.

Sano, Shinichiro, and Shigeto Kawahara. 2012. A corpus-based study of geminate devoicing in Japanese: The role of the OCP and external factors. Ms. International Christian University and Rutgers University.

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 (2): 206–211.

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

Sorace, Antonella, and Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115 (11): 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 (1): 155–167.

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

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

Sprouse, Jon, and 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, and 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, and Michigan State University.

Steriade, Donca. 2004. Projecting non-lexical phonology from phonetic knowledge. Talk presented at "Symposium on phonological theory: representations and architectures" (the City University

of New York, February 20th-21st, 2004).

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 (1): 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*. Cambridge, Mass., and Oxford, U.K.: 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 *Current issues in the history and structure of Japanese*, eds. Bjarke Frellesvig, Masayoshi Shibatani, and John Carles Smith, 153–170. Tokyo: Kuroshio.