Eurospeech '99 Proceedings
ESCA 7th European Conference on Speech
Communication and Technology

PRELEXICAL LOCUS OF AN ILLUSORY VOWEL EFFECT IN JAPANESE

Emmanuel Dupoux[*], *Takao Fushimi*[†], *Kazuhiko Kakehi*[‡], *and Jacques Mehler*[*]

* Laboratoire de Sciences Cognitives, EHESS/CNRS, 54 Bd Raspail Paris 75006, France
(e-mail: dupoux@lscp.ehess.fr, web: http://www.ehess.fr/centres/lscp/)
[†]Department of Language and Cognition,Tokyo Metropolitan Institute of Gerontology, Japan
[‡] Graduate School of Human Informatics, Nagoya University, Japan

ABSTRACT

Studies in vision have demonstrated that the visual system can induce the perception of illusory contours. In this study we document a similar phenomenon in the auditory mode: Japanese speakers report perceiving vowels that are absent in the acoustic signal. Such an illusion is due to the fact that in Japanese, succession of consonants are not allowed. Hence the linguistic system inserts an illusory vowel between adjacent vowels in order to conform to the expected pattern in this language. Here, we manipulate the lexical neighborhood of non-words that contain illegal consonant clusters and show that this illusion is not due to lexical influence. Rather, it arises before lexical knowledge is activated, suggesting that phonotactics impact perception routines at a very early processing stage.

1. INTRODUCTION

Previous research has shown that our perception of speech sounds is influenced by the linguistic properties of the maternal language. In particular, recent investigations have demonstrated that the perception of phonemes is influenced by phonotactic properties of the language [1,2,3]. For instance, Japanese syllables cannot have complex onsets (except for consonant-glide onsets) and cannot have codas (except for nasal consonants and the first half of geminates). As a result, Japanese tend to transform foreign words when they import them in the language by inserting "epenthetic" vowels between consonants. For instance, the word *festival* becomes /fesutibaru/[1].

Dupoux *et al*. [1] found that this effect arises during on-line speech perception. In a vowel detection experiment, Japanese subjects report the presence of an epenthetic vowel /u/ between consonants in non-words like /ebzo/, even in stimuli with no acoustic vowel correlate within the consonant cluster. In a speeded ABX discrimination experiment, they also found that Japanese subjects have problems discriminating between, for instance, /ebzo/ and /ebuzo/. This was found even in Japanese subjects who were quite proficient in French, a language which authorizes coda consonants and complex onsets. These results suggests that the linguistic system inserts an illusory vowel between adjacent consonants in order to conform to the expected pattern in this language. In other words, phonotactics plays a role so important as to create illusory perception of segments.

However, in these experiments, it remains difficult to assess the locus of the phonotactic effects. In particular, it is possible that some of the phonotactic effects are the result of top-down lexical influences on the perception of the individual phonemes [4]. Indeed, it could be that while no word in Japanese contains the sequence /bz/, many of them contain the sequence /buz/, /baz/ or /biz/. Thus, the possible existence of real Japanese words with an inserted vowel between the consonants might have induced subjects to report a vowel that was not present in the signal. In this paper, we address the issue of the locus of the phonotactic effects: do they arise early, i.e. before lexical access, or are they due to top-down lexical influences?

2. EXPERIMENT

We constructed lists of stimuli containing consonant clusters, controlling for the presence of lexical neighbors that would incite Japanese participant to insert an illusory vowel. Specifically, the consonant clusters stimuli were disyllabic items of the shape CVCCV such that there was only a single Japanese word that could be made by the insertion of a vowel within the consonant cluster. There were two sets of such stimuli: in the first set (the u-set) the inserted vowel was the default epenthetic vowel in Japanese, i.e., /u/, (e.g. *sokdo-sokudo*). In the second set (the non-u-set) the vowel is one that never participates in vowel epenthesis in this language, i.e., /a/, /i/, or /e/ (e.g. *mikdo-mikado*).

Subjects were tested on two tasks: a transcription task and a lexical decision task. In the transcription task, subjects were presented with stimuli and asked to transcribe them in the Roman alphabet. This task was performed in order to check that the

[1] In Japanese, the inserted epenthetic vowel is always an /u/ except after dental stops where it becomes an /o/ (as in Zeitgeist -> /tsaitogaisuto/).

epenthesis effect arose with these stimuli. In the lexical decision task, subjects had to decide whether the stimuli were actual words in Japanese. The predictions were as follows: if the epenthesis effect is lexically driven, we should expect a large effect of the lexical neighborhood on the perception of the items with an illegal consonant cluster. In particular, items in the u-set (e.g. *sokdo)* should be perceived with an epenthetic /u/, and items in the non-u set items should be perceived with a non-u vowel yielding a Japanese word (e.g. *mikdo* - *mikado*). Furthermore, items in both sets should be considered as potential Japanese words with in the lexical decision experiment. In contrast, if the vowel epenthesis arises before lexical access, both *sokdo* and *mikdo* items should yield an epenthetic /u/, and as a consequence, only *sokdo* will be considered as a Japanese word. Indeed, after /u/ epenthesis, *mikdo* becomes *mikudo* which is not a Japanese word. In order to have some baseline to assess the performance of the test items, we included in the experiment control words like *sokudo* and *mikado* and non-words like *sokado* and *mikudo*.

In this experiment, the stimuli were recorded by a phonetically trained native speaker of French, who imitated words and non-words produced by a Japanese speaker. This solution was adopted because the Japanese speaker could not produce the consonant clusters accurately.

2.1. Method

Materials

Two sets of 19 triplets were constructed. In the first set (the u-set) the triplets were of the type (*sokdo, sokudo, sokado*). The first element in the triplet was a disyllable containing an illegal consonant cluster in Japanese. The other two elements were identical to the first except for a vowel inserted between the consonants. The second element of the triplet was always a word and had the default epenthetic /u/ inserted, whereas the last element was a non-word and had one of the Japanese non-epenthetic vowels, /a/, /e/, or /i/. The second set (the non-u-set) contained triplets of the type (*mikdo, mikado, mikudo*). These triplets were similar to those of the u-set, except that the words were produced with the insertion of a non-epenthetic vowel, and the non-words were produced by the insertion of /u/. In all cases, there was only one possible way to make a word in Japanese through the insertion of a vowel in the consonant cluster of the first element of a triplet. An additional list of 78 filler items was also constructed, half words, half legal non-words.

The stimuli were recorded by a native speaker of Japanese and a native speaker of French in the following way: The Japanese speaker first read a given triplet, and the French speaker repeated the same triplet, trying to imitate the Japanese items in segmental and suprasegmental features, except for the cluster condition, where it was required to really produce the consonant cluster. Each triplet was recorded three times, and the best tokens were selected by a French and Japanese listener. The stimuli were then digitally recorded (16kHz). Of these, 5 triplets had to be removed because either the word or the non-word member of a triplet were considered ambiguous, badly pronounced, or of the wrong lexical category by a native Japanese phonetician. This left 14 triplets in the u-set and 14 triplets in the non-u-set (the items are listed in the Appendix). Similarly, 10 fillers were removed because of bad pronunciation, leaving a total of 68 fillers.

Procedure

For the phonetic transcription task, the items were presented in a randomized order over headphones. Subjects were instructed to type the transcription in the Roman alphabet of the items by typing them on the computer's keyboard.

For the lexical decision, the items were presented in a randomized order over headphones using the EXPE software [5] on a PC compatible. Subjects were instructed to classify the stimuli into real Japanese words versus non existent words as fast as they could. If no response was given within a 4 seconds deadline after the presentation of an item, the next stimulus was presented. Subjects were given a practice session of 10 trials with feedback, and then started in the main experiment with no feedback.

Participants

Fifteen native speakers of Japanese were recruited and ran on the lexical decision experiment. Five Japanese subjects performed the phonetic transcription task. They were all Japanese volunteers recruited in Paris.

2.2 Results

Two subjects had more than 50% errors in the non-words (both for distractors and test items) and were removed from subsequent analysis.

First, the phonetic transcription results for the consonant cluster items were analyzed separately for items in the u-set and items in the non-u-set. Consonant cluster items in the u-set produced 83% of transcription with an epenthetic /u/ inserted between the consonant. Only 1% of the responses contained a different vowel (/ou/ instead of /u/ in one item). In the remaining 16% of the cases, subjects transcribed a consonant cluster. In the non-u-set, there were 74% /u/ responses, 10% of a different vowel (the vowel /i/), and 16% of a consonant cluster response. The /i/ response was localized in two items: *rekSi* and *rikSi*. It is interesting to note that the /i/ response turns these two items into a real Japanese word (*rekiSi* and *rikiSi*, respectively). Note also that in these items, there were still 20% of the /u/ responses.

Table 1: percent "word" response to cluster items in the lexical decision experiment.

	Nonwords	Words	Clusters
u-set	6.0%	91.8%	70.4%
(example	*sokado*	*sokudo*	*sokdo)*
non-u-set	10.4%	96.1%	18.7%
(example	*mikudo*	*mikado*	*mikdo)*

Second, the lexical decision data was analyzed in terms of percent 'word' response. The data is displayed in Table 1. As one can see, words are labeled as words (93% of the cases), and non-words are not (8% of the cases). The results of the cluster words depends on the type of set. Cluster items in the u-set are labeled as words in 70% on average, a score significantly different from 50% ($p<.007$). Planned contrasts showed that this score was significantly different from the score with the control non-words ($F1(1,12)=103$, $p<.001$; $F2(1,13)=108$, $p<.001$) and with the control words ($F1(1,12)=12$, $p<0.003$; $F2(1,13)=13$, $p<0.001$). In contrast, the cluster items in the non-u set are classified as words in only 18.7% on average (significantly below 50%, $p<.001$). This score was significantly different than the score with the control words ($F1(1,12)=341$, $p<.001$; $F2(1,13)=109$, $p<.001$), and than the control non-words although the difference is only significant in the participants analysis ($F1(1,12)=11.7$, $p<.005$; $F2(1,13)=1.7$, $p>.1$).

Third, we analyzed the RTs for the responses in the dominant category, that is, responses as 'word' for cluster items in the u-set and responses as 'non-word' for items in the non-u-set. These responses are shown in Table 2, together with the RTs for the control words and non-words. An ANOVA analysis revealed that the control words yield significantly faster latencies than control non-words ($F1(1,12)=16$, $p<.002$; $F2(1,13)=91$, $p<.001$). Cluster items in the u-set did not differ significantly from control words (both $Fs<1$), but yielded significantly faster RTs than control non-words in the participant's analysis ($F1(1,12)=4.7$, $p<.05$; $F2(1,13)=2.7$, $p>.1$). In contrast, cluster items in the non-u-set yielded slower RTs than the control words ($F1(1,12)=28.0$, $p<.001$; $F2(1,13)=56.5$, $p<.001$), but did not differ significantly from the control non-words ($F1(1,12)=1.9$, $p>.1$; $F2(1,13)=3.0$, $p>.1$). In other words, the cluster items were responded to with latencies similar to that typical of the lexical category that they were assimilated to.

In the above analysis of the transcription task, we found two items in the non-u-set that yielded the identification of /i/ instead of /u/ (namely, *rekSi* and *rikSi*). A post-hoc analysis showed that the lexical decision responses for these two items were 77% and 54% of "word" response, respectively, whereas all the other items in the non-u-set yielded a majority of non-word responses. Could it be that these two items reflected a lexical influence on the epenthesis effect? The insertion if /i/ could indeed be caused by the presence of the Japanese words *rekiSi* and *rikiSi* which contain an /i/ between the two medial consonants. This would cause us to reconsider our above interpretation. However, note that none of the other items in the non-u-set showed this behavior. Despite the presence of lexical items with non-u vowels between the medial consonant, all the other items showed a consistent perception of /u/ and a non-word response. An alternative interpretation for these two items could rest on the observation that in fact cases of /i/ epenthesis in Japanese have been reported in Japanese. Shinohara [6] discusses some of these cases, and while she argues that /i/ epenthesis may not be a productive phenomenon in Japanese, it is worthwhile noting that the majority of the existing cases of loan words with /i/ insertion arise in the context of voiceless stop-fricative consonant clusters (i.e. *textile* -> /tekisutairo/), hence the same context as in *rikSi*. Further research will find out whether these cases are bona fide prelexical epenthesis effects or lexical influences.

3. DISCUSSION

In this experiment, we presented subjects with items containing illegal Japanese consonant clusters. Half of these items produced a word after the insertion of an epenthetic /u/ and half produced a non-word. In the transcription task, we found that the perception of an epenthetic /u/ was very similar in the two sets. Furthermore, in a speeded lexical decision, these items were consistently classified as words and non-words, respectively, and the insertion of the epenthetic /u/ did not slow down the reaction time compared to control items. In brief, the data obtained supports the interpretation of epenthesis as a prelexical effect, i.e., an effect that arises during the encoding of speech sounds before lexical access occurs. These results are consistent with an electrophysiological finding by Dupoux and colleagues [7] who reported that whereas French participants show an early MMN response to a change

Table 2: Reaction times (ms), standard error, and percent error to cluster and control items in the lexical decision experiment.

	Non-words			Words			Clusters		
	RT	SE	Err	RT	SE	Err	RT	SE	Err
u-set	1231	78	6.0%	1055	40	8.2%	1084	37	29.1%
non-u-set	1241	64	10.4%	949	34	3.8%	1323	86	18.7%

between ebzo and ebuzo, Japanese participants show no such MMN response. In brief, the obtained results support models that posit the presence of a language-specific representation that is intermediate betwen an acoustic representation and the lexicon. Such language-specific representation not only encodes phonetic categories [8], but also takes into account larger-scale properties the speech signal, such as syllable structure [9,10,11].

4. ACKNOWLEDGMENTS

We thank Nicolas Bernard for his help in preparing the stimuli, recruiting the participants and running the experiment. We thank Sharon Peperkamp and Christophe Pallier for useful comments.

5. APPENDIX

Items in the u-set

bakro bakuro bakaro
harka haruka haraka
kaksa kakusa kakasa
kokdo kokudo kokado
kokmu kokumu kokamu
kokso kokuso kokaso
kokti kokuti kokiti
komgi komugi komigi
magro maguro magaro
sokdo sokudo sokido
tukSi tukuSi tukeSi
yakba yakuba yakaba
yakza yakuza yakaza
yursi yurusi yurisi

Items in the non-u set

huksa hukasa hukusa
kikme kikime kikume
mikdo mikado mikudo
namda namida namuda
nikbi nikibi nikubi
reksi rekisi rekusi
sakba sakaba sabuka
riksi rikisi rikusi
sekri sekiri sekuri
sikti sikiti sikuti
takra takara takura
taksa takasa takusa
wakme wakame wakume
waksa wakasa wakusa

6. REFERENCES

[1] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (in press). Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology: Human Perception and Performance.

[2] 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.

[3] Massaro, D. W., & Cohen, M. M. (1983). Phonological constrainst in speech perception. Perception & Psychophysics, 34, 338-348.

[4] McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

[5] Pallier, C., & Dupoux, E. (1997). Expe: An expandable programming language for on-line psychological experiments. *Behavior Research, Methods, Instruments and Computers*, **29**, 322-327

[6] Shinohara, S. (1997). *Analyse phonologique de l'adaptation japonaise de mots étrangers.* Unpublished doctoral dissertation, Université de la Sorbonne Nouvelle.

[7] Dehaene-Lambertz, G., E. Dupoux and A. Gout (Submitted) Electrophysiological correlates of phonological processing: a cross-linguistic study.

[8] Best, C.T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model., In Goodman, J.C. and Nusbaum, H.C. (Ed), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, (pp. 167--224). Cambridge: MIT Press.

[9] Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, **29**, 316-335.

[10] Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 236-262). Mass: MIT Press.

[11] Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional Allocation within the syllabic structure of Spoken Words. *Journal of Memory and Language*, **32**, 373-389.