LISTENERS' REPRESENTATIONS OF WITHIN-WORD STRUCTURE: A CROSS-LINGUISTIC AND CROSS-DIALECTAL INVESTIGATION

Takashi Otake
Dokkyo University, 1-1 Gakuen-cho, Soka-shi, Saitama 340, Japan.
Sally M. Davis
Dept. of Psychology, University of Pennsylvania, Philadelphia 19104, USA
& Anne Cutler
Max-Planck-Institute for Psycholinguistics, PO Box 310,6500 AH Nijmegen, The Netherlands.

ABSTRACT

Japanese, British English and American English listeners were presented with spoken words in their native language, and asked to mark on a written transcript of each word the first natural division point in the word. The results showed clear and strong patterns of consensus, indicating that listeners have available to them conscious representations of within-word structure. Orthography did not play a strongly deciding role in the results. The patterns of response were at variance with results from on-line studies of speech segmentation, suggesting that the present task taps not those representations used in on-line listening, but levels of representation which may involve much richer knowledge of word-internal structure.

1. INTRODUCTION

There are cross-linguistic differences in the ways in which listeners segment continuous speech. A great deal of recent research has investigated this issue, which is of considerable practical importance: although human listeners appear effortlessly to recognise words in a natural utterance environment, the segmentation of an utterance into its lexical components still provides great problems for automatic speech recognition systems. The evidence from experiments with human listeners now suggests that processing procedures which work well for one language may not be of use for another. Whereas English-speakers, for example, exploit stress units in segmenting continuous speech into words [1,2,3,4], French listeners can use a syllabic segmentation procedure [5,6] and Japanese listeners a segmentation procedure based on the mora, a subsyllabic unit which may be a CV, a vowel, or a syllabic coda [7,8].

The above evidence is largely based on techniques which measure listeners' response time (RT) to perform some task (detection of a target sound, detection of a real word in a nonsense string, or similar). RT tasks are aimed at studying speech processing "on line", and the processes underlying patterns of response in such experiments, it is generally agreed, are not open to conscious inspection by the listeners who take part as experimental subjects. However, listeners can of course form conscious representations of potential within-word segmentations. We here examine whether the conscious segmentations generated by listeners are the same as the segmentations used in on-line listening. In other aspects of language processing, there is reason to believe that listeners' conscious representations may be different from - and perhaps richer than - the representations used in on-line processing. For example, the syllable appears not to play a role in on-line segmentations by English listeners [6], but language users' conscious manipulations of English words in language games and in laboratory tasks show clear evidence that syllables are represented in the lexicon [9].

In the three experiments described below, we use a segmentation task [10] which aims to capture listeners' immediate conscious perceptions of within-word structure. We compare two languages: Japanese and English, and within English, two dialects: British and American. The Japanese-English comparison allows us to compare the representations evident from the conscious responses elicited by this task with the known role of different structural regularities in RT results from both languages. It further allows us to examine effects of cross-language differences in permissible syllable structure; in particular, CVCV sequences in which the second vowel is reduced may include an ambisyllabic consonant in English but not in Japanese [11]. The British-American comparison allows us to test cross-dialect differences; ambisyllabicity, for instance, may be weaker in American than in British English [11].

2. METHOD

2.1 Subjects

The subjects in the Japanese experiment were 40 students at Dokkyo University, in the British experiment 33 students of Cambridge University, and in the American experiment 27 students at the University of Pennsylvania. None had any hearing deficit. The subjects were rewarded either with a small payment or with course credit for participating in the study.

2.2 **Materials**

The words chosen as stimuli varied in phonological structure, in the same way as the stimuli used in the reaction-time experiments cited above had varied. The

English materials used for both dialect groups comprised 64 test words, of which 20 began with a CVCV structure with a strong first and weak second syllable (e.g. *canopy, tonic),* 20 began CVCC with a strong first and weak second syllable (e.g. *cancel, destitute),* twelve had two strong syllables (SS, e.g. *canteen, pastime)* and twelve had a weak first and strong second syllable (WS, e.g. *contend, detract).* The Japanese materials comprised 72 words, of which 24 began CVCV (e.g. *norimono, tokorode, kamera),* 24 began CVCC in which the first syllable had a nasal coda (e.g. *kenri, nonbiri, tanku)* and 24 began CVCC in which the first syllable coda was a geminate consonant (e.g. *tosshin, katto, nokku;* in these words the intervocalic consonant - respectively [j], [t], [k] - is doubled and is effectively both coda of the initial syllable and onset of the second syllable). Within each set of 24, eight words had standard orthographic representation in kanji characters, eight words in hiragana characters (used for function words, in this instance adverbs) and eight words in katakana characters (used with loan words from foreign languages). Kanji, or Chinese characters, are not phonologically transparent; the two other (kana) orthographies directly represent mora structure. The three examples of each structure given above represent these three orthographic types.

2.3 Procedure

For each language/dialect, the words were recorded on tape in list sequence and presented to listeners through headphones. Listeners were tested individually or in pairs. The two English listener groups heard all the English stimulus words in a single randomised list. For the Japanese stimuli, however, there were three separate lists for the three phonological structures, and each list contained 24 filler words as well as the 24 experimental items. The CVCV- and geminate coda lists of Japanese stimuli were heard by 40 subjects, the nasal coda list by 17 subjects.

All listeners were provided with a transcript of the words in order of presentation; the words were printed in normal orthography for the English groups, and in Roman characters for the Japanese group. Listeners were instructed to mark, for each word, the first viable division within the word on the transcript. They were given free choice as to what that division might be. (In the English instructions, for example, subjects were told that they might think that the word *international* divides into two parts, or three, four, five or more, but that they were only to mark the first part.) Subjects were instructed first to listen to the word, then to decide upon their preferred segmentation, and only then to look at the transcript and make their response. The words were presented at a fairly rapid rate (one word every two seconds), allowing subjects no time to reflect over possible alternative responses, but encouraging a quick choice of a first available segmentation.

(a) Japanese

CVCV- words	CV	CVC	CVCV	Other
	.430	.002	.257	.311
CVCC- words	**CV**	**CVC**	**CVCC**	**Other**
Nasal	.255	.718	0	.027
Geminate	.336	.505	0	.158

(b) English

(i) *British*

SW CVCV- words	CV	CVC	CVCV	Other
	.053	.818	.047	.082
SW CVCC- words	**CV**	**CVC**	**CVCC**	**Other**
	.008	.688	.239	.065
WS words	.111	.513	.182	.194
SS words	.177	.581	.199	.043

(ii) *American*

SW CVCV- words	CV	CVC	CVCV	Other
	.087	.826	.035	.052
SW CVCC- words	**CV**	**CVC**	**CVCC**	**Other**
	.011	.885	.050	.054
WS words	.336	.506	.034	.124
SS words	.250	.611	.099	.040

Table 1. *Proportions of initial segmentations for each word type for each language group.*

3. RESULTS AND DISCUSSION

The proportions of choices for each initial segmentation, for each word type, are presented separately for each listener group in Table 1. Note that the labels in Table 1 are summary terms representing the majority case. Thus, for example, two Japanese and two English stimulus words were vowel-initial; a segmentation immediately following this initial vowel was scored as CV (thus *e/mbargo* fell in the same category as *co/ntend,* and *i/sasaka* in the same category as as *no/rimono),* a segmentation following the initial VC was scored as CVC, and so on. Likewise, three English stimulus words began with a consonant cluster; here the initial cluster (which was never split by any subject's choice of segmentation) was treated as a singleton (thus *trus/tee* was scored as *can/teen,* and so on). No separate category is given for single-phoneme segmentations, as only two responses in the entire set were of this type. These have been subsumed in the "Other" category, which also includes failures to respond, responses which failed to comply with the instructions (e.g. marks before or after the entire word, or marks through instead of before or after a grapheme) as well as some segmentations later in the word (e.g. *destit/ute).*

The patterns of response are quite clear, and our analysis is confined to simple non-parametric tests. The first feature to note is simply that the results are by no

means random: Chi-squared tests show that for each word type for each subject group the patterning of responses across categories is significantly different from that which would be expected by chance. The second feature of note is that one preferred segmentation dominates: The modal response for all word types and subject groups, with just one exception, is CVC. Sign tests across items show that a CVC response was significantly more likely (beyond the .001 level) than any other possible response for strong-weak CVCV- words *(canopy)* for both British and American subjects, for strong-weak CVCC- words *(cancel)* for both British and American subjects, and for CVCC- words with nasal coda *(kenri)* for Japanese subjects. Although still the modal response, it is however not significantly more likely than any other response for the weak-strong *(contend)* or strong-strong *(pastime)* English words, or for the Japanese CVCC- words with geminate consonants *(katto)*. In the one case where it is not the preferred response - for the Japanese CVCV- words *(kamera)* - the sign test shows that it is significantly less likely than any other response.

Recall that the Japanese materials also included a comparison of alternative orthographies (which is of course not possible in English). The results for the Japanese subjects are further presented in Table 2, as a function of characteristic orthography of the stimulus word. It can be seen that there are no great differences between the three orthographic conditions. A 3 x 3 chi-squared test (collapsing categories containing zero or a single response with the "Other" category) was carried out on the segmentation frequencies for each word type separately. For the CVCC- words with nasal coda, this test indeed showed no significant difference between conditions (x^2 [4] = 5.33, $p > 0.25$). However for the other two word types there was an effect of orthography: x^2 [4] = 44.55, $p < 0.001$ for the CVCV- words, and x^2 [4] = 28.12, $p < 0.001$ for the CVCC- words with geminate consonants. Although in both cases the kanji words received rather more segmentations consistent with kanji boundaries (CVCV for CVCV- words: *norimono,* CVC for the CVCC- words with geminate consonants: *tosshin),* by far the largest asymmetry contributing to the significant inter-condition difference occurred, as Table 2 shows, in the "Other" response category: in both cases there were more "Other" responses to katakana words *(kamera, nokku)* and fewer "Other" responses to kanji words *(norimono, tosshin)*.

The results of the orthographic comparison are consistent with the interpretation that subjects found the native content words written with kanji characters relatively easier to choose a segmentation for, but the foreign loan words rather less easy. What is most clear from these results, however, is that mora-based orthography (hiragana, katakana) did not increase the likelihood that subjects would choose the first mora of the word (in other words, the initial CV) as their initial segmentation.

CVCV- words	CV	CVC	CVCV	Other
kanji	.475	.003	.344	.178
hiragana	.434	0	.206	.359
katakana	.381	.003	.222	.394
CVCC-words	CV	CVC	CVCC	Other
Nasal				
kanji	.257	.735	0	.007
hiragana	.250	.728	0	.022
katakana	.257	.691	0	.051
Geminate				
kanji	.372	.556	0	.072
hiragana	.316	.497	0	.187
katakana	.322	.463	0	.215

Table 2. *Proportions of initial segmentations for each Japanese word type as a function of standard orthographic representation.*

Our predictions at the outset of this study were that language- and dialect-specific factors would play a role in listeners' conscious segmentations. An intervocalic consonant in Japanese words such as *kamera* is not ambisyllabic; and as we had predicted, the overall preference for CVC segmentations was suppressed for Japanese words beginning CVCV. Thus the most significant cross-linguistic difference was that these words elicited CV segmentations in Japanese but not in English (x^2 [1] = 395.37, $p < 0.001$). The weaker tendency to ambisyllabicity in American English also led, as further predicted, to a significant cross-dialectal difference (again involving words beginning CVCV): although both groups most often chose CVC segmentations with these as with other words, the frequency of CV segmentations was somewhat higher for American than for British English listeners (x^2 [1] = 4.45, $p < 0.05$). However, the frequency of CV segmentations was also significantly higher for American than for British English listeners with the WS words *(contend;* x^2 [1] = 54.07, $p < 0.001$) and the SS words *(pastime;*x^2 [1] = 5.77, $p < 0.02$).

4. CONCLUSION

The first conclusion is that deliberate choice of the first acceptable segmentation of a word is not a task which listeners find difficult or confusing. Just as on-line listening experiments reveal clear patterns of preferred segmentations, so does the present task which taps conscious representation of within-word structure. In no condition could subjects' responses be described as random.

The second conclusion is that listeners' representations of within-word structure as revealed by this task to a large extent do not correspond to the segmentations used in on-line listening as revealed by RT studies. English listeners, for instance, generally preferred a CVC initial segment, which represents for most of the stimulus words the earliest permissible

syllable, yet as we pointed out above, evidence from RT tasks does not provide strong support for the syllable as an on-line segmentation unit in English [6]. The syllable has, however, been well attested in previous work as an explicit unit in English-speakers' conscious structural representations of words [9]; in this respect the present results are firmly in line with preceding (non-RT) studies. This preference was somewhat suppressed for WS and SS words *(contend, pastime),* suggesting that when the second syllable is strong a segmentation point immediately before it is somewhat less highly favoured, in direct contradiction to the RT evidence for on-line segmentation at strong syllable onsets [1,2,3,4]. Moreover the fact that WS words and SS words did not pattern differently from one another contrasts with recent evidence for on-line sensitivity to stress pattern by American English listeners [12].

In Japanese, likewise, the preferred segmentation for two word types was CVC, which again could be the initial syllable of these two word types; again, RT evidence speaks against an on-line role for the syllable in the segmentation of Japanese [7]. The initial CVC of CVCC- words *(kenri, tosshin)* could however also be viewed as a bimoraic foot, which would lend further support to previous evidence for listeners' conscious representations of such a unit in Japanese [10]. (Note that although the CVC in CVCC- words is ambiguous between a syllabic and a bimoraic foot interpretation, in CVCV- words the two interpretations produce different results - the initial syllable is CV, the initial bimoraic foot CVCV. In these words the frequency of CV choices was somewhat higher, suggesting a greater preference for syllabic segmentation.) The greatest discrepancy between the present Japanese results and the results from RT experiments is that on-line evidence [7,8] suggests mora-based segmentation while the frequency of initial mora segmentations (CV) in the present results was relatively low - indeed, the single case in which relatively many CV choices were made (CVCV- words) was precisely also the single case in which a CV choice also corresponded to the choice of the initial syllable.

Our third conclusion concerns the role of orthography in listeners' representations of within-word structure. Although it is unlikely that orthography played no role at all in the present task - given that responses were made on a written transcript! - the results from the Japanese subject group clearly show that orthography is not the dominant factor determining subjects' choice of segmentation. The present task appears to be capable of tapping the structure within listeners' lexical representations without necessarily drawing upon their command of orthography. The internal structure of lexical representations is, as our earlier conclusions maintained, readily available to conscious inspection, and it appears furthermore to be richer than the structure implied by results from on-line listening tasks.

5. ACKNOWLEDGEMENTS

This research was supported by a grant from the Human Frontier Scientific Program (TO and AC), by grant number 06610475 from the Japanese Ministry of Education (TO) and by grant number NIH 1 R29 HD233385 to Michael H. Kelly (SMD). We are very grateful to James McQueen, Ruth Kearns, Kiyoko Yoneyama and Kazutaka Kurisu for assistance with the study and discussion of the results. Email addresses of the authors are: mycom09@jpndokyo.bitnet; sally@cattell.psych.upenn.edu; anne@mpi.nl.

6. REFERENCES

[I] Cutler, A. & Butterfield, S. "Rhythmic cues to speech segmentation: Evidence from juncture misperception." *Journal of Memory and Language,* 31, 218-236, 1992.

[2] Cutler, A. & Norris, D.G. "The role of strong syllables in segmentation for lexical access." *Journal of Experimental Psychology: Human Perception and Performance,* 14, 113-121, 1988.

[3] McQueen, J.M., Norris, D.G. & Cutler, A. "Competition in spoken word recognition: Spotting words in other words." *Journal of Experimental Psychology: Learning, Memory and Cognition,* 20, 621-638, 1994.

[4] Norris, D.G., McQueen, J.M. & Cutler, A. "Competition and segmentation in spoken word recognition." *Journal of Experimental Psychology: Learning, Memory and Cognition,* 21, 1995.

[5] Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. "The syllable's role in speech segmentation." *Journal of Verbal Learning and Verbal Behaviour,* 20, 298-305, 1981.

[6] Cutler, A., Mehler, J., Norris, D.G. & Segui, J. "The syllable's differing role in the segmentation of French and English." *Journal of Memory and Language,* 25, 385-400, 1986.

[7] Otake, T., Hatano, G., Cutler, A. & Mehler, J. "Mora or syllable? Speech segmentation in Japanese." *Journal of Memory and Language,* 32, 358-378, 1993.

[8] Cutler, A. & Otake, T. "Mora or phoneme? Further evidence for language-specific listening." *Journal of Memory and Language,* 33, 824-844, 1994.

[9] Treiman, R. "The structure of spoken syllables: Evidence from novel word games." *Cognition,* 15,49-74, 1983.

[10] Kurisu, K. "Further evidence for bimoraic foot in Japanese." Proceedings of the 1994 International Conference on Spoken Language Processing, Yokohama; Vol. 1, 367-370, 1994.

[11] Kahn, D. "Syllable-Based Generalizations in English Phonology." Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1976.

[12] Davis, S. M. "Knowledge and use of the English noun-verb stress difference in native and non-native English speakers." Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA, 1995.