HAL
open science

Epenthetic vowels in Japanese: A perceptual illusion?

Emmanuel Dupoux, Kazuhiko Kakehi, Yuki Hirose, Christophe C Pallier, Jacques Mehler

To cite this version:

Emmanuel Dupoux, Kazuhiko Kakehi, Yuki Hirose, Christophe C Pallier, Jacques Mehler. Epenthetic vowels in Japanese: A perceptual illusion?. Journal of Experimental Psychology: Human Perception and Performance, 1999, 25 (6), pp.1568-1578. 10.1037/0096-1523.25.6.1568 . hal-02341221

HAL Id: hal-02341221

https://hal.science/hal-02341221v1

Submitted on 31 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Epenthetic Vowels in Japanese: a Perceptual Illusion?

E. DUPOUX[1] K. KAKEHI[2] Y. HIROSE[3] C. PALLIER[4,5]

J. MEHLER[1]

January 6, 1998

Abstract

We report a set of experiments demonstrating that the number of phonemes perceived in a stimulus depends on the native language of the listener. Comparing French and Japanese subjects we found that the phonotactic properties of the native language can induce subjects to insert "illusory" segments. In Experiment 1, we varied the duration of an inter-consonantal vowel [u] in stimuli such as *ebuzo* and found that unlike the French, Japanese listeners report that the vowel [u] is present even in stimuli in which the vowel is absent. In Experiment 2 and 3 using an ABX task, we show that Japanese subjects have trouble discriminating stimuli that contain an [u] vowel from stimuli in which the vowel is absent, e.g., (*ebuzo* vs. *ebzo*). However, they can easily discriminate items that contain one versus two [u] vowels, e.g., *ebuzo* vs. *ebuuzo*, a distinctive contrast in Japanese. Results for French subjects are reversed.

Languages use different repertoires of distinctive sounds. Some languages use more than 50 distinct consonants, while others use only 5 (Crystal, 1987). In addition, languages differ in the way phonemes can be arranged in sequences, i.e., they respect different *phonotactic constraints*. For instance, some languages (French, English) allow complex strings of consonants, while others favor a quasi regular alternation of consonant and vowel (Japanese).

It is well known that differences in the phonemic repertoire of a language have a strong impact on the language processing by speakers of that language. For instance, it is well-known that Japanese speakers find it difficult to distinguish the sound of the English /r/ from /l/ (Goto, 1971; Mann, 1986; Miyawaki et al., 1981). Many studies have shown that people are strongly influenced by their own native phoneme categories when perceiving and producing foreign sounds(Best, McRoberts, & Sithole, 1988; Kuhl, 1992; Trehub, 1976; Werker & Tees, 1984a, 1984b; Flege, Munro, & MacKay, 1995).

The influence of phonotactic constraints on processing have been much less studied. However, the following observations suggest that they may play a role as important as that of the phonemic repertoire. In Spanish, there is no word starting with an /s/ followed by a consonant; in this language, /s/+consonant clusters are systematically preceded by a vowel. Accordingly, Spanish speakers of English typically say *especial* instead of *special*, *estimulus* instead of *stimulus*, *esport* instead of *sport*, etc. Spanish speakers even maintain that they *hear* an [e] vowel in /s/+consonant-initial English words. In this paper, we explore a similar, but more spectacular kind of phenomenon in Japanese.

Consonant clusters are not allowed in Japanese utterances which are composed of regular alternations of consonants and vowels [1]. This is certainly not the case in languages such as French or English in which words can have very complex onsets and almost equally complex codas, e.g., *spleen* and *facts*. Linguists have remarked that native speakers of Japanese tend to incorporate foreign words with consonant clusters into their own vocabulary after inserting a vowel between the consonants (see 1; Ito, 1993). This process of vowel insertion is referred to as *vowel epenthesis*.

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

Why does epenthesis occur? One obvious possibility is orthography. All kanji orthographic characters, by and large, are pronounced as either /n/, a V (vowel) or a 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. Thus, it may be that Japanese speakers modify foreign words by inserting epenthetic vowels so that they can be spelled in their language.

Undoubtedly, orthography has to be taken into account, although it seems unlikely that this factor alone explains vowel epenthesis. For instance, in the examples above, vowel epenthesis seems to follow a rather regular pattern: an [u] is inserted in all cases,

[1] The one exception to this regularity is when an /n/ appears after a vowel and before a consonant, in words like *Honda*, *kanji* and *tempura*.

except after a dental stop [t] or [d], in which case, an [o] is inserted. It is far from obvious how to explain this pattern on purely orthographic grounds. At any rate, if epenthesis is just an epiphenomenon of orthography, it should disappear (or greatly diminish) in a purely perceptual task. This is one of the questions that we will address in this paper.

Alternative factors that may determine epenthesis are based on phonology. Speakers of Japanese know that their language is a CV-language. When they are confronted with foreign words that contain consonant clusters, they may experience production and/or perceptual problems. Perhaps Japanese speakers have to some extent lost the ability to articulate consonant clusters, and therefore tend to insert vowels to trigger the more practiced CV motor programs. Another possibility, which is one that we explore in this paper, is that the problem originates in perception: namely, that Japanese speakers really *hear* illusory [u] vowels inside words such as *sphinx*. If this is true, it would support the view that the phonotactic constraints of one's native language play an important role in speech processing: indeed, one so strong that it produces illusory phonemes. The production- and perception-based accounts of epenthesis are not necessarily exclusive.

Do the phonological properties of one's native language, other than the phonemic repertoire influence speech perception? Psycholinguists have explored this question for the last fifteen years. For instance, studies have shown that listeners who perform various detection tasks are influenced by the syllabic structures of their language (Cutler, Mehler, Norris, & Segui, 1983; Kolinsky & Morais, 1993; Otake, Hatano, Cutler, & Mehler, 1993; Pallier, Sebastian-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastian-Galles, Dupoux, Segui, & Mehler, 1992; Zwitserlood, Schriefers, Lahiri, & Donselaar, 1993). Other studies have revealed that language-specific strategies are used to postulate word boundaries in continuous speech recognition (Cutler & Norris, 1988; Cutler, Mehler, Norris, & Segui, 1992; Suomi, McQueen, & Cutler, 1997). Such cross-linguistic studies are important for at least two reasons. First, if speech processing is influenced by phonological properties beyond the phonemic repertoire, models of speech perception have to be revised to incorporate such higher order structures, and, cross linguistic studies are necessary to establish a general model of speech perception that holds across all human languages. Second, psycholinguists will have to explain how the child acquires the perceptual adjustments necessary to perceive his or her native language (Best, 1994; Jusczyk, 1994; Mehler, Dupoux, & Segui, 1990; Polka & Werker, 1994).

What evidence exists that constraints on possible sequences of phonemes play a role in perception? Psychologists have established that humans are sensitive to statistical regularities in sequences of units (Miller, Heise, & Lichten, 1951; Miller, 1951). Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993), Jusczyk, Luce, and Charles-Luce (1994) have shown that at 9 months infants become sensitive to the phonotactic patterns of the words in their language. Indeed, some researchers have argued that such regularities could be useful in helping the child to discover words (Hayes & Clark, 1970; Brent & Cartwright, 1996).

Adults have rather clear intuitions about permissible sequences, e.g. English speakers know that "mba" is not a possible English word. However, the only study we are aware of, that has specifically investigated the influence of phonotactic constraints in phoneme perception is that of Massaro and Cohen (1983). 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 subjects in the /_li/ and

/_ri/ context. There was a significant shift in the identification functions between the two contexts demonstrating that subjects tend to hear segments that respect the phonotactics of their language (see also McClelland & Elman, 1986,Massaro & Cohen, 1991).

Notice, however, that this study only demonstrates an effect on ambiguous stimuli. We believe that it would be desirable to demonstrate the influence of phonotactics on endpoint (unambiguous) stimuli. Second, this 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.

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 two speeded ABX tasks (Experiments 2 and 3). We test the same stimuli on two populations: Japanese native speakers and French native speakers. As in English, French has complex syllabic structures and was used as a control because it typically does not exhibit epenthetic effects.

Experiment 1:

The aim of this experiment is to assess the extent of the epenthesis effect in a purely perceptual task. 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 clusters could always potentially yield an epenthetic [u] (that is, the first consonant of the cluster was never a dental stop). Subjects were then asked to decide whether or not the vowel [u] was present in the stimuli, so that no overt production of the stimuli was needed. If the epenthesis effect has a perceptual basis, Japanese subjects should be more reluctant than the French listeners to say that the [u] vowel is absent.

Method

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 either in French or 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 no vowel [u] at all (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 fout most extreme pitch periods (two on each side), and similarly, Stimulus 4, 6 pitch periods, and Stimulus 5, 8 pitch periods. Stimulus 6 was the original stimulus in which the number of pitch periods contained in the vowel [u] in each item varied from 10 to 13 (10.7 periods in average.) The average overall duration of one pitch period in the [u] vowels in each item was 9.06 msec. There were a total of 60 stimuli in one session.

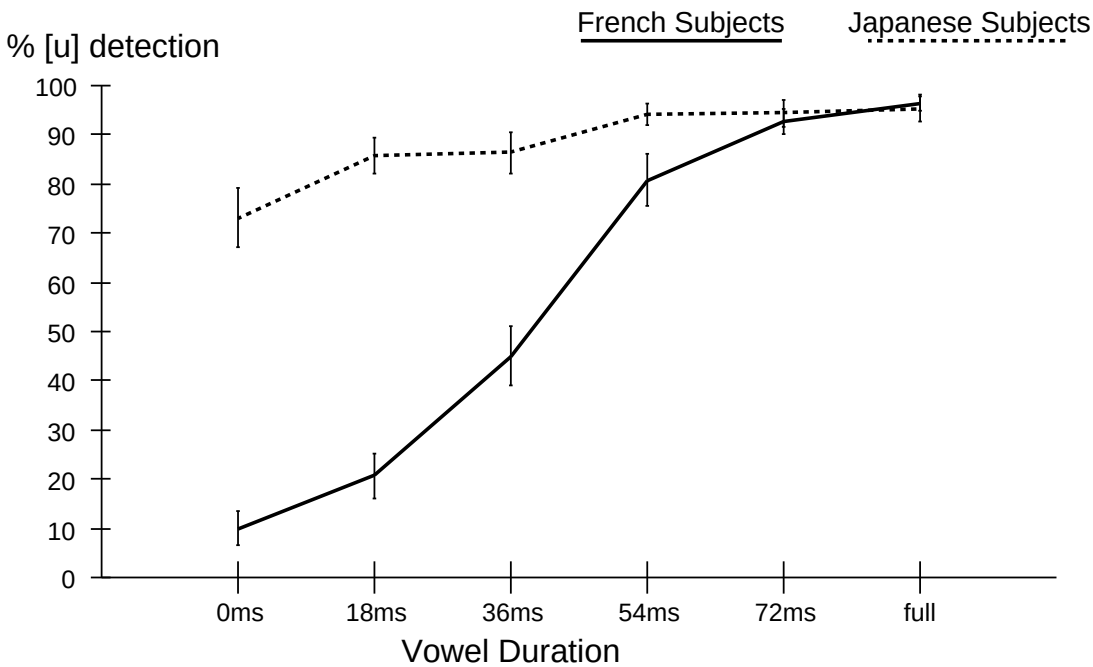

Figure 1: Percent [u] vowel judgments in stimuli like *ebuzo* in French and Japanese subjects as a function of vowel duration (Experiment 1).

Procedure Subjects 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 subjects 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 subjects in particular were told that the [u] target in the experiment was not meant to be equivalent to the kana syllabary " " which represents the single vowel [u], but rather to the phoneme [u] that was a part of another CV syllabary. Each subject heard the list three times with all the stimulus sets differently randomized each time.

The French subjects 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. Subjects were required to press the [O] key for yes ('oui') responses and the [N] key for no ('non') responses.

Subjects Ten Japanese and ten French native speakers volunteered to participate in Experiment 1. All the subjects were college students. None of the Japanese subjects had studied French and none of the French subjects had studied Japanese.

Results

The mean percentage of vowel responses as a function of language and vowel length is shown in Figure 1. We performed two analyses of variance on percentages of vowel responses, one with subjects and one with items as random variables. Language (Japanese or French) was a between-subject factor and Vowel Length a within-subject 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)=25.1, $p<.001$), with the Japanese subjects providing more Vowel responses than the French subjects. There was also a significant Vowel Length effect (MinF'(5,100)=56, $p<.001$), which had a significant linear component (MinF'(1,20)=152, $p<.001$), 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, $p<.001$), corresponding to the fact that the French subjects were much more influenced by vowel length than the Japanese. Yet, even in the Japanese subjects, the linear component of vowel length introduced a significant effect (MinF'(1,18)=11, $p<.005$).

We ran pairwise comparisons between the two languages for each vowel length. For the first three vowel lengths (0ms, 18ms, 36ms), Japanese subjects gave significantly more Vowel responses than French subjects (all MinF' $p<.001$)[2]. For the fourth vowel length (54ms), there was only a trend in the same direction (F1(1,18)=5.4, $p<.04$; F2(1,9)=3.67, p=.088). A significant difference between the two populations did not appear for the last two vowel lengths (72ms and full vowel).

Discussion

In this experiment, Japanese and French subjects judged the presence or absence of the vowel [u] in stimuli containing varying amounts of the acoustic correlate of the vowel. French subjects 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 subjects, 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 (38ms) of the vowel. In contrast, Japanese subjects predominantly judged that the vowel was present at all lvels 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 of the continuum where the vowel had been completely deleted, they still reported that the vowel was present in more than 70% of the cases.

This experiment establishes that, in a task involving no overt speech production, Japanese subjects consistently report vowel between two consonants in CC clusters. This experiment alone, however, cannot firmly establish the locus of the effect in speech perception for two reasons. First, the task requires subjects to make an explicit metalinguistic judgment: subjects 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

[2]These are significant, even if we used the Bonferoni correction and adjust the significance level to the more conservative value of .007.

can be manipulated in a metalinguistic task (see the collection of articles in Bertelson, 1986). Because the writing systems of Japanese and French differ, it is possible that it differentially affects vowels judgments in Japanese and French subjects. Second, the task did not use a speeded or on-line judgment. Therefore, it cannot identify which of the different sources of information (the orthographic code, covert production, explicit strategies) influenced the subjects' responses. For instance, it is possible that Japanese subjects were reluctant to give a vowel-absent response simply because they knew that such stimuli do not occur in Japanese.

The next two experiments were designed to address this issue. We used an ABX paradigm that only required identity judgments, thus involving no explicit or implicit mention of vowels. We also had subjects perform a speeded response, thereby reducing the use they could make of complicated response strategies.

Experiment 2:

In this experiment, we use a speeded ABX paradigm, whereby subjects hear three stimuli and have to decide whether the third stimulus is identical to the first or to the second. If the findings of Experiment 1 have no perceptual basis but are instead a by-product of metalinguistic limitations in segment manipulation, Japanese subjects should make few errors when discriminating between *ebuzo* and *ebzo*. In fact, their performance should be indistinguishable from that of French subjects. If, in contrast, the perceptual system inserts an epenthetic vowel to break up consonant clusters, Japanese subjects 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 subjects to rely on a more abstract/phonological representation rather than on an acoustic/phonetic one. Experiment 3 will specifically test the effect of talker change.

Note, however, that testing these hypotheses involves comparing the performance of different groups of subjects (that is, testing whether Japanese subjects are significantly better or worse than French subjects on a given task). Such comparisons raise many methodological issues since it is hard to match populations of subjects in all possible respects other than native language. This is why we introduced a within-population control.

This control 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. If this is so, Japanese subjects should have no problem in performing the ABX task on an *ebuzo-ebuuzo* contrast. In our stimuli, the *ebuuzo-ebuzo* contrast had the same difference in acoustic duration as the *ebuzo-ebzo* contrast (about 90ms). One can thus see our design as an extension of the continuum of vowel length used in Experiment 1.

An interesting feature of this design is that it predicts a cross-over interaction between contrast and native language. In French, vowel length has no contrastive function, i.e., no pairs of French words can be distinguished purely on the basis of the length

of one vowel. The hypothesis under examination is that language 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 subjects might have trouble in making the *ebuzo-ebuuzo* contrast whereas the Japanese should have no problem at all.

Method

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 morae. 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 store in a separate file using a waveform editor. It transpired 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. Four judges were then asked to decide whether the consonants and the vowels making up the stimuli sounded natural in their respective languages. The procedure was repeated until the stimuli were deemed fully satisfactory. The three classes of stimuli had a mean duration of 409ms for *ebzo*, 498ms for *ebuzo* and 593ms 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 in a male voice. A and B were taken from the same triplet but differed in the 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 either to A or B. The overall design was: $2 \times 2 \times 2$: Contrast $\times$ Order $\times$ 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), separated by a delay of 500ms. Subjects 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 foreign words. They were told that the third word (X) was identical to one of the first two (A or B). Their task was to press a button on their left or on their right to indicate whether X was identical to A or to B, respectively. Subjects were given 4000ms to respond. The next trial started 1000ms after each response (or after the 4000ms).

In the ten training trials, subjects 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”

Language	RT	SE	Err	RT	SE	Err
	Vowel Length Contrast			Epenthesis Contrast		
	ebuzo-ebuuzo			*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 subjects (Experiment 2).

(or B) when subjects failed to respond before the deadline. Feedback was displayed for 1000ms, and then was 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 subject. 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 (see Pallier, Dupoux, & Jeannin, 1997).

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

Results

Four ANOVAs were performed on the entire dataset: two on RT data by subject and by item and two on Error data, again by subject and by item. The ANOVAs had a 2×2 design: Language (French or Japanese) $\times$ 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, $p<.001$). This interaction was due to the fact that for the French subjects, the vowel length contrast yielded longer RTs than the epenthesis contrast (171ms, MinF'(1,19)=12, $p<.002$), whereas, for the Japanese subjects, there was a trend in the other direction (-105ms, F1(1,9)=4.5, p=.06; F2(1,15)=7.8, $p<.02$). There was no main effect of Language (F1(1,18)<1, $p>.1$; F2(1,15)=3.4, $.05<p<.1$),

and no main effect of Contrast ($F1(1,18)<1$, $p>.1$; $F2(1,15)=3.5$, $.05<p<.1$).

The analysis of the error data showed the same pattern of results. There was a highly significant interaction between Language and Contrast ($MinF'(1,26)=56$, $p<.001$). This interaction was due to the fact that for French subjects, the vowel length contrast was more difficult than the epenthesis contrast ($MinF'(1,16)=18$, $p<.001$), whereas the length contrast was easier for the Japanese ($MinF'(1,13)=35$, $p<.001$). Overall, Japanese subjects tended to make more errors than the French subjects, although this was only significant in the items analysis ($F1(1,18)=3.7$, $p=.07$; $F2(1,15)=20.1$, $p<.001$). Similarly, the epenthesis contrast tended to provoke more errors than the vowel length contrast, but again this was only significant in the items analysis ($F1(1,18)=4.1$, $p=.058$; $F2(1,15)=13$, $p<.002$).

Discussion

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

Interestingly, the difficulties of Japanese subjects with consonant clusters is mostly apparent in the Error data (32% of errors, and a 'trend' of 105ms in the latencies) whereas the Vowel Length contrast appears as difficult for the French, most apparently with the RT data (a significant 170ms slowing) and marginally in errors, (21% of errors). Whether such asymmetries reflect some kind of speed-accuracy tradeoff, or a deeper difference in the processing of these contrasts remains to be explored in further studies.

These results demonstrate that the phonotactics of a language influence speech perception with clear stimuli. That is, not only do Japanese subjects report "illusory" vowels in order to conform to their native phonology (Experiment 1), but this affects the perceived similarity (and hence the ability to discriminate) of two stimuli.

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 subjects to disregard low level acoustic characteristics and rely on a more abstract phonological representation. However, most studies of 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 subjects *can* use purely acoustic information? The next experiment addresses this issue.

Experiment 3:

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 2 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 subjects 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 consolidate the claim that, at a certain level, Japanese subjects are "deaf" to the difference between *ebuzo* and *ebzo*.

In this experiment, we will also perform analyses of two other variables in order to further characterize the robustness of the effects: the effect of training and the language background of the subjects.

The first variable we examine is the potential effect of training. Experiment 2 was rather short (15 minutes). It could be that the observed effects were due to subjects 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 266 trials, twice as many as Experiment 2. Furthermore, the lists are randomized and the blocks counterbalanced in such a way that potential sequential effects could 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 subjects should have similar results in the final part of the experiment.

The second variable is language background. We had subjects fill out a detailed biographical questionnaire concerning their language experience. We were particularly interested in the degree of fluency of our Japanese subjects 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

Materials The same materials as in Experiment 2 were used. We used the same 128 ABX experimental trials of Experiment 2 (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 2 was used.

Subjects Twenty Japanese subjects 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 subjects 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.

Subjects 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.

Side	RT	SE	Err	RT	SE	Err
	Vowel Length Contrast *ebuzo-ebuuzo*			Epenthesis Contrast *ebuzo-ebzo*		
	Japanese Subjects					
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 Subjects					
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 subjects and French Subjects (Experiment 3).

Results

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

The analysis of the RT data showed that there was a highly significant interaction between Language and Contrast (MinF'(1,53)=15, $p<.001$). This interaction was due to the fact that for French subjects, the vowel length contrast yielded significantly slower reaction times than the epenthesis contrast (117ms, MinF'(1,34)=14, $p<.001$), whereas for Japanese subjects, there was a nonsignificant trend in the other direction (-27ms, all $ps>.1$). No other interaction was significant, except the interaction between Language and Talker, which was only significant in the items analysis ($F1<1$; $F2(1,15)=16$, $p<.001$).

There was a main effect of Talker, with the same talker yielding faster RTs than the different talker (85ms, MinF'(1,52)=12, $p<.001$). There was also a main effect of Contrast, with the vowel length contrast on average yielding slower RTs than the epenthesis contrast (45ms, $F1(1,38)=7.8$, $p<.01$; $F2(1,15)=5.2$, $p<.04$). Finally, Japanese talkers tended to have longer RTs than French subjects, but this was only significant in the items

analysis (55ms, $F1<1$; $F2(1,15)=18$, $p<.001$).

The analysis of the error data showed similar results. There was a highly significant interaction between Language and Contrast ($MinF'(1,40)=34$, $p<.001$). This interaction was due to the fact that for Japanese subjects, the epenthesis contrast yielded significantly more errors than the vowel length contrast ($MinF'(1,31)=22$, $p<.001$), whereas for French subjects, there was a significant effect in the other direction ($MinF'(1,33)=8.6$, $p<.006$). No other interaction reached significance.

There was a main effect of Talker, with a different talker yielding more errors than the same talker ($MinF'(1,36)=6.5$, $p<.02$). There was also a main effect of Contrast, with the epenthesis contrast on average yielding more errors than the vowel length contrast, ($MinF'(1,35)=4.8$, $p<.04$). Finally, Japanese talkers tended to make more errors than French subjects, but this was only significant in the items analysis ($F1(1,38)=3.4$, $.05<p<.1$, $F2(1,15)=9.5$, $p<.01$).

Influence of training

We began studying the effect of training by using a correlation analysis. For each subject, the sequence of reaction times on experimental trials was cut into 16 successive bins of 16 datapoints. We found a significant negative correlation between sequential position and mean reaction time ($R^2=.67$, $F(1,14)=28$, $p<.001$). We also found a significant negative correlation between sequential position and error rate ($R^2=.67$, $F(1,14)=28$, $p<.001$). These effects show that training does have an impact, and that subjects improve their performance with time. We then computed the mean interaction between language and contrast for each sequential position. There was no significant correlation between sequential position and interaction size either in the reaction time ($R^2=.16$, $F(1,14)=2.6$, $p>.1$) or in the error analysis ($R^2=.17$, $F(1,14)=2.9$, $p>.1$).

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 data ($MinF'(1,36)=17.4$, $p<.001$).

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

We ran a new set of analyses including Response Type (A- or B-responses) as a within subject and within item factor. A-responses yielded longer RTs (127ms, $MinF'(1,52)=74$, $p<.001$) and more errors (12% vs. 5%, $MinF'(1,52)=32$, $p<.001$) than B-responses. B-responses showed a trend for less Language by Contrast interaction than A-responses, a trend which was significant for the error data (triple interaction between language, condition and Response Type: $MinF'(1,46)=22$, $p<.001$).

However, even with B-responses there was still a significant interaction between Language and Condition ($MinF'(1,50)=8.68$, $p<.005$ for the reaction times, $MinF'(1,31)=11.4$, $p<.002$ for the errors). This interaction was significant even for the same talker condition ($MinF'(1,52)=6.9$, $p<.015$ for the reaction times; $MinF'(1,29)=10.4$, $p<.003$, for the errors).

Influence of language background

Inspection of the questionnaire reveals that the Japanese subjects 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. Four Japanese subjects did not fill out the questionnaire. We separated the subjects into two groups, one labeled 'low proficiency' (7 subjects), the other labeled 'high proficiency' (9 subjects) based on the means of their own evaluation of fluency and pronunciation. 'High proficiency' subjects 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' subjects 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 best Japanese subjects with the greatest proficiency in English or French (both self-rated, and as evaluated by an external judge). The selected subjects 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 subjects, 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 subjects. They all knew English (all had learned it after age 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 separated these subjects into two groups (high and low proficiency) according to their evaluation of proficiency in these languages. We found no effect of fluency on the error data or on the reaction times ($p>.1$) which showed the same effects in the two groups.

More generally, every single Japanese subject that we tested in this experiment showed the epenthesis effect, that is, each subject showed more errors on the epenthesis contrast than on the vowel length contrast. Such regularity is also true of Experiment 2. In contrast, 18 out of 20 French subjects showed either no difference, or the opposite pattern (and 9 out of 10 in Experiment 2). In other words, the observed cross-over interaction of language and contrast in the error data is highly robust and reproducible from subject to subject, 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 2. 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 subjects 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 since 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 such cross-linguistic effects still obtained. Although training has a very powerful effect on both reaction time and error rate, it does not significantly modulate the size of the effects.

Moreover, we found that the interaction was still significant on B-responses, even in the same talker condition. This is noteworthy because subjects only need to judge the acoustic identity of two adjacent stimuli (B and X) in order to respond. Indeed, Response Type had a powerful effect on performance: A-responses generated both more errors and longer reaction times than B-responses. However, that variable did not interact with the cross-linguistic effect.

Finally, a post-hoc analysis in terms of linguistic background revealed no clear effect of fluency in a consonant cluster language 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 subjects in France or the US, we did not use extremely fluent bilinguals. Even our ”high proficiency” subjects 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 Table 1 and Table 2, the percentage of errors for the epenthesis contrast in Japanese subjects is smaller in Experiment 3 than in Experiment 2 (16% instead of 32%, a significant difference, $p<.05$). Given that, individually, neither training 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 2. 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 2, but drops to a value centered around 16% in Experiment 3. 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 3.

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 training *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 subject 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 subjects have difficulty discriminating between a stimulus that does not include a vowel (*ebzo*), and one that does (*ebuzo*). However, Japanese subjects, unlike 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 subjects discriminate (Experiment 3). 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 subjects actively use phonotactic knowledge in speech perception. This complements and extends the work by Massaro and Cohen (1993). Indeed, not only does phonotactic knowledge influence the perception of individual segments, but it can induce the perception of ”illusory” segments. Moreover, it does so in perfectly clear 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. How could such effects be accounted for? Different models provide possible answers.

Church (1987) has proposed that, in the course of speech perception, a hierarchical phonological representation is elaborated. This representation is made possible by relying on the language specific rules internalized during childhood. 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.

Norris (1994) has proposed a model of lexical recognition that uses phonemic units as input, lexical units as output, the two being linked by a recurrent network with hidden units. Such hidden units encode sequential statistical regularities of the segments. This model may have the potential to account for effects like the ones we reported in this paper. However, more studies are needed to explore the adequacy of the fit between the model and the data.

Finally, Mehler, Dupoux and Segui (1990) have proposed SARAH, a model based on an array of syllable detectors. In this model, speech sounds are chunked and categorized into syllable-sized units. The repertoire of syllables includes the totality of the syllables used in the language. In this view, an account of the epenthesis effect is quite straightforward. Indeed, Japanese subjects do not have syllable detectors that contain consonant clusters or coda consonants. Upon hearing a word like *ebzo*, they use the nearest syllabic candidate available, namely, *e+bu+zo*.

Obviously, more studies are necessary to choose among these specific models. However, our findings already allow us to pinpoint shortcomings in another class of models, namely, those that represent phonemes (McClelland & Elman, 1986; Norris, 1994), or subphonemic elements (Klatt, 1977, 1989; Marslen-Wilson & Warren, 1994) without any mention of higher order structures. In such models, no direct effect of the phonotactic organization of the language being used is expected in processing. Rather, such

effects have to arise indirectly from another information source, such as the statistical properties of the lexicon.

For instance, McClelland and Elman (1986) modeled the phonotactic effects found by Massaro and Cohen (1983) by allowing lexical feedback to bias the low-level representation of an ambiguous sound (but see Massaro and Cohen, 1991). In this top-down account, illegal sound patterns have to fight against the collective action of similar words that have a more frequent sound pattern. It is unlikely that the effects we report here can be modeled in a similar way. First, in TRACE, the sequence of phonemes is represented spatially, and thus, although lexical feedback could replace a segment by another one, it could not simply insert a segment. Second, even if this were possible, it would require that the majority of lexical items activated by the nonword *ebzo* contain a [u] sound between [b] and [z]. And this would have to be the case for each of the 16 items we used in the experiments since epenthesis was found to happen for each of these items. This is an unlikely situation.

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*), while French listeners are slow and error prone. Such effect arise, we believe, because Spanish uses accent contrastively (*bebé* vs. *bébe*), while French does not. More research is needed to understand how models can be modified to take into account such higher-order properties of signals.

Authors Notes

We thank Dianne Bradley, Susana Franck, Takao Fushimi, Peter Golato, and Takashi Otake 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 subjects and Alain Grumbach for providing access to French subjects. We also thank Franck Ramus and Evelyn Hausslein for additional help in recruiting subjects. This work was supported in part by a grant from the Human Frontiers Science Program and by a Human Capital and Mobility Program grant. During this work, C. Pallier was supported by a Lavoisier Grant from the French Ministry of Foreign Affairs and by a post-doctoral grant from the Fyssen foundation.

Mailing address: Emmanuel Dupoux, 54 Bd Raspail, 75006 Paris, France.

References

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.

Crystal, D. (1987). *The cambridge encyclopedia of language.* New York: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. *Nature*, *304*, 159–160.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, *24*, 381–410.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

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

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of english consonants. *16*(1), 1-26.

Goto, H. (1971). Auditory perception by normal japanese adults of the sounds 'r' and 'l'. *Neuropsychologia*, *9*, 317–323.

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.

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.

Jusczyk, P. W. (1994). Infant speech perception and the development of the mental lexicon. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception.* Cambridge MA.: MIT Press.

Klatt, D. (1989). Review of selected models in speech perception. In W. D. Marlsen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, Mass: MIT Press.

Klatt, D. H. (1977). Review of the ARPA speech understanding project. *Journal of the Acoustical Society of America*, *62*, 1345–1366.

Kolinsky, R., & Morais, J. (1993). Intermediate representations in spoken word recognition: A cross-linguistic study of word illusions. In *Proceedings of the 3rd european conference on speech communication and technology: Eurospeech'93* (pp. 731–734). Berlin.

Kuhl, P. (1992). Innate predispositions and the effets of experience in speech perception: the native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: speech and face processing in the first year of life* (pp. 259–274). The Netherlands: Kluwer.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from japanese listeners' perception of english [l] and [r]. *Cognition*, *24*, 169–196.

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.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives* (pp. 236–262). Cambridge, Mass: MIT Press.

Miller, G. A. (1951). *Language and communication.* New York, NY: McGraw-Hill.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*, 329–335.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1981). An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of japanese and english. *Perception & Psychophysics*, *18*, 331–340.

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

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? speech segmentation in japanese. *Journal of Memory and Language*, *32*, 258–278.

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., 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., & Werker, J. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 241–435.

Sebastian-Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in catalan and spanish. *Journal of Memory and Language*, *31*, 18-32.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in finnish. *Journal of Memory and Language*, *36*, 422–444.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*, 466–472.

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

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

Zwitserlood, P., Schriefers, H., Lahiri, A., & Donselaar, W. van. (1993). The role of the syllable in the perception of spoken dutch. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(2), 260–271.

Appendix

Materials used in Experiment 1

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

Materials used in Experiments 2, 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.