Vowel devoicing and the perception of spoken Japanese words

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and MARCS Auditory Laboratories, University of Western Sydney, Penrith South DC NSW 1797, Australia

Takashi Otake[b)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and E-Listening Laboratory, Tokorozawa, 359 0021, Japan

James M. McQueen[c)]
Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands and Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

(Received 6 August 2007; revised 20 November 2008; accepted 14 December 2008)

Three experiments, in which Japanese listeners detected Japanese words embedded in nonsense sequences, examined the perceptual consequences of vowel devoicing in that language. Since vowelless sequences disrupt speech segmentation [Norris *et al.* (1997). Cognit. Psychol. **34**, 191–243], devoicing is potentially problematic for perception. Words in initial position in nonsense sequences were detected more easily when followed by a sequence containing a vowel than by a vowelless segment (with or without further context), and vowelless segments that were potential devoicing environments were no easier than those not allowing devoicing. Thus *asa*, "morning," was easier in *asau* or *asazu* than in all of *asap, asapdo, asaf*, or *asafte*, despite the fact that the /f/ in the latter two is a possible realization of *fu*, with devoiced [u]. Japanese listeners thus do not treat devoicing contexts as if they always contain vowels. Words in final position in nonsense sequences, however, produced a different pattern: here, preceding vowelless contexts allowing devoicing impeded word detection less strongly (so, *sake* was detected less accurately, but not less rapidly, in *nyaksake*—possibly arising from *nyakusake*—than in *nyagusake*). This is consistent with listeners treating consonant sequences as potential realizations of parts of existing lexical candidates wherever possible. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3075556]

PACS number(s): 43.71.Es [PEI] Pages: 1693–1703

I. INTRODUCTION

In spoken Japanese, the high vowels /i/ and /u/ are devoiced when they follow a voiceless consonant and precede either another voiceless consonant or (in the case of fricatives and affricates) a pause, and do not bear accent (Vance, 1987). Thus *sashimi* "sashimi" (raw fish dish) and *sashiki* "cutting" begin with the same two-mora sequence; in writing, it is the same in both words. But in *sashiki* the /i/ of the medial mora, occurring before /k/, is devoiced, even in careful speech. The /i/ in *sashimi*, occurring before /m/, never devoices. Although devoicing is not obligatory, analyses of the Corpus of Spontaneous Japanese (Maekawa, 2003) show that it is highly probable (over 98% in some environments; Kondo, 2005; Maekawa and Kikuchi, 2005).

The effect of this devoicing is the creation of sequences of consonants not separated by the periodic articulation normally associated with vowels. Otherwise, though, Japanese phonology drastically restricts the occurrence of consonant sequences. Japanese has no consonant clusters, and allows only a very restricted range of simple syllable codas: nasals (*Hondo; Unzen*) or geminate consonants (*Hokkaido*; *Sapporo*—these are all place names). Loan-words that contain consonant sequences or un-Japanese codas are adapted by vowel insertion; e.g., *glove* becomes *gurabu* and *express* becomes *ekisupuresu*. This process of vowel insertion is so fundamental to Japanese phonology that in perceptual tasks Japanese listeners respond to nonsense VCCV strings such as *ebzo* as if they were VCVCV *ebuzo* (Dupoux *et al.*, 1999).

Japanese is not unusual in preferring consonants and vowels to alternate; such a preference appears across languages, and has a good perceptual foundation. It has long been known that a following vowel facilitates consonant identification (Liberman *et al.*, 1954; van Son and Pols, 1995). In line with this, the deletion of vowels, even where it occurs regularly in casual speech, makes words harder to recognize: for example, lexical decision responses can be slower for words with deleted vowels (e.g., *s'maine* for *semaine*; Racine and Grosjean, 2000). In contrast, insertion of a vowel into a consonant cluster (e.g., *fillum* for *film*) makes recognition easier, in part because the consonants in the cluster indeed become easier to identify if separated (van Donselaar *et al.*, 1999).

Vowel devoicing in Japanese could thus be perceptually disadvantageous. Further, it could complicate the parsing of continuous speech into its component words. All speech input is potentially consistent with alternative interpretations; *legacy* contains *leg* embedded within it, but *leg* itself con-

[a)] Present address: Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands. Electronic mail: anne.cutler@mpi.nl
[b)] Electronic mail: otake@e-listeninglab.com
[c)] Electronic mail: james.mcqueen@mpi.nl

tains *egg*, and whenever we hear the longer word spoken, we also receive input consistent with the shorter word or words. There are many powerful and efficient techniques which listeners can apply to reduce interference from unwanted embedded words that are accidentally present in the speech stream (and so might be activated by speech input). In the above example, for instance, interference from *egg* in recognition of *leg* is negligible. The process which achieves this is called the possible word constraint (PWC) (Norris *et al.*, 1997); it exploits the widespread rule that a vowel alone can be a word, but a consonant in general cannot.

The effect of the PWC can be seen in word-spotting (Cutler and Norris, 1988; McQueen, 1996), a psycholinguistic task for investigating segmentation of speech. In word-spotting, listeners detect any real word embedded in spoken nonwords. The task has exposed language-particular segmentation effects such as use of stress information in English (Cutler and Norris, 1988), of syllables in French (Dumay *et al.*, 2002), and of vowel harmony in Finnish (Suomi *et al.*, 1997), it has shown effects of native-language sequence constraints on segmentation in a second language (Weber and Cutler, 2006), it has confirmed that word onsets contribute relatively more in spoken-word recognition than offsets (McQueen, 1998), and it has revealed listeners' sensitivity to the likelihood of a word boundary in a given string of phonemes (McQueen, 1998; van der Lugt, 2001; Warner *et al.*, 2005).

It is very difficult to spot a word if accepting it leaves a vowelless residue of the input. Thus *egg* in *fegg* or *sugar* in *sugarth* are detected less easily than *egg* in *maffegg* or *sugar* in *sugarthig* (Norris *et al.*, 1997). The residues *maff* and *thig* are syllables, so although they are, in fact, not words, they might have been; *f* and *th*, however, as single consonants, are impossible word candidates in English. This finding appears in English (Norris *et al.*, 1997, 2001), in Dutch (McQueen and Cutler, 1998), in Japanese (McQueen *et al.*, 2001), in Sesotho (Cutler *et al.*, 2002a), in French (Spinelli *et al.*, 2003), and in Cantonese (Yip, 2004). These languages vary widely in the surface constraints on what may be a syllable of the language and what may be a stand-alone word, but the PWC difference between possible and impossible residues remains effectively constant. The PWC has thus been held (Norris *et al.*, 2001; Cutler *et al.*, 2002a) to express a universal constraint on syllabic viability: Across languages, vowels alone can be syllables, but consonants cannot. Syllables can be words; thus, because consonants alone cannot be syllables, they also cannot be words, though vowels can be (*eye*, *awe*). Applying the PWC radically reduces the effects of embedding in speech (Cutler *et al.*, 2002b), which makes it potentially very useful in everyday listening.

The Japanese experiment of McQueen *et al.* (2001) allowed a comparison between possible residues (with a vowel) and impossible residues (without a vowel) with the length in number of segments controlled; residues were a single vowel versus a single consonant. This was possible because Japanese allows sequences of vowels. Thus spotting of *uni* "sea urchin" was compared in *puni* versus *iuni*, and of *hiru* "noon" in *hiruk* versus *hirua*. Spotting the word was always hardest when the single-consonant residue remained, just as in the other languages. This shows that the findings from other languages could not have been due to length differences between impossible versus possible residues, and that Japanese listeners, like listeners with other native languages, are sensitive to the different viabilities of vowels versus consonants as residues in speech segmentation. Although vowels can be inserted into non-Japanese input (loanwords or nonwords such as those presented by Dupoux *et al.*, 1999), detection of real Japanese words is seriously hindered when the adjacent context contains no vowel.

The sequences created by the process of devoicing in Japanese thus have the potential to make speech segmentation and word recognition harder. Recall, however, that devoicing generally occurs only for /i/ and /u/ and only in voiceless contexts. It is possible that a single voiceless consonant that is heard often in devoicing environments will not hinder segmentation (and disrupt embedded word detection in consequence), because it is interpreted not as a vowelless consonant but as equivalent to a (devoiced) syllable. The test of the PWC by McQueen *et al.* (2001) in Japanese deliberately excluded potential devoicing contexts. In the present study, we focus specifically on these cases.

It is important to test both preceding contexts (which disrupt word onsets) and following contexts (which disrupt word offsets). In the PWC studies (Norris *et al.*, 1997; McQueen *et al.*, 2001), as in other studies (e.g., McQueen, 1998), preceding contexts exercised the strongest effects. Thus we here test both context positions. We begin with following contexts: in Experiment 1 we compare impossible single-consonant codas, which in the study of McQueen *et al.* (2001) made word-spotting difficult, with single-vowel contexts, which made word-spotting easy, and both of these with single-consonant devoicing environments. For instance, we compare how difficult it is to spot *asa* "morning" in *asap*, *asau*, or *asaf*; the stop /p/ could not occur at the end of a Japanese word, but /f/ is a voiceless fricative and so could occur followed by an underlying devoiced vowel. This comparison allows us to determine whether the devoiced case will pattern more like an impossible consonant or more like a possible vowel context.

Note that in the impossible coda case, what is being presented to listeners is an illegal sequence of the language. Phonotactic legality may affect both how sequences are perceived by listeners and how they are uttered by speakers. It was important, therefore, to rule out such speaker-related effects in our listening experiment. We did this by conducting two versions of the experiment, in one of which (Experiment 1B) the final consonants were produced as (illegal) codas, while in the other (Experiment 1C) they were produced as (legal) onsets of CV syllables, from which the final vowel was then digitally removed prior to the experiment. If speaker-related legality affects listener responses to the final consonants, Experiments 1B and 1C will return different response patterns. Since this comparison meant that the tokens of the embedded target words differed across contexts, a control lexical decision experiment first tested whether the versions differed in how recognizable they were (Experiment 1A).

TABLE I. Materials for Experiments 1 and 2.

		Experiment 1			Experiment 2		
Target noun	English gloss	V context	Possible devoicing C context	Impossible devoicing C context	CV context	Possible devoicing CCV context	Impossible devoicing CCV context
ase	sweat	asea	ases(u)	asep(u)	aseka	aseska	asepge
ani	brother	aniu	anits(u)	anit(u)	anigu	anitsse	anitmo
uso	lie	usoa	usoch(i)	usop(u)	usota	usochsi	usopzu
chizu	map	chizua	chizuf(u)	chizut(u)	chizuya	chizufte	chizutba
haru	spring	harua	haruts(u)	harup(u)	harupa	harutspu	harupbu
yuzu	citron	yuzua	yuzuch(i)	yuzut(u)	yuzupa	yuzushpi	yuzutge
saru	monkey	sarua	saruts(u)	sarup(u)	saruza	sarutsse	sarupze
motsu	giblets	motsua	motsuch(i)	motsug(u)	motsugu	motsuchka	motsugda
nasu	eggplant	nasua	nasuch(i)	nasub(u)	nasuza	nasushte	nasubda
hiru	noon	hirua	hiruch(i)	hiruk(u)	hiruha	hiruchhe	hirukbe
aki	autumn	akia	akich(i)	akib(u)	akiha	akichka	akibzo
ibo	wart	iboi	iboch(i)	ibop(u)	ibogi	ibochta	ibopzo
uzu	whirlpool	uzua	uzus(u)	uzut(u)	uzupa	uzuspe	uzutme
kinu	silk	kinua	kinuf(u)	kinup(u)	kinuza	kinufko	kinupgo
kuzu	trash	kuzua	kuzus(u)	kuzut(u)	kuzuga	kuzuska	kuzutge
naya	shed	nayau	nayaf(u)	nayat(u)	nayapu	nayafpe	nayatma
mitsu	honey	mitsua	mitsuch(i)	mitsud(u)	mitsupa	mitsufpo	mitsudba
kazu	number	kazua	kazus(u)	kazup(u)	kazuha	kazusko	kazupgo
rusu	absence	rusua	rusus(u)	rusug(u)	rusuha	rususke	rusugde
yoru	night	yorua	yoruts(u)	yorup(u)	yoruza	yorucha	yorupba
asa	morning	asau	asaf(u)	asap(u)	asazu	asafte	asapdo
usu	mortar	usua	usuf(u)	usud(u)	usuza	usufsha	usudba
uni	sea urchin	unia	units(u)	unip(u)	unika	unitska	unipge
eki	station	ekia	ekich(i)	ekid(u)	ekipa	ekichpi	ekidbi
matsu	pine tree	matsua	matsush(i)	matsud(u)	matsuha	matsuchta	matsudbi
moya	mist	moyau	moyach(i)	moyap(u)	moyazu	moyachse	moyapzu
gasu	gas	gasua	gasuch(i)	gasub(u)	gasupa	gasushpu	gasubgo
mesu	surgical knife	mesua	mesuch(i)	mesud(u)	mesuza	mesutsso	mesudzo
risu	squirrel	risua	risuf(u)	risud(u)	risupa	risufte	risudge
tsuru	crane	tsurua	tsuruch(i)	tsurut(u)	tsuruna	tsurushta	tsurutbe

II. EXPERIMENT 1

A. Method

1. Materials and design

Thirty vowel-initial two-mora target words (e.g., *asa*) were selected. Most had been used in the study of McQueen *et al.* (2001), where they had been chosen to contain as few embedded words as possible. These words were placed in five following contexts: (i) a single vowel (e.g., *asau*), (ii) a possible-devoicing consonant (e.g., *asaf*), (iii) the same consonant followed by a high vowel ([u] or [i], e.g., *asafu*), (iv) an impossible-devoicing consonant (e.g., *asap*), and (v) the same consonant followed by [u] (e.g., *asapu*). The devoicing contexts were fricatives or affricates; the impossible contexts were voiced stops, /t/ or /p/ (in native Japanese words, /t/ cannot precede /i/ or /u/ and a single /p/ cannot occur intervocalically). Appendix A lists the phonemes of Japanese, and Table I lists all target-bearing items. We further constructed 84 fillers. Four were words with following CV contexts which should be easy to spot (e.g., *biruta* with *biru* "building"). Eighty were not words and contained no embedded bimoraic words; 30 of these were matched to target-bearing items, with 10 each preceding a vowel (e.g., *dozao*), a possible-devoicing consonant (e.g., *zanuf*) or an impossible consonant (e.g., *bugep*). The remaining 50 fillers were all bimoraic (C)VCV nonwords. There were also 12 practice items (four with targets), modeled on the experimental items.

All materials were recorded by a phonetically trained native speaker of Tokyo dialect (the second author) in a sound-damped booth to digital audio tape, sampling at 48 kHz. Stops produced in final position were released and the targets' default accent patterns were preserved within the recorded items. The materials were transferred to computer (down-sampled to 16 kHz, 16 bits), examined, and labeled and the duration of each target word was measured using the XWAVES speech editor. Target-bearing items in contexts (i), (ii), and (iv) were used in Experiment 1B. The final vowels from contexts (iii) and (v) were digitally removed, cutting at a zero-crossing at the point at which no auditory trace of the final vowel remained. The resulting consonant-final items, plus those with vowel contexts (i), were used in Experiment 1C. A timing pulse was aligned with the onset of each target-bearing item. For the control Experiment 1A, the entire context was removed from each version of each target word, and from each of the 30 fillers matched to target-bearing items, to give nonwords such as *doza*, *zanu*, and *buge*.

For Experiment 1B three counter-balanced lists were made, with ten targets in each of the three context conditions per list (vowel, possible-devoicing consonant, impossible consonant); all targets appeared once on each list. These target-bearing items were mixed in pseudo-random order with all 84 fillers; there was always at least one filler between any pair of target-bearing items. The same lists were used in Experiment 1C (but with the different versions of the consonant-final target-bearing items). The design of Experiment 1A was similar, except that the 30 target words were rotated over five lists, with six words from each of the original five recorded contexts per list.

2. Participants

One hundred and thirty undergraduate members of Dokkyo University received course credits for participating in the study: 40 in Experiment 1A and 45 each in Experiments 1B and 1C. All were native Japanese speakers from the Tokyo area.

3. Procedure

In Experiment 1A, listeners were told that they would hear a list of words and fillers, and were asked to respond by pressing a button as rapidly as possible whenever they heard a real word, and then to say in a low voice what that word was into a microphone. In Experiments 1B and 1C, listeners were told that they would hear a list of nonsense words, some of which would contain real words embedded at their onset. Examples of bimoraic words in the three context conditions were provided. Listeners were asked to press a button as fast as possible if they spotted any real word and then to say what that word was. No prior information was provided as to the identity of the target words. In all experiments participants were tested in separate sound-attenuating carrels in a quiet room, either individually or in pairs. Prior testing ensured that listeners tested in pairs could not hear each other's spoken responses. The spoken responses were recorded. Participants were asked to press the button with their preferred hand. Each listener heard a practice list, and then one of the experimental lists (eight participants per list in Experiment 1A; 15 per list in Experiments 1B and 1C).

The experiments were run from a Sony TCD D10 DAT player and a computer running NESU experiment control software. The computer clock was started by each timing pulse and stopped by each button-press; responses were logged on the computer.

B. Results

In all experiments analyses of variance (ANOVAs) with participants ($F1$) and items ($F2$) as repeated measure were conducted on both reaction time (RT) and error data. Target durations were subtracted from the raw RTs prior to analysis, to obtain RTs from target offset.

Two control participants (Experiment 1A) who detected no words in one condition of the experiment were excluded from the analyses. No word-spotting participants (Experiments 1B and 1C) had to be excluded for this reason. Listeners' spoken responses were analyzed first. On a few trials (2.4% in Experiment 1A, 0.4% in Experiment 1B, and 0.2% in Experiment 1C), listeners misidentified target words. The button-press responses on these trials were treated as errors. Seven items (*haru*, *motsu*, *hiru*, *ibo*, *kuzu*, *naya*, *mesu*) were excluded from the final analyses because those items were missed by all participants in at least one condition in at least one of the three subexperiments. Mean RTs and error rates for each condition in Experiments 1B and 1C are shown in Tables II and III, respectively, along with the relevant control data from Experiment 1A (*N.B.* the control results for the vowel-context condition are therefore the same across tables).

TABLE II. Experiments 1A (control lexical decision) and 1B (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning; items in the consonant-context conditions were recorded with no following context).

	Context		
	Vowel	Possible devoicing consonant	Impossible devoicing consonant
	Lexical decision (Experiment 1A)		
Mean RT	565	677	655
Mean error	22%	25%	26%
Example	asa[u]	asa[f]	asa[p]
	Word-spotting (Experiment 1B)		
Mean RT	734	800	756
Mean error	21%	26%	36%
Example	asau	asaf	asap

1. Lexical decision control (Experiment 1A)

The RT ANOVA revealed a main effect of context (i.e., the five different contexts from which the target words had been excised): $F1(4,132)=6.41$, $p<0.005$; $F2(4,88)=3.42$, $p<0.05$. The sets of three conditions corresponding to Ex-

TABLE III. Experiments 1A (control lexical decision) and 1C (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning; items in the consonant-context conditions were recorded with following vowels, but those vowels were removed prior to the experiment).

	Context		
	Vowel	Possible devoicing consonant	Impossible devoicing consonant
	Lexical decision (Experiment 1A)		
Mean RT	565	618	606
Mean error	22%	28%	27%
Example	asa[u]	asa[f(u)]	asa[p(u)]
	Word-spotting (Experiment 1C)		
Mean RT	712	825	762
Mean error	15%	23%	31%
Example	asau	asaf(u)	asap(u)

periments 1B and 1C, respectively, were compared in pairwise t-tests. These showed the main effect of context to be due to responses in the vowel-context condition (e.g., to *asa* excised from *asau*) being faster than in the other four conditions. For the contexts tested in Experiment 1B and 1C, RTs were shorter to words from vowel contexts than to words from both possible-devoicing contexts [1B: $t1(37)=3.49$, $p<0.005$, $t2(22)=3.37$, $p<0.005$; 1C: $t1(37)=2.21$, $p<0.05$, $t2(22)=1.31$, $p>0.2$] and impossible contexts [1B: $t1(37)=4.24$, $p<0.001$, $t2(22)=2.22$, $p<0.05$; 1C: $t1(37)=1.65$, $p=0.11$, $t2(22)=2.09$, $p<0.05$], though note that the Experiment 1C results were not statistically significant across both participants and items. There were no significant differences between any consonant-context conditions. The overall mean error rate (25%) was relatively low given that these items had all been excised from context; the error ANOVAs showed no effects of context ($F1$ and $F2<1$). The RT effects, however, indicate that the target words for the word-spotting experiments were, irrespective of context, not equally easy to recognize. In all by-item ($F2$) analyses of the word-spotting data, therefore, the lexical decision data were entered in analyses of covariance (ANCOVAs) as covariates (control RTs in the RT analyses and control error rates in the error analyses). The by-participant analyses were standard ANOVAs.

2. Word spotting: Consonants recorded without following context (Experiment 1B)

In the overall analyses, there were effects of context in RTs [$F1(2,84)=3.46$, $p<0.05$; $F2(2,43)=2.18$, $p=0.125$] and, more strongly, in errors [$F1(2,84)=15.53$, $p<0.001$; $F2(2,43)=3.93$, $p<0.05$]. Pairwise comparisons of the context conditions showed that participants spotted words more rapidly in vowel contexts (e.g., *asa* in *asau*) than in possible-devoicing contexts (*asa* in *asaf*): $F1(1,44)=6.33$, $p<0.05$; $F2(1,21)=6.58$, $p<0.05$, and spotted words more accurately in vowel contexts than in impossible-devoicing consonantal contexts (*asa* in *asap*): $F1(1,44)=21.22$, $p<0.001$; $F2(1,21)=7.70$, $p<0.05$. No other pairwise comparisons of either RTs or errors were significant by both $F1$ and $F2$. There were thus no statistically reliable differences between the two-consonant-context conditions, and word-spotting in both of these contexts was more difficult than in vowel contexts (observed for the possible-devoicing contexts primarily in RTs, for the impossible-devoicing contexts primarily in errors).

3. Word spotting: Consonants recorded with following context (Experiment 1C)

The overall by-participant ANOVAs and by-item ANCOVAs revealed significant effects of context in RTs [$F1(2,84)=10.30$, $p<0.001$; $F2(2,43)=3.97$, $p<0.05$] and errors [$F1(2,84)=21.30$, $p<0.001$; $F2(2,43)=4.66$, $p<0.05$]. Word-spotting performance in the vowel contexts was faster [$F1(1,44)=20.66$, $p<0.001$; $F2(1,21)=7.59$, $p<0.05$] and more accurate [$F1(1,44)=9.40$, $p<0.005$; $F2(1,21)=5.82$, $p<0.05$] than in the possible-devoicing consonantal contexts. Word-spotting performance in the vowel contexts was also more accurate [$F1(1,44)=27.15$, $p<0.001$; $F2(1,21)=11.19$, $p<0.005$] than in the impossible-devoicing consonantal contexts. No other pairwise comparison was significant by both participants and items. Although there were thus small differences across Experiments 1B and 1C, the major pattern in the data was the same in both.[1] Finally, 2 × 2 ANOVAs combining the consonant-context data of Experiments 1B and 1C tested this directly; for both RTs and errors, the experiment by context interaction was insignificant ($F1$ and $F2<1$).

C. Discussion

The answer to the question posed in Sec. I is thus very clear: the results for the devoicing case are like those for single-consonant contexts, and not like those for vowel contexts. As in McQueen *et al.* (2001), vowel contexts made detection of embedded words easy, and all single-consonant contexts made detection hard. Differences of acoustic goodness could not underlie the results, since such differences were factored out by covarying the control lexical decision data. The difficulty of single-consonant contexts was constant even though some of the consonants, namely, those in the possible-devoicing condition, can effectively occur in final position when the vowel they precede is devoiced; word-spotting in this condition was as hard as in the condition where the consonants were impossible codas. The way in which the consonant contexts were produced (intentionally, or by excision of a vowel) did not influence performance.

Nonetheless, since Japanese allows no obstruent codas, all the consonant-final items in Experiment 1 were, as we have noted, illegal. In Experiment 2, we provided the devoicing cases with a potentially legal environment, thus providing a clearer and much more ecologically valid test of our hypothesis. A consonant sequence consisting of voiceless consonants (such as [ʃk] in a natural utterance of *sashiki*) is the canonical environment for devoicing in natural speech. We provided such environments by extending the contexts appended to the target words: each vowel context became CV, and each consonant context became CCV. For example, *asa* now occurred in *asazu* (which we predict to be very easy), *asapdo* (which we predict to be very hard), and *asafte*. The voiceless [f] and [t] in sequence could arise from devoicing in natural speech; in Experiment 2 we can assess whether this two-consonant sequence makes word-spotting hard too, or whether the potential for devoicing renders it easy.

III. EXPERIMENT 2

A. Method

1. Materials and procedure

For the 30 two-mora target words of Experiment 1, three new following contexts were constructed: a CV mora (e.g., *asazu*), a CCV in which devoicing was possible (fricative or affricate plus voiceless C plus V, e.g., *asafte*), and a CCV in which devoicing was impossible (the second C was voiced, e.g., *asapdo*). Fillers were also constructed as for Experiment 1, with again 30 matched to target-bearing items (e.g., *dozago*, *zanufte*, *bugepga*); 50 fillers were trimoraic

TABLE IV. Experiments 2A (control lexical decision) and 2B (word-spotting): Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*asa*=morning).

	Context		
	Consonant+vowel	Possible devoicing consonant cluster+vowel	Impossible devoicing consonant cluster+vowel
	Lexical decision (Experiment 2A)		
Mean RT	606	653	653
Mean error	18%	36%	29%
Example	asa[zu]	asa[fte]	asa[pdo]
	Word-spotting (Experiment 2B)		
Mean RT	715	841	846
Mean error	20%	38%	29%
Example	asazu	asafte	asapdo

(C)VCVCV non-words. The materials were recorded and measured as in Experiment 1. Experiment 2A was a lexical decision control, with truncation applied as in Experiment 1A. In Experiment 2A, the procedure was as for Experiment 1A, and in Experiment 2B as for Experiments 1B/C.

2. Participants

Sixty-nine Dokkyo University undergraduates participated, for course credit: 24 in Experiment 2A, and 45 in Experiment 2B. None had taken part in Experiment 1.

B. Results

All participants (eight per list) in the lexical decision control experiment (Experiment 2A) were included in the analyses. Three word-spotting participants (Experiment 2B) were excluded (one listener missed all targets in one condition; the other two, excluded to balance the sets, were the most erroneous on each of the other two lists). This left 14 participants per list. The button-press responses on 17 trials (2.1%) in Experiment 2A and one trial (0.1%) in Experiment 2B were accompanied by incorrect spoken responses and thus treated as errors. The data from six items were excluded from the final analysis (*haru*, *motsu*, *hiru*, *mitsu*, *matsu*, *moya*): Five words because all participants missed them in at least one condition in one of the subexperiments, the sixth because it had been recorded incorrectly in one condition. Table IV shows mean RTs and error rates.

1. Lexical decision control (Experiment 2A)

The effect of the context from which words had been excised was significant by participants [$F1(2,42)=4.73$, $p<0.05$] but not by items [$F2(2,46)=1.34$, $p>0.2$] in the RT analysis. Errors (overall mean 28%) were again reasonable for excised words; the context effect in errors was significant in both analyses [$F1(2,42)=16.75$, $p<0.001$; $F2(2,46)=4.20$, $p<0.05$]. No pairwise comparison was significant by both participants and items for RTs. Error rates, however, were lower for words taken from CV contexts (e.g., *asa* from *asazu*) than for words from CCV contexts where devoicing was possible [*asa* from *asafte*; $t1(23)=3.91$, $p<0.005$; $t2(23)=2.94$, $p<0.01$], and, less robustly, for words from CCV contexts where devoicing was impossible [*asa* from *asapdo*; $t1(23)=2.16$, $p<0.05$; $t2(23)=1.76$, $p<0.1$]. The difference between the two CCV contexts was not significant. As in Experiment 1, these results suggest that the target words differed across contexts in how easy they were to recognize. These control data were therefore again used as covariates in by-item ANCOVAs of the word-spotting data. The by-participant analyses of the word-spotting data were again ANOVAs.

2. Word spotting (Experiment 2B)

There was a main effect of context in RTs [$F1(2,78)=17.96$, $p<0.001$; $F2(2,45)=11.44$, $p<0.001$] but not in errors [$F1(2,78)=24.76$, $p<0.001$; $F2(2,45)=1.95$, $p>0.15$]. Pairwise comparisons on the RT data showed that participants spotted words faster in CV contexts than in either type of CCV context (possible-devoicing environment: $F1(1,41)=25.14$, $p<0.001$; $F2(1,22)=18.98$, $p<0.001$; impossible environment: $F1(1,41)=28.05$, $p<0.001$; $F2(1,22)=14.81$, $p<0.005$). There was no latency difference between the two CCV context conditions ($F1$ and $F2<1$). No pairwise comparisons on the error data were significant by both participants and items. Word spotting was thus easier in CV contexts than in CCV contexts, with no effect of whether the CCV sequence was or was not a possible-devoicing environment.

C. Discussion

Again, the contexts which were potential devoicing environments made it very hard for listeners to spot the words embedded in the nonwords they heard. The results were very similar to those of Experiment 1; providing a potentially legal and hence more natural environment for devoicing to occur did not increase the acceptability of consonant sequences in Japanese speech segmentation. Both *asapdo* and *asafte* consist of three full morae (*a*, *sa*, and *do* or *te*) plus a single consonant ([p], [f]); it makes no difference that [f] before *te* could possibly have arisen from *fu*, whereas [p] before *do* cannot have arisen from *pu*. Vowelless sequences, it seems, are not treated as if they might be hiding a vowel; just as in other languages, such sequences are impossible word candidates and hence they resist being segmented from the adjacent speech stream.

Experiments 1 and 2 allow us to dispense with the possibility that Japanese listeners always treat sequences of voiceless consonants as if they contained a vowel; clearly, they do not. But results with following contexts, which potentially combine with word offsets, do not force us to conclude that devoicing will always disrupt the segmentation of normal Japanese speech. It is still necessary to assess devoicing in preceding contexts, which potentially combine with word onsets and thus exercise a more powerful effect in word segmentation. Preceding contexts are processed before targets, thus affecting target recognition differently than following contexts; their effect can be so strong that a target word is not even recognized at all (as occurred in the study of McQueen *et al.*, 2001, for example). In our third experiment, we therefore examined devoicing sequences attached as preceding context to an embedded word.

Again we compared the devoicing sequences to other sequences which the previous findings had indicated as hard or as easy. We used two types of item: VCV words such as *asa*, preceded by clearly easy CCVCV versus hard CCVC contexts (see McQueen *et al.*, 2001), and CVCV words such as *sake* "salmon," preceded by clearly easy CCVCV contexts and potential devoicing CCVC contexts. Thus detection of *asa* was compared in *myojiasa* (easy, because the context consists of two full morae: *myo*, *ji*) versus *myochasa* (hard, because the context is a mora *myo* plus a vowelless affricate). The prohibition of devoicing before vowels (or any voiced segment) means that these vowel-initial words can never be preceded by devoicing; they are therefore the baseline against which we can compare the voiceless-initial words like *sake*. Detection of *sake* was compared in *nyagusake* (easy: *nya*, *gu*) versus *nyaksake* (hard: *nya*, [k], but potentially a rendition of *nyakusake* with devoicing). The results of McQueen *et al.* (2001) lead us to expect a large difference between easy and hard contexts for the VCV words like *asa*; the crucial question is whether there is also a large difference between easy and hard contexts for CVCV words like *sake*.

If there is an equivalently large difference in the *sake* case, we will have to conclude that devoicing indeed makes Japanese speech segmentation more difficult. The vowelless affricate in *myochasa* can never be licensed as a possible word, because devoicing cannot occur before vowels. Thus if the context effect for CVCV words is as large as for VCV words, we would have to conclude that the vowelless stop in *nyaksake* likewise cannot be a possible word.

On the other hand, if the context effect for CVCV words is significantly less than for VCV words (i.e., if *sake* is as easy or almost as easy to spot in *nyaksake* as in *nyagusake*), we may conclude that Japanese listeners interpret /ks/ (and like sequences) as containing an underlying vowel, and hence licensed as a possible word. Sequences such as /ks/ will often have been heard; the Japanese lexicon contains many words in which the first mora allows vowel devoicing, including words beginning *kusa*- (e.g., *kusabi* "wedge" and *kusari* "chain"), which will be pronounced with a devoiced first vowel, so effectively with an initial /ks/. If such experience licenses the devoiced sequence /ks/ in *nyaksake* as a possible word, it should not interfere with segmentation.

IV. EXPERIMENT 3

A. Method

1. Materials and procedure

Fifty-two high-frequency two-mora target words were selected: 26 VCV words (e.g., *asa*) and 26 CVCV words beginning with voiceless consonants (e.g., *sake*). CCVC (e.g., *bya*+C) and CCVCV (e.g., *bya*+CV) nonsense sequences were constructed as preceding contexts for both types of target. For the CVCV targets, the CCVC contexts created possible-devoicing environments between the context and target word onset (e.g., *nyaksake* is a potentially devoiced rendition of *nyakusake*). This was not the case for the CCVC contexts followed by the vowel-initial VCV targets, because devoicing cannot occur before vowels.

We further constructed 106 nonsense fillers with no embedded bimoraic words in offset position. Of these, 52 were matched to target-bearing items: 26 ending with CVCV after either a CCVC or CCVCV sequence like those in the target-bearing items (e.g., *chaksomi* and *kyagukeni*), and 26 ending with VCV sequence after a CCVC or CCVCV (e.g., *gyopagi* and *myoguige*). The remaining 54 fillers were all trimoraic CVCVCV nonwords. There were 24 practice items modeled on the experimental materials, including eight with embedded target words (four CVCV, four VCV).

The materials were recorded as in Experiments 1 and 2, by the same speaker. The accent pattern of the targets was preserved in the way the items were recorded. The items with CVCV targets in CCVC contexts were recorded with the potential underlying vowel fully devoiced (e.g., the [u] in *nyak(u)sake* was not realized). The target-bearing stimuli were then digitally cross-spliced, cutting at zero-crossings and using auditory criteria to determine the excision points. Tokens of the CVCV targets recorded in the CCVCV context (e.g., *sake* from *nyagusake)* were spliced onto a CCVC context sequence (e.g., *nyak* from *nyaksake*) and onto a CCVCV sequence from a second recording of a CCVCV item (e.g., *nyagu* from a different token of *nyagusake* than that used for the target token). A similar cross-splicing procedure was used for the VCV targets (e.g., the *asa* in the final experimental stimuli came from a recording of *myojiasa*). All target-bearing stimuli were thus cross-spliced from two recordings, and the target word realization in all contexts was constant. Six items with VCV targets could not be cross-spliced without audible discontinuities and were therefore excluded (along with six matched fillers) from the experiment. There was no detectable trace of splicing in the remaining items (all listed in Tables V and VI).

Two counter-balanced lists were constructed, with 13 CVCV and 10 VCV targets in each context condition per list, and with all 46 targets appearing once per list. As in the earlier experiments, target-bearing and filler items were presented in pseudo-random order. The procedure was as in Experiments 1B/C and 2B.

2. Participants

Thirty-two undergraduates (16 per list) from the same population took part in return for course credits. None had participated in Experiments 1 or 2.

B. Results and discussion

No participants were excluded from the analyses, but one item was. This word (*soro*, "solo," from the CVCV set) was missed by all participants in the devoiced vowel context and by all but one participant in the surface vowel context. No incorrect spoken responses were recorded. Mean word-spotting RTs and error rates are shown in Table VII. Note that since the materials in this experiment were controlled by the cross-splicing, control lexical decision data, as required for Experiments 1 and 2, are here unnecessary.

TABLE V. Materials for Experiment 3: CVCV targets.

Target	English gloss	Voiceless consonant+devoiced vowel	Voiced consonant+vowel
		Many word candidates	
kachi	value	gyats(u)kachi	gyazukachi
kado	corner	kyah(u)kado	kyabukado
kako	past	nyach(i)kako	nyajikako
kashu	singer	byas(u)kashu	byazukashu
kata	shoulder	shof(u)kata	shobukata
kare	he	byos(u)kare	byozukare
saru	monkey	gyuk(u)saru	gyugusaru
kumo	cloud	nyos(u)kumo	nyozukumo
tsuru	crane	nyosh(i)tsuru	myojitsuru
kazu	number	shas(u)kazu	shazukazu
sake	salmon	nyak(u)sake	nyagusake
kaba	hippopotamus	nyach(i)kaba	nyajikaba
kage	shadow	shos(u)kage	shozukage
kechi	stinginess	byaf(u)kechi	byabukechi
		Few word candidates	
fugu	blowfish	byas(u)fugu	byazufugu
hamu	ham	nyak(u)hamu	nyaguhamu
haru	spring	gyash(u)haru	gyajuharu
hage	baldness	chosh(u)hage	chojuhage
hada	skin	nyots(u)hada	nyozuhada
kamo	duck	myosh(u)kamo	myojukamo
sora	sky	kyof(u)sora	kyobusora
shuwa	sign language	nyok(i)shuwa	nyogishuwa
shugo	subject	pyach(i)shugo	pyajishugo
chibi	kid	ryosh(u)chibi	ryojuchibi
sobo	grandmother	myak(u)sobo	myagusobo
soro	solo	myok(u)soro	myogusoro

TABLE VI. Materials for Experiment 3: VCV targets.

Target	English gloss	Voiceless consonant+vowel	Voiced consonant+vowel
ase	sweat	gyasase	gyazuase
aka	red	myachaka	myajuaka
ani	brother	gyachani	gyajuani
eki	station	nyaheki	nyabueki
aki	autumn	shochaki	shojuaki
ama	nun	ryochama	ryojiama
aji	horse mackerel	chochaji	chojuaji
ibo	wart	nyahibo	nyabuibo
umi	sea	shochumi	shojiumi
aku	badness	nyoshaku	nyojuaku
asa	morning	myochasa	myojiasa
ato	mark	kyachato	kyajuato
ane	sister	shasane	shazuane
uni	sea urchin	byachuni	byajiuni
ego	ego	gyosego	gyozuego
oku	inside	myahoku	myabuoku
ine	rice plant	chuchine	chujuine
imi	meaning	kyochimi	kyojuimi
obi	a sash	ryuhobi	ryubuobi
ima	now	gyuchima	gyujuima

The most striking result is in the error rates. As in McQueen *et al.* (2001), Japanese listeners found it almost impossible to spot VCV target words in a CCVC context where the initial vowel of the target was not aligned with a mora boundary (e.g., *asa* in *myochasa*). Error ANOVAs revealed main effects of target type (VCV words were harder to spot than CVCV words: $F1(1,30)=107.84$, $p<0.001$; $F2(1,43)=19.00$, $p<0.001$) and context type [words were harder to spot in CCVC contexts than in CCVCV contexts: $F1(1,30)=302.30$, $p<0.001$; $F2(1,43)=134.72$, $p<0.001$], and these two factors interacted [$F1(1,30)=90.97$, $p<0.001$; $F2(1,43)=54.76$, $p<0.001$]. Pairwise comparisons showed that the context effect was significant for both types of target: CVCV words (*sake*) were harder to spot in CCVC contexts (possible devoicing environments, e.g., *nyaksake*) than in CCVCV contexts (e.g., *nyagusake*), $F1(1,30)=15.13$, $p<0.001$, $F2(1,24)=9.90$, $p<0.005$; as already noted, VCV targets (*asa*) were much harder to spot in CCVC contexts (e.g., *myochasa*) than in CCVCV contexts (e.g., *myojiasa*): $F1(1,30)=368.79$, $p<0.001$, $F2(1,19)=163.75$, $p<0.001$.

Given the high error rates for the VCV control words, the context effect in RTs was analyzed only for the CVCV words. CVCV targets (e.g., *sake*) were detected equally rapidly in the two contexts ($F1$ and $F2<1$); there was in this case no increased difficulty for devoicing (*nyak*-) over vowel (*nyagu*-) contexts. Within the CCVCV contexts, the VCV words were spotted as quickly as the CVCV words ($F1$ and $F2<1$).

The results for the VCV baseline condition show that the listeners here were behaving exactly as the listeners in the study of McQueen *et al.* (2001). Preceding context without a vowel (e.g., a single affricate) makes segmentation and hence word-spotting hard. The crucial results are for the CVCV words. Here the devoicing contexts were clearly less problematic for listeners than they had proven to be in Experiments 1 and 2. Although the error rate was raised by a vowelless context, it was not raised to the heights observed for the VCV words (or by McQueen *et al.*, 2001). And in the RTs, no delay of word-spotting as a function of context could be observed at all. This striking finding suggests that Japanese listeners are indeed sensitive to the potential presence of a devoiced vowel in voiceless obstruent sequences such as /ks/.

TABLE VII. Experiment 3: Mean correct RTs (in milliseconds, from target word offset) and mean error rates (in percent) by context condition (*sake* =salmon; *asa*=morning; no reliable estimate could be computed of the mean RT for VCV targets in CCVC contexts).

		Context	
	Target type	CCVC	CCVCV
	CVCV		
Mean RT		739	755
Mean error		43%	29%
Example		nyaksake	nyagusake
	VCV		
Mean RT		⋯	750
Mean error		93%	29%
Example		myochasa	myojiasa

We conducted two further analyses of the data, neither of which explained away this finding. We first examined the phonetic structure of the devoicing contexts we had tested. Maekawa and Kikuchi's (2005) analyses of the corpus of spontaneous Japanese showed devoicing to vary as a function of the surrounding consonants. The manner of articulation of the following consonant has the strongest effect: the likelihood of the vowel being devoiced is far greater before a stop or affricate than before a fricative, both for /i/ and for /u/ and across all types of preceding consonants. According to Kondo (2005), devoicing is virtually obligatory in the most favored environments, and is only inhibited for potential sequences of devoiced syllables. If listeners are sensitive to these probabilities, they may find the most likely cases least difficult. Accordingly, we divided the CVCV items into two sets, varying in frequency of devoicing occurrence. One set contained 11 words beginning with fricatives (in Maekawa and Kikuchi's (2005) data, devoiced in about 61% of cases), while the other set had 14 targets beginning with stops and affricates (about 96% devoiced). An analysis of the error data including this factor revealed a significant context effect [$F1(1,30)=17.92$, $p<0.001$; $F2(1,23)=9.73$, $p<0.005$]: CCVC contexts made word-spotting harder than CCVCV contexts. But there was no effect of devoicing probability ($F1$ and $F2<1$) and no interaction of this factor with the context effect ($F1$ and $F2<1$). The context effect was present both where devoicing is very likely [$F1(1,30)=6.60$, $p<0.05$; $F2(1,13)=6.74$, $p=0.05$], and where it is less likely [$F1(1,30)=13.48$, $p<0.001$; $F2(1,10)=3.74$, $p<0.1$]. Thus this factor appeared not to have affected our results.

The second analysis examined the potential lexical support for devoiced vowels in our sequences. From Sugito (1995) we computed the number of words consistent with the bimoraic sequence linking CCVC context and each CVCV target word (e.g., *kusa* given the target *sake* in *nyaksake*, if the devoiced [u] were realized). The materials formed two sets (see Table V), one of 14 items where relatively many words (≥ 10) either matched the sequence or had onsets the same as the sequence (e.g., *kusa* in *nyaksake*) and one of 11 items (minus *soro*) with few words (≤ 5) matching the sequence (e.g., *shuha* given the target *haru* "spring" in the CCVC+target sequence *gyashharu*). The mean number of competitors for the former set was 13.6, for the latter 2.0. ANOVAs on the error data for the CVCV words split in this way revealed a main effect of word set size by participants but not by items [$F1(1,30)=6.52$, $p<0.05$; $F2(1,23)=1.10$, $p>0.3$]: Targets in the many-words set were spotted more accurately, across context conditions, than those in the few-words set. There was also still a main effect of context [$F1(1,30)=22.06$, $p<0.001$; $F2(1,23)=10.16$, $p<0.005$]. In the by-participant analysis only, the context effect varied with word set size: Words were harder to spot in CCVC than in CCVCV contexts, but more so when fewer lexically consistent words were available [$F1(1,30)=6.97$, $p<0.05$; $F2<1$]. As pairwise comparisons confirmed, the context effect was less robust in the many-words set [$F1(1,30)=2.50$, $p>0.1$; $F2(1,13)=4.52$, $p=0.05$] than in the few-words set [$F1(1,30)=32.91$, $p<0.001$; $F2(1,10)=5.20$, $p<0.05$]. This suggests that greater lexical support may strengthen licensing of devoicing environments, but is not the sole rationale for it.

V. GENERAL DISCUSSION

Our first conclusion must be that vowel devoicing in Japanese certainly does not make speech processing easier. The basic perceptual difficulty of consonant sequences holds as much for Japanese listeners as for listeners to any other language: consonants are easier to process if they are adjacent to vowels. The differences in phonological functionality of vowels and consonants likewise hold as strongly for Japanese as for other languages: a vowel can stand alone as a syllable and hence as a word, but a consonant in general cannot. Finally, the role of this vowel-consonant asymmetry in segmenting continuous speech is also parallel for Japanese and for other listeners.

Research on spoken-word recognition has amassed abundant evidence that speech input concurrently activates many word candidates which it fully or partially supports; because all languages construct very large vocabularies from relatively few phonemes, words in all languages exhibit a great deal of embedding and overlap, so that many such candidates will be unintended competitors which need to be rejected if the real message is to be recognized. Many mechanisms exist to deal with this competition (see McQueen, 2007, for a review). The PWC, which allows competitors to be rejected if accepting them would strand a vowelless residue of the input, enables English listeners to suppress unwanted activation of *egg* when they hear *leg* or *legacy*, and by the same token it enables Japanese listeners to suppress unwanted activation of *asa* when they hear *kasa* "umbrella." The operation of the PWC runs parallel in English (Norris *et al.*, 1997, 2001) and Japanese (McQueen *et al.*, 2001). Our new experiments have shown that the widespread phenomenon of devoicing in spoken Japanese does not modulate the power of this effect at all. Even consonants which could accompany a devoiced vowel made recognition of an adjacent word hard, to effectively the same extent as consonants which could not have preceded devoicing. Though recognition of *asa* was relatively easy before a vowel, it was hard before a vowelless consonant, even one which might have supported devoicing (e.g., the voiceless fricative /f/).

These segmentation effects operate upon words activated by speech input. The equivalence of the vowelless environments which do support devoicing and those which do not implies that during prelexical processing vowels are not automatically restored or inserted into either environment. In this our results match with those from a recent study by Mash *et al.* (2006), who tested the effect of vowel devoicing on the compensation for coarticulation which shifts identification of a /t-k/ continuum following /s/ versus /ʃ/ (Mann and Repp, 1981). Japanese has no cluster onsets and no coda obstruents, so such sequences are not lexically possible, but /sk/, /st/, /ʃk/, and /ʃt/ sequences can indeed arise from devoicing. The necessary environments for compensation for coarticulation thus exist in practice. But if the underlying vowel had been perceptually restored at the prelexical level, there would be no such compensation because the consonants would not be sequential, but would be interrupted by a vowel. Mash *et al.* (2006) found, however, that Japanese listeners performing /t/-/k/ categorization in nonwords produced more /k/ responses in *rusko* than in *rushko*—exactly the compensation effect that Mann and Repp (1981) had found. This suggests, in agreement with our own findings, that the vowels are not there in any sense which would affect prelexical processing of the auditory signal.

Our second conclusion, however, is that devoiced vowels will not in practice cause word recognition difficulty. In real speech, vowel devoicing will be encountered in known words. By presenting nonwords in an experiment, we can show that vowelless sequences are not automatically furnished with vowels. But in normal listening, listeners are rarely presented with nonwords. The results of Experiment 3 suggest that vowelless sequences arising by devoicing, though nominally illegal in Japanese phonology, will be effectively licensed as possible words. This may be especially so when stored lexical representations are activated—words with a devoiced initial syllable (e.g., *sukiyaki*, "sukiyaki" beginning /sk/) activate the intended lexical candidates. In Experiment 3, spotting *sake* was as fast in *nyaksake* as in *nyagusake*, and the error difference was much smaller than for the corresponding VCV targets. This may have been because the /ksa/ sequence was consistent with words beginning *kusa-,* i.e., with words which would be pronounced with initial /ksa/. Note that these candidates would, of course, compete for recognition with *sake*, but because *sake* was fully supported by the input while the other candidates were not, *sake* would win this competition. The competition offered by /ksa/ would presumably be no lesser or greater than the competition offered by the voiced-consonant context; /gusa/ would also have activated competitors, such as *gusaku* "rubbish." Note also that, in contrast to Experiments 1 and 2, the potentially devoiced sequences in Experiment 3 preceded the target words. There was thus time for words consistent with those sequences to be retrieved, and hence for the vowelless sequences to be licensed. This result contrasts particularly strongly with that for the VCV words which support no such licensed candidates; here the finding of McQueen *et al.* (2001) was replicated, in that the preceding context effectively blocked access to the embedded words completely. Words such as *asa* embedded in contexts such as *myochasa* were just as hard to detect as items in McQueen *et al.* (2001) such as *uni* in *gyabuni*; in both cases the PWC operated effectively to inhibit a parse of the input which would leave a vowelless residue ([tʃ], [b]). That this did not happen with *sake* in *nyaksake* thus constitutes powerful evidence of how listeners cope with the effects of devoicing in practice. In Experiments 1 and 2, where the potentially devoiced sequences followed the target words, no comparably licensed interpretation was available while the targets were being heard, so the underlying equivalence of a consonantal sequence arising from devoicing and any other (illegal) consonantal sequence emerged; the PWC again played its universal role, and segmentation was interfered with.

Other word recognition evidence from Japanese also suggests that devoiced forms effectively activate lexical representations. Ogasawara and Warner (2009) found that lexical decisions were faster for words such as *hashika* "measles" if the potentially devoiced vowel was reduced than if it was fully articulated. Further, although in nondevoicing environments phoneme detection responses were slower to reduced /i/ than to fully articulated /i/, in a devoicing environment the shorter, less clear, reduced vowel was responded to no more slowly than the longer, clearer full vowel. The lack of disadvantage for the acoustically less clear vowels in Ogasawara and Warner's phoneme detection study strongly suggests that the responses were not based on prelexical processing alone, but drew on lexical evidence (as the phoneme detection task allows; Norris *et al.*, 2000) to support the vowel interpretation.

Thus Japanese listeners appear not to restore devoiced vowels prelexically; the spoken forms with devoicing are perfectly functional in word recognition. Encountering a completely new word with a devoiced vowel could, of course, produce segmentation difficulty. Moreover, another indirect effect of devoicing might cause recognition delay. Consider that an important source of information for word recognition in Japanese is word accent pattern; whether a syllable is accented or not is perceived from only a fraction of the vowel (Cutler and Otake, 1999), and the accent pattern of a spoken fragment allows rapid rejection of alternative words with different accent (Sekiguchi and Nakajima, 1999). If a vowel is devoiced, the pitch information necessary for this efficient use of accent is no longer available, and this has been held to be the reason why devoicing is disfavored in accented syllables (Vance, 1987). However, devoicing of accented syllables does indeed occur, and indeed is becoming more common (Sugito, 1982; Kitahara, 1998). For a single devoiced syllable, accentedness may be accurately apprehended from compensatory pitch modification in an immediately following syllable (Sugito, 1982; Sugito and Hirose, 1988; Maekawa, 1990). If, however, one of a sequence of devoiced syllables is accented, listeners can tell that there was an accented syllable (i.e., they accurately distinguish unaccented sequences from sequences containing an accent), but they cannot reliably tell on which syllable the accent fell (Maekawa, 1990). Thus, as well as possibly delaying segmentation, devoicing could also interfere with word recognition via disruption of accent information.

Our results thus indicate that although devoiced sequences may effectively access lexical representations, without any prelexical restoration of vowels being necessary, the cross-linguistically observed disadvantage for vowelless sequences in speech segmentation holds in Japanese as strongly as it does in other languages.

ACKNOWLEDGMENTS

This research was supported by the NWO-SPINOZA project "Native and Non-Native Listening" (AC) and by Japan Society for the Promotion of Science Grant No. 11610566 (TO). We thank Aoju Chen, Maarten Jansonius, and Bram de Kruijff for technical assistance, and Natasha

Warner, Kikuo Maekawa, and an anonymous reviewer for helpful comments.

APPENDIX A: PHONEMES OF JAPANESE

Vowels: /a,e,i,o,u/; Stops: /p,t,k,b,d,g/; flap: /ɽ/; nasals: /m,n,ɴ/; approximants: /w,j/; affricates: /c/; fricatives: /s,z,h/

[1] Additional analyses compared item subsets. No differences between the sub-groups of the items in which the consonant in the "impossible" environment was voiced (e.g., *gasub*) versus voiceless (e.g., *asap*) appeared in either Experiment 1B or 1C. The sub-group of items in which the impossible-devoicing context contained [t] or [d] were somewhat harder than items containing other consonants in this context, in both Experiments 1B and 1C, but the context effects were the same across sub-groups (no context by consonant interactions were significant).

Cutler, A., Demuth, K., and McQueen, J. M. (**2002a**). "Universality versus language-specificity in listening to running speech," Psychol. Sci. **13**, 258–262.

Cutler, A., McQueen, J. M., Jansonius, M., and Bayerl, S. (**2002b**). "The lexical statistics of competitor activation in spoken-word recognition," *Proceedings of the Ninth Australian International Conference Speech Science and Technology*, edited by C. Bow (Australian Speech Science and Technology Association, Canberra), pp. 40–45.

Cutler, A., and Norris, D. (**1988**). "The role of strong syllables in segmentation for lexical access," J. Exp. Psychol. Hum. Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1999**). "Pitch accent in spoken-word recognition in Japanese," J. Acoust. Soc. Am. **105**, 1877–1888.

Dumay, N., Frauenfelder, U. H., and Content, A. (**2002**). "The role of the syllable in lexical segmentation in French: Word-spotting data," Brain Lang **81**, 144–161.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (**1999**). "Epenthetic vowels in Japanese: A perceptual illusion?," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1568–1578.

Kitahara, M. (**1998**). "The interaction of pitch accent and vowel devoicing in Tokyo Japanese," in *Japanese-Korean Linguistics*, Vol. **8** (CSLI & SLA, Stanford, CA), pp. 303–315.

Kondo, M. (**2005**). "Syllable structure and its acoustic effects on vowels in devoicing environments," in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter, Berlin), pp. 229–246.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," Psychol. Monogr. **68**, 1–13.

Maekawa, K. (**1990**). "Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese," Proceedings of the First International Conference on Spoken Language, ICSLP 90, Kobe, Japan, pp. 517–520.

Maekawa, K. (**2003**). "Corpus of spontaneous Japanese: Its design and evaluation," Proceedings of ISCA/IEEE Workshop Spontaneous Speech Processing and Recognition, SSPR2003, Tokyo.

Maekawa, K., and Kikuchi, H. (**2005**). "Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report," in *Voicing in Japanese*, edited by J. van de Weijer, K. Nanjo, and T. Nishihara (Mouton de Gruyter, Berlin), pp. 205–228.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

Mash, D., Kawahara, S., Kingston, J., Brenner-Alsp, K., and Chambless, D. (**2006**). "Sequential contrast versus compensation for coarticulation in Japanese versus English," Paper presented to the Acoustical Society of America, Providence, RI.

McQueen, J. (**1996**). "Word-spotting," Lang. Cognit. Processes **11**, 695–699.

McQueen, J. M. (**1998**). "Segmentation of continuous speech using phonotactics," J. Mem. Lang. **39**, 21–46.

McQueen, J. M. (**2007**). "Eight questions about spoken-word recognition," in *The Oxford Handbook of Psycholinguistics*, edited by G. Gaskell (Oxford University Press, Oxford), pp. 37–53.

McQueen, J. M., and Cutler, A. (**1998**). "Spotting (different types of) words in (different types of) context," Proceedings of the Fifth International Conference on Spoken Language Processing, Sydney, Vol. **6**, pp. 2791–2794.

McQueen, J. M., Otake, T., and Cutler, A. (**2001**). "Rhythmic cues and possible-word constraints in Japanese speech segmentation," J. Mem. Lang. **45**, 103–132.

Norris, D., McQueen, A., Cutler, J. M., Butterfield, S., and Kearns, R. (**2001**). "Language-universal constraints on speech segmentation," Lang. Cognit. Processes **16**, 637–660.

Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–325.

Norris, D., McQueen, J. M., Cutler, A., and Butterfield, S. (**1997**). "The possible-word constraint in the segmentation of continuous speech," Cogn. Psychol. **34**, 191–243.

Ogasawara, N., and Warner, N. (**2009**). "Processing missing vowels: Allophonic processing in Japanese," Lang. Cognit. Processes. In press.

Racine, I., and Grosjean, F. (**2000**). "The influence of schwa deletion on the recognition of words in continuous speech," Année Psychol. **100**, 393–417.

Sekiguchi, T., and Y., Nakajima, (**1999**). "The use of lexical prosody for lexical access of the Japanese language," J. Psycholinguist. Res. **28**, 439–454.

Spinelli, E., McQueen, J. M., and Cutler, A. (**2003**). "Processing resyllabified words in French," J. Mem. Lang. **48**, 233–254.

Sugito, M. (**1982**). *Nihongo akusento no kenkyuu (Studies on Japanese accent)* (Sanseido, Tokyo).

Sugito, M. (**1995**). *Osaka-Tokyo Akusento Onsei Jiten (Osaka-Tokyo Accent Pronunciation Dictionary)* (Maruzen, Tokyo).

Sugito, M., and Hirose, H. (**1988**). "Production and perception of accented devoiced vowels in Japanese," Annual Bulletin Research Institute of Logopedics and Phoniatrics **22**, 19–37.

Suomi, K., McQueen, J. M., and Cutler, A. (**1997**). "Vowel harmony and speech segmentation in Finnish," J. Mem. Lang. **36**, 422–444.

van der Lugt, A. H. (**2001**). "The use of sequential probabilities in the segmentation of speech," Percept. Psychophys. **63**, 811–823.

van Donselaar, W., Kuijpers, C., and Cutler, A. (**1999**). "Facilitatory effects of vowel epenthesis on word processing in Dutch," J. Mem. Lang. **41**, 59–77.

van Son, R. J. J. H., and Pols, L. C. W. (**1995**). "The influence of local context on the identification of vowels and consonants," Proceedings of Eurospeech95, Madrid, pp. 967–970.

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).

Warner, N., Kim, J., Davis, C., and Cutler, A. (**2005**). "Use of complex phonological patterns in processing: Evidence from Korean," J. Linguist. **41**, 353–387.

Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.

Yip, M. C. (**2004**). "Possible-word constraints in Cantonese speech segmentation," J. Psycholinguist. Res. **33**, 165–173.