Ha! Linguistic Studies in Honor of Mark R. Hale

Ha! Linguistic Studies in Honor of Mark R. Hale

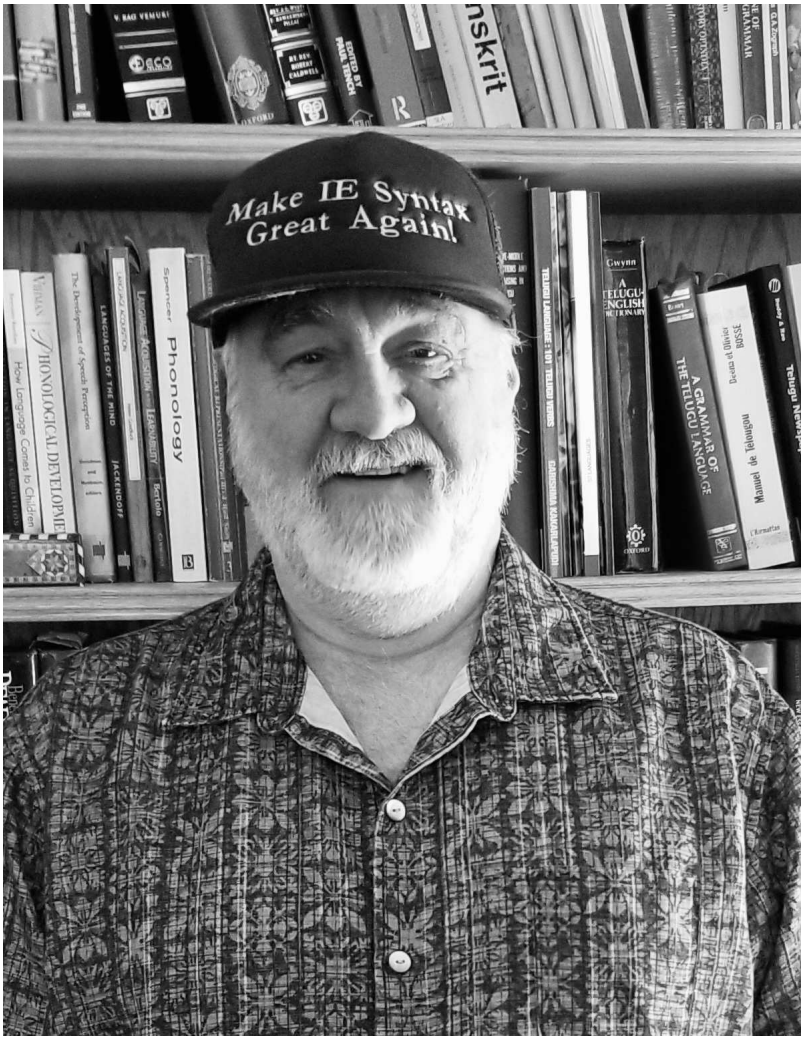
Laura Grestenberger, Charles Reiss,
Hannes A. Fellner and Gabriel Z. Pantillon (eds.)

WIESBADEN 2022
DR. LUDWIG REICHERT VERLAG

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten
sind im Internet über http://dnb.dnb.de abrufbar.

ISBN: 978-3-7520-0606-3 (Print)
eISBN: 978-3-7520-0085-6 (E-Book)
https://doi.org/10.29091/9783752000856
www.reichert-verlag.de

Gedruckt auf säurefreiem Papier (alterungsbeständig pH7 –, neutral)
Printed in Germany


Make IE Syntax Great Again!
Phonology
A GRAMMAR OF THE TELUGU LANGUAGE
Manuel de Telougou

Contents

Phonological Knowledge and Perceptual Epenthesis*

Korlin Bruhn, Bridget Samuels & Bert Vaux

1 Introduction

It has long been known that one's native (L1) phonemic inventory can influence the perception of non-native (L2) contrasts (e.g., Sapir 1933). While infants initially display the ability to distinguish both native and non-native phonemic contrasts (cf. Saffran et al. 2006), by around 10–12 months (e.g., Werker and Tees 1984a, Werker 1989) this ability is significantly attenuated as the child becomes attuned more specifically to the L1 inventory. More recently, the difficulty of perceiving sequences that are phonotactically illicit in L1 has also been widely discussed: studies have claimed that French-speaking listeners perceptually assimilate illicit [tl], [dl] onset clusters to licit /kl/, /gl/ onsets (Hallé et al. 1998); that English listeners epenthesize a schwa to repair onset clusters that violate the language's preferred rising sonority profile (Berent et al. 2007, Davidson et al. 2007); and that Japanese speakers use epenthesis to break up illicit consonant clusters and resyllabify coda consonants in loanwords (e.g., Dupoux et al. 1999).

In the present work we focus on this phenomenon of epenthesis in Japanese and its relevance for models of speech perception. We present evidence that Japanese speakers can still access phonetic details necessary to learn a non-native contrast, in spite of their native phonology. Japanese disallows all consonants other than nasals and the first part of geminate consonants in coda position. Consequently, foreign loanwords violating these phonotactic restrictions are repaired via epenthesis of /u/ (typically pronounced [ɯ̟], as will become relevant later) or /o/ (data from Itô and Mester 1995):

(1) a. 'Sphinx' → *sufiŋkusu*
b. 'Zeitgeist' → *tˢaitogaisuto*
c. 'fight' → *faito*

* The present work owes a great deal both directly and indirectly to Mark Hale's thinking on perception, acquisition, the initial state of the language acquisition device, and what performance effects do and do not reveal about competence. The third author remembers vividly discussions with Mark in the basement of Grays Hall and the ground floor of 77 Dunster Street in the 1990s about contrast-driven feature acquisition versus full perceptual access, ideas which were eventually published as Hale and Kissock (1998), Hale et al. (2007) and Hale and Reiss (2008), and (along with Janet Werker's findings and theories, whose significance Mark and Charles Reiss first presented to us) significantly shaped the thinking reflected in the present chapter. We appreciate helpful suggestions from Laura Grestenberger, Charles Reiss, and Markus Pöchtrager, which greatly improved the paper.

doi./org/10.29091/9783752000856/001 eISBN: 978-3-7520-0085-6

Dupoux et al. (1999) draw a connection between perception and loanword adaptation, suggesting that adaptation happens at the perceptual level: That is, when Japanese listeners are confronted with, e.g., [sfɳks], they perceive it as [sɯ̥.ɸiɳ.kɯ̥.sɯ̥]. This seems to be confirmed by the results of two of their experimental studies (Dupoux et al. 1999, Dupoux et al. 2001), in which French and Japanese subjects indicated whether they heard a medial /u/ in nonword stimuli taken from a VCCV-VCuCV continuum. Japanese listeners indicated the presence of the vowel around 70% of the time even in the VCCV condition. In two speeded ABX tasks—one with and one without speaker change—Japanese listeners had difficulty distinguishing stimuli like *ebzo* and *ebuzo* reliably. Japanese listeners made fewer errors in the same-speaker task than in the different-speaker task but still made significantly more errors than the French listeners. Dupoux et al. concluded that phonotactic knowledge influences listeners so strongly that it creates a perceptual illusion: the Japanese listeners judge there to be a speech segment (/u/) despite the absence of acoustic correlates in the signal, simply because their L1 phonology insists it should be there.

Subsequent studies by Dupoux and colleagues pursued this so-called *ebzo*-effect. In a 2001 study, they evaluated lexical influences (cf. Ganong 1980): In a speeded lexical decision task, participants judged whether a CVCCV speech token was a real word or not. The stimuli were designed so that insertion of /u/ or a different vowel would yield an existing word: e.g., *sokdo* is not a word of Japanese, but *sokudo* means 'speed'; *mikdo* is not a word but *mikado* means 'emperor.' The stimuli that needed an epenthetic /u/ to form a word were largely judged to be real words despite the surface consonant cluster. The stimuli that required a vowel other than /u/ to form a word were mostly judged to be non-words. The authors thus concluded that the illusion of epenthesis cannot be caused by top-down influences from lexical neighbors, but rather must be a pre-lexical effect; perceptual insertion of the vowel happens before the lexicon is consulted.

Dehaene-Lambertz et al.'s (2000) ERP study using an oddball paradigm with stimuli such as *ebzo* and *ebuzo*, which contrast after their first consonant, appears at first blush to confirm that phonology influences speech perception at very early processing stages. French but not Japanese listeners displayed a mismatch negativity (MMN) at a latency of 140–280 ms after the offset of the oddball's first consonant. As MMNs are usually elicited when a change in the stimulus is noticed, the lack of MMN in the Japanese listeners in response to an *ebzo-ebuzo* change implies that they do not perceive these stimuli to be different from one another. This suggests that epenthesis of the vowel happens at a stage so early in processing that it prevents an MMN from being elicited.

Why does this perceptual epenthesis occur? The answer to this question must be formulated within the larger context of Japanese phonology. While Dupoux et al. (1999) and Dehaene-Lambertz et al. (2000) only tested perceptual epenthesis of /u/, vowels other than /u/ can be found inserted in loanwords: /o/ is typically inserted after dental stops (e.g., 'fight' borrowed as *faito*), presumably because when a /u/ follows a dental stop, the stop becomes affricated, e.g., /kat-/ 'win' → [katanai] 'win - negated' but [kat^{s}ɯ] 'win - present tense' (Itô and Mester 1995). Only /u/ appears to be perceptually inserted, though, as a study by Monahan et al. (2009) suggests: nonword stimuli like *etma*, which offer the environment for an /o/ to be inserted,

could be distinguished from both *etuma* and *etoma* by Japanese listeners, whereas participants had trouble with the classic *ebzo-ebuzo* contrast. This suggests that /u/-insertion is blocked in environments where /u/ is not phonologically licensed, yet nonetheless /o/ is not perceptually inserted. One explanation suggested in the literature (Dupoux et al. 1999, Dupoux et al. 2001, Dupoux et al. 2011) is that high vowel devoicing, which is optional in many dialects (cf. Vance 1987, chapter 6), plays a role. This means that both /i/ and /u/ can be realized as anything from a whispered vowel with visible formants to simply frication without formant information, or even total deletion (Tsuchida 1987, Varden 1998). On the other hand, /o/ does not typically undergo devoicing,[1] so it might be sufficiently distinguishable from [Ø] that its presence cannot be posited without strong acoustic evidence in the signal—cf. also Steriade's (2001, 2009) P-map theory, in which the perceptually minimal deformation of the input is chosen in phonotactic repairs, and Samuels and Vaux (2020) on the silence-cued perception of epenthetic stops. A similar argument could rule out the perception of /u/ in stimuli like *etma*, since /u/ triggers affrication of a preceding dental stop: [t^{s}u] and even [t^{s}u̥] could be too perceptually different from [tØ] to be perceived here.

Another important question is why /u/, but not the other Japanese high vowel, /i/, is subject to perceptual epenthesis.[2] We return to this issue in Section 4.1. For now, we raise the following question: if the only reason for perceptual epenthesis is that these vowels can sometimes be realized as [Ø] and thus that listeners are used to interpreting CC-sequences as underlyingly containing a high vowel, then why is /u/ preferred over /i/? Further, high vowel devoicing usually happens between two voiceless consonants, yet the stimuli used in Dupoux's studies mainly involve environments that do not license devoicing. The key could be that /u/ in Japanese is the closest to [Ø] in that it is the shortest vowel (Beckman 1982) and the one which allows the most formant variability (Keating and Huffman 1984).[3] This leads Dupoux et al. (1999, 2001) to a modified version of Best's (1994) Perceptual Assimilation Model, wherein they suggest that the perceptual unit in Japanese is the syllable. Therefore, when a foreign sequence is perceived, it is mapped syllable-by-syllable onto the perceptually closest native category, which for CØ would be C/u/ for the reasons set out above.

Peperkamp and Dupoux (2003) and Peperkamp (2005) refine this theory into what we will refer to as the Phonetic Decoder model. In this model, a phonetic decoding module maps the speech signal, one word at a time, into a discrete phonetic representation that conforms to the L1 phonology; a phonological decoding module then maps this surface form onto an underlying representation. Accordingly, a sequence that is phonotactically illicit in the L1 cannot be mapped accurately: The phonetic decoder for L1 Japanese speakers cannot accommodate two consonants next to each

1 Nonhigh vowels sometimes devoice, too, but less often and less systematically (Vance 1987: 48f.) and will therefore not be considered here.

2 It should of course be noted that epenthetic—and specifically perceptually epenthetic—vowels differ cross-linguistically. For example, Dupoux et al. (2011) demonstrate that in a context where Japanese listeners indicate the presence of a perceptually epenthetic /u/, Brazilian Portuguese listeners perceive an epenthetic /i/.

3 Alternatively or in addition, this variability could indicate that /u/ is featurally underspecified. We set this issue aside.

other, so an empty vowel segment intervenes and is carried over to the phonological mapping, where the empty vowel slot is interpreted as the closest phonetic match.

However, in Dupoux et al.'s (1999) study an effect of condition (i.e., different responses to *ebzo* and *ebuzo*) was observed in Japanese listeners for an ERP at a latency of 290–400 ms, suggesting that speakers perceive the difference between stimulus types on some level. Moreover, Tremblay et al. (1997) have shown that a weak MMN response to non-native contrasts like the one Dehaene-Lambertz et al. (2000) found for Japanese listeners can be strengthened by training. Tremblay et al. trained native English speakers on the non-native category of pre-voiced labial stops and found electrophysiological evidence that these subjects could generalize the newly learned VOT boundary to pre-voiced alveolar stops.

Ample behavioral evidence also points to the conclusion that some ability to perceive non-native phonological patterns remains into adulthood. Werker and Tees (1984b) found that under certain circumstances, even difficult foreign contrasts can be distinguished by adults. Furthermore, when listeners are not in "speech mode" or when auditory linguistic stimuli are altered to a sufficient degree that renders them non-speech-like, the ability to discriminate non-native contrasts reveals itself as intact (Best et al. 1981, Remez et al. 1981, Liberman 1982, Werker and Tees 1984b). Davidson et al. (2007) found that a picture-matching task which teaches participants to distinguish minimal pairs can help English listeners to overcome epenthetic perceptual repair of illegal onset clusters. By associating each stimulus with a meaning—e.g., *zemagu* denotes a picture of a dragon while *zmagu* refers to a fish—performance was enhanced relative to an AX discrimination task using the same stimuli without associated meanings. Even in Dupoux et al.'s (1999) study, while Japanese listeners underperformed relative to the French listeners, they performed significantly better than chance in the VCCV vs. VCuCV perception task, and their error rate was also lower than would be expected if they were simply guessing in the ABX task. In sum, the evidence suggests that the perception of non-native contrasts persists on some level in adults and can be improved. Thus, perceptual illusions of the *ebzo-ebuzo* type may not be so inevitable after all.

The question therefore arises: Can Japanese listeners be taught to perceive faithfully consonant clusters such as the one in *ebzo*? If they are indeed able to avoid or override perceptual epenthesis, it would show that Japanese listeners are not "'deaf' to the difference between *ebuzo* and *ebzo*" (Dupoux et al. 1999: 1574) and imply that phonological influence is not as inevitable as Dupoux and colleagues assert. If Japanese listeners, like the English listeners studied by Davidson et al. (2007) (cf. also Berent et al. 2007), can be taught not to activate perceptual epenthesis, it would indicate that even in those cases where phonotactic knowledge has been claimed to interfere with speech perception at very early stages (e.g., Dehaene-Lambertz et al. 2000), phonetic detail must still be perceived accurately initially before phonology alters the percept. Otherwise, Japanese listeners would not be able to use the phonetic detail to learn that there is a meaningful difference between *ebzo* and *ebuzo*.

If Japanese speakers are indeed able to learn such non-native contrasts, it would provide some evidence against the Phonetic Decoder model. The present study therefore investigates whether perceptual epenthesis can be overcome by Japanese listeners by adapting Davidson et al.'s (2007) study to the sound patterns in question. We also

replicate Dupoux et al.'s (1999) ABX task. This allows for a more reliable measure of improvement while at the same time allowing a direct comparison to Dupoux et al.'s study. We demonstrate that Japanese speakers are able to learn the *ebzo-ebuzo* contrast, and provide an explanation of these results that does not depend on the Phonetic Decoder model.

1.1 Predictions

Both possible outcomes of this study—the contrast between *ebzo* and *ebuzo* can or cannot be taught to Japanese L1 speakers—have interesting implications and each outcome is predicted by a different class of models of speech perception.

1. Dupoux and colleagues (e.g., Dupoux et al. 1999, 2001, 2011; Dehaene-Lambertz et al. 2000; Peperkamp and Dupoux 2003) predict that Japanese listeners cannot be taught to perceive the *ebzo-ebuzo* contrast reliably. That is, a contrast can only be perceived if L1 phonotactics allow it.[4] If this prediction is correct, then the triggering factors for English onset epenthesis (e.g., *lbif-lebif* in Berent et al. 2007) must be different from those involved in the Japanese case to explain the fact that English listeners can learn to override perceptual epenthesis (Davidson et al. 2007) but Japanese listeners apparently cannot. This is consistent with the Phonetic Decoder model—L1 categories must be available for the speech signal to be mapped onto. The relevant underlying difference between English and Japanese may be represented in terms of CV sequences (cf. also Clements and Keyser 1983). Japanese listeners would thus map a stimulus like *ebzo* onto V.CV.CV, because no VC.CV template is available. English listeners, however, have an appropriate template available to parse the CC sequences in both *ebzo* and *lbif* accurately: while they may not be used to mapping *lb* to an onset sequence, their task under this interpretation is merely to map *lbif* onto the CCVC template available in English, rather than onto the CVCVC template which would create a percept of *lebif*.
2. The alternative hypothesis holds that Japanese listeners can be taught to overcome perceptual epenthesis reliably. This outcome would not be consistent with the Phonetic Decoder model. Instead, it would suggest that, in certain tasks, Japanese listeners can access the phonetic details which provide the crucial information to distinguish *ebzo* from *ebuzo*, especially given that the phonemes involved in the contrast are already in the L1 inventory of Japanese listeners.

1.2 Approach

The current study applies Davidson et al.'s (2007) paradigm to Japanese speakers, testing whether discrimination between the types of stimuli that Dupoux et al. (1999 et seq.) have used in their studies can improve when the crucial phonetic difference is

4 Note that this does not mean that all consonant clusters are treated equally: some are misperceived more often than others. For example, Berent et al. (2007) found that word-initial sonority falls are misperceived more often than sonority plateaus. Accidental gaps in a language's inventory of allowed sequences should be relatively unproblematic for perception, by virtue of being accidental rather than systematic.

highlighted. A brief explanation of the paradigm is necessary in order to understand the motivation behind the changes that we introduced. The methods are reviewed in detail in Section 2.

Davidson et al.'s (2007) experiment is a three-phase picture-masking task (PMT). In the familiarization phase, participants are introduced to picture-name pairs (see Figure 2). In the following training phase, participants receive training on contrasts that are difficult for them to perceive correctly; the training involves matching minimal pair stimuli to the pictures they denote. Participants hear a name and have to match it to the correct picture (Figure 3). Once they have correctly matched all of the names in one trial without a mistake, participants move onto the test phase in which they see a picture and hear the minimal pair together and are asked to indicate which one of the two minimally contrasting stimuli denotes the picture (Figure 4). To ensure that the contrast learned in the training phase for one speaker can also be carried over to another speaker, a different voice is used in the test phase.

The present study adds two ABX tasks to the PMT, for two reasons. One motivation was to monitor the effect of training more reliably, as the original PMT design does not include a measure of performance before training on the contrast has occurred. Davidson et al. (2007) instead refer to a previous AX-discrimination experiment with the same stimuli (Davidson 2007). Participants performed at chance in that task, which compared CC onset stimuli with their counterparts containing actual schwas (e.g., [vtɑke]∼[vətɑke]). Because participants chose the correct token 60.5%[5] of the time in the PMT (Davidson et al. 2007), the authors hypothesized that their learning paradigm was successful.

There are several issues with using a different task from a different experiment as a baseline for performance improvement. Firstly, this comparison interprets a performance difference between two different groups of participants as 'improvement'. Secondly, the AX experiment (Davidson 2007) and the PMT (Davidson et al. 2007) are not directly comparable, because they test different abilities. The AX task required participants to indicate whether two stimuli heard consecutively were "exactly the same" or different, thus placing emphasis on acoustic identity rather than phonological category membership of the two items. However, in order to perform well in the PMT test phase, participants had to build phonological representations of minimal pairs, since they had to carry over the learned contrast to a new speaker. Consequently, the difference in performance between those two tasks could stem from the difference in tasks. The present study eliminates these confounds by running the same ABX task before and after the PMT. The 'before' task provides a baseline against which performance in the 'after' task can be evaluated; an increase in performance is thus likely to be a result of the training.

Given the methodological concerns just discussed, we felt it was necessary to replicate the training paradigm used by Davidson et al. (2007) with English-speaking listeners to confirm its efficacy. This group was tested on word-initial sequences that typically trigger schwa epenthesis for English listeners. Consequently, if the

5 This figure was not explicitly given in Davidson et al. (2007). Their analysis involved dividing the participants into "high performer" and "low performer" groups as well as splitting the scores according to whether the target word had a CC or a CəC onset. When these groups/conditions are collapsed, the overall percentage of correct responses is 60.5%.

English group showed performance improvement, a negative result for Japanese listeners would not indicate a faulty training method but could be attributed to the insurmountable influence of phonology on perception.

The second motivation for the ABX tasks in the present study was to enable comparison to parts of Dupoux et al. (1999). Since the present study's intention is to test whether Japanese listeners can learn to suppress perceptual epenthesis and thus improve performance on the *ebuzo/ebzo*-type contrast, it is advisable to use the more challenging task as a baseline in order to maximize possibility for improvement. Of the two ABX designs that Dupoux et al. (1999) used, the more challenging one (with an error rate of 32%) was Experiment 3, which used one speaker for A and B and a second speaker for X. The speaker change is designed to prevent participants from using only acoustic cues; they have to build a more abstract, phonological representation of the stimuli, so that X can be compared to A and B despite the speaker difference. Given that it has been argued especially for the Japanese case (e.g., Dupoux et al. 1999, 2001; Dehaene-Lambertz et al. 2000) that phonology interferes with speech perception at even very low levels of acoustic processing, it can be assumed that a task in which phonological representations are built and compared allows even more phonological interference to distort the acoustic input. If, as predicted by the Phonetic Decoder model, both stimuli A and B map onto the same phonological form (i.e., VCuCV), participants should have trouble matching X with the correct stimulus. On the other hand, if Davidson et al.'s (2007) PMT training paradigm is successful, then Japanese participants should be able to learn to suppress the epenthesis percept.

In addition to the fundamental question of whether Japanese listeners can overcome perceptual vowel epenthesis, the present study also addresses the basis for vowel epenthesis. Dupoux and colleagues have frequently evoked the notion of high vowel devoicing as a possible factor in perceptual vowel epenthesis (Dupoux et al. 1999, Dupoux et al. 2001, Dupoux et al. 2011) but did not examine this possibility systematically, instead only analyzing their results retrospectively. Consequently, their stimuli were not balanced for contexts that are known to be unfavorable to high vowel devoicing, which is unfortunate because this bears on whether the Phonetic Decoder model adequately describes the observed results. If favorable contexts yield more epenthesis, this could be taken as confirmation that acoustic proximity (cf. Peperkamp and Dupoux 2003) does indeed play a role in the mapping of the acoustic stream onto native syllables. For example, with a stimulus like *ebzo*, we would not expect much perceptual epenthesis because high vowels do not devoice between voiced consonants. On other hand, with a stimulus that presents a more favorable context for /u/-epenthesis, such as *etma*, we would expect at least some Japanese listeners to accept CØ as a variant of C/u/ such that *etma* is not as perceptually distinct from *etuma* as *ebzo* is from *ebuzo*.

To investigate this, 50% of the present study's stimuli contain a phonetic context favorable for high vowel devoicing, i.e., both surrounding Cs are voiceless obstruents, and the other half contains contexts which are unfavorable to high vowel devoicing, i.e., at least one of the Cs is voiced. If high vowel devoicing plays a role in vowel epenthesis, we predict that the phonetic contexts triggering devoicing should elicit significantly more perceptual epenthesis than the other environments, and hence that

Japanese listeners will make more mistakes distinguishing the consonant cluster from its epenthetic counterpart.

2 Materials and Methods

2.1 Participants

2.1.1 English

Fifteen monolingual native speakers of British English (3 men and 12 women) participated in all three tasks (ABX1, PMT, ABX2). None reported any hearing impairments. The average age was 22.8 years (SD = 2.27). Two participants had received brief training in phonetics while attending university but their performance did not deviate from that of participants without phonetic training; their scores were therefore included. Four participants had experience with languages allowing the onset clusters used in this experiment, namely Hebrew and Russian, but began study relatively late (at ages 16, 18 and 22).[6] One of these participants also had received Welsh lessons in early childhood but claimed that it had not been extensive enough to learn the language; in any case, Welsh does not allow the onsets used in this experiment. None of these participants were outliers with regard to their performance, suggesting that experience with those languages was not an influencing factor.

2.1.2 Japanese

Ten Japanese native speakers (5 men and 5 women) raised monolingually in Japan participated in all three tasks (ABX1, PMT, ABX2). None reported any hearing impairments. The mean age was 25.6 (SD = 4.84). Except for one participant, who started learning English at the age of 11, none had learned a foreign language extensively before the age of 12. One other participant began studying French at the age of 8 but claimed that it had not been extensive enough to learn the language properly. One participant reported being diagnosed with Asperger's Syndrome but both his results and the time he needed to complete the experiment were within the range of the other participants, so we decided not to exclude him. The average time that participants had lived in an English-speaking country was 4.1 years (SD = 3.59, range = 10.5, mode = 2).

2.2 Materials

All recordings were made in a sound-attenuated booth onto a Nagra Ares-M II handheld digital recorder with a cardioid Sennheiser ME64 microphone at a sampling rate

6 Pallier et al. (1997) found that even adult Spanish-Catalan bilinguals who had been exposed to L2 on a daily basis before the age of six did not show categorical perception on the Catalan phoneme contrast /e/-/ɛ/. While this study was concerned with phonemes rather than phonotactics, it nevertheless suggests that even language acquisition in early childhood does not necessarily result in complete native-like language competence. Therefore, language learning during or after adolescence should not influence the results. Note also that our participant pool is similar to that of Dupoux et al. (1999), which included participants who had begun learning a foreign language after the age of 12.

of 22050 Hz. Recordings were edited as described below in Praat, and any recording artifacts (clicks) were excised from the sound files. Every stimulus was assigned to a picture of a unique cartoon character (16 experimental pictures and 6 practice pictures), so that participants learned to associate the auditory stimuli with meanings.

2.2.1 Practice set

Three German minimal pairs differing in the initial consonant were recorded once by a female native German speaker and once by a male native British English speaker who is also fluent in German: *Scholle-Wolle, Sonne-Tonne, Mund-Hund*. These stimuli were used in the practice sessions of all three tasks to familiarize participants with the task. This is the same approach used by Davidson et al. (2007) in their PMT, using Spanish words rather than German ones.

2.2.2 English set

For the English stimulus set, the same nonce words, recorded anew for this experiment, were used in all three tasks and were a subset of those used by Berent et al. (2007). Since the Japanese experimental design included investigation of two different conditions (contexts favorable and unfavorable to devoicing high vowels), the English experimental design also included two different conditions of differing favorability to epenthesis, namely one with onsets with sonority plateaus and a second with falling sonority, as the latter trigger schwa epenthesis more often than the former (Berent et al. 2007).

A set of four tokens with sonority plateau (= SP) onsets (e.g., *kpim*) plus their epenthetic counterparts (e.g., *kəpim*) and four with sonority falls (= SF) (e.g., *mkag*) as well as their counterparts with inserted schwas (e.g., *məkag*) were used. A male native British English speaker recorded the eight CəC-onset tokens and the eight corresponding CC-onset tokens were created by excising the schwa from the recorded CəC set. In addition, the entire set of nonce words was recorded by a female native British English speaker as detailed below. A native British English speaker and trained phonetician listened to all of the stimuli (male and female) and deemed them natural-sounding.

To ensure that listeners were distinguishing the stimuli within a pair on the basis of the contrast under investigation (presence versus absence of schwa), CC-onset stimuli were made from recordings of CəC-initial stimuli, to generate pairs of stimuli that were acoustically identical apart from the schwa. This set was used to train participants on the contrast: A and B in the ABX tasks were drawn from it, and it was used in the training phase of the PMT. CC onset stimuli were created using Praat (Boersma and Weenink, 2011) to digitally remove the schwas from the CəC recordings at zero crossings. Care was taken not to create abrupt transitions that could result in unnatural-sounding stimuli. Since the length of the schwa plays an important role in the discriminability between CC and CəC, the CəC stimuli were edited to have schwas of average length, as measured by Davidson (2007) and employed by Davidson et al. (2007), namely 68 ms. Some of the schwas in the CəC stimuli thus had to be reduced in length slightly, by cutting single periods at zero crossings. If more than

one period had to be removed, non-neighboring periods were selected, so as not to obscure formant transitions.

Stimuli recorded by the female speaker were used for the X in the ABX tasks and in the test phase of the PMT. Acoustic identity was not an issue for this set, since it was not used in the training phase. Therefore, naturalness of the stimuli was favored over consistency with the male set. Further, we wished to avoid the possibility of participants simply exploiting the length difference of the stimuli if both sets of stimuli used near-identical CC-CəC pairs. By using different recordings for the female speaker, we sought to ensure that participants would have to learn the abstract contrast in order to perform well. Thus, CC stimuli were recorded naturally. Schwas were reduced or lengthened to around 68 ms (M = 63 ms) by removing or inserting single periods at zero crossings. In some cases the onset Cs of the CəC recordings had to be spliced and used as the onset C of the CC-stimuli because they differed strongly in quality. For example, the speaker produced very strong aspiration in /tpʌg/, so the [t] from [təpʌg] was used to replace the original one, and since the [m] in [mkʌg] was longer than the entire [mə] in [məkʌg], a few periods were excised from it.

2.2.3 Japanese set

For the Japanese stimulus set (recorded anew for this study), the same tokens were used in all three tasks and were a subset of the stimuli used by Dupoux et al. (1999) except for the nonwords *aksa-akusa*, which were added since otherwise an insufficient number of stimuli included phonetic environments that trigger high-vowel devoicing. Four tokens with phonetic contexts favorable to high-vowel devoicing (= FC) and their epenthetic counterparts (e.g., *ekshi-ekushi*) and four tokens with unfavorable contexts (= UC) and their epenthetic partners (e.g., *ebza-ebuza*) were recorded once by a male native British English speaker and once by a female native British English speaker.

While making the Japanese stimuli, the main concern was to eliminate the possibility of coarticulatory cues betraying the former presence of a vowel after digitally removing the [u] from a VCuCV token. (Note that Dupoux et al. 1999: 1573 used Japanese speakers who "could not be prevented from inserting a short vowel [u] [. . .] in some of the *ebzo* stimuli.") Dupoux et al. (2011) showed that Japanese listeners are highly sensitive to such coarticulatory traces, and it has further been shown that even 10 ms of a stop burst can contain enough information to identify a following vowel (Blumstein and Stevens 1980). In order to avoid these problems, both VCuCV and VCCV tokens were recorded naturally.

To stay as close as possible to Dupoux et al.'s (1999) study, the average vowel length of that study—89 ms[7]—was used, and vowels were accordingly reduced by removing single periods at zero crossings or lengthened by inserting single periods: the average length was 89 ms for the male set and 90 ms for the female set.

7 Dupoux et al. (1999) do not explicitly state the length of the /u/s in their stimuli. They only indicate that the mean difference between *ebzo* and *ebuzo* items was 89 ms, suggesting that the difference is due to the presence vs. absence of the vowel. However, it is entirely possible that other elements in the stimuli varied in length as well. Nevertheless, since this is the only indication the authors provide regarding the crucial length of the vowel, we decided to use this as target length for the /u/s in our study.

2.3 Procedure

All three experiments (ABX1, PMT, ABX2) were programmed using E-PRIME (Psychology Software Tools, Inc., Pittsburgh, PA, USA).[8] Participants were tested individually in a sound-attenuated booth. The stimuli were played over Sennheiser HD 555 headphones and the pictures were shown on a computer screen. Instructions were given at the beginning by the experimenter and also appeared on the screen before each task.

Participants took around 30 minutes to conclude the entire experiment, with the English participants needing slightly less time on average (mean = 27 mins) than the Japanese speakers (mean = 32 mins).

2.3.1 ABX

The first ABX task (ABX1) took place directly before the PMT and the other ABX task (ABX2) took place directly after the PMT. ABX1 and ABX2 used the same stimuli and differed only in the order in which the stimuli were presented, as the stimuli were randomized for each test run.

The eight pairs of stimuli (CC onset and [ə]-epenthetic counterpart for English listeners, VCCV and [u]-epenthetic counterpart for Japanese participants) yielded 32 trials: two possible orders within a pair and two options for the identity of X, which repeated either A or B (2x2: Order x Identity). A and B were taken from the male stimulus set, X from the female set. The interstimulus interval was 450–500ms.[9]

Participants were told that they were taking part in a study about language intuitions and were going to hear triplets of foreign words, the first two spoken by one voice and differing from one another, the third one spoken by a different voice but repeating either the first or the second word. They were advised that they only had a few seconds to respond; after about four seconds "no response" was logged and the next sound triplet was announced on the computer screen with the words "Next triplet ...". Further, participants were told that if they were unsure about the correct match they should answer based on their intuitions. Participants were instructed to press 'A' on the keyboard if X was like the first word they had heard or 'B' if X was a repetition of the second word of the triplet.

The trial session was preceded by a short practice session consisting of four trials created from two German practice stimulus pairs with partial permutation to demonstrate that items could be repeated.

The experimental design and number of participants mirrored those employed by Dupoux et al. (1999), to facilitate comparison. Note that there are a few divergences from Dupoux et al.'s set-up; their 1999 study includes an additional condition in which a vowel-length contrast was tested, using *ebuzo-ebuuzo*-type stimuli. This condition

8 Many thanks to Lisa Davidson and Jason Shaw for use of their PMT script (Davidson et al. 2007) for E-PRIME, which we modified to suit this study.

9 This variation was due to the fact that E-PRIME inexplicably clipped the beginnings of sound files upon playback. In order to prevent the beginnings of the stimuli from being truncated, we added 200–250 ms silence to the beginning of each sound file using clips taken from silent intervals in the recordings. Variations in length are therefore due to the necessity of inserting the silent intervals at zero crossings.

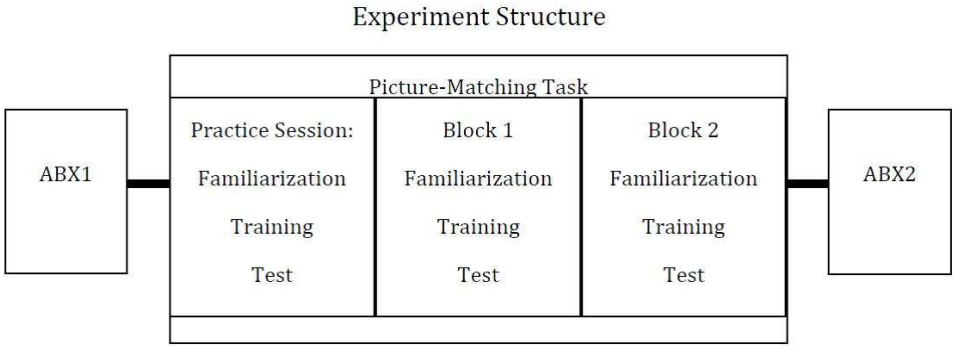
was introduced to achieve a complete crossover design in which the opposite predictions held for the two conditions for the French and Japanese groups. Since different stimulus sets were used for the English and Japanese groups in the present study, this condition was omitted. Further, Dupoux et al.'s (1999) epenthesis contrast condition comprised 16 stimulus pairs (*ebzo-ebuzo*); given different A, B and X permutations this yielded 64 trials instead of this study's 32. Since the present study used the same stimuli for all three experiments (ABX1, PMT, ABX2), using more than eight pairs would have excessively extended the time needed to complete the experiment.

Note also that participants in this study did not receive feedback during the practice session. However, this did not affect the results negatively: As will be evident in Section 3.3, this study's participants performed much better than those of Dupoux et al. (1999), despite this disadvantage.

2.3.2 Picture-Matching Task

The PMT consisted of three main parts, schematized in Figure 1, each with a familiarization phase, a training phase, and a test phase.

Figure 1: General structure of the experimental sequence



In the familiarization phase of each block, participants were introduced to the pictures (cartoon characters) and their respective names. In the training phase they had to learn those names, and in the test phase they had to choose the correct name for a given picture. This procedure was introduced through a practice session in which participants completed each of the stages once with the practice stimuli. The practice session, using German stimuli, preceded the first block of the main component and served to introduce participants to the procedure described in the remainder of this section. Participants were led through the procedure step by step via instructions on the screen before each phase.

The main part of the task was divided into two blocks. Sixteen words in total (eight per block) were learned during the main component. Each block consisted of four minimal pairs, two for each condition (SP/SF for English listeners, FC/UF for Japanese). The test phase was the only phase in which responses from participants were logged for later analysis.

The familiarization phases of the main component used the male stimulus set. Each picture was shown while playing the name associated with that picture twice (see Figure 2), and pictures were presented in random order such that each one appeared twice. Listeners were told that they would hear words in a foreign language that were the names of the cartoon characters. They were asked to memorize the names, although unbeknownst to the participants, this stage just served as an introduction to the names and pictures.

Figure 2: Excerpt from the familiarization phase. The cartoon characters appeared individually on the screen while their name was played twice. Each picture was shown twice in random order.

The training phase used the male stimulus set again. This phase presented participants with the eight characters arranged on the screen (Figure 3). When a word was played, subjects had to use the mouse to click on the corresponding picture. If they were correct, positive feedback was given and they were told how many matches in a row they got right. If they made a mistake, they received negative feedback and were told that the count was set back to zero, then the sound was played again while the correct cartoon was shown alone on the screen. Participants also had the option to practice the names further if needed: a 'Practice' button led them to a screen on which they saw the characters again. Whenever they clicked on a picture, the name was played. Participants were told that they could spend as much time as they needed on the practice page. Using the 'Practice' button reset the count to zero. Additionally, participants were brought automatically to the practice page when they failed to match any names correctly for 15 minutes.

Once participants correctly matched all words to the corresponding pictures without any mistakes, they automatically moved on to the test phase. Getting to this stage thus implied that participants had mastered the minimal pair contrast for at least the male speaker. The test phase used the female stimulus set, and participants were informed that the voice would be a different one from the one they had heard in the previous phases. A picture appeared on the screen and two words were played: the character's name and its minimal pair partner (e.g., *kpim-kepim, ebza-ebuza*) (see Figure 4). Participants were asked to press the 'A' key on the keyboard if the first word they heard was the character's name and 'B' if it was the second word. All characters were shown twice in random order with the order of the minimal pair switched. This yielded 16 trials per block and 32 in total.

Figure 3: Screenshot of the training phase of the two blocks. Listeners heard single words spoken and had to click on the correct picture denoted by the word.

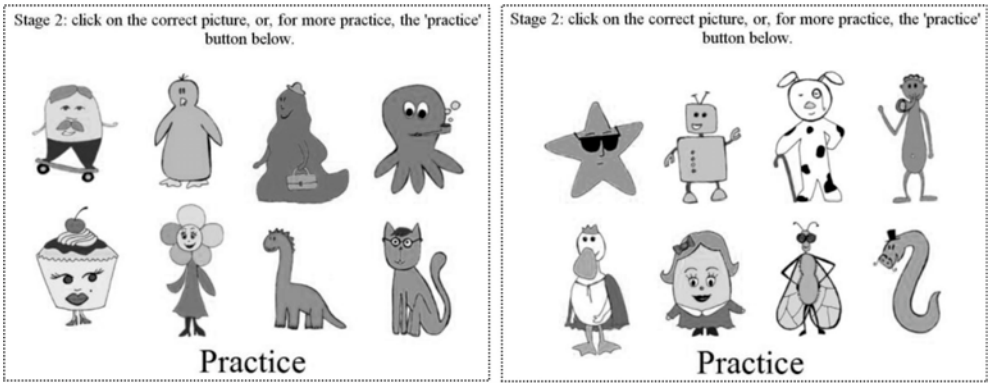
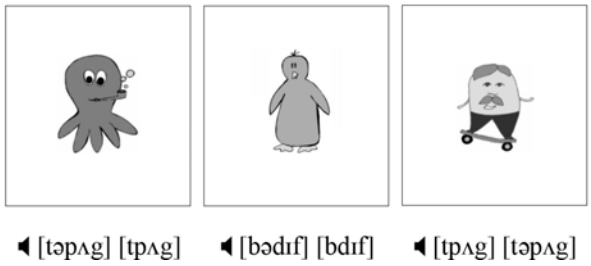


Figure 4: Excerpt from the test phase. The cartoon characters appeared individually on the screen while their name and its minimal pair partner were played. Each picture was shown twice in random order with the order of the target word and minimal pair partner counterbalanced.



3 Results

The results of the ABX tasks were subjected to a three-way ANOVA with time of measurement (before the PMT = ABX1, after the PMT = ABX2), sonority profile of the stimuli's onset cluster (sonority plateau = SP, sonority fall = SF) for the English group and phonetic environment for high vowel devoicing (unfavorable environment for devoicing = UC, favorable environment = FC) for the Japanese group, respectively, as within-subject factors. The third within-subject factor was sequence, i.e., the structure of the target word: whether the target word included a consonant cluster (CC) or the epenthetic vowel (CəC for English or CuC for Japanese). For the PMT, a two-way ANOVA was performed with sonority profile/phonetic environment and sequence as the two within-subject factors. ANOVAs were chosen because of their relative robustness against violations of assumptions (e.g., Box 1953, Srivastava 1959). All effects are reported as significant at $p \leq 0.05$.

3.1 English group

The results of the English listeners are summarized in Table 1.

Table 1: Mean percentages of correct responses for the ABX1, ABX2 and PMT tasks. ABX total collapses responses over time. Responses are divided into the two sonority conditions and sequences.

	ABX1	ABX2	ABX total	PMT
Sonority plateau (SP)	71.3	80.0	75.6	80.0
Sonority fall (SF)	77.5	80.0	78.8	75.8
CC-sequence in target word	67.5	80.0	73.8	77.1
CəC-sequence in target word	81.3	80.0	80.6	78.8
Total	74.3	80.0	77.2	77.9

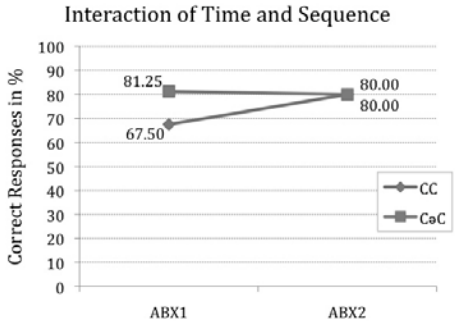
ANOVAs showed a significant main effect of time of measurement (ABX1/ABX2) $[F(1,14) = 5.06, p < 0.05]$ and of sequence (CC/CəC) $[F(1,14) = 5.7, p < 0.05]$. Sonority (SP/SF) as a main effect was not significant ($p > 0.05$). Significant interactions were between time and sequence $[F(1,14) = 7.8, p < 0.05]$ and sonority (SP/SF) and sequence $[F(1,14) = 6.17, p < 0.05]$. All other interactions were non-significant ($p > 0.05$).

Main Effect: Time
Performance before the PMT (ABX1) differed significantly from performance after the PMT (ABX2) $[F(1,14) = 5.06, p < 0.05]$. Table 1 sums up the mean values of correct responses for the various conditions; performance was generally better in ABX2, i.e., after the contrast training. Collapsed across all conditions, performance before the PMT was 74.3% correct responses compared to 80% after the PMT.

Main Effect: Sequence
Participants' performance varied significantly depending on whether the target word contained the illicit consonant cluster or the sequence repaired with an epenthetic schwa $[F(1,14) = 5.7, p < 0.05]$. Collapsed across all other conditions, performance was significantly better for epenthetic target words (80.6% correct answers) than for CC-onset target words (73.8%; see Table 1).

Interaction: Time and Sequence
The interaction of time and sequence was significant $[F(1,14) = 7.8, p < 0.05]$ and shows that participants benefited most from the contrast training in the CC onset target word condition: the number of correct responses increased significantly for CC onsets after the PMT (in ABX2), whereas the performance on CəC onsets was already high before the contrast training (ABX1).

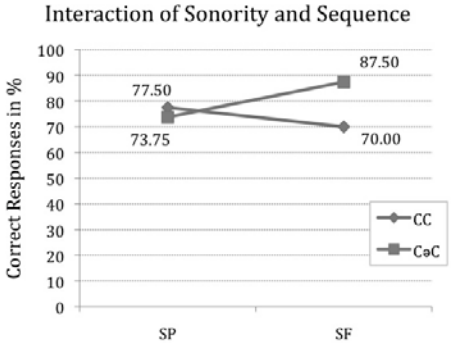
Figure 5: Mean percentages of correct responses for the two target word sequences (CC/CəC onsets) before and after the PMT (ABX1/ABX2).



Interaction: Sonority and Sequence

The interaction between sonority and sequence was also significant [$F(1, 14) = 6.17$, $p < 0.05$]. Sequence interacted with sonority such that the number of correct responses for sequences involving a sonority fall was higher when a schwa was present, but the number of correct responses for sonority falls was lower than for sonority plateaus where schwa was absent. However, it should be borne in mind that there are technically no sonority falls or plateaus in onset sequences involving a schwa, as the schwa breaks up the onset cluster. That is to say, this interaction should properly be interpreted as CəC onset clusters being identified correctly more often when their non-epenthetic counterpart would have constituted a sonority fall than in the condition where their non-epenthetic counterpart would have constituted a sonority plateau. Illegal onset clusters, on the other hand, were perceived correctly more often in the SP condition than in the SF condition.

Figure 6: Mean percentages of correct responses for the two target word sequences (CC/CəC onsets) in the two sonority conditions (SP/SF).



There were no significant main effects and no significant interactions in the PMT; performance was high overall (see Table 1). This result is not necessarily surprising,

since all responses were recorded after training on the difficult contrast had occurred. Rather, particularly in light of the difference in performance between ABX1 and ABX2, it confirms that training had an effect.

3.2 Japanese group

The results for the Japanese-speaking test subjects are summarized in Table 2.

Table 2: Mean percentages of correct responses for the ABX1, ABX2 and PMT tasks. ABX total collapses responses over time.

	ABX1	ABX2	ABX total	PMT
Unfavorable context for devoicing (UC)	85.6	93.8	89.7	92.5
Favorable context for devoicing (FC)	90.0	90.0	90.0	89.4
CC-sequence in target word	87.5	92.5	90.0	90.6
CuC-sequence in target word	88.1	91.3	89.7	91.3
Total	87.8	91.9	89.9	90.9

Performance on the ABX tasks was very good and uniform overall. There were neither any significant main effects nor any significant interactions in the Japanese data analysis ($p > 0.05$).

In the PMT, there were no significant main effects and no significant interactions. The responses in this task represent a stage after training on the crucial contrast has occurred, much like ABX2. However, since the PMT lacks a direct comparison to a comparable task prior to training, the comparison of ABX1 and ABX2 is more informative.

4 Discussion

The results of the English listener group are consistent with previous studies, e.g., Davidson et al. (2007). First, we confirmed that training with a minimal pair contrast helps draw attention to the crucial phonetic detail and improves performance. Second, performance before training is worse when the target word contains an illegal onset cluster compared to its epenthetic counterpart. Third, performance before training is worse for CC onsets when these constitute a sonority fall compared to a sonority plateau.

Although the statistical analysis of the Japanese listener group yielded no significant results, the results are nevertheless enlightening and warrant discussion of several issues, especially the points where this study's results do not replicate those of Dupoux et al. The following issues will be discussed in turn: (i) high vowel devoicing as a factor in vowel epenthesis; (ii) the exceptionally high rate of correct responses compared to what Dupoux and colleagues found (e.g., Dupoux et al. 1999, 2000, 2011); (iii) the robustness of perceptual epenthesis.

4.1 High vowel devoicing

As discussed in Section 1, previous studies of perceptual epenthesis in Japanese (e.g., Dupoux et al. 1999, 2001, 2011) failed to adequately counterbalance experimental stimuli for contexts that were (dis)favorable for high vowel devoicing. Dupoux et al. (2001) briefly discuss reasons why high-vowel devoicing cannot be the sole basis for the epenthesis effect, and while this may be correct, there is an indirect mechanism by which it may be responsible.

Dupoux et al. (2001) point out that if high-vowel devoicing were the sole basis for the epenthesis effect, one would expect to find a reasonable amount of /i/-epenthesis. And indeed, as Dupoux et al. (2001, 2011) demonstrate, /i/ can be epenthesized perceptually under very limited circumstances, too. However, /u/ seems to be the favored epenthetic vowel in Japanese in the absence of coarticulatory cues for /i/. Specifically, two items caused identification of /i/ instead of /u/ in Dupoux et al.'s (2001) study: *rekshi* and *rikshi*. *Rekishi* 'history' and *rikishi* 'sumo wrestler, strong man' are both words in Japanese, but the other items in that set, which would all yield a valid word with a vowel other than /u/ inserted, did not trigger epenthetic percepts other than /u/. Dupoux et al. (2001) suggested therefore that pre-lexical effects may be at work. Moreover, Shinohara (1997) claims that the extant cases of /i/-epenthesis in Japanese are not productive. Where they do occur in loanwords, they are mostly in the context of a voiceless stop and fricative, as for example in 'textile' → *tekisutairo*, which is the same context as in the stimuli in question in Dupoux et al. (2001). It can be argued that this is not true illusory epenthesis: rather, in the transition from [k] to [s] or [ʃ], there is a brief interval between the release of the stop and the formation of the constriction during which the air is pressed through an oral configuration which is close to the one employed for [i]. Other stop/fricative clusters can yield similar transitions depending on their respective places of articulation. Japanese listeners may interpret these transitions as an intended vowel, especially when top-down influences such as phonotactic knowledge reinforce this interpretation. The /i/-percept may be further influenced by the surrounding (mid-)close vowels coloring the transition (cf. Hawkins 2010).

The most convincing argument brought forward by Dupoux et al. (2001) against high vowel devoicing—and thus possibly against phonetics as the basis for perceptual epenthesis—is their finding during *post hoc* analysis that there were not significantly more /u/-epenthesis responses in contexts that favor high-vowel devoicing (i.e., between two voiceless obstruents) versus contexts that do not allow high vowel devoicing: voiceless Cs do not seem to provide the only context in which listeners accept a vowel realization that is close to zero. However, Dupoux et al. (2011) report in a footnote that *post hoc* analyses of their results showed that more cases of epenthesis were found in environments favoring high vowel devoicing [$F(1, 26) = 12.6, p < 0.001$]. Despite this significance, the authors did not pursue the effect further. Moreover, they did not take it to indicate that high vowel devoicing may lead listeners to accept CØ as underlyingly C/u/. On the contrary, they expected that those items should yield less epenthesis, since high vowels may be devoiced to the point of deletion, and therefore Japanese listeners should have experience with the resultant consonant clusters.

The present study's results do not confirm Dupoux et al.'s (2011) finding, as no significant effect for environment (UC/FC) was found. However, there was a

trend concerning the interaction between time of measurement (ABX1/ABX2) and environment which went in the opposite direction: the UC condition was the one in which performance was the lowest initially (85.6%) and where listeners benefited from training the most (93.8% in ABX2); performance in the FC condition was very high to begin with (see Table 2) and remained high. Since the present study—analogous to Dupoux et al. (1999)—only tested ten Japanese participants and used eight stimulus pairs per environment which were repeated only once, the data points collected per condition are relatively few. It is possible that this trend would become significant in a larger study.

These conflicting results between Dupoux et al. (2001, 2011) and the present study indicate that a vowel's ability to devoice in a voiceless C_C environment is not the only factor affecting the epenthetic percept. Phonology tells listeners that a vowel should be present in the C_C environment, and since this expectation is relatively strong, only the slightest of acoustic cues is needed to license the percept of a vowel. Massaro and Cohen (1983) observed that in ambiguous contexts, listeners need more convincing acoustic evidence for the illicit sequence than for the licit one. Since the illicit sequence in this case is CØ, it is easier to find weak evidence for the faint presence of something vowel-like in that position than strong evidence that another C follows the first C, especially if Japanese listeners are used to interpreting even frication noise as a (devoiced) vowel. Thus, noise from neighboring fricatives or stop bursts may be sufficient to trigger the percept of a voiceless vowel. Likewise, voicing in neighboring segments could constitute sufficient acoustic evidence to license perception of a very short vowel. Under this view, the characteristics of Japanese /u/—being the shortest and the least sonorous, allowing for most formant variability, and having the strongest propensity to devoice—could bias listeners to accept even very weak acoustic cues as hinting at the presence of this vowel. Thus, high vowel devoicing in voiceless environments is not directly responsible for vowel epenthesis but indirectly affects illusory vowel perception in FC and UC conditions because it is part of listeners' experience with that vowel.

4.2 Performance of Japanese listeners in ABX tasks

The relatively high rates of correct responses given by the Japanese group in all tasks and all conditions in the present study stand in striking contrast to previous research concluding that Japanese native speakers have problems perceiving the *ebzo/ebuzo* contrast reliably. In this section we review this research in order to determine what the reasons for the discrepancy might be.

Dupoux et al. (1999) asked Japanese listeners to indicate whether VCCV/VCuCV-type stimuli taken from a continuum ranging from 'no vowel' to a full vowel contained a vowel; the continuum was created from VCuCV stimuli uttered by a Japanese speaker. Even in the no-vowel condition, listeners judged a vowel to be present in ∼70% of the tokens. A second continuum was created with a French speaker, this time also including a true zero condition (i.e., containing naturally produced CC clusters, rather than CuC stimuli from which the vowel was digitally removed). Again, Japanese listeners perceived a vowel ∼60% of the time in both sets of VCCV stimuli. French listeners did not perceive vowels in these conditions. In an ABX task where X was spoken by a different voice from A and B, Japanese listeners matched X correctly 68%

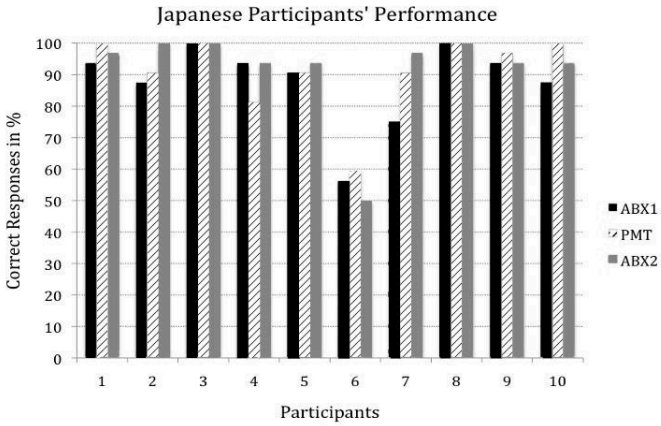
of the time. Repeating the experiment but adding a same-speaker condition yielded correct responses 80.9% of the time for the different-speaker condition and 86.3% for the same-speaker condition.

Dupoux et al.'s (2001) transcription task revealed that Japanese listeners judged /u/ to be present in the C-cluster in CVCCV items 77% of the time when uttered by a Japanese speaker and 62% when produced by a French speaker. In a lexical decision task, CVCCV items that become a real word through /u/-insertion between C_C were judged to be words 71.4% of the time for the Japanese speaker and 70.4% for the French speaker, whereas items that required a different vowel to become a valid word were judged to be words 8% of the time for the Japanese speaker and 18.7% for the French speaker.

Dupoux et al.'s (2011) classification task had Japanese listeners judge whether a vowel was present in stimuli taken from a VCCV-VCuCV continuum spoken by a French speaker. Listeners judged a /u/ to be present 72.4% of the time in the naturally produced VCCV stimuli and 63% of the time in VCCV-items where the /u/ was removed digitally. In an ABX task comparing VCCV and VCuCV produced by French speakers and involving a speaker change, Japanese listeners matched X correctly 58.8% of the time.

The figure of most interest in relation to the present study is the percentage of correct responses reported in Dupoux et al.'s (1999) ABX task with speaker change, since the present study sought to replicate that experiment as closely as possible. Our study's 87.8% correct responses even prior to training stands in stark contrast to Dupoux et al.'s (1999) 68% in their Experiment 3.

Figure 7: Individual performance by the Japanese participants in all three experiments (ABX1, PMT, ABX2).



Both this study and Experiment 3 in Dupoux et al. (1999) used Japanese participants who were raised monolingually in Japan but later moved outside the country. Dupoux et al. (1999) investigated foreign language proficiency as a factor and concluded that not even the performance of the most proficient speakers—an English teacher, a student of phonetics, and two university students, all living in France or the United

States—deviated from that of the rest of the participants.[10] Similarly, most of the speakers in the current study fell within the same range of performance, irrespective of the duration of their stay in a foreign country. In fact, only one participant deviated noticeably from the rest in her overall performance (participant 6 in Figure 7). Performance on the ABX tasks would have been even closer to that of Dupoux et al.'s French control group if participant 6 had been excluded from the analysis: 91.32% in ABX1 and 96.53% in ABX2 vs. 94.2% for the French control group.

The significantly less frequent perception of a /u/ in VCCV-stimuli in the present study versus Dupoux et al.'s (1999) Experiments 1 and 2 and Dupoux et al.'s (2011) Experiment 1 could be attributed to differences in the experimental design. Dupoux et al.'s (1999) Experiment 1 presented 50 VCuCV-stimuli that contained at least two glottal pulses, ranging up to a full vowel (approximately 91–118 ms; we refer to these as "/u/ stimuli" in the following discussion), pitted against a mere 10 stimuli which were deemed to contain "*little* or no vowel" (1999:1570) (emphasis added) and no betraying transitions. However, all stimuli were digitally created from ten original recordings (using ten different vowel/consonant combinations), so listeners were exposed to six stimuli per token which were identical except for the duration of the vowel. If, as discussed earlier, Japanese listeners are particularly sensitive to any cues that might license the percept of /u/, it becomes clear why Japanese listeners might have been biased to perceive /u/s even in the cluster-condition. The fact that /u/-responses jumped from just over 70% in the 0 ms condition to almost 90% in the 18 ms condition confirms that even just two glottal pulses suffice to validate a robust /u/-percept for Japanese speakers.

Japanese listeners faced with these stimuli may have become desensitized to the absence of /u/ and/or might have shifted the perceptual boundary towards accepting more variations of VCuCV as containing an intended /u/, consistent with Massaro and Cohen's (1983) finding that listeners need more evidence to believe they heard an illicit sequence than to believe that what they heard was in line with their native phonotactics. In a way, phonotactics set expectations regarding what structures/sequences to expect when listening to a particular language. Behavioral experiments, however, can create their own expectations: for example, Dell et al. (2000) found that participants' production errors adhered to the phonotactics of a toy grammar they had been taught as part of the experiment—native phonotactics were overridden. Therefore, the experimental design could have caused Japanese listeners to expect more stimuli containing /u/ after they were exposed to the first trials, from which it became clear that stimuli clearly containing a /u/ outnumbered the cluster stimuli by a ratio of 5:1.[11] French, on the other hand, allows both VCuCV and VCCV structures, so French listeners would require stronger cues for the presence of a vowel and would not experience such a strong expectation bias in favor of perceptual epenthesis.

10 Cf. also Moreton (2002: 67), who concludes that "listeners' acceptance of [bw] was not increased by up to 9 years of explicit training in languages in which [bw pw] onsets are common."

11 Compare McQuade's (1981) finding that pseudohomophone effects are affected by the proportion of pseudohomophones in the stimulus set, and the findings of Coltheart et al. (1991) and Jared and Seidenberg (1991) that homophony effects are affected by the percentage of homophones in the stimulus set.

The same effect may be at work in Dupoux et al.'s (1999) Experiment 2, in which 50 /u/ stimuli were pitted against 10 artificial cluster stimuli, 10 natural cluster stimuli and 10 /i/ stimuli. Due to the increase of non-/u/ stimuli, the effect might be weaker here, as reflected in the slightly decreased percentage of /u/-responses (∼60% in the cluster-conditions). Dupoux et al. (2011) used a similar design, so the same explanation holds.

Dupoux et al.'s (1999) ABX tasks tested whether participants were sensitive to an *ebuzo/ebuuzo* distinction, in effect evaluating whether participants were able to distinguish a relatively long vowel from what they might perceive as a shorter vowel, not whether they were able to distinguish whether a vowel was present or absent. In performing this task, the Japanese listeners could call upon their experience with /u/, their sensitivity to fine phonetic cues, and the fact that Dupoux et al. (1999) used Japanese speakers for A, B and X. Of particular relevance is the method by which the stimuli were constructed; Dupoux et al. note that the "two Japanese speakers [. . .] could not be prevented from inserting a short vowel [u] within the consonant clusters in some of the *ebzo* stimuli" and that "the vocalic part was progressively removed until a French listener found that he or she could no longer hear the [u] vowel" (1999:1573). This by no means entails that a Japanese speaker would also judge there to be no /u/. Further, coarticulatory cues may have still been present after the removal of part(s) of the vowel. Studies such as Dupoux et al. (2001) confirm that while coarticulatory cues might not be the source of /u/-epenthesis, they may nevertheless enhance it: in the transcription task, French recordings elicited 10% more cluster responses ($p < 0.001$). Overall, Dupoux et al.'s (1999) ABX tasks compared (i) what Japanese listeners might class as VCuCV containing very subtle cues for a /u/ and (ii) VCuCV containing a full /u/, whereas the present study—using English speakers—might be closer to comparing VCCV with VCuCV.

Since Japanese has no phonological contrast between 'barely perceptible' and 'full' /u/s, all of these tokens would be mapped to the same category. What's more, the vowel that is actually encountered in Japanese is phonetically [ɯ]; thus, both the full vowel used in Dupoux et al. (1999) and the faint vowel perceived in VCCV-stimuli in the same study both represent non-prototypical exemplars of this phonological category. Nevertheless, it may be that these exemplars sound quite distinct to Japanese listeners, since the language has a phonemic vowel length contrast. As Kuhl and Iverson (1995) have shown, discriminability between two exemplars is decreased when both are close to the category prototype. However, it may also be the case that two bad exemplars from very different areas of the perceptual space encompassed by a category are also distinguishable. Specifically, [u] and [u̥ɯ] could represent such exemplars. We hypothesize that they may be distinguishable at an early stage of processing, but may be conflated once they are mapped onto a phonological form. This departs from the Phonetic Decoder model's assumption that such conflation already happens at the stage where the acoustic signal is mapped onto surface phonetic categories, not during the mapping onto the phonological underlying form.

The analysis we propose implies that the acoustic and perceptual space for the phonological /u/ category is relatively large for Japanese speakers and thus allows for a great deal of variation. This would also explain why non-words in the /u/-set of

Dupoux et al. (2001) (e.g., *sokdo*) were so frequently identified as words[12] (*sokudo*), yet at the same time, this account does not entail that listeners are unable to hear the difference between *sokdo* and *sokudo*.

We suggest that the high error rates in the prior studies discussed here may be largely due to the issues highlighted above. Moreover, ABX tasks are very cognitively demanding as they require listeners to build representations of three very similar words and remember the exact sequence in which they were heard. Even if Japanese listeners are able to perceive the difference between *ebzo* and *ebuzo*, it is still possible that some errors were made since these stimuli map onto the same underlying form. More errors were likely made in Dupoux et al.'s (1999) ABX tasks than in the present study's ABX tasks because Dupoux et al.'s study had the additional confounding factors discussed in this section, leading to a stronger /u/ percept in the VCCV stimuli.

In sum, the responses in the previous studies may have been biased towards the conclusion that Japanese listeners cannot distinguish VCCV and VCuCV-type stimuli. Some studies claimed to test whether Japanese listeners differentiate these two (e.g., Experiment 1 and 2 in Dupoux et al. 1999; Dupoux et al. 2001; Experiment 1 in Dupoux et al. 2011) when in fact they only tested whether Japanese listeners stipulated the presence of a vowel or not. This does not entail that listeners do not hear a difference between VCCV, where they might perceive a faint/short vowel, and VCuCV with a full vowel. The ABX task designs came closer to examining the real issue, but still suffered from design flaws that we hypothesize put listeners at a disadvantage (how the stimuli were created and the ratios in which they were presented, using Japanese speakers, and not controlling for subtle cues that could bias Japanese listeners).

On the other hand, even though Japanese listeners may well be able to perceptually distinguish VCCV and VCuCV, this does not exclude the possibility that they interpret VCCV as containing an epenthetic vowel other than /u/. We have suggested that the acoustic and perceptual space for /u/ may be exceptionally large for Japanese speakers and that they are accustomed to accepting very subtle phonetic cues as evidence for the presence of a /u/. This, of course, may be aided by phonology: if phonotactics cause the listener to expect a (certain type of) segment in a certain location, s/he might be more likely to reinterpret the cues in the signal to match the expectation. That is to say, the vowel perceived by Japanese speakers in these experiments may not be entirely illusory, but rather motivated by both the acoustic signal and the phonology.

Our view is supported by independent evidence. For example, a neutralization rule in Korean renders word-final /s/ as [t] (Kim and Jongman 1996). English loanwords, however, undergo epenthesis rather than neutralization (Kenstowicz and Sohn 2001), even though Korean has no native words ending in [ɨ] (Yoonjung Kang, personal communication with Sharon Peperkamp; see Peperkamp 2005), resulting in

12 Note that one could argue that a Ganong effect (Ganong 1980) shifts the responses towards "word." However, /u/ is also perceived in stimuli that remain nonwords even with the /u/; a Ganong effect should result in a significantly higher /u/-response for the /u/-set than for the non-/u/-set. There were only 4% more /u/-responses for the Japanese speaker and 6% more for the French speaker.

borrowings such as [posɨ] 'boss.' We suggest that this epenthesis arises because part of the fricative is interpreted as an unstressed, short devoiced vowel; if this is correct, then the Japanese interpretation of frication as intended /i/ is not an isolated case. More support comes from Uyghur, where devoiced /i/ sounds like [ʃ̟] (Kaisse 1992). Other transitions between consonants may well be interpreted as containing acoustic cues to the presence of a /u/. The broader idea of languages exploiting the speech signal differently and segmenting the stream differently is also supported by informal reports of Spanish learners of English who, even after years of experience, insist that they hear an [e] before /sC/-clusters (mentioned in Dupoux et al. 1999): part of the frication may be interpreted as a voiceless vowel. Undoubtedly, this interpretation is partly motivated by the knowledge that Spanish has no initial /sC/ clusters; they are always preceded by a vowel, e.g., *escuela* 'school,' *estructura* 'structure.' Again, the percept of a vowel may be phonologically motivated but phonetically validated.

To put it succinctly, we argue here that speech processing is a process of weighing the odds of what is likely within the frame of interpretation—which usually is one's native language—combined with what is licensed by the signal according to that language's rules about interpreting or mapping acoustic cues. Naturally, the trust in the acoustic signal varies according to how clear or degraded it is (cf. Massaro and Cohen 1983). On this view, a training task such as the PMT affects speech processing by teaching listeners to alter their segmentation and interpretation of the speech signal: "meaning and experience affect the interpretation of fine phonetic detail" (Davidson et al. 2007: 3707).

4.3 Consequences for models of speech perception

This study has found that Japanese listeners are quite capable of differentiating VCCV and VCuCV stimuli even without training, contrary to the influential claim of Dupoux et al. (1999, 2001, 2011) and Dehaene-Lambertz et al. (2000). We have also suggested that this does not necessarily entail that CC-sequences are perceived as such; the presence of a short epenthetic vowel between these consonants may be perceived. However, we have argued that this percept is not entirely unmotivated by the acoustic signal. When phonotactics strongly suggest that there should be a vowel in a given position, listeners may be willing to accept unusually subtle cues in the speech signal in order to match that interpretation. In the case of Japanese, listeners' experience with an exceptionally variable /u/ may push the acceptance of small cues to an extreme; voicing in surrounding segments, transitional frication, stop bursts, etc. may suffice to legitimize the percept of a vowel. Thus, the epenthetic percept in such cases is not truly illusory: it is phonetically motivated and phonologically licensed.

In normal speech situations, listeners expect the speech signal to conform to the phonotactics of the language being spoken. Attention is focused on the larger picture, i.e., on successful communication instead of on phoneme monitoring or acoustic detail. Thus, sequences that violate expectation and/or native phonotactics may go unnoticed, or rather, become perceptually repaired, especially when the speech signal offers cues that can be interpreted accordingly (cf. Samuel 1987; Bashford et al. 1992; Repp 1992 on perceptual displacement). And since there is no one-to-one mapping between the speech signal and abstract phonological categories, there is always room for interpretation. Trading relations ensure that while one phoneme can be cued by

several auditory and visual properties, they need not all be present together. Conversely, properties that typically cue a certain phoneme may not always relate to that phoneme. Furthermore, the acoustic signal is often degraded, i.e., portions are masked, reduced, unclear, etc., and many properties vary widely on both an inter- and intra-speaker basis.

Given this fundamental discrepancy between the acoustic signal and the abstraction necessary for communication to succeed, it is not surprising that listeners allow expectations to limit their search space when attempting to map a continuous percept to a discrete phonemic string. As Hawkins (2010: 77) puts it, "[w]hen the listening situation is normal, in that the listener tries to make sense of the speech signal, and the signal is globally consistent with a particular interpretation, then a part of the signal that provides only low-certainty evidence for an expected distinctive feature can be sufficient to trigger construction of the percept for that feature, in the absence of strong negative evidence." This can be taken to describe the Japanese epenthetic effect. Phonotactic knowledge influences the listener to expect a V after a C. The expectation that the sequence adheres to native phonotactics can be strong enough to license the interpretation of transitions as an intended vowel. We propose that expectation is another type of filter that emerges with language experience—cf. Hume and Mailhot's (2013) notion of 'expectedness,' which they define as the inverse of surprisal, as a source of phonological change.

We suggest that speech perception in this case proceeds as follows, in general terms. First, the perceived acoustic signal is segmented into syllables. This process involves wo key components: syllable position identity (Kabak and Idsardi 2007) and sonority profiles (Berent et al. 2007). Aberrant forms trigger perceptual reorganization of the acoustic cues so that segments appear in allowed positions. These two syllable structure guidelines may conflict, and in this case we suggest it is possible that information about the sonority profile of syllable structure is subordinated to information about possible positional identities of single phonemes. In general, listeners may latch onto whatever segmentation information is the most salient in their language.

Once word forms have been picked out of the speech stream, the words themselves are analyzed. The Phonetic Decoder model assumes whole-word analysis (Peperkamp and Dupoux 2003, Peperkamp 2005). Although speech processing can begin even before a whole word is perceived (e.g., Reinisch et al. 2010), once word boundaries are identified, the analysis can be revised. In this way, speech perception is an incremental process (cf. Heinrich et al. 2010, Hawkins 2010). Our proposed inclusion of expectations about the speech signal could be formulated to include probabilities of likely sequences, taking into account phonotactics and the cues in the speech signal. A good example of how expectations and probabilities come into play is provided by the findings by Monahan et al. (2009), which were discussed in Section 1. They tested the extent of phonology's influence on speech perception and found that phonological environments which do not license /u/-epenthesis in normal contexts do not admit perceptual epenthesis of /u/ either. Even in an interconsonantal position where simple phonotactics should lead listeners to expect a vowel in that position, neither /u/, which phonetics would favor, nor /o/, which phonology would favor, is heard in [etma]. Monahan et al. (2009) suggest that something like [etVma] is heard, and that

this percept is distinct from both [etoma] and [etuma]; however, it may instead be that listeners simply heard [etma]. Whichever the case, this example shows that the acoustic signal has to match the expectation to a certain degree. In cases like these, when the acoustic signal provides strong evidence against phonological expectations, an illicit sequence is nevertheless perceived (cf. Best et al. 1988).

The notion of 'expectation' discussed here is crucial for allowing variation in interpretation of the speech signal according to task demands and language ability, and may play a role in explaining the different outcomes of different studies of Japanese /u/-epenthesis. It is possible that, if the participants identified that it was a Japanese voice speaking the stimuli, they would expect conformity to Japanese phonotactics in parts of Dupoux et al. (1999), in Dupoux et al. (2001) and Dehaene-Lambertz et al. (2000). Further, telling participants they would be hearing Japanese nonwords, as in Dehaene-Lambertz et al.'s (2000) ERP study, could have influenced participants towards expecting Japanese phonetics and phonotactics. A study by Werker and Tees (1984b) demonstrates the powerful influence of expectations on perception: When English adult listeners were tested on a foreign contrast from Nthlakampx, /k'i/-/q'i/, they failed to make the distinction reliably. However, when they were played only the ejective part of that syllable and were told that they were listening to water dropping into a bucket, listeners were able to distinguish the sounds reliably. Native phonetic categories presumably interfered with faithful perception when the listeners classified the sounds as speech, whereas in a non-speech context they were unconstrained by these expectations.

5 Conclusions

In this study we investigated why Japanese listeners have been reported to fail at differentiating *ebzo* and *ebuzo*-type stimuli (Dupoux et al. 1999 et seq.; Dehaene-Lambertz et al. 2000), while English listeners succeed at distinguishing *lbif* and *lebif*-type stimuli (Davidson et al. 2007) after minimal pair training. We reported that both the Japanese and English studies suffer from design flaws, compromising their ability to test the contrasts under investigation and to evaluate the effects of training.

The present study produced evidence that Japanese listeners are in fact able to distinguish *ebzo*-type stimuli from *ebuzo*-type stimuli even without training on this contrast (cf. Werker and Tees 1984b; Polka 1995; Hale and Kissock 1998, 2007). However, even this strong initial performance distinguishing VCCV from VCuCV stimuli could be improved through a picture-matching task designed to highlight minimal pair contrasts, similar to the one employed by Davidson et al. (2007) in their English study. We argue that Japanese listeners' experience with the high variability of /u/ may be a factor in their readiness to accept very faint cues in the acoustic signal, e.g., voicing from surrounding segments or frication from segment transitions, as signaling /u/.

Prior to the present study, the hypothesis that high vowel devoicing between voiceless obstruents plays a role in perceptual epenthesis was never systematically examined. We argue that while this phenomenon is not the sole basis for perceptual epenthesis, experience with the high variability of /u/ due to devoicing adds to the acceptance of faint cues as referring to /u/. This claim is supported by the observation

that perceptual epenthesis in the present study as well as in others (Dupoux et al. 1999, 2001, 2011; Dehaene-Lamertz et al. 2000) was not restricted to contexts that favor high vowel devoicing. As a result, we hypothesized that the perceptual and acoustic space for Japanese /u/, which is normally realized as [ɯ], is exceptionally large and encompasses [ɯ̟] (or something even closer to [Ø]) as well as [u]. Since two deviant exemplars within a single category are easier to distinguish than two near-prototypical exemplars (Kuhl and Iverson 1995), it is likely that Japanese listeners generally have no problem discriminating them. Nevertheless, in situations where cognitive load is particularly high, such as in ABX tasks where the perceived order of similar items is crucial, confusion is possible as both exemplars map onto the same phonological category.

Overall, we argue that in cases where there is perceived epenthesis in VCCV stimuli, the vowel is not an illusion, but rather is both phonetically and phonologically motivated: phonotactic expectations encourage this interpretation of subtle phonetic cues in the acoustic signal. Nevertheless, if the phonetic cues in the speech stream are wildly incompatible with the interpretation suggested by phonotactics, the expectations made on the basis of the phonotactics can be overridden, as Monahan et al. (2009) show. The Phonetic Decoder model cannot account for the present study's results—or for Dupoux et al.'s (1999) results in Experiment 4. We argue instead for a model of adult speech perception that includes expectations based on linguistic experience, while still allowing access to phonetic detail if a particular task requires it. Within such a model, it is still possible to maintain Best's (1994) assumption that foreign phones that are similar to native ones are assimilated into L1 categories, making the perception of a contrast between two L2 phones that fall within the same L1 category much more difficult. This is in line with Kuhl and Iverson's (1995) Perceptual Magnet Effect. Nevertheless, such contrasts may still be possible to perceive, even for untrained speakers, and the ability to distinguish between L2 phones can be improved with training, as our study confirmed.

References

Bashford, James A., Keri R. Riener, and Richard M. Warren. 1992. Increasing the intelligibility of speech through multiple phonemic restorations. *Perception and Psychophysics* 51:211–217.

Beckman, Mary E. 1982. Segmental duration and the "mora" in Japanese. *Phonetica* 39:113–135.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104:591–630.

Best, Catherine, Barbara Morrongiello, and Rick Robson. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics* 29:191–211.

Best, Catherine T. 1994. The emergence of native-language phonological influence in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Language*, ed. Judith C. Goodman and Howard C. Nusbaum, 167–224. Cambridge, MA: MIT Press.

Best, Catherine T., Gerald W. McRoberts, and Nomathemba M. Sithole. 1988. Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 13:345–360.

Blumstein, Sheila E., and Kenneth N. Stevens. 1980. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America* 67:648–662.

Box, George E.P. 1953. Non-normality and tests on variances. *Biometrika* 40:318–335.

Clements, G. Nick, and Samuel Jay Keyser. 1983. *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press.

Coltheart, Veronika, S. E. Avons, Jacqueline Masterson, and Veronica J. Laxon. 1991. The role of assembled phonology in reading comprehension. *Memory and Cognition* 19:387–400.

Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24:261–286.

Davidson, Lisa, Jason Shaw, and Tuuli Adams. 2007. The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America* 122:3697–3709.

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux, and A. Gout. 2000. Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience* 12(4):635–647.

Dell, Gary S., Kristopher D. Reed, David R. Adams, and Antje S. Meyer. 2000. Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26:1355–1367.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25:1568–1578.

Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi, and Jacques Mehler. 2001. New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16:491–505.

Dupoux, Emmanuel, Erika Parlato, Sónia Frota, Yuki Hirose, and Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64:199–210.

Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Evolutionary Psychology* 6:110–125.

Hale, Mark, and Madelyn Kissock. 1998. Nonphonological triggers for renewed access to phonetic perception. In *Proceedings of the GALA '97 Conference on Language Acquisition*, ed. Antonella Sorace, Caroline Heycock, and Richard Shillcock, 229–234. Edinburgh: University of Edinburgh.

Hale, Mark, and Madelyn Kissock. 2007. The phonetics-phonology interface and the acquisition of perseverant underspecification. In *The Oxford Handbook of Linguistic Interfaces*, ed. Charles Reiss and Gillian Ramchand, 81–101. Oxford: Oxford University Press.

Hale, Mark, Madelyn Kissock, and Charles Reiss. 2007. Microvariation, variation, and the features of universal grammar. *Lingua* 117(4):645–665.

Hale, Mark, and Charles Reiss. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.

Hallé, Pierre A., Juan Segui, Ulrich H. Frauenfelder, and Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24:592–608.

Hawkins, Sarah. 2010. Phonological features, auditory objects, and illusions. *Journal of Phonetics* 38:60–89.

Heinrich, Antje, Yvonne Flory, and Sarah Hawkins. 2010. Influence of English r-resonances on intelligbility of speech in noise for native English and German listeners. *Speech Communication* 52:1038–1055.

Hume, Elizabeth, and Frédéric Mailhot. 2013. Entropy and surprisal in phonologization and language change. In *Origins of Sound Change*, ed. Alan C. L. Yu, 29–47. Oxford: Oxford University Press.

Itô, Junko, and Armin Mester. 1995. Japanese phonology. In *The Handbook of Phonological Theory*, ed. John A. Goldsmith, 817–838. Oxford: Blackwell.

Jared, Debra, and Mark S. Seidenberg. 1991. Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General* 120:358–394.

Kabak, Baris, and William J. Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50:23–52.

Keating, Patricia, and Marie K. Huffman. 1984. Vowel variation in Japanese. *Phonetica* 41:191–207.

Kenstowicz, Michael, and Hyang-Sook Sohn. 2001. Accentual adaptations in North Kyungsan Korean. In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 239–270. Cambridge, MA: MIT Press.

Kim, H., and A. Jongman. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics* 24(3):295–312.

Kuhl, Patricia K., and Paul Iverson. 1995. Linguistic experience and the "perceptual magnet effect". In *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, ed. Winifred Strange, 121–154. Timonium: York Press.

Liberman, Alvin. 1982. On finding that speech is special. *American Psychologist* 37(2):148–167.

Massaro, Dominic W., and Michael M. Cohen. 1983. Phonological constraints in speech perception. *Perception and Psychophysics* 34:338–348.

McQuade, Debra V. 1981. Variable reliance on phonological information in visual word recognition. *Language and Speech* 24:99–109.

Monahan, Philip J., Eri Takahashi, Chizuru Nakao, and William J. Idsardi. 2009. Not all epenthetic contexts are equal: differential effects in Japanese illusory vowel perception. In *Japanese/Korean Linguistics 17*, ed. Shoishi Iwasaki, Hajime Hoji, Patricia M. Clancy, and Sung-Ock Sohn, 391–405. Stanford: CSLI.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84:55–71.

Pallier, Christophe, L. Bosch, and Núria Sebastián-Galles. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64:B9–B17.

Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptations. In *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society: General Session*, ed. Marc Ettlinger, Nicholas Fleisher, and Mischa Park-Doob, 341–352. Berkeley: Berkeley Linguistics Society.

Peperkamp, Sharon, and Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: The role of perception. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 3–9 August 2003*, ed. Maria-Josep Solé, Daniel Recasens, and Joaquin Romero, 367–370. Adelaide: Causal Productions.

Polka, Linda. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America* 97(2):1286–1296.

Reinisch, Eva, Alexanra Jesse, and James M. McQueen. 2010. Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology* 63(4):772–783.

Remez, Robert, Philip Rubin, David B. Pisoni, and Tom D. Carrell. 1981. Speech perception without traditional speech cues. *Science* 212:947–950.

Repp, Bruno. 1992. Perceptual restoration of a missing speech sound: Auditory induction or illusion? *Perception and Psychophysics* 11:799–813.

Saffran, Jenny R., Janet F. Werker, and Lynne A. Werner. 2006. The infant's auditory world: Hearing, speech, and the beginnings of language. In *Handbook of Child Psychology,* vol. 2: *Cognition, Perception, and Language.* 6th ed., ed. Deanna Kuhn, Robert S. Siegler, William Damon, and Richard M. Lerner, 58–108. Hoboken: Wiley & Sons.

Samuel, Arthur G. 1987. Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language* 26:36–56.

Samuels, Bridget D., and Bert Vaux. 2020. Silence-cued stop perception: Split decisions. In *Bauta: Janne Bondi Johannessen in Memoriam*, ed. Kristin Hagen, Arnstein Hjelde, Karine Stjernholm, and Øystein A. Vangsnes, 393–409. Oslo Studies in Language 11(2), University of Oslo.

Sapir, Edward. 1933. La réalité psychologique des phonèmes. *Journal de Psychologie Normal et Pathologique* 30:247–265.

Shinohara, Shigeko. 1997. Analyse phonologique de l'adaptation japonaise de mots étrangers. Doctoral Dissertation, Université de la Sorbonne Nouvelle, Paris.

Srivastava, A. B. L. 1959. Effects of non-normality on the power of the analysis of variance test. *Biometrika* 46:114–122.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In *The Role of Speech Perception in Phonology*, ed. Elizabeth Hume and Keith Johnson, 219–250. San Diego: Academic Press.

Steriade, Donca. 2009. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In *The Nature of the Word: Studies in Honor of Paul Kiparsky*, ed. Kristin Hanson and Sharon Inkelas, 151–179. Cambridge, MA: MIT Press.

Tremblay, Kelly L., Nina Kraus, Thomas D. Carrell, and Therese McGee. 1997. Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America* 102(6):3762–3773.

Tsuchida, Ayako. 1987. The Phonetics and Phonology of Japanese Vowel Devoicing. Doctoral Dissertation, Cornell University, Ithaca, NY.

Vance, Timothy J. 1987. *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Varden, John K. 1998. On High Vowel Devoicing in Standard Modern Japanese: Implications for Current Phonological Theory. Doctoral Dissertation, University of Washington, Seattle.

Werker, Janet F. 1989. Becoming a native listener: A developmental perspective on human speech perception. *American Scientist* 77:54–59.

Werker, Janet F., and Richard C. Tees. 1984a. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49–63.

Werker, Janet F., and Richard C. Tees. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75(6):1866–1878.