Epenthetic Vowels in Japanese: a Perceptual Illusion?

E. Dupoux
EHESS–CNRS, Paris, France

K. Kakehi
Nagoya University, Nagoya, Japan

Y. Hirose
Graduate Center, CUNY, NewYork, U.S.A

C. Pallier and J. Mehler
EHESS–CNRS, Paris, France

— Authors' version printed on October, 14th, 1998 —
To appear in *Journal of Experimental Psychology: Human Perception and Performance*

In four cross-linguistic experiments comparing French and Japanese hearers, we found that the phonotactic properties of Japanese (very reduced set of syllable types) induce Japanese listeners to perceive "illusory" vowels inside consonant clusters in VCCV stimuli. In Experiments 1 and 2, we used a continuum of stimuli ranging from no vowel (e.g. ebzo) to a full vowel between the consonants (e.g. ebuzo). Japanese, but not French participants, reported the presence of a vowel [u] between consonants, even in stimuli with no vowel. A speeded ABX discrimination paradigm was used in Experiments 3 and 4, and revealed that Japanese participants had trouble discriminating between VCCV and VCuCV stimuli. French participants, in contrast had problems discriminating items that differ in vowel length (ebuzo vs. ebuuzo), a distinctive contrast in Japanese but not in French. We conclude that models of speech perception have to be revised to account for phonotactically-based assimilations.

Human languages differ in the sound contrasts used to distinguish words. A contrast between two phones (e.g. a bilabial voiced stop and a bilabial unvoiced stop) may signal a difference in meaning in one language but not in another. It has been known for a long time that this has an influence on the perception of speech sounds: speakers of a given language often have trouble distinguishing non-distinctive phones (e.g. Sapir, 1921). For example, Japanese listeners map the American [l] and [r] phones onto their own, single, [R] category, and, as a result, have lots of trouble discriminating them. However, not all foreign contrasts are difficult: in fact, they vary in the degree of perceptual difficulty (Polka, 1991; Best, McRoberts, & Sithole, 1988). It is only recently that the perception of non-native speech sounds has been systematically explored and that theories to account for it have been advanced. For instance, in Best's Perceptual Assimilation Model (Best, 1994), a foreign sound can be processed in one of two ways. If the phonetic characteristics of that sound are close to those of an existing phoneme category in the maternal language, the sound will be *assimilated* to that category. In this case, listeners will only be able to judge whether it is a good or a bad examplar of that category, but will not have access to its detailed phonetic characteristics. (In particular, two equally bad examplars but phonetically distinct of a category will be very difficult to discriminate.) In contrast, if the foreign sound is too distant from any of the available categories, it will not be assimilated at all and listeners will have conscious access to its fine phonetic characteristics.

The Perceptual Assimilation Model is only meant to account for the effects of the phonemic repertoire. However, human languages also differ in the rules that govern what sequences of phonemes are allowed in an utterance. For instance, some languages (e.g., French or English) allow rather complex clusters of consonants, while others (e.g., Japanese) disallow them. One may expect that language-specific constraints play a role in speech perception and that language-specific influences may be demonstrated that go beyond phonemic categorization. For instance, in Spanish, /s/+consonant clusters are always preceded by a vowel and we have informally heard reports by Spanish speakers who maintain that they hear the vowel [e] preceding English words that begin with an /sC/ cluster. Accordingly, many

We thank Dianne Bradley, Susana Franck, Takao Fushimi, Peter Golato, Takashi Otake and Sharon Peperkamp for useful comments on the paper and discussion. We thank Stanka Fitneva, Olivier Crouzet and Laurent Somek for experiment preparation and running. We are especially grateful to Hideko Yamashki for her invaluable help in recruiting Japanese participants and Alain Grumbach for providing access to French participants. We also thank Franck Ramus and Evelyn Hausslein for additional help in recruiting participants. This work was supported by grants from the Fyssen fondation, the Human Frontiers Science Program, the Human Capital and Mobility Program grant, and the Direction de la Recherche et des Techniques. Mailing address: Emmanuel Dupoux, 54 Bd Raspail, 75006 Paris, France.

Spanish speakers of English sometimes produce *especial* instead of *special*, *estimulus* instead of *stimulus*, *esport* instead of *sport*, etc. This has nothing to do with the phonemic categories of [s] and [e] in Spanish versus English but rather seems to depend on a Spanish-specific phonotactic property.

In this paper, we will focus on a similar phenomenon in Japanese. As we indicated earlier, the Japanese language disallows complex consonant clusters. This is a consequence of the language's simple syllable structure: indeed, the complete syllable inventory of Japanese consists of V, VV, CV, CVV, CVN and CVQ (where Q is the first half of a geminate consonant). This can be illustrated by loan words, that is, words of foreign origin that were changed to conform to the Japanese pattern (see 1; Itô & Mester, 1995):

(1) 'fight' → *faito*
'festival' → *fesutibaru*
'sphinx' → *sufiNkusu*
'Zeitgeist' → *tsaitogaisuto*

As we can see, [u] or [o] is inserted after every syllable-final consonant (with the exception of nasal consonants)[1]. Why do the Japanese insert vowels in loan words? A first possibility is that this phenomenom (called "vowel epenthesis") arises in speech production. Perhaps Japanese speakers have, to some extent, lost or fail to develop the ability to articulate consonant clusters, and therefore tend to insert vowels to trigger the more practiced CV motor programs. A second possibility is orthography: Kanji orthographic characters, by and large, are pronounced as either [n], V (vowel) or CV (consonant-vowel). Hence, there is no Japanese character or combination of characters (in the kanji system) that can spell an item like /sfinks/ or any other item with a consonant cluster that does not include nasals. In contrast, /sufinkusu/ can easily be spelled in Japanese. Could it be that Japanese speakers modify foreign words by inserting epenthetic vowels so that they can be spelled in their language? In this paper, we will explore a third possible account, according to which vowel epenthesis can occur at the *perceptual* level. Assessing the perceptual reality of epenthesis is important because it bears on the role of phonotactics in speech perception.

What evidence exists that phonotactic constraints play a role in perception? Adults have rather clear intuitions about permissible sequences. For example, English speakers know that "mba" is not a possible English word. Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993), Jusczyk, Luce, and Charles-Luce (1994) have shown that nine month old infants are sensitive to the phonotactic patterns of the words in their language and some researchers have argued that such regularities could be useful in helping the child to discover words (Hayes & Clark, 1970; Brent & Cartwright, 1996). Massaro and Cohen (1983) investigated the influence of phonotactic constraints on phoneme perception. They used the fact that /sri/ and /ʃli/ are not allowed in English while /sli/ and /ʃri/ are allowed. They synthesized a series of stimuli ranging from [s] to [ʃ] and presented them to participants in the /_li/ and /_ri/ context. There was a significant shift in the identification functions between the two contexts, demonstrating that participants tend to hear segments that respect the phonotactics of their language[2] .

Notice, however, that the Massaro and Cohen study only demonstrates an effect on ambiguous stimuli. It would be desirable to demonstrate the influence of phonotactics on end-point (unambiguous) stimuli. Second, their study was conducted with a single language, leaving open the possibility that some of the effects might be found in all speakers regardless of their native language. Hallé, Segui, Frauenfelder, and Meunier (1998), using natural stimuli and various tasks, showed that illegal French syllables such as /dla/ tend to be assimilated to legal ones, such as /gla/. Again, this study was conducted within a single language, leaving open the possibility that part of the observed effect might have been due to universal properties of phonetic perception.

Here, we further explore the role of phonotactics on perception by using a methodology that involves non-degraded speech stimuli and a cross-linguistic design. We investigate the perceptual reality of epenthesis using an off-line phoneme detection task (Experiment 1 and 2), and two speeded ABX tasks (Experiments 3 and 4). We test the same stimuli on two populations: native Japanese speakers and native French speakers. French has complex syllabic structures and hence should not trigger epenthetic effects. Comparing the performances of French and Japanese participants on exactly the same materials allows to assess how language experience influences the perception of these stimuli.

Experiment 1

The aim of this experiment was to assess the extent of the epenthesis effect. We created nonword stimuli that formed a continuum ranging from trisyllabic tokens like *ebuzo* to disyllabic tokens like *ebzo* by progressively removing acoustic correlates of the vowel from the original stimuli. We selected our materials in such a way that the word internal consonant clusters would always yield an epenthetic [u] in Japanese (that is, the first consonant of the cluster was not a nasal nor a dental stop). French and Japanese participants were then asked to decide whether or not the vowel [u] was present in the stimuli. No overt production of the stimuli was needed. If the epenthesis effect has a perceptual basis, Japanese participants should report the presence of [u] more often than French listeners.

[1]The inserted vowel is most often [u], except after a dental stop, in which case it is an [o] (see Shinohara, (1997) for a more complete discussion).

[2]McClelland and Elman (1986), claim that such effects are not due to phonotactics per se, but rather to top down word to phoneme activation during perception. See Massaro and Cohen (1991) for a reply.

Method

Participants. Ten Japanese and ten French native speakers volunteered to participate in Experiment 1. All the participants were college students. The French participants were recruited in Paris and the Japanese at Nagoya university. None of the Japanese participants had studied French, and none of the French participants had studied Japanese.

Materials. Ten sequences of VC_1uC_2V (V: four Japanese vowels excluding [u], C_1: voiced and voiceless stops, C_2: nasals and voiced obstruents) uttered by a male Japanese speaker were used as stimulus items (see the Appendix). None of the stimulus items constituted a meaningful word in French or in Japanese.

The stimuli were digitized on a PC Compatible computer using an OROS AU22 A/D board. Five different files were then created from each original item by splicing out pitch periods of the medial vowel [u] at zero crossings. Stimulus 1 contained little or no vowel [u] (most of the transitions in and out of the vowel were also removed). Stimulus 2 contained the two most extreme pitch periods of the vowel (i.e., one from the transition of the first consonant to the vowel [u], and another from the end part of [u] into the following consonant). Stimulus 3 contained the four most extreme pitch periods (two on each side), and similarly, Stimulus 4 six pitch periods, and Stimulus 5 eight pitch periods. Stimulus 6 was the original stimulus in which the number of pitch periods varied from 10 to 13 across items (10.7 periods in average.) The average overall duration of one pitch period in the [u] vowels in each item was 9.06 ms. There were a total of 60 stimuli in one session.

Procedure. Participants were instructed to listen to the stimuli through headsets and make a judgment as to whether there was a [u] vowel in the middle of each stimulus word. The stimuli were played on a SONY DAT player. The participants were provided with an answer sheet and asked to draw a circle for "Yes" and a cross for "No". We emphasized that the experiment was not aimed at measuring their linguistic skills, and that the number of Yes and No answers need not be balanced. Japanese participants were additionally told that the [u] target in the experiment was not meant to be equivalent to the kana character " " which represents the whole syllable "u", but rather to the sound [u] as it appears inside syllables like "bu", "pu" or "mu". Each participant heard the list three times with all the stimulus sets differently randomized each time.

The French participants were given a similar setup and instructions, except that stimulus presentation and response recording was performed on a PC Compatible with a Proaudio Spectrum 16 D/A Board. Participants were required to press the [O] key for yes ('oui') responses and the [N] key for no ('non') responses.

Results

The mean percentages of vowel responses as a function of language and vowel length are shown in Figure 1. We performed two analyses of variance on percentages of vowel responses, one with participants and one with items as random variables. Language (Japanese or French) was a between-participant factor and Vowel Length a within-participant factor (with 6 levels). In the following, and in all subsequent analyses, we report the MinF' statistics when they are significant ($p<.05$), and the F1 and F2 statistics otherwise.

Overall, there was a significant Language effect (MinF'(1,25)=25.10, $p<.001$), with the Japanese participants providing more Vowel responses than the French participants. There was also a significant Vowel Length effect (MinF'(5,100)=56.18, $p<.001$), which had a significant linear component (MinF'(1,20)=152.32, $p<.001$)[3], in that longer vowels yielded more Vowel responses than shorter vowels. There was an interaction between Language and the linear component of Vowel Length (MinF'(1,26)=128.62, $p<.001$), corresponding to the fact that the French participants were much more influenced by vowel length than the Japanese. However, even in the Japanese participants, the linear component of vowel length was significant (MinF'(1,18)=11.03, $p<.005$).

We ran pairwise comparisons between the two languages for each vowel length. For the first three vowel lengths (0 ms, 18 ms, 36 ms), Japanese participants gave significantly more Vowel responses than French participants (all MinF' Bonferroni corrected $p<.006$). For the fourth vowel length (54 ms), there was only a trend in the same direction ($F1(1,18)=5.40$, $p<.04$; $F2(1,9)=3.67$, $p=.088$). A significant difference between the two populations did not appear for the last two vowel lengths (72 ms and full vowel).

Discussion

In this experiment, Japanese and French participants judged the presence or absence of the vowel [u] in stimuli containing varying extents of the acoustic correlates of the vowel. French participants were able to judge that the vowel was absent in the *ebzo* case, and present in the *ebuzo* case, with a monotonic function for the intermediate cases. The cutoff point for the French participants, that is, the point at which they judged the vowel to be present in 50% of the cases, can be estimated at just over 4 pitch periods (38 ms) of the vowel. In contrast, Japanese participants predominantly judged that the vowel was present at all levels of vowel length. Like the French, Japanese vowel responses show a steady decrease as a function of decreasing vowel length – which shows that they are sensitive to manipulation in vowel length– but the slope is much less sharp. Even at the extreme

[3]The linear component is defined as a zero sum linear set of coefficients applied across the six levels of vowel length, see Winer, Brown and Michels (1991), pp 198-210, and 148.

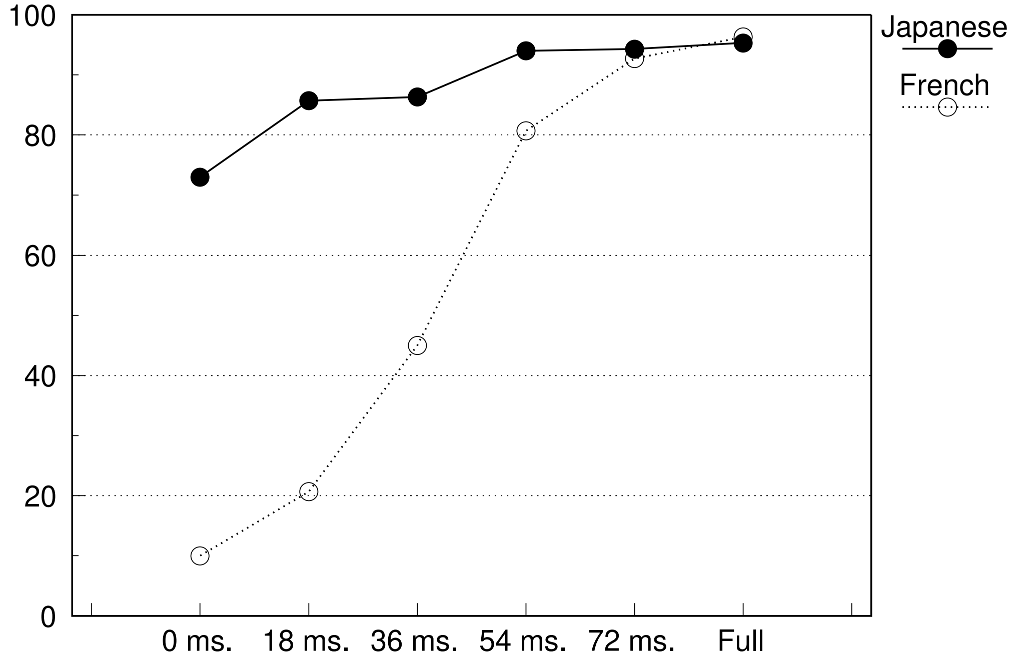

Figure 1. Percent [u] vowel judgments in stimuli like *ebuzo* in French and Japanese participants as a function of vowel duration (Experiment 1).

of the continuum where the vowel had been removed, they still reported that the vowel was present in more than 70% of the cases.

At this point, we would like to raise the following caveat. Even though we digitally removed the vocalic [u] portions of VCuCV stimuli, it is unlikely that it was completely deleted. Indeed, coarticulatory information for the rounded vowel is likely to be present throughout the portion of acoustic signal for the surrounding consonants. It may be that unlike French, Japanese hearers have developed a very fine capacity for perceiving very short vowels. One reason might be that in Japanese, the [u] vowel gets occasionnaly devoiced (Keating & Hoffman, 1984; Beckman, 1982). Another one is that the speaker that we used in Experiment 1 was Japanese, and so, he might have introduced coarticulation cues in the adjacent consonants that are especially salient to Japanese hearers.

So, even though we have identified a perceptual difference between members of two language communities, the difference may not be due to phonotactics. The next experiment is designed to address this issue.

Experiment 2

Experiment 2 used a paradigm similar to Experiment 1 with the following modifications: we recorded a French speaker and digitally generated similar continua to those in the previous experiment (ebuzo-ebzo). In addition, we recorded two extra conditions: one condition with no vowel, that is, a naturally produced consonant cluster (ebzo), and one condition with a vowel different from [u] (ebizo). This last condition was introduced to measure baseline performance.

If the results of Experiment 1 were due to coarticulation information about the vowel on the adjacent consonants, then we should expect that Japanese [u] responses on the naturally produced ebzo should drop to baseline level. If, in contrast, the obtained effects are genuinely due to phonotactics, then naturally produced clusters (ebzo), should produce at least as many [u] responses as the artificially produced clusters.

Method

Participants. Ten Japanese and 10 French native speakers volunteered to participate in Experiment 2. All the participants were college students. The French participants were recruited in Paris and the Japanese at the Nagoya university. None of the Japanese participants had studied French, and none of the French participants had studied Japanese.

Materials. We used the same ten sequences of $V_1C_1uC_2V_2$ stimuli as in Experiment 1. To these stimuli, we added 10 correponding $V_1C_1C_2V_2$ and $V_1C_1V_3C_2V_2$ stimuli (with V_3 vowels different from [u] and from V_1 and V_2). The stimuli were uttered by a male French speaker. None of the stimulus items constituted a meaningful word in French or in Japanese.

The stimuli were digitized on a PC Compatible computer using an OROS AU22 A/D board. As in Experiment 1, we digitally generated five extra stimuli from the $V_1C_1uC_2V_2$

stimuli by splicing out pitch periods of the medial vowel [u] at zero crossings. Stimulus 1 contained little or no vowel [u] (most of the transitions in and out of the vowel were also removed). Stimulus 2 contained the two most extreme pitch periods of the vowel (i.e., one from the transition of the first consonant to the vowel [u], and another from the end part of [u] into the following consonant). Stimulus 3 contained the four most extreme pitch periods (two on each side), and similarly, Stimulus 4, six pitch periods, and Stimulus 5, eight pitch periods. Stimulus 6 was the original production.

Procedure. Participants were instructed to listen to the stimuli through headsets and make a judgment as to whether a [u] vowel was present in the middle of each stimulus word. The whole stimuli set was played three times, each time in a different pseudo-random order from a PC compatible computer with a Proaudio Spectrum 16 D/A Board using the EXPE program (Pallier, Dupoux, & Jeannin, 1997). The participants had to press one key if the [u] vowel was present, and another if it was absent. Otherwise, the same procedure as in Experiment 1 was used.

Results

The mean percentages of vowel responses as a function of language and vowel length are shown in Figure 2. We performed two sets of analyses. The first set is identical to that used in Experiment 1 and analyzes the effect of the six levels of vowel length that were produced from digitally editing the $V_1C_1uC_2V_2$ stimuli. The second set of analyses tests more directly the effect of coarticulation and compares the natural $V_1C_1C_2V_2$, the digital $V_1C_1C_2V_2$ and the baseline $V_1C_1C_3C_2V_2$ stimuli with one another.

Effect of Vowel Length. We performed two analyses of variance on percentages of vowel responses, one with participants and one with items as random variables. Language (Japanese or French) was a between-participant factor and Vowel Length a within-participant factor (with 6 levels).

Overall, there was a significant Language effect (MinF'(1,47)=10.25, $p<.004$), with the Japanese participants providing more Vowel responses than the French participants. There was also a significant Vowel Length effect (MinF'(5,127)=36.50, $p<.0001$), which had a significant linear component (MinF'(1,26)=74.99, $p<.0001$), in that longer vowels yielded more Vowel responses than shorter vowels. There was an interaction between Language and the linear component of Vowel Length (MinF'(1,25)=15.48, $p<.001$), corresponding to the fact that the French participants were much more influenced by vowel length than the Japanese. However, even in the Japanese participants, the linear component of vowel length was significant (MinF'(1,12)=10.19, $p<.008$).

We ran pairwise comparisons between the two languages for each vowel length. For the first three vowel lengths (0 ms, 15 ms, and 29 ms), Japanese participants gave significantly more Vowel responses than French participants (all MinF' Bonferroni corrected $p<.02$). For the fourth vowel length (44 ms), there was only a trend in the same direction (Bonferroni corrected p=.10). A significant difference between the two populations did not appear for the last two vowel lengths (58 ms and full vowel).

Effect of Coarticulation. We performed two analyses of variance on percentages of vowel responses, one with participants and one with items as random variables. Language (Japanese or French) was a between-participant factor and Stimulus Type a within-participant factor (with 3 levels: natural cluster, digital cluster, and different vowel).

We found an overall effect of Language (MinF'(1,25)=19.77, $p<.001$), Stimulus Type (MinF'(2,53)=17.99, $p<.001$) and an interaction between these two variables (MinF'(2,54)=13.34, $p<.001$). Individual post-hoc contrasts revealed that this interaction was due to the fact that in French participants, the "natural cluster"condition was not different from the baseline condition (Fs<1), whereas the "digital cluster" condition elicited slightly but significantly more [u] responses than either baseline or natural clusters (ps<.03). In contrast, in Japanese, stimuli in both natural and digital clusters condition elicited considerably and significantly more [u] responses than baseline stimuli (ps<.0001), and the two kinds of clusters did not differ from each other (Fs<1).

Discussion

In this experiment, we replicated the pattern found in Experiment 1. Moreover, we found that this pattern of results cannot be attributed to coarticulatory cues left in the original Japanese tokens. We used tokens produced by a French speaker, and compared digitally produced clusters (that might have residual coarticulation information) and naturally produced clusters (that have no coarticulatory information for a vowel). We found that Japanese participants did not perceive more [u] vowels in digital than in natural clusters (in fact, there was a nonsignificant trend in the other direction). Hence, we observe that a majority of [u] responses do arise in Japanese participants even in the total absence of [u] information in the signal.

Experiments 1 and 2 establish that, in a task involving no overt speech production, Japanese participants consistently report a vowel between two consonants in CC clusters. These experiments alone, however, cannot firmly establish the perceptual locus of the effect for two reasons. First, the task requires participants to make an explicit *metalinguistic* judgment: participants have to know what a vowel is in order to do the task. It is known that learning to read influences the way in which individual phonemic segments can be manipulated in a metalinguistic task (see the collection of articles in Bertelson, 1986). Given that the writing systems of Japanese and French differ, it is possible that they differentially affect vowel judgments in Japanese and French participants. Second, the task did not use a speeded or on-line judgment. Therefore, it cannot identify which of

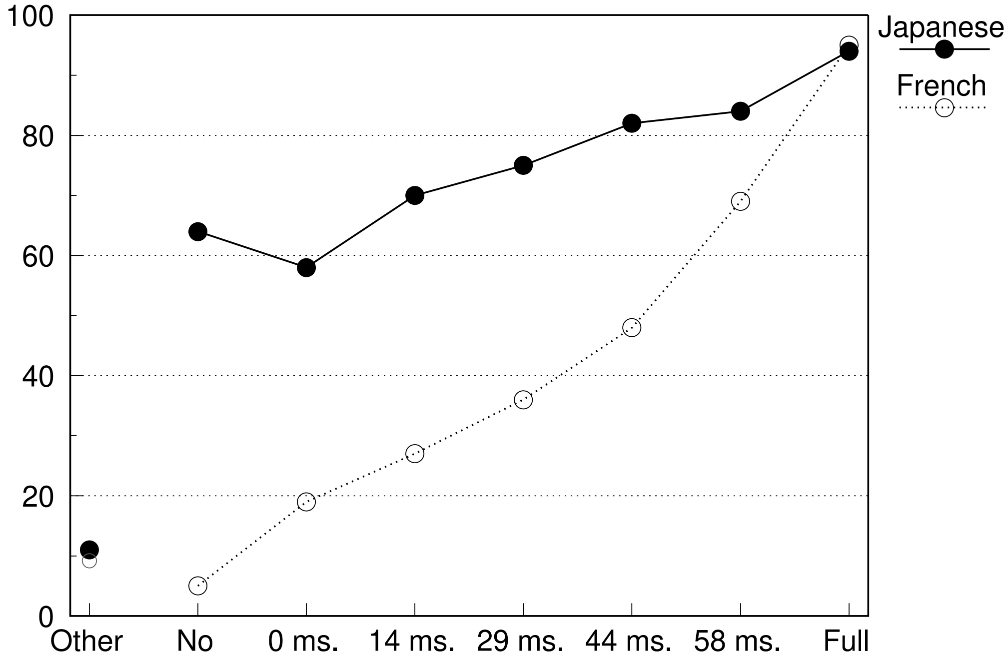

Figure 2. Percent [u] vowel judgments in stimuli like *ebuzo* in French and Japanese participants as a function of vowel duration (Experiment 2).

the different sources of information (the orthographic code, covert production, explicit strategies) influenced the participants' responses. For instance, it is possible that Japanese participants were reluctant to give a vowel-absent response simply because they knew that such stimuli do not occur in Japanese.

In the next two experiments, we use an ABX paradigm that only requires identity judgments, thus involving no explicit or implicit mention of vowels. We also had participants perform a speeded response, thereby reducing the likelihood of them using complicated response strategies.

Experiment 3

This experiment uses a speeded ABX paradigm in which participants hear three stimuli in a row and have to decide whether the third stimulus is the same as the first or the second. If the findings of Experiments 1 and 2 have no perceptual basis but are instead a by-product of metalinguistic limitations in segment manipulation, Japanese participants should make few errors when discriminating between *ebuzo* and *ebzo*. In fact, their performance should be indistinguishable from that of French participants. If, in contrast, the perceptual system inserts an epenthetic vowel to break up consonant clusters, Japanese participants should have trouble distinguishing stimuli such as *ebuzo* from stimuli such as *ebzo*, because they will in fact "hear" the same thing twice. However, *ebzo* may be "heard" as containing a vowel with different acoustic/phonetic characteristics from the [u] in *ebuzo*. For this reason, in this experiment we chose to have different talkers produce the X stimuli and the other two stimuli (A and B), thereby forcing participants to rely on a more abstract/phonological representation rather than on an acoustic/phonetic one. Experiment 4 will specifically test the effect of talker change.

Note, however, that comparing the mean performances of different groups of participants (that is, testing whether Japanese participants are significantly better or worse than French participants on a given task) raises a methodological problem: it is difficult to match populations of participants in all possible respects other than native language. This is why we introduced a complete crossover design in which we make the opposite predictions across the two language groups.

This design was achieved by considering another property of the phonology of Japanese: in Japanese, vowel length is contrastive, for instance, *tokei* (watch) vs. *tookei* (statistics). The long vowel is, in fact, perceived as two adjacent vowels. Therefore, Japanese participants should have no problem in performing the ABX task on an *ebuzo*-*ebuuzo* contrast. In our stimuli, the *ebuuzo*-*ebuzo* contrast had a difference in acoustic duration (95 ms) within the same range as the *ebuzo*-*ebzo* contrast (89 ms).

By contrast, in French, vowel length is not contrastive. That is, no pairs of French words can be distinguished purely on the basis of the length of one vowel. The hypothesis under examination is that listeners impose the phonology of their native language on unfamiliar linguistic stimuli, regardless of whether the stimuli are native or foreign. Hence, we predict that French participants might have trouble in making the *ebuzo-ebuuzo* contrast whereas the Japanese should have no problem at all.

Method

Participants. Ten Japanese and ten French participants participated in the experiment. All were recruited in Paris. The age of the Japanese participants varied from 20 to 48 years of age (median 36). Two had no knowledge of French and knew some English. All had begun the study of foreign languages after 12 years of age. There were 4 men and 6 women in the group. The age of the French participants varied from 20 to 50 years of age (median 24). None spoke Japanese, but they all had studied English at school. Like the Japanese participants, the French participants had started studying a foreign language after the age of 12. There were 9 men and 1 woman in the French group. The Japanese and French participants were all right handed; they volunteered for the experiment, and no one was paid for his or her participation.

Materials. Sixteen triplets of the form (*ebzo, ebuzo, ebuuzo*) were constructed (see the Appendix). All triplets conformed to the model $V_1C_1C_2V_2 - V_1C_1UC_2V_2 - V_1C_1UUC_2V_2$. The first consonants were from the set [b, k, g, ʃ], the initial and final vowels were from the set [e, i, a, o], and the second consonants were from the set [z, d, g, n, m, ʃ, t]. All stimuli were non-words in both French and Japanese. All stimuli consisted of phonologically valid French syllables and, with the exception of the first member of the triplets, of valid Japanese syllables. Four additional triplets with the same phoneme range constraints as for V_1, C_1, C_2, and V_2 were used in the training set.

The materials consisting of the twenty triplets were recorded twice: once by a male Japanese speaker and once by a female Japanese speaker. The recordings were made in a sound attenuated room, and digitized at 16kHz/16 bits on an OROS AU22 D/A board. Each stimulus was stored in a separate file using a waveform editor. It appeared that although our two Japanese speakers were fluent in French and had some training in phonetics, they could not be prevented from inserting a very short vowel [u] within the consonant clusters in some of the *ebzo* stimuli. These *ebzo* stimuli were therefore edited with a waveform editor, and the vocalic part was progressively removed, until a French listener found that he/she could no longer hear the [u] vowel. The three classes of stimuli had a mean duration of 409 ms for *ebzo*, 498 ms for *ebuzo* and 593 ms for *ebuuzo*, respectively.

One hundred and twenty eight experimental trials were constructed using the 16 experimental triplets. Each experimental trial consisted of three stimuli: A, B, and X, where the first two were spoken in a female voice, and the last one in a male voice. A and B were taken from the same triplet but differed in the intermediate vowel duration. There was an Epenthesis contrast (*ebzo-ebuzo*), and a Vowel Length contrast (*ebuzo-ebuuzo*). Each contrast could appear in 2 different possible orders resulting in 4 A-B combinations for each triplet. The X stimulus was identical to either A or B. The overall design was: 2 × 2 × 2: Contrast × Order × X-identity. By partial counterbalancing, 16 training trials using the four training triplets were obtained. These contained the same conditions as in the experimental trials.

The 128 experimental trials were split into two blocks, with each condition and item equally represented in each block.

Procedure. Each experimental trial consisted of the presentation of the three stimuli (A, B and X), with an inter-stimuli interval of 500 ms. Participants were told that the stimuli were words from a foreign language and that the purpose of the experiment was to test their intuitions about the sounds of these words. They were told that the third word (X) was the same as one of the first two (A or B). Their task was to press a button on their left or right to indicate whether X was the same as A or B.. Participants were given 4 seconds to respond. The trial ended immediatly after response or after the four seconds had elasped; The next trial started one second later.

In ten training trials, participants received feedback as to whether their response was correct or not. Feedback consisted of the word “Correct” or “Incorrect”, or the string “The response is A” (or B) when participants failed to respond before the deadline. Feedback was displayed for one second, and was then erased from the screen. For incorrect responses, the same trial was presented again immediately until the response was correct. In the two experimental blocks of 64 trials, no feedback was presented. The blocks were randomized separately for each individual participant. A short pause was introduced between the two experimental blocks. Responses were recorded and reaction times measured from the onset of the X stimuli with the EXPE software package (Pallier et al., 1997).

Results

Four ANOVAs were performed on the results: two on RT data by participant and by item and two on error data, again by participant and by item (reaction times were analyzed only for correct responses). The ANOVAs had a 2×2 design: Language (French or Japanese) × Contrast (epenthesis or vowel length contrast). The means, standard error and error rates are displayed for each condition in Table 1.

The analysis of the RT data showed a highly significant interaction between Language and Contrast (MinF’(1,29)=14.16, p<.001). This interaction was due to the fact that for the French participants, the vowel length contrast yielded longer RTs than the epenthesis contrast (RT difference: 171 ms, MinF’(1,19)=12.01, p<.002), whereas, for the Japanese participants, there was a trend in the other direction (RT difference: -105 ms,

Language	RT	SE	Err	RT	SE	Err
	Vowel Length Contrast *ebuzo-ebuuzo*			Epenthesis Contrast *ebuzo-ebzo*		
Japanese	1082	45	7.5%	1187	75	32%
French	1173	73	21%	1002	54	5.8%

Table 1
Mean reaction time (ms), standard error, and error rate in ABX judgments on an epenthesis contrast and a vowel length contrast in French and Japanese participants (Experiment 3).

F1(1,9)=4.52, p=.06; F2(1,15)=7.84, p<.02). There was no main effect of Language (F1(1,18)<1, p>.1; F2(1,15)=3.41, .05<p<.1), and no main effect of Contrast (F1(1,18)<1, p>.1; F2(1,15)=3.53, .05<p<.1).

The analysis of the error data showed the same pattern of results. There was a highly significant interaction between Language and Contrast (MinF'(1,26)=56.27, p<.001). This interaction was due to the fact that for French participants, the vowel length contrast was more difficult than the epenthesis contrast (MinF'(1,16)=18.11, p<.001), whereas the length contrast was easier for the Japanese (MinF'(1,13)=35.48, p<.001). Overall, Japanese participants tended to make more errors than the French participants, although this was only significant in the items analysis (F1(1,18)=3.71, p=.07; F2(1,15)=20.17, p<.001). Similarly, the epenthesis contrast tended to provoke more errors than the vowel length contrast, but again this was only significant in the items analysis (F1(1,18)=4.10, p=.058; F2(1,15)=13.18, p<.002).

Discussion

In this experiment, French and Japanese participants had to perform an ABX discrimination task on two contrasts: an epenthesis contrast (*ebzo-ebuzo*) and a vowel length contrast (*ebuzo-ebuuzo*). We found a cross-over interaction: the Japanese participants had relatively more difficulty with the epenthesis contrast, whereas the French had more difficulty with the vowel length contrast.

These results demonstrate that the phonotactics of a language influence speech perception, even with naturally produced speech stimuli. That is, not only do Japanese participants tend to *report* more vowels than are really present in the signal (Experiments 1 and 2), but their ability to *discriminate* two stimuli, one that has a vowel and one that does not have one, is also strongly affected.

Note that in this experiment, we introduced a change in talker between stimulus X and the two preceding A and B stimuli. This was done to induce participants to disregard low level acoustic characteristics and rely on a more abstract phonological representation. However, most studies on the perception of nonnative contrasts have used a more conventional ABX paradigm with no such change in talker. Would our results still hold without a talker change, that is, in a situation in which participants *can* use purely acoustic information? The next, and final, experiment addresses this issue.

Experiment 4

The present experiment was designed to evaluate the effect of a change in talker on the robustness of the language-specific pattern of previously obtained results. In this experiment, we replicate the conditions of Experiment 3 and add a new set of conditions with no change in talker. In this condition, one of the two stimuli, A or B, is acoustically *identical* to the X stimulus. This should strongly induce participants to use a rather low level of representation, since in principle it is possible to accomplish this task on a purely acoustic basis. If the epenthesis effect is still present in the same-talker condition, this will consolidate the claim that, at a certain level, Japanese participants are "deaf" to the difference between *ebuzo* and *ebzo*.

In this experiment, we will also look more closely at two factors that may affect the size of the epenthesis effect: (1) practice with the stimuli (2) influence of experience with foreign languages by the participants.

The first variable we examine is the potential effect of practice. Experiment 3 was rather short (15 minutes). It could be that the observed effects were due to participants not being very familiar with the stimuli and the task. Does the effect disappear or diminish with more extensive exposure to the contrasts? The present experiment contains 256 trials, twice as many as Experiment 3. Furthermore, the lists are randomized and the blocks are counterbalanced in such a way that potential sequential effects can be evaluated. If the epenthesis effect is labile, we should find a negative correlation between effect size and sequential position. In addition, Japanese and French participants should have similar results in the final part of the experiment.

The second variable is experience with foreign languages. We had participants fill out a detailed biographical questionnaire concerning their language experience. We were particularly interested in the degree of fluency of our Japanese participants in a language that includes consonant clusters (such as English or French). It could be that with exposure to such languages, speakers of Japanese learn to overcome the epenthesis effect. If so, we should find that the more proficient bilinguals show less effect (or no effect) compared to less proficient bilinguals or monolinguals.

Method

Participants. Twenty Japanese participants were recruited (10 in Paris, 8 in New York and 2 in Nagoya) and tested individually in a quiet room. None of them had participated in the previous experiments. Their ages ranged from 22 to 40 (median 29). There were 14 women and 6 men in the group.

Twenty French participants recruited in Paris were tested on the same materials. None of them had participated in the previous experiments. Their ages ranged from 19 to 50 (median 21.5). There were 4 women and 16 men in the group.

Participants filled out a detailed biographical questionnaire about their experience with foreign languages. They also rated their own fluency and pronunciation in these languages on a 10 point scale. The questionnaire for the Japanese participants was in English. Moreover, their fluency in either French or English (or both) was subjectively assessed by a native speaker of French or English, respectively. Four Japanese participants did not fill out the questionnaire.

Materials.

The same materials as in Experiment 3 were used. We used the same 128 ABX experimental trials of Experiment 3 (A and B stimuli spoken by the female talker, and X stimuli by the male talker) and created another 128 trials with the stimuli A, B and X all spoken by the same male talker. In these last trials, X was acoustically identical to either A or B. The overall design was: 2 x 2 x 2 x 2 : Contrast x Order x X-identity x Talker.

The 256 experimental trials were split into four blocks of 64 trials, with each condition and each item equally represented in each block.

Procedure.

The same procedure as in Experiment 3 was used.

Results

The means, standard error and error rates are displayed in Table 2 for each condition. As in Experiment 3, we ran four ANOVAs, two by participants and two by items, on reaction times and error rates, respectively, with Language, Talker, and Contrast as experimental factors.

The analysis of the RT data showed that there was a highly significant interaction between Language and Contrast (MinF'(1,53)=14.81 $p<.001$). This interaction was due to the fact that for French participants, the vowel length contrast yielded significantly slower reaction times than the epenthesis contrast (117 ms, MinF'(1,34)=14.33, $p<.001$), whereas for Japanese participants, there was a nonsignificant trend in the other direction (-27 ms, all ps>.1). No other interaction was significant, except the interaction between Language and Talker, which was only significant in the items analysis (F1<1; F2(1,15)=16.10, $p<.001$).

There was a main effect of Talker, with the same talker yielding faster RTs than the different talker (85 ms, MinF'(1,52)=12.00, $p<.001$). There was also a main effect of Contrast, with the vowel length contrast on average yielding slower RTs than the epenthesis contrast (45 ms, F1(1,38)=7.83, $p<.01$; F2(1,15)=5.20, $p<.04$). Finally, Japanese talkers tended to have longer RTs than French participants, but this was only significant in the items analysis (55 ms, F1<1; F2(1,15)=18.19, $p<.001$).

The analysis of the error data showed similar results. There was a highly significant interaction between Language and Contrast (MinF'(1,40)=34.11, $p<.001$). This interaction was due to the fact that for Japanese participants, the epenthesis contrast yielded significantly more errors than the vowel length contrast (MinF'(1,31)=22.05, $p<.001$), whereas for French participants, there was a significant effect in the other direction (MinF'(1,33)=8.62, $p<.006$). No other interaction reached significance.

There was a main effect of Talker, with the different talker condition yielding more errors than the same talker condition (MinF'(1,36)=6.50, $p<.02$). There was also a main effect of Contrast, with the epenthesis contrast on average yielding more errors than the vowel length contrast, (MinF'(1,35)=4.80, $p<.04$). Finally, Japanese talkers tended to make more errors than French participants, but this was only significant in the items analysis (F1(1,38)=3.41, $.05<p<.1$, F2(1,15)=9.51, $p<.01$).

Influence of practice

We began our investigation of the effect of practice by using a correlation analysis. For each participant, the sequence of reaction times on experimental trials was partitioned into 16 successive bins of 16 datapoints. We found a significant negative correlation between sequential position and mean reaction time (R^2=.67, F(1,14)=28.10, $p<.001$). We also found a significant negative correlation between sequential position and error rate (R^2=.67, F(1,14)=28.02, $p<.001$). These effects show that practice does have an impact, and that participants improve their performance with time. We then computed the numerical size of the interaction between language and contrast (i.e. $epenthesis\_in\_Japanese + vowel\_length\_in\_French - epenthesis\_in\_French - vowel\_length\_in\_Japanese$) for each sequential position. There was no significant correlation between sequential position and interaction size either in the reaction time (R^2=.16, F(1,14)=2.60, $p>.1$) or in the error analysis (R^2=.17, F(1,14)=2.91, $p>.1$).

In a second step, we ran ANOVAs similar to the ones reported above, but restricted the analysis to the final block of 64 trials (after 202 trials). In this analysis, the interaction between Language and Contrast was still significant, both for the reaction times (MinF'(1,51)=4.65, $p<.04$) and the error

Participants		Contrast					
		Vowel Length *ebuzo-ebuuzo*			Epenthesis *ebuzo-ebzo*		
Japanese		RT	SE	Err	RT	SE	Err
	Same Talker	1008	41	3.1%	1032	48	13.7%
	Different Talker	1058	45	5.6%	1089	46	19.1%
	Mean	1033	30	4.4%	1060	33	16.4%
French							
	Same Talker	1095	76	8.9%	991	55	4.1%
	Different Talker	1225	72	10.8%	1095	58	5.4%
	Mean	1160	53	9.8%	1043	40	4.7%

Table 2
Mean reaction time, standard error, and error rate in ABX judgments on an epenthesis contrast and a vowel length contrast in Japanese participants and French Participants (Experiment 4).

data (MinF'(1,36)=17.40, $p<.001$)[4].

Influence of language background

Inspection of our questionnaire revealed that the Japanese participants mostly had experience with French or English (one reported having studied some Italian, and one some Russian). They had all begun to study these foreign languages in school after the age of 12. We divided the participants in two groups, one labeled 'low proficiency' (7 participants), the other labeled 'high proficiency' (9 participants) based on the means of both their self-evaluation and our evaluation of fluency and pronunciation. 'High proficiency' participants could all understand spoken English or French and sustain a conversation in these languages with good fluency and a moderate foreign accent, as assessed by the experimenters. 'Low proficiency' participants had trouble both understanding and being understood in English or French; some of them could not express themselves in either of these languages.

We found that the Proficiency factor introduced no significant effect nor any interaction in the analysis of errors ($p>.1$). In fact, the 'high proficiency' group displayed roughly the same pattern of errors as the 'low proficiency' group (both showed 16% of errors in the epenthesis contrast).

In a further analysis, we selected the four Japanese participants with the greatest proficiency in English or French (both self-rated, and as evaluated by an external judge). The selected participants had all lived in France or the US for more than 4 years (one is an English teacher, another a student of phonetics, and two others are university students in the US), and were very fluent in French or English. For these participants, the percent error on the epenthesis contrast was in the same range as that of the other Japanese listeners (15.9% on average vs. 4.7% for the vowel length contrast).

We also analyzed the linguistic background of the French participants. They all knew English (all had learned it after the age of 6). Some also knew German, Italian, Spanish, or Arabic. Note that none of these languages use vowel length contrastively. However, English, Spanish, and Italian use stress contrastively, and vowel length is used as a cue for stress. We then tentatively divided these participants into two groups (high and low proficiency) according to their evaluation of their proficiency in these languages: we found no effect of fluency on the error data or on the reaction times ($p>.1$).

More generally, every Japanese participant that we tested in this experiment showed the epenthesis effect, that is, each participant showed more errors on the epenthesis contrast than on the vowel length contrast. Such regularity is also true of the Japanese participants tested in Experiment 3. In contrast, 18 out of 20 French participants (9 out of 10 in Experiment 3) showed either no difference, or the opposite pattern of behavior. In other words, the observed cross-over interaction of language and contrast in the error data is highly robust and reproducible from participant to participant, at least in the sample we tested.

Discussion

In this experiment, we studied the effect of a talker change on the size of the language-specific effects reported in Experiment 3. We found that even though the same talker condition elicited significantly shorter reaction times and fewer errors than the different talker condition, this variable had a very small effect on the previously reported interaction between language and contrast. We found that Japanese

[4]The effects on the last block were very similar in the same talker and different talker conditions both for the reaction times and the errors, although there was a nonsignificant trend toward a smaller magnitude of the effect for the same talker condition (10 ms in the reaction times, and one percent on the error data).

participants had more difficulty with the epenthesis contrast than with the vowel length contrast, and the French vice-versa, regardless of whether the ABX task involved tokens produced by the same talker or not. This is all the more remarkable because in the same talker condition, a judgment of acoustic identity alone was sufficient to perform the task.

In addition, we found that after more than 200 trials the cross-linguistic effects still obtained. Although practice has a very powerful effect on both reaction time and error rate, it does not significantly modulate the size of the effects.

Finally, a post-hoc analysis in terms of linguistic background revealed no clear effect of fluency in languages allowing consonant clusters, such as English or French. That is, both fluent and nonfluent Japanese speakers showed an epenthesis effect of about the same size. Of course, although we used Japanese participants in France or the US (in Experiments 1, 3 and 4), we did not use extremely fluent bilinguals. Even our "high proficiency" participants had learned English or French after age 12 and had a noticeable Japanese accent in these languages. It remains an open question as to whether extremely proficient bilinguals or more early bilinguals might have a reduced epenthesis effect.

Finally, we have to address a minor caveat. When we compare Tables 1 and 2, the percentage of errors for the epenthesis contrast in Japanese participants is smaller in Experiment 4 than in Experiment 3 (16% instead of 32%, a significant difference, $p<.05$). Given that, individually, neither practice nor nature of talker significantly reduce the epenthesis effect, why should such a difference obtain?

This apparent discrepancy may be due to the fact that two weak variables can nonetheless conjointly have a significant effect. Indeed, if we look at the first experimental block in the present experiment, we find that the epenthesis contrast yielded 28% errors for the different talker condition, which is not significantly different from the 34% score of the equivalent first block in Experiment 3. At the very onset of both experiments, comparable effect sizes were thus found for the different talker conditions. In the next trials, however, divergences appear, as the score stays at 31% in block 2 of Experiment 3, but drops to a value centered around 16% in Experiment 4. Such a drop is not found for the same talker condition which yields an initial score of 12% and stays around this value throughout Experiment 4.

In other words, there is an initial difference between same and different talker conditions ($p<.05$), but after the first block, the different talker condition drops to the same value as the same talker condition. This suggests that it is only practice *in a same talker condition* that reduces the size of the epenthesis effect in the different talker condition. One might think that the same talker condition should allow the participant to focus his/her attention on the right acoustic/phonetic cues, a strategy that can be used on subsequent trials. But we have not demonstrated that, with more extensive training, even better performance cannot be achieved. So, more research would be needed to explore this point.

General discussion

The present series of experiments has shown that Japanese listeners, in contrast to French listeners tend to perceive illusory epenthetic [u] vowels within consonant clusters. Indeed, Japanese participants have difficulty discriminating between a stimulus that does not include a vowel (*ebzo*), and one that does (*ebuzo*). However, Japanese participants, unlike the French, easily discriminate stimuli that contain one versus two successive [u] vowels. The epenthesis effect we have established is robust. It was present in each of the Japanese volunteers that we tested and was still significant even when the experimental setting was designed to help participants discriminate (Experiment 4). Moreover, we found very little evidence that proficiency in English or French changes the pattern of data. Needless to say, no tendency toward epenthesis was present in our French volunteers.

These results buttress the hypothesis that speech perception is heavily influenced by phonotactic knowledge. This complements and extends the work by Massaro and Cohen (1983). Indeed, not only does phonotactic knowledge influence the classification of individual phonemes, but it can also induce the perception of "illusory" phonemes that have no acoustic correlates. Moreover, it does so in nondegraded stimuli. This shows that the way in which the continuous speech stream is segmented into discrete phonemes is not universal, but depends on what the typical pattern of alternation between consonants and vowels is in the language in question. In brief, when we perceive nonnative sounds, not only do we assimilate them to our native categories, but also we may invent or distort segments so as to conform to the typical phonotactics of our language. How could such effects be accounted for? We foresee two possibilities.

The first possibility would be to amend the Best's Perceptual Assimilation Model by stipulating that native categories are not (or not only) categories of single phonemes but rather categories that span larger chunks of signal. For example, Mehler, Dupoux and Segui (1990) have proposed SARAH, a model based on an array of syllable detectors. In this model, speech sounds are categorized into syllable-sized units. The repertoire of syllables includes the totality of the syllables used in the language. Similar proposals have been made for triphones (Wicklegren, 1969), diphones (Klatt, 1979) and semi-syllables (Fujimira, 1976; Dupoux, 1993). In such a view, an account of the epenthesis effect is quite straightforward. For the sake of illustration, let us endorse syllable-sized categories. Faced with a foreign language, our perceptual system tries to parse the signal using the available native syllabic categories. However, in Japanese, there are no syllable categories containing consonant clusters or coda consonants. A stimulus like /ebzo/ therefore activates categories for "e" and "zo". It also activates to a lesser degree all syllables that start with /b/: "bu", "ba", "be", "bi" and "bo". Why is the "bu" interpretation favored? One possibility is that in Japanese, the [u] vowel is frequently shortened or devoiced and shows considerable allophonic variation (see Keating & Hoffman, 1984; Beckman, 1982). Hence the prototype for "bu" is rather compliant and is likely to emerge as the best match. Note that such a model could help to account for the preference for [o] epenthesis after a dental stop ("batman" – "batoman"). In Japanese, dental stops become affricates in front of high vowels. Hence, there is no available "du" or "tu" syllable, only "dsu" or "tsu" syllables. In that case, one might then expect that the best match will be rather a syllable like "do" or "to" for which the first consonant is not affricated and hence closer to the signal. Of course, this interpretation would have to be backed up by further experiments.

A second and quite different possibility would be to keep phonemes as the basic level in the Perceptual Assimilation Model, but to add an extra layer of processing that is allowed to modify the output of the phoneme detectors. For instance, Church (1987) has proposed a parser that yields a syllabified representation based on language-specific constraints. Indeed, in a study by Pallier, Sebastian-Gallés, Felguera, Christophe, and Mehler (1993), evidence was found that listeners build such a syllabified representation on-line (see also Pallier & Mehler, 1994). In order to accommodate our data, such models would have to stipulate that incorrect or deviant phonological forms are automatically regularized by the parsing device. The exact nature of the regularization routines, however, needs to be further specified. Note that such a proposal would predict a time course difference and, maybe, also a brain localization difference, between phoneme-based assimilations and phonotactically-based assimilations.

Obviously, more studies are necessary to choose among these possibilities. However, our findings already allow us to pinpoint shortcomings of models that represent phonemes or subphonemic elements without any mention of higher order structures (McClelland & Elman, 1986; Norris, 1994; Marslen-Wilson & Warren, 1994). In such models, no direct effect of the phonotactic organization of the language being used is expected in processing, and our present results are difficult to interpret.

Finally, we would like to highlight the difficulty that French participants experienced in dealing with duration differences. The French results are interesting because they suggest that not only the succession of C and V but also the precise timing of these elements is important. In Japanese, vowel length is contrastive, and words can contain up to four consecutive identical vowels (e.g. *to, too, tooo, toooo*). In French, by contrast, there are no pairs of words that differ only in vowel length. That a simple acoustic dimension such as duration has different functions cross-linguistically is also shown by Takagi and Mann (1994) who studied the perception of English words and nonwords by Japanese speakers. In particular, they show that in English CVC syllables, tense vowels are perceived by Japanese listeners as long vowels (e.g. [gip] yields [giipu]), whereas lax vowels yield the perception of a geminate consonant (e.g. [gIp] yields [gippu]). Hence, the mapping between the phonetic and the phonological level involves more than a set of phonetic (or even syllabic) detectors, but also relies upon rhythmic properties of adjacent phonemes over a rather large time window.

On a broader perspective, our research is consistent with other studies showing that it will be difficult to build a realistic model of speech perception that only relies on linear strings of phonemes. For instance, as Dupoux, Pallier, Sebastian, and Mehler (1997) already showed, the way in which suprasegmental information is perceived depends on the accentual regularities in the language of the hearer. Spanish listeners have no difficulty in swiftly responding to a difference in accentual pattern (*vásuma* vs. *vasúma*), whereas French listeners are slow and error prone. Such an effect arises, we believe, because Spanish uses accent contrastively (*bebé* vs. *bébe*), whereas French does not. More research is needed to understand how models can be modified to take into account such higher-order properties of signals.

References

Beckman, M. E. (1982). Segmental duration and the 'mora' in japanese. *Phonetica*, *39*, 113–135.

Bertelson, P. (Ed.). (1986). *The onset of literacy.* Cambridge, MA: MIT Press.

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

Best, C. T., McRoberts, G. W., & Sithole, M. N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by english-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 345–360.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*, 93–125.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, *25*, 53–70.

Dupoux, E. (1993). Prelexical processing: the syllabic hypothesis revisited. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing* (pp. 81–114). Hove, East Sussex, UK: LEA.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in french? *Journal of Memory and Language*, *36*, 406–421.

Fujimira, O. (1976). Syllables as concatenated demisyllables and affixes. *Journal of the Acoustical Society of America*, *59*(S1), S55.

Halle, 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*(2), 592–608.

Hayes, J. R., & Clark, H. H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). New York: Wiley.

Ito, J., & Mester, A. (1995). Japanese phonology. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 817–838). Oxford: Blackwell.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound pattern of native language words. *32*, 402-420.

Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *33*, 630-645.

Keating, P., & Hoffman, M. (1984). Vowel variation in japanese. *Phonetica*, *41*, 191–207.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representations and process in lexical access: words, phonemes and features. *Psychological Review*, *101*, 653–675.

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

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, *23*(4), 558–614.

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

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

Pallier, C., Dupoux, E., & Jeannin, X. (1997). Expe5: an expandable programming language for on-line psychological experiments. *Behavior Research, Methods, Instruments and Computers*, *29*, 322–327.

Pallier, C., & Mehler, J. (1994). Language acquisition: Psychobiological data. In *4th refresher course of the esnr. nancy 1994* (pp. 23–26). Udine: Edizioni del Centauro.

Pallier, C., Sebastian, N., Dupoux, E., Christophe, A., & Mehler, J. (in press). Perceptual adjustment to time-compressed speech: a cross-linguistic study. *Memory and Cognition*.

Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within syllabic structure of spoken words. *Journal of Memory and Language*, *32*, 373–389.

Polka, L. (1991). Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*(6), 2961–2977.

Sapir, E. (1921). *Language.* New York: Harcourt Brace Jovanovich. (Trad. française : *Le Langage* Paris, Payot, 1967.)

Shinohara, S. (1997). *Analyse phonologique de l'adaptation japonaise de mots trangers.* Unpublished doctoral dissertation, Universite de la Sorbonne Nouvelle.

Takagi, N., & Mann, V. (1994). A perceptual basis for the systematic phonological correspondences between japanese loan words and their english source words. *Journal of Phonetics*, *22*, 343–356.

Wicklegren, W. A. (1969). Context-sensitive coding, associative memory and serial order in (speech) behavior. *Psychological Review*, *76*, 1–15.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.

Appendix

Materials used in Experiment 1 and 2

abge–abuge, abno–abuno, agmi–agumi, akmo–akumo, ebza–ebuza, egdo–egudo, ibdo–ibudo, igna–iguna, obni–obuni, ogza–oguza.

Materials used in Experiments 3 and 4

abge–abuge–abuuge, agmi–agumi–aguumi, akmo–akumo–akuumo, aʃmi–aʃumi–aʃuumi, ebza–ebuza–ebuuza, egdo–egudo–eguudo, ekʃi–ekuʃi–ekuuʃi, eʃmo–eʃumo–eʃuumo, ibdo–ibudo–ibuudo, igna–iguna–iguuna, ikma–ikuma–ikuuma, iʃto–iʃuto–iʃuuto, obni–obuni–obuuni, ogza–oguza–oguuza, okna–okuna–okuuna, oʃta–oʃuta–oʃuuta.