PHONOTACTICS IN THE PERCEPTION OF JAPANESE VOWEL LENGTH: EVIDENCE FOR LONG-DISTANCE DEPENDENCIES

Elliott Moreton
NTT Communication Science Laboratories and
University of Massachusetts, Amherst
elliott@linguist.umass.edu

Shigeaki Amano
NTT Communication Science Laboratories
amano@av-hp.brl.ntt.co.jp

ABSTRACT

These experiments investigate whether the perceptual boundary between [a] and [a:] is affected by the differing phonotactics of the Sino-Japanese and Foreign strata of the Japanese lexicon. We presented a range of edited natural vowels from [a] to [a:] at the end of carrier nonwords of the form [CoC'__] and asked subjects to judge whether the final vowel was long or short, while C and C' were varied to make 9 carriers ranging from very Sino-Japanese-like to very Foreign-like. The perceptual boundary between [a] and [a:] was affected by both C and C', shifting through about 20 msec—a larger and more robust effect than was obtained in a word-superiority experiment with the same paradigm and subjects. These results cast doubt on explanations of phonotactic effects based on lexical-activation spreading. The influence of C at a distance of three phonemes from the ambiguous segment cannot be explained by simple segment-to-segment transitional probabilities.

1. INTRODUCTION

1.1 Background

This paper is concerned with the effect discovered by Massaro and Cohen [1]: when a sound acoustically intermediate between two phonemes is placed in a phonetic context where one of them is phonotactically illegal, it tends to be heard as the legal one.

The source of this effect is a question of current interest. Proposals tend to fall into three main groups. *Lexical models* like TRACE [2] attribute it to the influence of lexical knowledge, attested sequences being facilitated by partial activation of the lexical items in which they occur. *Statistical models* hold that the mechanisms of speech perception are sensitive to frequency, and that effects of phonotactic "impossibility" can be analyzed as effects of zero frequency. Proposals as to the nature of the frequency sensitivity include segment-to-segment transitional probabilities [3, 4] and sublexical phoneme sequences [5]. Unlike these models, which take phonotactics to be an emergent phenomenon, *grammatical models* claim that competing categorical, rule-like prohibitions of varying strengths are part of a speaker's linguistic knowledge [6] encode phonotactics directly, and are consulted by perceptual mechanisms [7].

This paper reports two experiments designed to test these claims. Experiment 1 replicates a well-established word-superiority effect originally due to Ganong [8] in order to establish the sensitivity of our stimulus-presentation paradigm and to estimate the size of an undoubted lexical effect which we can then compare with that of the second experiment.

Experiment 2 uses the same paradigm and subjects to measure the ambiguous-phoneme perception effect of phonotactics. Previous investigations of the Massaro-Cohen effect have studied the effects of immediately adjacent phonetic context on the perception of an ambiguous sound. Here, we also consider the effect of a context which is not immediately adjacent to the ambiguous sound, but which is still linked to it indirectly through phonotactics.

1.2. Lexical strata and Japanese phonotactics

Owing to its historical development, the Japanese lexicon is structured in such a way that a word which contains certain sounds is virtually sure to lack certain others.

The vocabulary of the Japanese language is divided into four "strata", each corresponding to a different historical source of lexical items. These strata are shown in Table 1.

Table 1. The lexical strata of Japanese

Stratum	Origin	Examples
Foreign	Recent borrowing (usually European)	*kompyuuta, sutoraiku*
Sino-Japanese	Chinese languages	*byooki, heiwa*
Yamato	Antiquity	*miya, kami*
Mimetic	Onomatopoeia	*girigiri, goñ*

The strata are synchronically identifiable by two phenomena in the spoken language. First, they are morpheme co-occurrence classes; that is, in forming compounds, morphemes tend to combine only with others of the same stratum [9]. Second, the strata are different phonologically; there are sounds and sound sequences which occur only in one stratum, or which are lacking only in one stratum. For example, [a:] is found in Foreign, but not in Sino-Japanese (vowel length is distinctive in Japanese) [10, 11].

The absolute phonotactics of the strata can be viewed as a "core-periphery" phenomenon: strictest for the indigenous Yamato stratum, less so for the newer Sino-Japanese words, and most permissive for the recent Foreign loans [12, 13].

However, there are many configurations in Sino-Japanese which, though theoretically permitted in the Foreign stratum, are virtually never actually found there [14]. For instance, the palatalized onsets [ry] and [hy] are almost nonexistent outside of Sino-Japanese.

The result is to create long-distance statistical dependencies. For example, a word containing [ry] or [hy]

is almost certain to be Sino-Japanese, and hence to lack [a:]—irrespective of position in the word.

Stratum phonotactics provide redundancy which listeners could in principle exploit in speech perception. These experiments are intended to see whether they in fact do, and, if so, how.

2. EXPERIMENT 1

It was shown by Ganong [8] that, when a sound acoustically in between two phonemes is presented in a phonetic context where only one of the two phonemes makes a word, it tends to be heard as the one that makes the word. Ganong's experiment was done in English with ambiguous stop consonants; we replicate it in Japanese with vowels of ambiguous length.

2.1 Stimuli

Stimuli were the three pairs of real Japanese nouns shown in Table 2. All were judged to be of very high familiarity by the subjects of Amano, Kondo, & Kaheki [15].

Table 2. Stimuli for Experiment 1

Word	Stratum	Gloss
foroo	Foreign	'follow-up'
puro	Foreign	'professional'
posutaa	Foreign	'poster'
pasuta	Foreign	'pasta'
syuugo	Sino-Japanese	'meeting'
ringo	Sino-Japanese	'apple'

One member of each pair ended in a long vowel, the other in the short version of the same vowel. Both members of each pair had the same accent pattern.

Each word pair thus provided one context expected to bias listeners to report a long vowel, and one expected to bias them towards a short vowel. Following Ganong [8], we expected the [a]-[a:] boundary to be closer to [a] following *posut-* , since *posutaa* is a real word (and *posuta* isn't), and closer to [a:] after *pasut-*.

A male native speaker of Japanese recorded several tokens of each word in isolation. These were digitized at 48kHz with 16-bit resolution, downsampled to 44.1kHz, and normalized for peak amplitude.

Using a waveform editor, a single token each of the final syllables—[ro:], [ta:], and [go:]—was excised and cross-spliced with each of the initial syllables to get three pairs of stimuli with each pair ending in the same long vowel.

2.2 Participants and procedure

Experimental subjects were 24 native speakers of Japanese, post-secondary students in the Tokyo area, with equal numbers of each sex. Subjects reported normal speech and hearing. They were paid for their participation.

Subjects were tested 8 at a time in a sound-treated room in the presence of the experimenters. Stimuli were presented diotically through Sennheiser headphones at a peak intensity of 70dB SPL. Subjects responded by mouse-clicking on one of two buttons displayed on the screen. The buttons were labelled in the *katakana* syllabary (used primarily for writing Foreign loanwords); one showed the stimulus word with a short final vowel, the other with a long one (e.g., "ringo" and "ringoo"). Subjects were asked to choose the button which better matched what they had heard. The next stimulus followed after 0.8 seconds.

To vary the length of the final vowel, a quarter-wave cosine filter was applied. This caused the final vowel to be gradually reduced to zero amplitude during the 50 ms following the specified starting point of the filter, providing a natural-sounding end to the vowel..

The boundary between the long and short vowel was established using the adaptive method PEST [16]. One of the 6 stimuli was randomly selected, and presented with either the longest or the shortest possible final vowel. On the next trial, the opposite endpoint was presented. Thereafter, the stimulus presented depended on the subject's responses to previous stimuli, as specified by the PEST rule, so as to zero in on the boundary. The series ended when the step size was reduced to 1/256 of the continuum—i.e., about 1 ms—or after 80 trials. To reduce the dependence of each trial on the previous one, two PEST series for the same context were interleaved with each other, switching back and forth randomly. Once both series finished, the screen cleared and a button appeared with the message "Click to go on to the next sound". This procedure took place once for each of the 6 stimuli.

2.3 Results and discussion

Three subjects failed to converge after 80 trials on 6 or more series (in both experiments together) and were excluded from the analysis, leaving 21 valid subjects, who converged on 96% of all series after an average of 20.4 trials. Both experiments together, plus the post-experiment questionnaire, took about two hours.

For each subject, an [a]-[a:] or [o]-[o:] boundary was computed for each series. The boundary was the stimulus that would have been presented by PEST if there had been one more trial. Since two interleaved series were presented for each carrier context, the subject's boundary for that stimulus was taken as the average of the two.

For each subject and each pair of words, we calculated the difference between the boundary in the long-bias context and that in the short-bias context. Results are shown in Table 3.

Table 3. Boundary difference in long- and short-biased contexts (in milliseconds), Experiment 1 (N = 21 Ss).

Long-bias context	Short-bias context	Boundary difference		se
for-	pur-	-4.22		4.92
posut-	pasut-	12.21	*	5.45
syuug-	ring-	9.04		4.92

Note: Asterisk represents 5% significance.

There were large amounts of individual variation in the differences, leading to large variances and low significance levels. The difference was significantly different from zero in the *posut-/pasut-* context (one-tailed t-test, $t(20) = 2.086$, $p < 0.05$). The *syuug-/ring-* difference just misses significance at the 5% level ($t(20) = 1.725$). No effect could be found for the *for-/pur-* pair.

This experiment partially replicated the lexical bias in ambiguous-phoneme perception of Ganong [8]. A robust word-superiority effect, however, was not found. The weak showing of this well-established lexical effect is all the more striking when compared with the strong results of the next experiment.

3. EXPERIMENT 2

We use the Massaro-Cohen paradigm to test whether stratum phonotactics can shift the phonetic boundary between two sounds—specifically, whether the different phonotactics of the Foreign and Sino-Japanese strata can shift the boundary between word-final [a] and [a:].

The idea is to present an [a]-[a:] continuum at the end of a carrier word containing two consonants, C and C'. If C and C' are sounds that occur almost exclusively in Sino-Japanese words, then they do not naturally co-occur with [a:]. We therefore expect that listeners will require exceptionally strong acoustic evidence (i.e., a longer vowel) before reporting [a:]. On the other hand, if C and C' are sounds found only in Foreign words, then it is unsurprising if the word contains [a:], and we expect that listeners will be readier to report the long vowel. In other words, we predict that the [a]-[a:] boundary will be shifted towards [a:] when C and C' are highly valid cues for Sino-Japanese, and that it will be shifted towards [a] when they are highly valid cues for Foreign.

3.1 Stimulus design

Cue validity and co-occurrence statistics were computed based on a preliminary (pre-release) updated version of the Japanese-language database of Amano, Kondo, and Kakehi [15], by exploiting a feature of the Japanese writing system: the lexical stratum of a word is reflected in its orthography. Foreign words are written in the *katakana* syllabary. Chinese characters are used to write both Sino-Japanese and Yamato words, but the dictionary indicates which pronunciations of a given character are Sino-Japanese and which are Yamato. There are some exceptions, but they are not numerous. (Details of the classification procedure are given in [14].)

Cues to stratum membership were chosen on the basis of these statistics. For the Foreign stratum we chose as C the singleton [p] (i.e., [p] neither geminated nor preceded by a nasal coda) and as C' the voiceless bilabial fricative [□] (orthographically <f>), which outside of the Foreign stratum is found only before [u]. For the Sino-Japanese stratum we chose [ry] and [hy]. For neutral contexts we chose [r] and [t]. The statistics are shown in Table 4.

Table 4. Nouns containing each cue, by stratum

Cue	Nouns in database containing cue: Sino-Japanese	Foreign	*Other*
Foreign (F) cues			
C [p] singleton	1	812	*83*
C' [□] before [i e a o]	0	214	*16*
Neutral (N) cues			
C [r]	2917	2922	*6530*
C' [t]	4068	1683	*4336*
Sino-Japanese (SJ) cues			
C [ry]	959	19	*109*
C' [hy]	273	7	*38*

Note: "Other" includes Yamato, Mimetic, mixed compounds, and words not classifiable by the orthographic algorithm. Nouns make up 82% of the database.

3.2 Stimulus manufacture

The procedure was almost identical to that of Experiment 1. The same speaker (naive to the purpose of the experiment) was recorded on digital audio tape producing the cues embedded in the context [CoC'a:], with a high pitch on the [o] and a low one on the [a:]. This accent pattern was chosen because it is common in both the Sino-Japanese and Foreign strata, and because it rules out the possibility of a word boundary inside the stimulus. The 9 possible combinations of C and C' yielded the following nonwords:

[po□a:]	[ro□a:]	[ryo□a:]
[pota:]	[rota:]	[ryota:]
[pohya:]	[rohya:]	[ryohya:]

These were digitized and normalized as before. Using a waveform editor, single tokens of [po], [ro], and [ryo] were selected and excised. One of the [o]s was chosen and spliced in to replace the original [o] of the other two, resulting in [po], [fo], and [ryo] tokens with identical vowels.

A single token of [a:] was selected and lengthened to 250ms by repeating medial pitch periods. One token each of [fa:], [ta:], and [hya:] was chosen, and the [a:] was cut out at the first zero crossing following the fourth pitch period of the vowel (early in the steady state). The 250-ms [a:] token was spliced onto the end of each one to produce [fa:], [ta:], and [hya:] tokens with identical extra-long vowels.

3.3 Participants and procedure

The subjects and procedure were as in Experiment 1. A short break separated the two experiments.

At the end of the experiment, subjects received written questionnaires. They listened as many times as they wished to the longest [a:] and the shortest [a] in each carrier context (in a random order for each subject) and transcribed the resulting pseudo-word in *katakana*, then answered questions about it, including whether they knew it as a real word of Japanese.

3.4 Results and discussion

The same subjects were excluded as in Experiment 1. For each valid subject, an [a]-[a:] boundary was computed for each series as in Experiment 1. The boundary for each context, averaged across subjects, is shown in Figure 1.

It is apparent that the boundary stimulus tends to become longer as the consonantal cues go from Foreign (F) to Sino-Japanese (SJ). A by-subjects two-factor ANOVA shows that manipulations of both C and C' had very significant effects on the boundary location. The C manipulation caused a 7.25ms shift ($F(2, 40) = 6.473$, $p < 0.01$); the C' manipulation caused a 12.6ms shift ($F(2, 40) = 11.529$, $p < 0.01$). Their interaction did not reach significance ($F(4, 80) = 0.714$).

Questionnaire results confirmed that the subjects heard the stimuli correctly; no stimulus was misheard by more than 3 subjects. One stimulus, [pota], was judged to be a real word by 15 subjects (it is part of the reduplicated Mimetic *potapota*). One other was judged a real word by 2 subjects, and 5 others by 1 subject.

The long-distance effect of the C manipulation is unexpected under the statistical theories based on segment-to-segment transitional probabilities, since C is three segments away from the ambiguous sound. Such theories can be adapted to keep track of longer-distance dependencies, but at the cost of greatly enlarging the component which keeps track of the probabilities. If this

component is to keep a record of, e.g., the probability that a word beginning with [p] ends with [a:], it begins to resemble a second lexicon—favoring the lexical models.

However, the combined effect of the phonotactic manipulations in this experiment was numerically larger and statistically more robust than the word-superiority effect of Experiment 1, obtained with different stimuli but the same subjects and paradigm. This is unexpected under lexical theories like TRACE: the activation spreading from a highly activated word should be much more than that from a few partially activated overlapping lexical items.

The simplest grammatical account would be that C and C' cue stratum membership, and the Sino-Japanese stratum phonotactics absolutely forbid [a:], thus shifting the boundary towards [a:] when the cues are Sino-Japanese. However, since [po] and [□a] are actually illegal (not just rare) in Sino-Japanese, they should be perfect cues to Foreign stratum membership. All carriers containing either one should therefore produce the same results—contrary to what was actually found.

4. CONCLUSION

We have shown that phonotactics of the Japanese lexical strata can have strong, robust effects on the perception of vowel length, even when the triggering segment is three segments away from the ambiguous segment. The combined phonotactic effect is both numerically larger and statistically more robust than a word-superiority effect obtained with the same subjects and paradigm. This suggests that the Massaro-Cohen effect cannot be entirely explained by lexical activation spreading.

The present results are also hard to accommodate under the simplest theories based on pure corpus statistics or grammar. Further research is planned to disentangle the contributions of these components to the overall syndrome of effects which have traditionally been grouped together as just "phonotactics".

The most pressing need is a direct attempt to dissociate effects of long-range co-occurrence probability from those of stratum membership. There are consonants that are equally good stratum cues, but differ in the probability that a following [a] will be long; conversely, there are those which differ in validity as stratum cues but are similar in their purely statistical effect on the likelihood of a following [a] being long. Thus, these factors can be varied independently to test the statistical view against the grammatical. The next step is to test the lexical theory more directly, using a pair of experiments like these in which the two stimulus sets (words and nonwords) are made from the same materials. This would allow direct comparison of the effect sizes between the two experiments, and an analysis of how the effect sizes in the two experiments correlate across individual subjects.

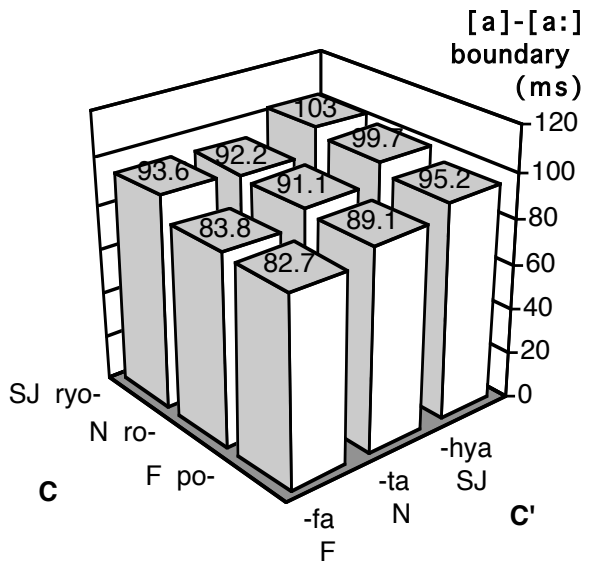

Figure 1. The [a]-[a:] boundary (in ms) for each carrier context [CoC'_]. Standard errors range from 4.17ms to 7.45ms (N=21 Ss). SJ = Sino-Japanese cue, N = neutral cue, F = Foreign cue.

REFERENCES

[1] D. W. Massaro & M. M. Cohen (1983). Phonological context in speech perception. *Perception & Psychophysics* 34(4):338-348.

[2] J. L. McClelland & J. L. Elman (1986). Interactive processes in speech perception: The TRACE model. In: D. E. Rumelhart & J. L. McClelland (eds.): *Parallel Distributed Processing*, Vol. II. Cambridge, MA: MIT Press.

[3] R. W. Brown & D. C. Hildum (1956). Expectancy and the perception of syllables. *Language* 32:411-419.

[4] M. A. Pitt & J. M. McQueen (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39(3):347-370.

[5] M. S. Vitevich & P. Luce (to appear). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*.

[6] A. Prince & P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. MS, Rutgers University Department of Linguistics.

[7] E. Moreton (1999). Evidence for phonological grammar in speech perception. *Proceedings of the 14th International Congress of Phonetic Sciences*.

[8] W. F. Ganong (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1):110-125.

[9] M. Shibatani (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.

[10] K. Tateshi (1990). Phonology of Sino-Japanese morphemes. In: G. Lamontagne & A. Taub (eds.), University of Massachusetts Occasional Working Papers in Linguistics (UMOP) 13:209-235. Amherst, MA: Graduate Linguistic Students' Association.

[11] S. E. Martin (1952). Morphophonemics of standard colloquial Japanese. *Language* 28(3, Part 2):1-113.

[12] J. Ito & R. A. Mester (1994). Japanese phonology. In: J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Cambridge, MA: Blackwell.

[13] J. Ito & R. A. Mester (1995). The core-periphery structure of the lexicon and constraints on reranking. In: J. N. Beckman, L. Walsh Dickey, & S. Urbanczyk (eds.), University of Massachusetts Occasional Working Papers in Linguistics (UMOP) 18:181-209.

[14] E. Moreton, S. Amano, & T. Kondo (1998). Statistical phonotactics of Japanese: Transitional probabilities within the word. *Transactions of the Technical Committee on Psychological and Physiological Acoustics, Acoustical Society of Japan* H-98-120.

[15] S. Amano, T. Kondo, & K. Kakehi (1995). Modality dependency of familiarity ratings of Japanese words. *Perception and Psychophysics* 57(5):598-603.

[16] M. M. Taylor & C. D. Creelman (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America* 41:782-787.