Journal of Memory and Language 64 (2011) 199–210

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Where do illusory vowels come from?

Emmanuel Dupoux [a,b,c,*], Erika Parlato [d], Sonia Frota [e], Yuki Hirose [f], Sharon Peperkamp [b,c]

[a] Ecole des Hautes Etudes en Sciences Sociales, France
[b] Département d'Etudes Cognitives, Ecole Normale Supérieure, France
[c] Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, France
[d] Universidade Federal de Minas Gerais, Brazil
[e] Linguistics Department, Universidade de Lisboa, Portugal
[f] Department of Language and Information Sciences, University of Tokyo, Japan

ARTICLE INFO

Article history:
Received 3 September 2010
revision received 16 December 2010
Available online 31 January 2011

Keywords:
Speech perception
Perceptual epenthesis
Context effects
Phonotactic constraints
Coarticulation

ABSTRACT

Listeners of various languages tend to perceive an illusory vowel inside consonant clusters that are illegal in their native language. Here, we test whether this phenomenon arises after phoneme categorization or rather interacts with it. We assess the perception of illegal consonant clusters in native speakers of Japanese, Brazilian Portuguese, and European Portuguese, three languages that have similar phonological properties, but that differ with respect to both segmental categories and segmental transition probabilities. We manipulate the coarticulatory information present in the consonant clusters, and use a forced choice vowel labeling task (Experiment 1) and an ABX discrimination task (Experiment 2). We find that only Japanese and Brazilian Portuguese listeners show a perceptual epenthesis effect, and, furthermore, that within these participant groups the nature of the perceived epenthetic vowel varies according to the coarticulation cues. These results are consistent with models that integrate phonotactic probabilities within perceptual categorization, and are problematic for two-step models in which the repair of illegal sequences follows that of categorization.

Introduction

Decades of research on speech perception have demonstrated that listeners tend to misperceive sounds that are not part of their native language. The perceptual system is said to become attuned to the native language during the first years of life (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Kuhl et al., 2008; Werker & Tees, 1984). Two types of mechanisms have been proposed to account for this attunement process. According to the first one, nonnative segments that are sufficiently close to a native category are assimilated to that category (Best, 1994). According to the second one, phonetic dimensions that are not used contrastively are processed less efficiently (Iverson et al., 2003; Dupoux, Sebastián-Gallés, Navarete, & Peperkamp, 2008).

Yet, a growing body of observations shows that the influence of the native language on perceptual processes goes beyond the effects described above: segmental sequences that are illegal in a given language tend to be 'repaired' by listeners of that language, often by insertion of a so-called epenthetic segment, such that the resulting sequence is legal. In Japanese, for instance, all consonant sequences except some starting with a nasal are illegal. Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) showed that native speakers of Japanese perceive an illusory, epenthetic, vowel /u/ when they listen to illegal consonant sequences: the nonword /ebzo/ is perceived as /ebuzo/. Insertion of epenthetic vowels inside illegal clusters during perception has also been reported in Korean (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Kabak & Idsardi, 2007) and English (Berent, Lennertz, Smolensky,

* Corresponding author at: Laboratoire de Sciences Cognitives, 29 rue d'Ulm, Pavillon Jardin, 75005 Paris, France. Fax: +33 680581687.
E-mail address: emmanuel.dupoux@gmail.com (E. Dupoux).

0749-596X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.12.004

& Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007).

These effects present a challenge to current psycholinguistic models, in which the basic units for speech recognition are segments or sub-segmental features (TRACE: McClelland & Elman, 1986; Native Language Magnet: Kuhl, 1993; Kuhl et al., 2008; Perceptual Assimilation Model: Best, 1994; Shortlist: Norris, 1994; Norris & McQueen, 2008; Featurally Underspecified Lexicon: Lahiri & Reetz, 2002). In these models, the recognition of one phoneme is essentially independent from that of the adjacent ones.[1] In particular, language-dependant sequential information, i.e. *phonotactics*, is not taken into account, and therefore, perceptual epenthesis effects cannot be explained. There are two possible strategies to modify standard psycholinguistic models to account for perceptual epenthesis. One is to introduce a second processing step, which repairs illegal sequences after categorization takes place. The other one is to have phonotactic constraints interact with categorization within a single processing step.

There are several reasons why one-step models are a priori more likely than two-step models. First of all, the effects of sequential information on speech perception concern not only illegal but also legal clusters. Specifically, frequent sequences of segments (diphones or triphones) are processed faster than infrequent ones in shadowing and same-different matching tasks (Vitevitch & Luce, 1998, 1999; see also Onishi, Chambers, & Fisher, 2002). Second, sequential information interacts with segmental categorization. Massaro and Cohen (1983), for instance, found that the classification of a phonetic continuum is biased by the legality of the sequence in which it appears. Similarly, Pitt and McQueen (1998) reported that identification of ambiguous segments is biased towards structures with higher phonotactic probability (see also Coetzee, 2008). These effects can only be accounted for by models in which segmental categorization and repair of illegal sequences take place simultaneously. Third, a strict two-step model predicts that the repair of illegal phonotactics follows that of illegal segments during on-line processing. However, as far as we know the two types of repair occur equally early: In an ERP study, Dehaene-Lambertz, Dupoux, and Gout (2000) found that Japanese listeners, contrary to French ones, do not elicit a Mismatch Negativity (MMN – a neural component arising 150–250 ms after a deviant stimulus in an oddball paradigm) for sequences in which /ebuzo/ is the standard and /ebzo/ the deviant. This finding complements similar ones concerning the reduction of the MMN for segmental category mismatches in nonnative listeners (Näätänen et al., 1997) and suggests that both illegal segments and illegal sequences are repaired prior to mismatch detection, compatible with a one-step view.[2]

In this paper, we test the empirical validity of one-step approaches by focusing on their core prediction, namely that segmental categorization and phonotactic processing *interact*. Specifically, the epenthetic vowel inserted to repair illegal clusters in perception should not be fixed once and for all across languages, but depend simultaneously on phonotactic probability and the match with the spectral information present in the transition part of the cluster. More precisely, the vowel that is inserted, if any, should be the one that optimizes the product of the sequence probability on the one hand and the acoustic match between the segment and the relevant part of the signal (i.e., the transition between the consonants) on the other hand. One way to test this is to maintain the acoustic signal constant and vary the phonotactic and phonetic properties of the test languages. Another way is to maintain the test language constant and manipulate the phonetic properties of the stimuli. Here, we use both strategies. First, speakers of three languages – Japanese, European Portuguese, and Brazilian Portuguese, the last two being dialectal variants of each other – are tested on the same stimuli. These three languages share a phonotactic constraint against obstruent consonants in syllable codas; obstruent-obstruent and obstruent-nasal clusters like in /ebzo/ and /agnu/, respectively, are therefore not found in any word in the language. However, as we detail below, these languages vary in their phonetic properties in ways that could influence perceptual epenthesis. Thus, we assess perceptual epenthesis across languages as a function of their phonetic properties. Second, we construct different coarticulatory variants of illegal consonant clusters by digitally excising a vowel from in between consonants (e.g. /ebuzo/ → /ebzo/ and /ebizo/ → /ebzo/). Coarticulation traces of the excised vowels on the consonants are predicted to influence the identity of the epenthetic vowel within listeners of the same language. Thus, we also assess perceptual epenthesis within languages as a function of phonetic properties of the stimuli.

Concerning the phonetic properties of the three test languages, both Japanese and Brazilian Portuguese have a process of high vowel devoicing, turning /i/ and /u/ into unvoiced segments in certain environments (for Japanese, see Han, 1962; Vance, 1987; for Brazilian Portuguese, see Camara & Mattoso, 1979; Cristófaro Silva, 1998). This makes /i/ and /u/ good candidates for epenthetic repairs in perception, since these categories already contain exemplars that are phonetically minimized, and they can therefore be matched to the non-vocalic transition between, say, /b/ and /z/ in /ebzo/. Moreover, /u/ is the shortest vowel in Japanese (Han, 1962), and /i/ is the shortest vowel in Brazilian Portuguese (Escudero, Boersma, Schurt Rauber, & Bion, 2009). Therefore, if the epenthetic vowel is the phonetically minimal element of the language, it should be /u/ in Japanese and /i/ in Brazilian Portuguese. In

[1] Several models of phonetic perception integrate context effects like compensation for coarticulation or compensation for assimilation, but view them in terms of language-general processes like vector analysis (Fowler & Smith, 1986), auditory contrast effects (Lotto & Kluender, 1998), or perceptual feature parsing mechanisms (Gow, 2003). These models have difficulties to account for language-specific context effects. Other models introduce context effects via lexical top-down influences (TRACE), or post-perceptual lexical biases (Shortlist). For results showing that perceptual epenthesis is, however, not due to lexical influences, see Dupoux, Pallier, Kakehi, and Mehler (2001).

[2] Of course, this is not a knock-down argument, since a more detailed comparison of MMN latencies, amplitudes and source localization would be necessary to assess whether illegal segments and illegal sequences are processed by the same mechanism.

Japanese listeners, the presence of perceptual /u/[3]-epenthesis has already been extensively documented (Dupoux et al., 1999; Dupoux et al., 2001). In Brazilian Portuguese, the presence of epenthetic /i/ in loanwords strongly suggests the existence of perceptual /i/-epenthesis in listeners of this language, but so far this has not been tested experimentally. European Portuguese presents an interesting contrast with Brazilian Portuguese: The two languages are very similar at the phonological level, but differ phonetically. Specifically, European Portuguese has the same phonotactic structure as Brazilian Portuguese, disallowing obstruent-obstruent and obstruent-nasal clusters in the lexicon. However, it has a phonetic process that optionally deletes unstressed vowels in running speech (Vigário, 2003). As a consequence, coda obstruents are common at the phonetic surface, and sequences like /bz/ have a non-zero probability, (for instance, *obesidade* [obzidad] 'obesity', and *besuntar* [bzuntar] 'smear'). If the type of phonotactics that matters for perception is defined at the level of the surface distribution of sounds (instead of being defined lexically), clusters like /bz/, then, need not be repaired at all in European Portuguese listeners. We would therefore expect little if any vowel epenthesis. If, by contrast, it is defined lexically, European and Brazilian Portuguese should not differ in terms of the amount of perceptual epenthesis.

As to the within-language comparisons, we test the effect of coarticulation by using three variants of each illegal cluster, a neutral one and two coarticulated ones. Neutral clusters are constructed by recording naturally produced stimuli like /ebzo/ in a language in which these stimuli are legal. For these clusters, we predict /u/-epenthesis in Japanese, /i/-epenthesis in Brazilian Portuguese, and little or no epenthesis in European Portuguese listeners. Coarticulated clusters are constructed by excising the vowels /u/ and /i/ from naturally produced stimuli containing /u/ and /i/, respectively, in between consonants (for instance, /ebuzo/, /ebizo/). For both Japanese and Brazilian Portuguese, we predict that although the three types of clusters are identical in terms of their segmental make-up, the coarticulation-based acoustic traces present in the adjacent consonants induce significantly more /i/-epenthesis for /i/-co-articulated clusters and more /u/-epenthesis for /u/-co-articulated clusters compared to neutral clusters.

The predictions of two-step models are quite different. Recall that in these models, illegal sequences are repaired by a process that takes as its input the output of segmental categorization, i.e. the process that transforms the signal into a sequence of discrete segments. Therefore, two-step models have little to say about the pattern of results across languages, except that the variables predicting the presence or absence of epenthesis as well as the nature of the epenthetic vowel should be defined at the level of abstract phonemes and phonological rules, not fine-grained phonetic information. Concerning the within-language comparisons, one-step models predict that vowel epenthesis should *not* be affected by coarticulation, since the output of segmental categorization does not contain coarticulatory information. Hence, the three types of clusters – neutral, /i/-coarticulated and /u/-coarticulated – should give rise to identical patterns of epenthesis within a given language.

In the following, we present the results of two experiments with monolingual participants of Japanese, Brazilian Portuguese, and European Portuguese. As in Dupoux et al. (1999), we use both a vowel categorization task with artificial continua in which vowel duration is manipulated (Experiment 1), and a more implicit ABX discrimination task (Experiment 2). In both experiments, we vary the presence of coarticulation cues in the illegal segment sequences, and compare the perception of naturally produced clusters in disyllables (e.g., /ebzo/), to that of clusters that are obtained by removal of the central vowel out of naturally produced trisyllables (e.g., /ebuzo/ and /ebizo/).

Experiment 1

In this experiment, we tested the perception of illegal clusters in Japanese, European Portuguese, and Brazilian Portuguese listeners in a vowel classification task. Participants listened to two sets of artificial continua: /u/-continua, involving stimuli intermediate between, for instance, /ebuzo/ and /ebzo/, and /i/-continua, with stimuli intermediate between, for instance, /ebizo/ and /ebzo/. They were asked to decide whether the stimuli did or did not contain a vowel between the two consonants, and if yes, which vowel it was: /a/, /e/, /i/, /o/, or /u/. This design was very similar to that used in the first and second experiments of Dupoux et al. (1999). Note that the /ebzo/ endpoints of the two sets of continua make it possible to assess the effect of coarticulation on perceptual epenthesis. Indeed, these stimuli were obtained by the complete removal of the vocalic portion of either /ebuzo/ or /ebizo/. Even though they do not contain any vocalic part, traces of the original /u/ and /i/ can be assumed to be still present in the form of coarticulatory information in the consonants.

Method

Stimuli

Thirteen $V_1C_1C_2V_2$ items were selected, with V_1 and V_2 vowels in the set /a,e,i,o,u/, C_1 a stop consonant, and C_2 a stop or a nasal consonant (see Appendix). For each item, three stimuli were recorded by a male native speaker of French; these stimuli were of the form $V_1C_1C_2V_2$, V_1C_1/u/C_2V_2 and V_1C_1/i/C_2V_2. They all had stress on the first syllable.

In the stimuli that contained a full medial vowel, the mean duration of /u/ was 112.9 ms (SD = 15.3) and that of /i/ was 112.8 ms (SD = 20.3). For each of these stimuli, seven additional stimuli were created by removing successively larger portions of the medial vowel at zero crossings, starting at the endpoints and moving inwards; the left endpoint was defined as the first zero crossing after the burst and the right one as the point in the signal where no vowel

[3] Strictly speaking, the epenthetic vowel is /ɯ/, the Japanese unrounded counterpart of the more common vowel /u/. Here, we note /u/ for convenience, since this is the vowel used in Portuguese. Note also that the stimuli used by Dupoux et al. (1999) contained /u/, not /ɯ/ (and so will the stimuli in the present study). The difficulty that Japanese listeners had to discriminate /ebzo/ from /ebuzo/ shows that they easily confound /u/ and /ɯ/.

formants could be seen or heard (see Dupoux et al., 1999, for details, and Fig. 1 for an illustration). Hence, for each item there was one continuum ranging from $V_1C_1/u/C_2V_2$ to $V_1C_1C_2V_2$ and another one ranging from $V_1C_1/i/C_2V_2$ to $V_1C_1C_2V_2$.

For each item, one additional filler stimulus was recorded by the same speaker; it contained a medial vowel that was either /a/ or /o/. Overall, then, there were 18 stimuli per item: two 8-step continua plus one stimulus with a natural cluster and one filler stimulus, for a total of $13 \times 18 = 234$ stimuli. They were all nonwords in the three test languages.

Procedure

The 234 stimuli were concatenated in random order into a single audio file, with an ISI of 3s. Participants listened to this audio file and had to fill in a response sheet, where the stimuli were presented orthographically in the form $V_1C_1?C_2V_2$ (e.g. *eb?zo*), and a forced choice with six orthographically presented alternatives was given for each stimulus ("*a*", "*e*", "*i*", "*o*", "*u*", "*no vowel*"). The roman alphabet was used for all participants.

Participants was explained that the stimuli were all of the form shown on the answer sheet, and that they had to make a decision about the presence and identity of a vowel in the position indicated by the question mark. They were thus instructed to concentrate on the middle portion of the stimuli. The experiment lasted about 20 min.

Participants

Twenty-eight Japanese participants were tested in Tokyo (mean age 22, 14 males, 14 females), 15 Brazilian Portuguese participants were tested in Sao Paulo (mean age 21, two males, 13 females), and 18 European Portuguese participants were tested in Lisbon (mean age 20, two males, 16 females). All of the participants were recruited at universities, and none of them had had extensive exposure to a language that allows the relevant clusters, like French or English.

Results

The percentages of responses to each of the six alternatives ("*a*", "*e*", "*i*", "*o*", "*u*", "*no vowel*") were tabulated

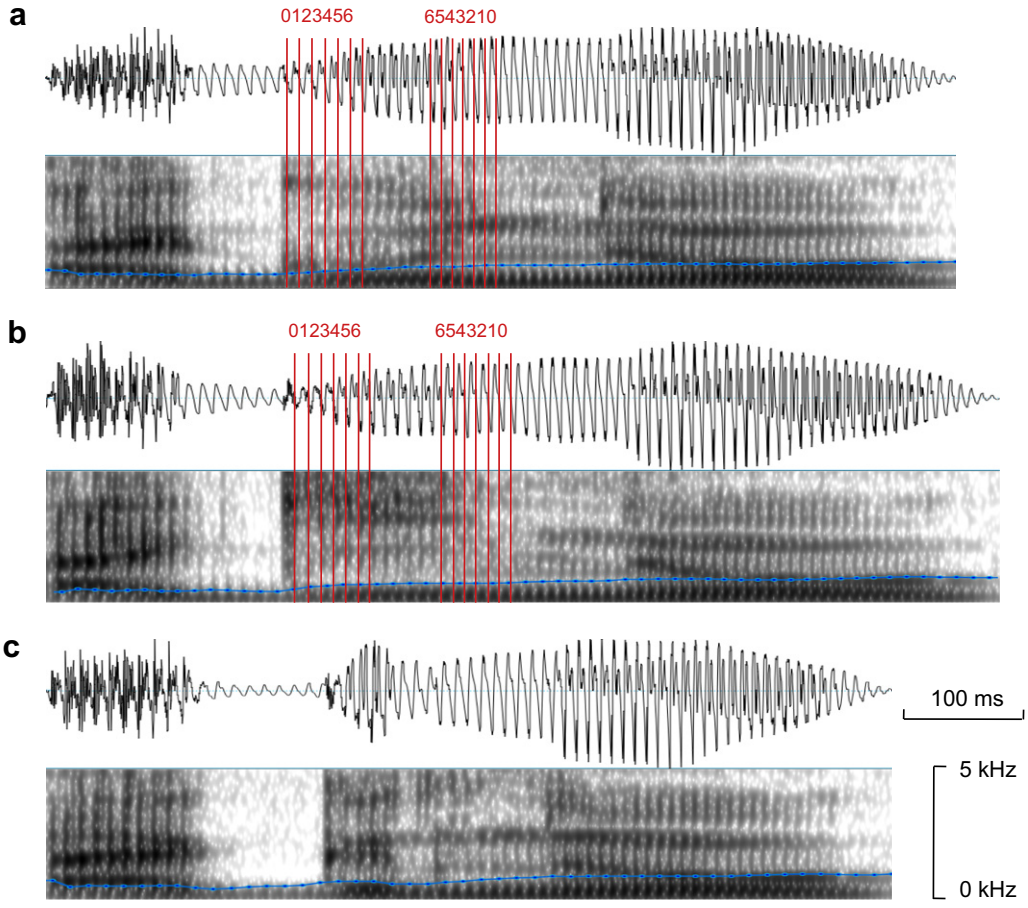

Fig. 1. Waveforms and spectrograms of sample stimuli used in Experiments 1 and 2: /aguno/ (a), /agino/ (b) and /agno/ (c). The vertical lines indicate the points in the waveforms used for removing successive portions of the signal. All vertical lines appear at a zero crossing; successive lines are separated from one another by a single pitch period. To construct the artificial clusters (/ag(u)no/ and /ag(i)no/), the entire stretch of signal between the lines labeled '0' is removed. For step 1 of the vowel length continuum, the stretch of signal between the lines labeled '1' is removed, and so on until step 6. Finally, for the full vowel endpoint, nothing is removed.

across the stimulus types and averaged across items and participants. In Table 1, we display only the three alternatives that received the highest percentages of responses across conditions and languages, namely, "*u*", "*i*", and "*no-vowel*". These are also the responses that we analyze subsequently.

Qualitatively, our results in the Japanese group for the /u/-continua replicate almost exactly the findings in Dupoux et al. (1999): as the vowel gets shorter, the proportion of *u*-responses decreases, but remains the dominant response (56%), even when there is no vowel (/u/-coarticulated cluster). Similarly, natural clusters yield a dominant proportion of *u*-responses (59%). Very similar results are obtained with the Brazilian Portuguese participants, with the vowel /i/ taking the role of /u/. That is, for the /i/-continua, the proportion of *i*-responses is dominant even when there is no vowel (/i/-coarticulated cluster: 71%), and the natural cluster elicits a majority of *i*-responses (66%). Conversely, the presentation of the /i/-continua to Japanese listeners and the /u/-continua to Brazilian Portuguese listeners yielded the following results. As the vowel gets shorter, the perception of this vowel drops sharply, to 34% and 10%, respectively. Towards the end of the continuum, moreover, the language's primary epenthetic vowel starts to rise in the responses; that is, we find high percentages of *u*-responses for the Japanese participants (20% at the end of the continuum) and of *i*-responses for the Brazilian Portuguese participants (58%). Finally, in European Portuguese participants, the response pattern is very different from those obtained with the other two groups, and it resembles that found with the French listeners in Dupoux et al. (1999): For both the /u/- and the /i/-continua, the percentages of *u*- and *i*-responses, respectively, decrease as the vowel gets shorter, yielding predominantly *no_vowel* responses at the end of the continuum (77% and 75%); the natural consonant clusters likewise yield a majority of *no_vowel* responses (69%).[4]

Below, we analyze separately the cross-linguistic epenthesis effect and the coarticulation effect.

Cross-linguistic epenthesis effect

This analysis explores the extent to which participants report hearing a vowel in the stimuli containing no vowel (i.e. the three types of consonant clusters: natural, /u/-coarticulated, /i/-coarticulated). The percent 'vowel' responses was obtained by summing the percent responses in all vowel categories (*a*, *e*, *i*, *o*, *u*) and was subjected to two ANOVAs (one with participants and the other with items as random variable), with the factors Language and Cluster-type (see Fig. 2).

We found an effect of Language ($F_1(2, 57) = 23.6$, $p < .001$, $F_2(2, 24) = 145$, $p < .001$). Post-hoc *t*-tests showed that European Portuguese participants yielded less 'vowel' responses than either the Japanese ($p < .001$) or the Brazilian Portuguese participants ($p < .001$). These latter groups did not differ from each other ($p > .05$). The effect of Cluster-type was significant in the participants analyses only ($F_1(2, 114) = 10.5$, $p < .001$; $F_2(2, 24) = 2.8$, $p < .08$), with natural clusters yielding slightly more vowel response than artificial clusters (on average, 62% vs. 53%). There was no interaction between Language and Cluster-type ($F_1(4, 114) = 2.2$, $p < .08$; $F_2(4, 48) = 1.8$, ns). An effect of Cluster-type was also reported in Dupoux et al. (1999) and corresponds to the fact that some natural clusters may contain what amounts to a very small schwa, whereas any traces of inter-consonantal vowel have been removed in the artificial clusters.

Coarticulation effect

This analysis explores the quality of the epenthetic vowel in the responses of the Japanese and Brazilian Portuguese participants (*i* vs. *u*) as a function of coarticulation. In Japanese, the dominant response to cluster stimuli is *u* (69% of the 'vowel' responses), and in Brazilian Portuguese, it is *i* (82% of the 'vowel' responses). We therefore defined a *i* minus *u* difference score, by computing, for each participant and condition, the percent response *i* minus the percent response *u* (see Fig. 3). This difference score was subjected to ANOVAs by participant and by item with the factors Language and Cluster-type. Note that since European Portuguese participants had very low percentages of *i*- and *u*-responses (less than 2% across all clusters), we did not include this language in this analysis.

We found an effect of Language ($F_1(1, 40) = 170.2$, $p < .001$; $F_2(1, 12) = 290.6$, $p < .001$), corresponding to the fact that Brazilian Portuguese participants had an overall positive *i*–*u* difference score (60, $SE = 4.0$), whereas the Japanese participants had a negative *i*–*u* difference score (-29, $SE = 4.6$). In addition, there was an effect of Cluster-type ($F_1(2, 80) = 75.9$, $p < .001$; $F_2(2, 24) = 15.1$, $p < .001$). This was due to the fact that the *i*–*u* difference scores were negative for the natural clusters (-10, $SE = 9.0$) and the /u/-co-articulated clusters (-16, $SE = 8.9$) but positive for the /i/-co-articulated clusters (34, $SE = 5.4$). Finally, there was an interaction between the two factors ($F_1(2, 80) = 18.1$, $p < .001$; $F_2(2, 24) = 16.1$, $p < .001$), corresponding to the fact that the nature of the coarticulation effect was not identical in the two groups. As can be seen in Fig. 3, the presence of /i/-coarticulation had a strong effect on Japanese listeners, reverting their predominant responses from /u/ to /i/. In contrast, in Brazilian Portuguese listeners, the effect of coarticulation was more modest and did not affect the predominance of /i/ responses over /u/ responses.

Discussion

In this experiment, we replicated the finding of Dupoux et al. (1999) that Japanese listeners report hearing a vowel /u/ in stimuli containing an illegal cluster. More importantly, we found that the same stimuli yielded a similar effect in Brazilian Portuguese listeners, with /i/ taking the role of /u/. This is consistent with the claim that Japanese and Brazilian Portuguese listeners exhibit a perceptual

[4] Interestingly, in European Portuguese listeners, the three cluster conditions yielded a small proportion of *e*-responses (for the no-vowel end of the /u/-continua, /i/-continua, and for the natural clusters: 14%, 19%, and 29% of responses, respectively). This vowel corresponds to the most common orthographic transcription of schwa, i.e., the neutral vowel that surfaces in unstressed positions in European Portuguese. This indicates that, as has been reported in English, there is a small tendency for schwa epenthesis in illegal clusters, but this does not constitute the dominant response, unlike in Japanese or Brazilian Portuguese.

Table 1
Mean percentages of responses *u*, *i*, and *no vowel* for a natural cluster and two vowel continua in Japanese, Brazilian Portuguese, and European Portuguese participants (Experiment 1).

	Japanese (*N* = 27)			Brazilian Portuguese (*N* = 15)			European Portuguese (*N* = 18)		
	u	*i*	*no vowel*	*u*	*i*	*no vowel*	*u*	*i*	*no vowel*
Natural cluster	58.7	9.4	27.6	5.6	65.6	20.0	0.4	1.3	69.2
/u/-continuum									
0	56.1	4.3	37.0	10.3	58.5	24.1	4.7	0.0	77.4
1	61.5	3.8	32.8	23.1	53.3	14.4	20.5	0.9	65.0
2	74.1	1.1	21.1	30.8	48.7	9.7	50.4	2.1	30.8
3	83.8	0.9	12.8	46.2	33.8	8.7	74.4	1.7	11.1
4	86.0	0.3	10.0	70.3	16.4	2.6	80.8	0.0	7.7
5	90.0	0.6	6.0	79.5	11.3	2.1	88.9	0.4	1.3
6	89.2	0.9	4.8	79.0	10.3	2.1	85.5	0.0	1.7
Full	91.7	0.6	2.9	79.0	9.2	1.5	84.6	0.0	0.0
/i/-continuum									
0	20.2	34.5	41.0	0.5	71.3	19.5	0.9	3.4	75.6
1	10.8	64.1	18.2	1.0	81.0	11.8	0.0	9.8	67.5
2	4.8	78.6	10.3	0.0	86.7	6.7	0.9	25.6	50.0
3	1.4	91.2	3.7	0.0	89.2	5.6	0.0	61.5	21.4
4	0.6	92.6	2.9	0.0	91.8	5.1	0.0	82.9	8.1
5	0.6	94.0	1.7	0.0	95.9	1.5	0.9	92.7	0.9
6	0.6	95.4	1.1	0.0	96.4	1.5	0.0	97.4	0.4
Full	0.6	95.7	0.9	0.0	94.4	3.1	0.0	98.3	0.0

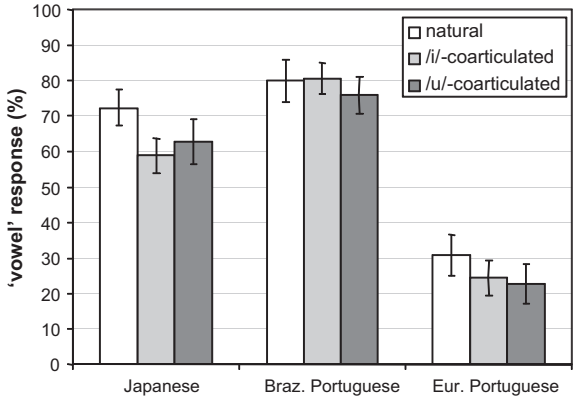

Fig. 2. Mean percentages 'vowel present' responses as a function of cluster type in Japanese, Brazilian Portuguese, and European Portuguese participants (Experiment 1). Error bars represent standard errors.

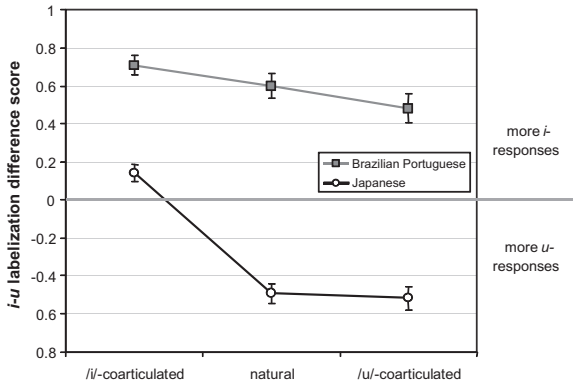

Fig. 3. Mean *i* minus *u* difference scores as a function of cluster type in Japanese and Brazilian Portuguese participants (Experiment 1). Error bars represent standard errors.

epenthesis effect, and that the epenthetic vowel is /u/ in the former and /i/ in the latter. This has to be qualified for the Japanese participants, where we found that the dominant response in /i/-co-articulated clusters was *i*, even though there was a substantial amount of *u*-responses. Finally, we found evidence for no or little perceptual epenthesis in the European Portuguese participants, a result that is similar to that obtained previously in French speakers (Dupoux et al., 1999).

Before drawing too strong conclusions regarding these data, it is important to confirm them using a technique that does not ask the participants to explicitly segment the stimuli into consonant and vowels and classify them into linguistic categories. Indeed, participants' reports may only partially reflect what they really heard. They had to perform a forced choice with discrete alternatives regarding the identity (or absence) of vowels. This requires the metalinguistic skill of being able to segment and label phonemes from the speech stream, a skill that has been shown to be intimately linked to the acquisition of reading (Morais, Cary, Alegria, & Bertelson, 1979). This is a possible concern, since the participants of the different languages may vary in reading skills, and, moreover, the Japanese writing system differs from Portuguese one in that it is syllabic (and ideographic) rather than alphabetic. In order to address this possible confound, Experiment 2 uses a task that requires participants only to judge if two stimuli are identical or different. This task involves no explicit segmentation; the potential effects of metalinguistic and orthographic factors are thus minimized.

Experiment 2

In this experiment, we tested the perception of consonant clusters using an ABX discrimination paradigm. ABX discrimination can be performed on the basis of overall

similarity in spectral and temporal features without making individual decisions regarding each of the segments' identity. Yet, the standard ABX paradigm enables participants to base their responses purely on physical identity. This is why, as in Dupoux et al. (1999), we used different voices for the stimuli A, B, and X. Thus, it is only at a more abstract level that X can be deemed identical to either A or B. This cross-talker version of the ABX paradigm taps into a phonetic similarity space, which is abstract enough to capture generalizations across speakers but does not require explicit segmentation into consonants and vowels.

We defined nine experimental conditions and one control condition. Two experimental conditions involved contrasts between tokens containing /u/ or /i/ on the one hand and the corresponding token with a natural cluster on the other hand (i.e., /ebuzo–ebzo/, /ebizo–ebzo/); four more conditions likewise involved contrasts between the vowel tokens and tokens with a coarticulated cluster (i.e., /ebuzo–eb(u)zo/, /ebizo–eb(u)zo/, /ebuzo–eb(i)zo/, and /ebizo–eb(i)zo/; here, the vowel that was removed and that hence defines the type of coarticulation is denoted in parentheses); the last three conditions involved contrasts among the three cluster types (i.e., /ebzo–eb(u)zo/, /ebzo–eb(i)zo/, and /eb(u)zo–eb(i)zo/). Finally, the control condition involved the contrast between the two vowel tokens (i.e., /ebuzo–ebizo/); this condition should be easy for listeners of all three languages and hence provides a baseline performance.

The predictions were derived from the results of Experiment 1: whenever two stimulus types received the same transcription, we predicted that the discrimination between them should be difficult. For instance, for the Japanese participants, since /ebzo/ was transcribed in the same way as /ebuzo/, the contrast between /ebzo/ and /ebuzo/ should be more difficult (yielding more errors and slower reaction times), compared to, say, the contrast between /ebuzo/ and /ebizo/.

Method

Stimuli

For each of the 13 items of Experiment 1, five stimuli were selected: the one with the natural cluster and the two endpoint stimuli of both the /u/- and the /i/-continua (i.e., /ebzo/, /ebuzo/, /ebizo/, /eb(u)zo/, and /eb(i)zo/). Additional recordings of the base triplets (/ebzo/, /ebuzo/, /ebizo/) were made by two female speakers of French. The new recordings were processed in the same way as in Experiment 1 to create the artificial clusters with /u/- and /i/-coarticulation.

Procedure

An ABX discrimination trial was constructed for each possible pairing of two out of the five token types (20 possibilities per item). Two counterbalanced lists of 260 ABX trials were then constructed, where an ABA trial in list 1 was matched to an ABB trial in list 2. The A and B tokens of the ABX trial were spoken by two different female speakers, and the X token was spoken by a male speaker. The ISI was 150 ms. Additionally, a training session of 13 trials was constructed. These trials used only VCVCV tokens and involved discrimination of /u/ vs. /o/, /u/ vs. /a/, /i/ vs. /o/, and /i/ vs. /a/.

Participants were randomly assigned to one of the two experimental lists. They were told that they would listen to sequences of three words in a foreign language, that in each sequence the third word was the same as either the first or the second one, and that their task would be to indicate which of the first two words was the same as the third one. Participants had 2500 ms to press a button on their left or right to indicate whether X was the same as A or B. The trial ended immediately after the response was given or after the 2500 ms had elapsed, whichever came first. The trials were presented in a pseudo-randomized order, with a different order for each participant. During the training session, participants received repetitive feedback; that is, in the case of an incorrect response or no response within 2500 ms the trial was repeated until the correct response was given. No feedback was given during the test session. The experiment lasted about 20 min.

Participants

Twenty-four Japanese participants were tested in Tokyo (mean age 22, 11 men and 13 women), 21 Brazilian Portuguese participants were tested in Sao Paulo (mean age 21, 2 men and 19 women), and 21 European Portuguese participants were tested in Lisbon (mean age and 20, 2 men and 19 women). All participants were recruited at universities, and none had had extensive exposure to a language with obstruent clusters like French or English. Participants making more than 40% errors in the control condition were excluded and replaced (Japanese: $N = 0$, Brazilian Portuguese: $N = 17$, European Portuguese: $N = 3$).

Results

The overall reaction times and error rates for the ten conditions are presented in Table 2.

We performed three analyses. The first analysis explored the extent to which participants in the three groups exhibited an epenthesis effect by testing the contrasts involving a natural cluster and a full vowel. The second analysis examined the coarticulation effect, and tested the three types of clusters. The third analysis directly evaluated the extent to which the results of Experiment 2 can be predicted from the results of Experiment 1.

Cross-linguistic epenthesis effect

To analyze the cross-linguistic effect independently of the coarticulation effects, we focused the analysis on the discriminations involving the natural cluster (/ebzo/) and the two vowel endpoints (/ebuzo/ and /ebizo/). We analyzed both the error rates and the reaction times by means of two ANOVAs, one by participant and one by item, with the factors Language and Contrast.

For the error rates, we found an effect of Language ($F_1(2, 63) = 8.4$, $p < .001$; $F_2(2, 24) = 24.0$, $p < .001$), an effect of Contrast ($F_1(1, 63) = 16.5$, $p < .001$; $F_2(1, 12) = 9.3$, $p < .01$), and an interaction between these two factors ($F_1(2, 63) = 35.8$, $p < .001$; $F_2(2, 24) = 21.6$, $p < .001$).

Table 2
Mean error rates and reaction times for 10 contrasts involving the endpoint stimuli and the natural cluster stimuli of Experiment 1 in Japanese, Brazilian Portuguese and European Portuguese (Experiment 2). Eb(u)zo denotes a cluster obtained by digitally excising the vowel /u/ from /ebuzo/; similarly for eb(i)zo.

	Japanese ($N = 24$)				Brazilian Portuguese ($N = 21$)				European Portuguese ($N = 21$)			
	% Err	*SE*	RT	*SE*	% Err	*SE*	RT	*SE*	% Err	*SE*	RT	*SE*
Vowel vs. cluster												
/ebuzo–ebzo/	41.2	2.2	1070	32	35.5	3.0	1088	29	23.4	3.7	1067	29
/ebizo–ebzo/	19.6	2.2	984	27	42.3	3.1	1183	37	22.9	3.5	1090	37
/ebuzo–eb(u)zo/	45.0	1.9	1131	44	35.7	2.4	1130	32	26.9	3.1	1077	34
/ebizo–eb(u)zo/	23.6	1.8	951	33	34.4	3.1	1117	33	21.4	3.6	1057	31
/ebuzo–eb(i)zo/	29.5	2.4	1024	39	27.3	2.8	1090	30	25.1	3.1	1061	34
/ebizo–eb(i)zo/	38.5	1.9	1041	37	39.0	2.9	1149	32	27.1	3.7	1114	35
Cluster vs. cluster												
/ebzo–eb(u)zo/	47.0	1.8	1084	39	52.9	1.9	1125	35	50.6	2.6	1142	36
/ebzo–eb(i)zo/	40.5	1.9	1015	33	51.5	2.4	1143	36	51.5	1.7	1166	30
/eb(u)zo–eb(i)zo/	44.9	1.9	1061	39	46.7	1.8	1142	42	48.5	2.7	1172	36
Vowel vs. vowel												
/ebuzo–ebizo/	12.5	2.1	937	28	19.6	2.0	1114	35	10.3	2.1	1045	38

Separate tests ran in each population revealed an effect of Contrast for the Japanese participants, with /ebuzo–ebzo/ generating more errors than /ebizo–ebzo/ ($F_1(1, 23) = 71.7$, $p < .001$; $F_2(1, 12) = 36.3$, $p < .001$), an effect in the opposite direction for the Brazilian Portuguese participants in the participants analysis ($F_1(1, 20) = 5.8$, $p < .03$; $F_2(1, 12) = 4.2$, $p = .062$), and no effect for the European Portuguese participants (Fs < 1).

For the reaction times, we found an effect of Language ($F_1(2, 63) = 3.5$, $p < .04$; $F_2(2, 24) = 22.1$, $p < .001$), no effect of Contrast (Fs < 1), and an interaction between these two factors ($F_1(2, 63) = 13.3$, $p < .001$; $F_2(2, 24) = 10.7$, $p < .001$). Separate tests revealed an effect of Contrast for the Japanese participants, with /ebuzo–ebzo/ yielding longer reaction times than /ebizo–ebzo/ ($F_1(1, 23) = 13.0$, $p < .002$; $F_2(1, 12) = 13.1$, $p < .005$), a reverse effect for the Brazilian Portuguese participants ($F_1(1, 20) = 11.2$, $p < .003$; $F_2(1, 12) = 17.3$ $p < .001$), and no effect for the European Portuguese participants (Fs < 1).

Coarticulation effect

This analysis tests the effect of coarticulation on the epenthetic vowel for the two languages with a strong vowel epenthesis effect, i.e., Japanese and Brazilian Portuguese. Similarly to what we did in Experiment 1, we defined *i* minus *u* difference scores for each participant and condition (vowel vs. natural cluster, vowel vs. /u/-coarticulated cluster, and vowel vs. /i/-coarticulated cluster). For error rates, these difference scores were defined as the percent error on the contrast involving /i/ minus the percent error on the contrast involving /u/; likewise for the reaction times (see Fig. 4). These difference scores were subjected to ANOVAs by participant and by item, with the factors Language and Cluster-type.

For the error rates, we found an effect of Language ($F_1(1, 43) = 72.7$, $p < .001$; $F_2(1, 12) = 33.3$, $p < .001$), corresponding to the fact that Brazilian Portuguese participants had an overall positive *i*–*u* difference score (5.7, $SE = 1.6$), whereas the Japanese participants had a negative *i*–*u* difference score (−11.4, $SE = 2.3$). In addition, there was an effect of Cluster-type ($F_1(2, 86) = 38.8$, $p < .001$; $F_2(2, 24) = 24.9$, $p < .001$). This was due to the fact that the *i*–*u* difference scores were negative for the natural clusters (−8.4, $SE = 2.8$) and the /u/-co-articulated clusters (−12.1, $SE = 2.2$), but positive for the /i/-co-articulated clusters (10.2, $SE = 2.0$). Finally, there was an interaction between the two factors ($F_1(2, 86) = 11.6$, $p < .001$; $F_2(2, 24) = 7.3$, $p < .003$), indicating that the nature of this effect was not identical in the two groups. As seen in Fig. 4, the asymmetry between Japanese and Brazilian Portuguese regarding this effect is very similar to the asymmetry in Experiment 1. The results for the reaction times were very similar: there was an effect of Language ($F_1(1, 43) = 37.5$, $p < .001$; $F_2(1, 12) = 39.3$, $p < .001$), an effect of Cluster-type ($F_1(2, 86) = 17.3$, $p < .001$; $F_2(2, 24) = 20.3$, $p < .001$), and an interaction between these two factors ($F_1(2, 86) = 4.8$, $p < .01$; $F_2(2, 24) = 5.2$, $p < .02$).

Correlation between Experiments 1 and 2

In order to check the consistency of the results across the two experimental paradigms, we derived a measure of perceptual distance from Experiment 1 and tried to predict the outcome of the discrimination task in Experiment 2. A perceptual distance measure between two stimuli was defined using the categorical responses obtained in Experiment 1. For each of the five stimulus types used in Experiment 2 (/ebzo/, /eb(u)zo/, /eb(i)zo/, /ebuzo/, and /ebizo/), and for each participant group, we computed a 6-dimensional numerical vector of the shape $x = [x_1, \ldots x_6]$, corresponding to the percent responses to the response categories *a*, *e*, *i*, *o*, *u*, and *no_vowel*, respectively. For a given condition involving stimulus types *x* and *y*, we defined the perceptual distance $d(x, y)$ as the Euclidian distance between the associated vectors:

$$d(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

Since each of the three participant groups was tested in 10 conditions, we obtained a set of 30 data points. The correlation between this set of distances and the correspond-

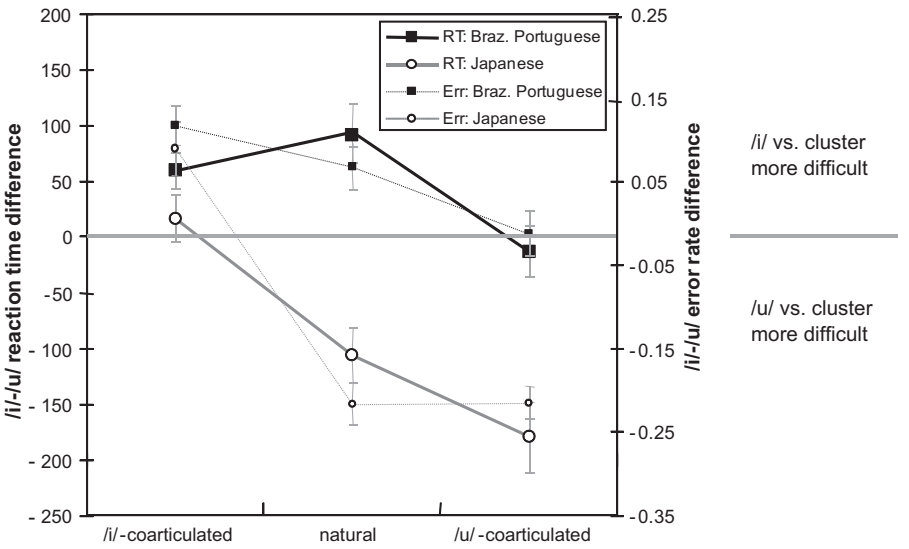

Fig. 4. Mean *i* minus *u* difference scores for errors and reaction times as a function of cluster type in Japanese and Brazilian Portuguese participants (Experiment 2). Error bars represent standard errors.

ing error rates in the ABX task was very high ($R = .94$, $F(1, 28) = 227.1$, $p < .001$), showing that the categorical responses in Experiment 1 accurately predict the error rates in Experiment 2 (see Fig. 5).[5]

Discussion

The results of the present ABX discrimination task are highly similar to those obtained with a vowel categorization task in Experiment 1. First, Japanese and Brazilian Portuguese listeners have difficulty perceiving nonnative clusters, whereas European Portuguese listeners perceive such clusters more accurately. Second, for Brazilian Portuguese listeners, clusters are perceptually closer to /i/ than to /u/, and vice versa for Japanese listeners. The coarticulation effects likewise replicate those of Experiment 1. Specifically, Japanese participants exhibit a strong coarticulation effect (shown in both their error rates and reaction times), in that they perceive /i/ rather than their dominant epenthetic vowel /u/ in /i/-co-articulated clusters; Brazilian Portuguese listeners, by contrast, continue to perceive their primary epenthetic vowel /i/ in /u/-co-articulated clusters. This shows that the exploitation of coarticulation cues is not due to universal auditory processes, but rather, is integrated within the language-specific process of segmental categorization.

The excellent correlation across groups and conditions between the ABX error rates and the labeling distances derived from Experiment 1 illustrates the fact that the former can be predicted with a high accuracy by the responses in the vowel categorization task. In other words, the results of Experiment 1 are not due to group differences in the metalinguistic interpretation of the notion of 'vowel' or 'consonant', but rather reflect on-line phonological processes.

General discussion

Using two experimental paradigms, one with an explicit metalinguistic task (vowel categorization), the other one with an implicit discrimination task (ABX), we probed for the presence of perceptual epenthesis in response to illegal consonant clusters in native speakers of Japanese, Brazilian Portuguese, and European Portuguese. Our results, which were highly correlated across the two paradigms, can be summarized as follows. First, we replicated the existence of perceptual epenthesis in illegal consonant clusters in Japanese listeners, and observed the same effect in Brazilian Portuguese but not in European Portuguese listeners. Second, the dominant epenthetic vowel is /u/ in Japanese and /i/ in Brazilian Portuguese listeners. Third, in these two populations there is a significant effect of coarticulation; for Japanese listeners, this effect is strong enough to overcome the primary epenthetic /u/-response and yield predominant /i/-epenthesis in the presence of /i/-coarticulation.

All of these results are compatible with one-step models of speech perception, in which illegal stimuli are repaired by simultaneously taking into account the proximity to the perceptual categories of the language and the phonotactic probability. In such models, speech perception involves a tradeoff between favoring probable sequences on the one hand and a good match with the acoustic signal on the other hand. The sequence /bz/ is illegal in Japanese and Brazilian Portuguese and cannot be recognized as such; indeed, the transition probability from /b/ to /z/ is zero in these languages, giving a total probability of zero for an interpretation including this sequence. A more probable interpretation involves inserting a phonetically minimal vowel between /b/ and /z/, thus maximizing both

[5] The correlation between perceptual distance and RTs was also significant, although with a lower value for R ($R = .61$, $F(1, 28) = 16.9$, $p < .001$).

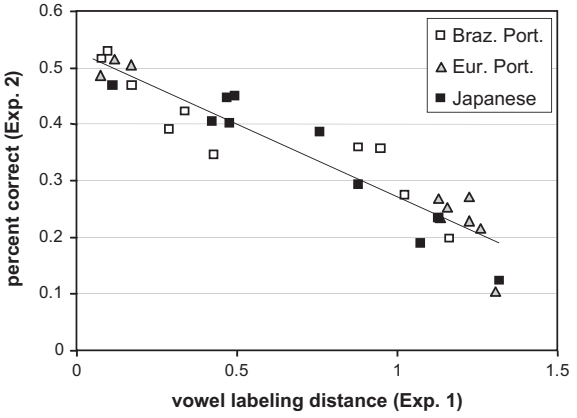

Fig. 5. Correlation between the perceptual distance derived from the forced choice responses in Experiment 1 and the ABX error rates of Experiment 2, for the 10 conditions across the three groups of participants (Japanese, Brazilian Portuguese, and European Portuguese).

its acoustic match with the signal and the total transition probability.[6] As to the identity of this vowel, recall that high vowels undergo devoicing in both Japanese and Brazilian Portuguese and that, furthermore, /u/ is the shortest vowel in Japanese (Han, 1962) and /i/ in Brazilian Portuguese (Escudero et al., 2009). In other words, /u/ and /i/ are the phonetically minimal vowels in Japanese and Brazilian Portuguese, respectively; the presence of /u/-epenthesis in Japanese listeners and that of /i/-epenthesis in Brazilian Portuguese listeners thus follows straightforwardly. The effects of coarticulation that we observed are also readily accounted for, because coarticulation of /i/ and /u/ will tend to make the transitions via /i/ and /u/, respectively, more likely.

By contrast, the coarticulation effects are difficult to account for in two-step models, which parse the input speech into discrete categories in the first step and repair illegal phonotactic structure in the second step (Berent et al. 2007; Church, 1987). Hence, repairs of illegal sequences occur after the speech stream has been categorized into discrete segments; given that after this categorization process, coarticulation of /i/ or /u/ is no longer represented, these models fail to explain the effect of coarticulation on the identity of the epenthetic vowel.

Finally, we found that European Portuguese listeners show little or no perceptual epenthesis, despite the fact that the phonological and lexical structures of their language are very similar to those of Brazilian Portuguese. Contrary to Brazilian Portuguese, though, consonant clusters have a non-zero transition probability in European Portuguese, due to a pervasive phonetic process of high back vowel deletion (Vigário, 2003). This process results in consonant clusters on the phonetic surface that are of the type used in our experiments. Thus, the difference between European and Brazilian Portuguese listeners strongly suggests that phonotactic probabilities are computed on the basis of the surface distributions of speech segments, not on the basis of the underlying representation of words. This conclusion has to be qualified, however, because some researchers have argued that the process of high vowel devoicing in Japanese can actually yield complete vowel deletion, at least in some phonological contexts (Tsuchida, 1997, Varden, 1998). We have no data concerning the frequency of complete vowel deletion, but it potentially makes Japanese similar to European Portuguese for those contexts in which is applies. It would be interesting to test whether the vowel epenthesis effect is absent or less strong in these particular contexts.[7]

Our results have potential implications for other studies on perceptual epenthesis. In particular, several studies proposed to account for perceptual epenthesis in terms of abstract grammatical principles. For instance, Moreton (2002) found that /dl/ gives rise to more perceptual repairs in English listeners than /bw/, despite the fact that both clusters are illegal in English. He argued that this is due to the fact that coronal clusters such as /dl/ are phonologically more *marked* than labial ones such as /bw/. Similarly, Berent et al. (2007, 2008) found that in English and Korean listeners, falling sonority clusters (e.g., /lb/) yield more epenthesis than rising sonority clusters (e.g. /bn/), with level-sonority clusters (e.g. /bd/) yielding intermediate performance (for a similar study with nasal-initial clusters, see Berent et al., 2009). All these clusters being illegal in English and Korean, the authors likewise argue that these results are due to differences in the universal markedness of consonant clusters as a function of their *sonority profile*. Finally, Kabak and Idsardi (2007) compared two types of heterosyllabic clusters in Korean: clusters that are illegal because the first consonant cannot occur in a syllable coda (for instance /cm/; in Korean, an underlying /c/ obligatorily surfaces as [t] in coda position), and clusters that are illegal because they undergo a phonological assimilation process (e.g. /km/, which undergoes obligatory nasal assimilation in Korean, yielding [ŋm]). They found that only the former give rise to perceptual epenthesis in Korean listeners, and argued that epenthesis is driven by *syllable structure constraints* in the phonological grammar. In all of the above-mentioned studies, crucial comparisons are made among *different* clusters. The problem with such between-cluster comparisons is that even though the sequential probabilities of the clusters are matched (i.e.,

[6] Phonotactic repairs other than segment insertion have also been attested. In French, for instance, words cannot start with the clusters /tl/ and /dl/, and French listeners tend to perceive these clusters as the legal onset clusters /kl/ and /gl/, respectively (Hallé, Segui, Frauenfelder, & Meunier, 1998). It is theoretically possible to obtain such segment substitution, as well as segment deletion, instead of segment insertion. The type of repair will depend on the balance between the sequence probability and the match with the acoustic signal.

[7] High vowel devoicing occurs in between two voiceless consonants. If the vowel is completely deleted, this thus yields surface sequences of two voiceless consonants. It turns out that four out of our 13 items contained such a sequence (/akpa/, /apti/, /epta/ and /epto/). A reanalysis of the natural cluster condition in Experiment 1, however, reveals that these four items did not yield less *u*-responses in Japanese listeners, but rather more (voiceless: 71%; voiced: 53%; $F(1, 26) = 12.6$; $p < .001$). This would need to be confirmed with items containing fricatives (e.g., /aski/), where vowel devoicing is arguably stronger than in our four items.

close to zero), their acoustic makeup is not, making it possible that they vary with respect to their degree of matching to a consonant–vowel–consonant sequence. For instance, if markedness is (partly) grounded in articulatory difficulty, it is possible that the more marked clusters – being harder to produce – contain more covert schwa-like phonetic material (Peperkamp, 2007). Further research is necessary to test this hypothesis.

In brief, we found that the perceptual epenthesis effect varies as a function of native language and that it is influenced by coarticulation. Crucially, the present data can be accounted for by two-step models but not by one-step models of speech perception. This result also speaks to the broader debate about how information is processed during on-line speech recognition. On the one hand, there is the notion that 'crisp' decisions are made at early points in speech perception, before information is passed onto higher processing levels. On the other hand, there is the notion that fuzzy or probabilistic information is passed on and evaluated at the latest possible decision stage. Within such a broad perspective, the present study points to the fact that if an early decision is made regarding phoneme identity, it has to be late enough to incorporate at least local phonotactic information.

Before closing, we must recognize that a fully specified one-step model remains to be elaborated. Hidden Markov Models developed for speech recognition use diphone or triphone transition probabilities to model phonotactics, and Gaussian mixtures over acoustic parameters to model segmental categories. In these models, phoneme recognition is a process of probability maximization which takes into account simultaneously the acoustic match with the segmental categories and transition probabilities. Such an architecture could be amended to take syllable boundaries into account. Alternatively, one could use recurrent connectionist models like the ones by Levy, Shillock, and Chater (1991) and Gaskell, Hare, and Marslen-Wilson (1995), which have been shown to extract sequential constraints in the speech input and to undo, in perception, the effects of assimilatory processes. Still another possibility would be to use constraint-based formalisms like Optimality Theory to model on-line speech perception (Berent et al., 2009; Boersma & Hamann, 2009). Hidden Markov Models, recurrent connectionist models, and constraint-based models differ with respect to the representations they use (e.g., segments or sub-segmental features) and how they model phonotactic constraints (e.g., local transition probabilities or feature sharing in abstract phonological tiers). They therefore make distinct predictions on the profile of perceptual repairs within and across languages. Further cross-linguistic work is needed to tease them apart.

Acknowledgments

We would like to thank Anne Christophe and Paul Smolensky for comments and discussion. This work was supported by a grant from the Agence Nationale pour la Recherche (ANR-05-BLAN-0065-01), the European Commission (FP6 Neurocom project), the Universidade Federal de Minas Gerais, and Fapemig.

Appendix

Items used in Experiments 1 and 2:

/abda/, /abdo/, /adgi/, /agno/, /akpa/, /apti/, /ebdo/, /ebna/, /epta/, /epto/, /ibna/, /igba/, /igna/.

References

Berent, I., Lennertz, T., Jun, J., Moreno, M., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences, 105*, 5321–5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology, 26*, 75–108.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*, 591–630.

Best, C. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In L. Wetzels & A. Calabrese (Eds.), *Studies in loan phonology* (pp. 11–53). Amsterdam: Benjamin.

Camara, J. M., Jr., & Mattoso, J. (1979). *The Portuguese language*. University of Chicago Press.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition, 25*, 53–69.

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

Cristófaro Silva, T. (1998). *Fonética e Fonologia do Português*. São Paulo: Editora Contexto.

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

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568–1578.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes, 5*, 491–505.

Dupoux, E., Sebastián-Gallés, N., Navarete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106*, 682–706.

Escudero, P., Boersma, P., Schurt Rauber, A., & Bion, R. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America, 123*, 1–15.

Fowler, C., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123–136). Hillsdale, NJ: Erlbaum.

Gaskell, M., Hare, M., & Marslen-Wilson, W. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science, 19*, 407–439.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics, 65*, 575–590.

Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24*, 592–608.

Han, M. (1962). Unvoicing of vowels in Japanese. *Onsei no kenkyuu, 10*, 81–100.

Iverson, P., Kuhl, P. K., Akahane-yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47–B57.

Kabak, B., & Idsardi, W. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech, 50*, 23–52.

Kuhl, P. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht: Kluwer Academic Publishers.

Kuhl, P., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B, 363*, 979–1000.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, N. Werner, & T. Rietveld (Eds.). *Papers in laboratory phonology* (vol. 7, pp. 637–676). Berlin: Mouton.

Levy, J., Shillock, R., & Chater, N. (1991). Connectionist modelling of phonotactic constraints in word recognition. In *Proceedings of the international joint conference on neural networks*, Singapore, pp. 101–106.

Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics, 60*, 602–619.

Massaro, D., & Cohen, M. (1983). Phonological constraints in speech perception. *Perception & Psychophysics, 34*, 338–348.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331.

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

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*, 357–395.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13–B23.

Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition, 104*, 631–637.

Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing*. Doctoral dissertation, University of Cornell.

Vance, T. (1987). *Introduction to Japanese phonology*. Albany, N.Y.: State University of New York Press.

Varden, J. K. (1998). *On high vowel devoicing in standard modern Japanese: implications for current phonological theory*. Unpublished Ph.D. dissertation, University of Washington.

Vigário, M. (2003). *The prosodic word in European Portuguese*. Berlin: Mouton de Gruyter.

Vitevitch, M., & Luce, P. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

Vitevitch, M., & Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374–408.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.