Rhythmic Similarity Effects in Non-native Listening?

Anne Cutler[‡], Lalita Murty[‡] and Takashi Otake[†]

[‡]Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands

† Dokkyo University, 1-1 Gakuen-cho, Soka-shi, Saitama 340, Japan

E-mail: Anne.Cutler@mpi.nl, Lalita.Murty@mpi.nl, otake@dokkyo.ac.jp

ABSTRACT

Listeners rely on native-language rhythm in segmenting speech; in different languages, stress-, syllable- or mora-based rhythm is exploited. This language-specificity affects listening to non-native speech, if native procedures are applied even though inefficient for the non-native language. However, speakers of two languages with similar rhythmic interpretation should segment their own and the other language similarly. This was observed to date only for related languages (English-Dutch; French-Spanish). We now report experiments in which Japanese listeners heard Telugu, a Dravidian language unrelated to Japanese, and Telugu listeners heard Japanese. In both cases detection of target sequences in speech was harder when target boundaries mismatched mora boundaries, exactly the pattern that Japanese listeners earlier exhibited with Japanese and other languages. These results suggest that Telugu and Japanese listeners use similar procedures in segmenting speech, and support the idea that languages fall into rhythmic classes, with aspects of phonological structure affecting listeners' speech segmentation.

1. INTRODUCTION

Word recognition is the heart of speech recognition; but it requires listeners to turn a continuous stream of speech into a percept consisting of discrete, non-overlapping individual lexical units. This process of segmentation works wonderfully well in the native language, but not always so well for languages learned later in life.

One reason for this is that segmentation procedures are in part language-specific. There are in fact many ways in which listeners tailor their speech processing, and in particular the processes by which they segment continuous speech, such that great efficiency is achieved with native-language input; any aspect of phonological structure can be exploited in this way. In the present symposium the focus is on language rhythm. This is an aspect of phonology which definitely plays a role in listening.

That languages differ in characteristic rhythm has long served as an inspiration in phonetic research [1,2], even though no consensus has emerged on how the undoubted differences in rhythmic structure should be captured [3]. A new type of approach in which rhythmic structure is expressed via a formulation of segmental and syllabic patterning [4,5] offers perspectives of innovative advance in this area. Moreover, this approach offers at last the possibility of a more solid foundation for the notion of rhythmic classes into which languages may be organised. This notion can be traced to Abercrombie [1], who distinguished stress-timed from syllable-timed languages. The class of mora-timed languages was later added [e.g. 6]. The idea that there is a small and finite number of possible classes along the rhythm dimension is attractive, but has so far not been really subject to extensive test.

Phonetic work on rhythmic classes prompted studies of the role of rhythm in speech segmentation. Syllabic segmentation effects appeared in French [7] but not in English [8], which in contrast showed stress-based segmentation [9, 10]. The proposal that rhythm underlay these cross-language differences was supported by mora-based segmentation effects in Japanese [11, 12].

Importantly, the segmentation effects observed in these studies patterned with listener group rather than with language input. English, Dutch and Japanese listeners presented with French did not show syllabic segmentation [8, 13, 14], and English and French listeners presented with Japanese showed no moraic effects [11, 12]. Instead, listeners inappropriately applied their native procedures to non-native input, so that syllabic effects were observed when French listeners heard English or Japanese [8, 11] and moraic effects appeared when Japanese listeners heard English, French and Spanish [12, 14]. The inefficiency of speech segmentation in second-language listening could therefore rest in good part on the inappropriate application of native procedures in non-native listening.

The hypothesis of rhythmic classes, however, suggests that not all non-native listening should involve inefficient segmentation. Two languages of the same class should encourage use of the same rhythmic segmentation procedures, and thus listening to one with the listening procedures encouraged by the other should not result in a loss of efficiency. Similarities have been attested across languages – thus effects observed for Dutch are effectively the same as those in English [15], and Spanish and Catalan likewise resemble French [16]. But these languages come respectively from the same etymological families. In order to test the hypothesis, it is desirable to examine language pairs which are rhythmically similar but etymologically unrelated. In the present study, we compare Japanese with Telugu, a Dravidian language. Telugu resembles Japanese

ISBN 1-876346-48-5

in just those aspects of syllable structure which principally determine rhythm [17, 18] – e.g. it has a preference for open syllables, but allows nasal codas. From a wider study we here report two non-native listening experiments, in which Japanese listeners hear Telugu, and Telugu listeners hear Japanese.

The experiments use the methodology of Otake et al. [11]. In their study, Japanese listeners found it hard to detect target sequences in speech when target boundaries mismatched mora boundaries; the target *ta-* was detected equally easily in *tanishi* and *tanshi*, i.e. irrespective of syllable boundary placement; the target *tan-* could be detected in *tanshi*, but proved very hard to detect in *tanishi*. Since the mora structure of these words is *ta-ni-shi* and *ta-n-shi*, it is clear that only targets of which the boundaries correspond to mora boundaries proved detectable. Figure 1 displays Otake et al.'s results. Importantly, this pattern was not replicated when the same materials were heard by English or French listeners [11].

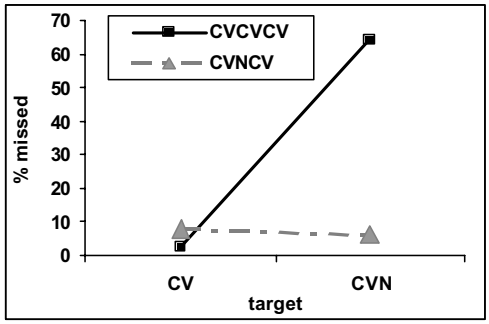

Fig. 1 Mean proportion missed responses as a function of target size (*TA-* vs *TAN-*) and word structure (*tanishi* versus *tanshi*).

2. EXPERIMENT 1

2.1 METHOD

2.1.1 *Materials*: 32 meaningful words were selected from a standard dictionary of the Telugu language: *penamu, penku, janudu, janta, winati, winta, kunuku kunda, chemata, chembu, tamaru, tampi, minapa, minchu, punugu, punju, manawi, manta, panasa, panga, gunupu, gunta, manishi, mandu, kanaru, kanchu, chinuku, chinta, jinudu, jinka, kamaru, kampu*. These words form 16 pairs containing a word-medial mora in the form in the first word of the pair of a Nasal+V sequence and in the second of a single Nasal homorganic with the following segment. The words in each pair were unrelated in meaning (e.g., *penamu* 'frying pan', *penku* 'tile'). A male native speaker of Modern Standard Telugu read these pairs along with a further 278 words, and the words were then arranged into 64 sequences. Each sequence could contain from 4 to 7 words, with the target word in third, fourth, fifth or sixth position. Type of target was counterbalanced across order for each pair of items in two test versions. Thus in version A subjects monitored for MA in a sequence containing the target word *manishi,* and for MAN in a sequence with *mandu*. Subjects in version B monitored for MAN in *manishi* and for MA in *mandu.* The target specifications that preceded each sequence were spoken separately by the same speaker.

2.1.2 *Subjects*: 38 undergraduate students, aged 18 to 21, from Dokkyo University participated in this experiment in return for course credit. 19 subjects heard each test version.

2.1.3 *Procedure*: Subjects were tested individually in a quiet room. They were instructed to listen first for the target (a CV or a CVN) to be followed by a sequence of words and then to press a response button as soon as they heard a word beginning with the target sound.

2.2 RESULTS AND DISCUSSION

Two target items which were wrongly recorded, and their respective pairs, were omitted from the analysis. Mean response time (RT) in milliseconds (ms.) and mean number of missed responses for each subject on each item were calculated, and separate analyses of variance conducted on each measure both for subjects and items. For comparison with the earlier results, in which the significant effects were observed in the miss rate analyses, we present the same analyses here (Fig.2).

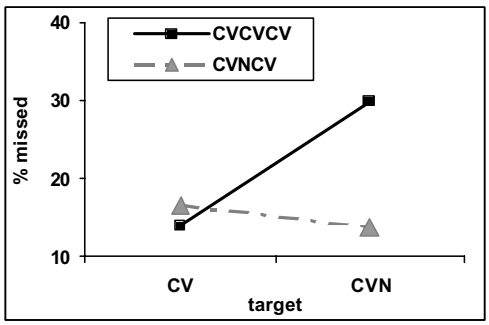

Fig. 2 Mean proportion missed responses as a function of target size and phonological structure of the word.

Analysis of variance revealed a significant interaction of word and target type for both subjects and items (F1 [1, 22] = 10.08, p< .005; F2[1, 25] = 10.87, p< .005). Separate *t* tests examined the basis of this interaction. No significant differences appeared for target type in CVNCV words. However, there was a significant difference in the miss rates for the two target types in CVCVCV words (*t*1 [24] = 3.05, p< .01; *t*2 [13] = 3.25, p< .01).

When listening to a non-native language, Japanese listeners do not show miss rates as high as in the mismatching condition in their own language; this was earlier observed in experiments in which Japanese listeners heard French, English and Spanish [14]. The results shown in Fig. 2 nonetheless show a remarkable similarity to the earlier results for Japanese subjects listening to their own language (Fig. 1). Miss rates on CVN targets for CVCVCV words were double the miss rates in the other 3 conditions. CVN targets in CVCVCV words constitute the only case in which the target cannot be mapped to a portion of the word

where it is bounded by mora boundaries – at least in Japanese phonological structure. Thus this study has supported the hypothesis that the moraic listening strategy that Japanese speakers employ in listening to their own language would also be used in listening to Telugu, an unrelated language. As described above, Telugu and Japanese phonology exhibit strong similarities in the aspects of syllable structure which determine rhythm [17, 18]. We might thus expect that the appropriate listening strategy would be applied in the reverse direction as well. In our second experiment, we therefore reverse the language-listener mapping, and allow Telugu listeners to hear the stimuli of Otake et al. [11].

3. EXPERIMENT 2

3.1 METHOD

3.1.1 *Materials*: The materials used in this experiment were those of Otake et al. [11], namely 16 meaningful words, in eight pairs: *tanishi, tanshi, monaka, monka, kanoko, kanko, sanaka, sanka, nanoka, nanka, kinori, kinri, haneda, handa, shinigao, shingao.* Unlike the Telugu materials, the words within each of these Japanese pairs had the same initial and final mora. Thus *tanishi* and *tanshi* have *ta* and *shi* as their initial and final moras respectively, though they differ in the medial mora, *tanishi* having a CV mora beginning with a nasal, *ni,* and *tanshi* a nasal consonant /n/. A further 250 words were selected and arranged into 64 sequences. Each target word occurred twice, once in one of the first 32 sequences and again in one of the last 32 sequences. Sequence length varied from 3 to 6 words, with the target words in second, third, fourth or fifth position. In other design respects, this experiment was identical to Experiment 1. The words were spoken by a male native speaker of Standard Tokyo Japanese.

3.1.2 *Subjects*: 41 undergraduate students at a degree college in Hyderabad, India, participated in the experiment for a small payment. All the subjects were between the ages 19 and 22. Their mother tongue was Telugu and none had any experience with the Japanese language. 20 heard one version of the experiment and 21 the other.

3.1.3 *Procedure*: Subjects were tested indvidually in a quiet room in the college. They were told that they were going to be listening to Japanese, and were then instructed (in Telugu and English) to listen for a word beginning with the sounds specified as target for each sequence and to press a button as soon as they heard the target sound.

3.2 RESULTS AND DISCUSSION

Mean RT in ms. and mean number of missed responses for each subject and item were calculated and separate analyses of variance again conducted as for Experiment 1. The miss rate results are displayed in Fig. 3. Once again, the pattern is very similar to that of the original Japanese results; the miss rate for the CVN targets in CVCVCV words can be seen to be double that in the other three conditions.

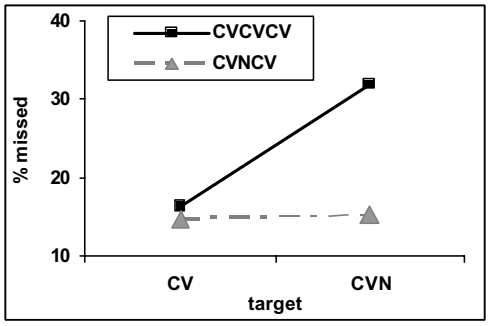

Fig 3. Mean proportion missed targets as a function of target size and phonological structure of the word.

Analyses of variance for the miss rates showed a pattern similar to that in the previous experiment, although here the interaction between word type and target size was significant only in the items analysis (F1 n.s, F2 [1, 27] = 4.68, $p < .04$). Separate *t* tests showed no significant differences in the miss rates for target type in CVNCV words, but a significant difference in the miss rates for target type in CVCVCV words ($t2\ [15] = 4.16, p < .002$).

Thus the results for Telugu listeners to Japanese have revealed patterns very similar to those produced by the original Japanese listeners. Telugu listeners, just like Japanese listeners, found it hardest to detect the targets that were not aligned with a mora boundary, i.e. the CVN targets in the CVCVCV words.

4. CONCLUSION

A number of studies in the last two decades have demonstrated that efficient procedures in native listening involve exploiting the intrinsic rhythmic patterns of the language. This study began with the hypothesis that speakers of languages with similar rhythmic patterns would extend these listening strategies to each others' language even if the languages are typologically unrelated.

The two listener groups in our study, the Japanese listeners and the Telugu listeners, had no prior knowledge of each others' languages. Nonetheless, the listening results they produced in Experiments 1 and 2 were strikingly similar. Moreover, these results strongly resembled the pattern that was earlier observed when Japanese listeners performed the same task with their own language. Difficulty appeared when targets mismatched with mora boundaries within the word, and the difficulty showed up in the proportion of targets which listeners missed in such a condition. This is strongly in line with the previous findings that listeners extend their native listening strategies to non-native languages, whether or not they are familiar to them.

When unfamiliar languages are nevertheless rhythmically comparable, an interesting consequence follows, namely that the native listening strategies should prove efficient for listening to the non-native language, irrespective of questions of typological relatedness. It can thus be

ISBN 1-876346-48-5

concluded from our results that the way Telugu listeners segment Japanese and Japanese listeners segment Telugu should prove relatively efficient. Since this is not true for French and English listeners to Japanese [11], our findings further imply that Telugu listeners should perform better at segmenting Japanese than English or French listeners, and Japanese listeners should likewise segment Telugu more efficiently than English or French listeners. We look forward to extensions of this research to second-language learning situations.

The relevance of our findings for the wider study of rhythm consists in the similarity we have observed between the patterns of results produced by these two listener groups. Telugu and Japanese do not belong to the same language family. Nevertheless it appears that they encourage very similar patterns of speech segmentation performance, just as performance similarity has previously been observed for related languages with stress rhythm [9, 10, 15] and for related languages with syllabic rhythm [8, 16]. This is consistent with the proposal that rhythmic similarity is a broader phenomenon than just an aspect of etymological relatedness; the hypothesis of rhythmic classes thus receives a further quantum of support.

The exploitation of language-specific rhythmic structure in speech segmentation is, we assume, a universal strategy, with a major role in streamlining the process of efficient word recognition in the native language. Because rhythmic units differ across languages, using a native strategy can lead to inefficiency in listening to a non-native language. However, as our results show, rhythmic similarity of the two languages encourages a more native-like efficiency.

ACKNOWLEDGEMENTS

This research was supported by NWO SPINOZA award "Native and Non-native Listening" to the first author. We thank Maarten Jansonius for technical assistance.

REFERENCES

[1] D. Abercrombie, "A phonetician's view of verse structure," *Linguistics*, vol. 6, pp. 5-13, 1964.

[2] K.L. Pike, *The Intonation of American English*. Ann Arbor: The University of Michigan Press, 1945.

[3] A. Cutler, "Linguistic rhythm and speech segmentation," in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord and R. Carlson, Eds., pp. 157-166. London: Macmillan, 1991.

[4] E.L. Low, E. Grabe and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," *Language and Speech*, vol. 43, pp. 377-402, 2000.

[5] F. Ramus, M. Nespor and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 71, pp. 1-28, 1999.

[6] C.E. Hoequist, "Syllable duration in stress-, syllable- and mora-timed languages," *Phonetica*, vol. 40, pp. 203-237, 1983.

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

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

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

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

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

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

[13] A. Cutler, "The syllable's role in the segmentation of stress languages," *Language and Cognitive Processes*, vol. 12, pp. 839-845, 1997.

[14] T. Otake, G. Hatano and K. Yoneyama, "Speech segmentation by Japanese listeners," in *Phonological Structure and Language Processing: Cross-linguistic Studies*, T. Otake and A. Cutler, Eds., pp. 183-201. Berlin: Mouton de Gruyter, 1996.

[15] J. Vroomen, M. van Zon and B. de Gelder, "Cues to speech segmentation: Evidence from juncture misperceptions and word spotting," *Memory & Cognition*, vol. 24, pp. 744-755, 1996.

[16] N. Sebastián-Gallés, E. Dupoux, J. Seguí and J. Mehler, "Contrasting syllabic effects in Catalan and Spanish," *Journal of Memory and Language*, vol. 31, pp. 18-32, 1992.

[17] Bh. Krishnamurti, "Telugu," in *The Dravidian Languages*, S.B. Steever, Ed., pp. 202-240. London: Routledge, 1998.

[18] P. Sailaja, "Syllable structure of Telugu," in *Proceedings of the Fourteenth International Congress of Phonetic Sciences*, vol. 4, pp. 743-746. San Francisco, USA, 1999.