Perceptual Tests of Rhythmic Similarity: II. Syllable Rhythm

Jeesun Kim[1], **Chris Davis**[1], **Anne Cutler**[1,2]

[1] *MARCS Auditory Laboratories, University of Western Sydney, Australia*
[2] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Key words

French

Korean

rhythm

segmentation

word recognition

Abstract

To segment continuous speech into its component words, listeners make use of language rhythm; because rhythm differs across languages, so do the segmentation procedures which listeners use. For each of stress-, syllable- and mora-based rhythmic structure, perceptual experiments have led to the discovery of corresponding segmentation procedures. In the case of mora-based rhythm, similar segmentation has been demonstrated in the otherwise unrelated languages Japanese and Telugu; segmentation based on syllable rhythm, however, has been previously demonstrated only for European languages from the Romance family. We here report two target detection experiments in which Korean listeners, presented with speech in Korean and in French, displayed patterns of segmentation like those previously observed in analogous experiments with French listeners. The Korean listeners' accuracy in detecting word-initial target fragments in either language was significantly higher when the fragments corresponded exactly to a syllable in the input than when the fragments were smaller or larger than a syllable. We conclude that Korean and French listeners can call on similar procedures for segmenting speech, and we further propose that perceptual tests of speech segmentation provide a valuable accompaniment to acoustic analyses for establishing languages' rhythmic class membership.

1 Introduction

Rhythm is one of the first attributes of the native tongue which language users acquire; the rhythm of speech is audible to hearing infants in the womb, and, even in the earliest days of life, infants prefer to listen to speech with the rhythm of the mother's

Acknowledgments: This research was supported by the NWO-SPINOZA project "Native and non-native listening", with additional support from ARC grant DP0453143. We thank Natasha Warner for help in preparing materials, and Hyounjoo Jeung, Jayoung Sim and Youngnim Pyon for assistance in organizing and running experiments. At http://www.mpi.nl/world/persons/private/anne/materials.html, further information on the experimental materials is available.

Address for correspondence. Anne Cutler, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands; email: <anne.cutler@mpi.nl>

Language and Speech; 0023-8309; Vol 51(4): 343–359; 099069; 10.1177/0023830908099069
www.sagepublications.com

language rather than to speech with another linguistic rhythm (Condon & Sander, 1974; Demany, McKenzie, & Vurpillot, 1977; Nazzi, Bertoncini, & Mehler, 1998; Nazzi & Ramus, 2003).

Unsurprisingly, rhythm then remains important for listening to language throughout life. In particular, adult language users exploit the rhythm of their language to assist in the task of segmentation: resolving heard speech from a continuous stream into a string of separate words.

This means that segmentation procedures used by speakers of different languages will differ, simply because rhythmic structure differs across languages (Abercrombie, 1967; Bloch, 1950; Classe, 1939; Hoequist, 1983; Pike, 1946). Differences in rhythmic structure are unavoidable, because rhythm, though it basically concerns the timing of linguistic events, is a complex regularity which is necessarily modulated by other aspects of the language's phonology, such as syllable structure or intonation. These factors thus in turn determine the best way in which listeners can segment the speech stream—across languages, different structures encourage different procedures.

Perceptual experiments over the past two decades have abundantly documented the language-specificity of speech segmentation procedures. French listeners were shown to use a syllabic procedure (Cutler, Mehler, Norris, & Seguí, 1983, 1986, 1992; Kolinsky, Morais, & Cluytens, 1995; Mehler, Dommergues, Frauenfelder, & Seguí, 1981). English listeners were shown to use something quite different: segmentation based on stress (Cutler & Butterfield, 1992; Cutler & Norris, 1988; Gow & Gordon, 1993). This listening difference corresponded to the rhythmic difference between English (a so-called "stress-timed" language) and French (so-called "syllable-timed"), which had long been documented in the linguistic literature (Abercrombie, 1967; Classe, 1939; Pike, 1946). A third type of rhythmic structure, mora-timing, postulated of Japanese (Bloch, 1950; Hoequist, 1983; Ladefoged, 1975), also proved to be accompanied by moraic segmentation in Japanese listeners (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993; Otake, Yoneyama, Cutler, & Van der Lugt, 1996).

A lively linguistic interest in language rhythm has also arisen in recent years, and one suggestion that has attracted particular interest is known as the Rhythmic Class Hypothesis (RCH; see, e.g., Grabe & Low, 2002; Ramus, Nespor, & Mehler, 1999). According to this proposal, languages fall into groupings defined by rhythmic structure, and there are only a limited number of dimensions defining these groupings and hence most probably a limited number of possible classes. Besides differences across languages in rhythmic structure, similarities of rhythmic structure have also been established, with corresponding perceptual similarities. Thus not only French but also Spanish and Catalan encourage syllabic segmentation (Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992), and not only English but also Dutch encourages stress-based segmentation (Vroomen, Van Zon, & De Gelder, 1996).

Of course, Spanish and Catalan are languages with a close historical relation to French, and Dutch is likewise closely related to English. The RCH, however, postulates rhythmic similarity independently of such typological relationship; class membership is defined by phonetic features alone. Indeed, different varieties of the same language can, if the phonetic features differ appropriately, be assigned to differing rhythmic classes (Deterding, 2001; Low, Grabe, & Nolan, 2000; White, Mattys, Series, & Gage, 2007). Thus to test the RCH, it is essential that metrics be found for comparing rhythmic structure across languages.

The most commonly used metrics are based on measurement of segmental durations. The measure developed by Ramus (1999; 2002; Ramus et al., 1999) represents the standard deviation of intervocalic intervals and the proportion of speech time taken up by vowels. The Pairwise Variability Index (nPVI, where the "n" signifies normalized; Grabe, 2002; Grabe & Low, 2002; Low et al., 2000) is based on the average variability in the normalized duration of successive vowels and the average variability in the duration of successive intervals between vowels. Each of these measures clearly captures the fact that syllable structure (e.g., permissibility of cluster onsets and degree of coda complexity) and intonation can affect speech rhythm. The measures are being widely applied, to ever more languages (e.g., Hawaiian: Parker Jones, 2006; Tamil: Keane, 2006). However, since results on these measures are known to vary with speech rate (Barry & Russo, 2003), across register (Keane, 2006) and across speakers (Keane, 2006), a very large and representative database must be collected for results to be definitive. Since the measurement of individual segment durations required for computation of these metrics is quite labor-intensive, this is not a short-term prospect.

Converging evidence from other metrics is therefore very useful. The present report arises from a project aimed at acquiring converging evidence from perceptual performance across languages to test the RCH. Most of the studies of language-specificity in speech segmentation described above employed quite simple empirical techniques, in particular the fragment detection task (Frauenfelder & Kearns, 1996), in which listeners monitor speech input for the occurrence of a specified target fragment (e.g., listeners could be instructed to listen for the fragment ba- in initial position). This task can be successfully performed without comprehension of the meaning of the input—thus, it can be used with artificially constructed nonsense input, or with real speech in the native or in a non-native language. This makes it applicable for any combination of listener language and input language. Moreover, the task also usefully allows two separate views of the effects, via speed of response times (RTs) or accuracy (miss rates). This is a particularly useful characteristic of the task in the situations in which it is used in this line of research. With listeners hearing native input, especially listeners who are experienced (i.e., have previously participated in laboratory experiments of this type), target detection is reliable so that response accuracy may be near 100%. Then, miss rates are too low to show any significant differences due to manipulation of the independent variable(s). With these listeners and this type of input, response speed thus provides the best view of any operative effects. When listeners have no prior experience of laboratory listening tasks, however, or are confronted with very unfamiliar (e.g., non-native) input, response speed can be extremely variable and many targets may simply be missed. In these cases, response accuracy usually provides the best view of the data. This double sensitivity is one reason why the method has been applied so often in these studies.

Given that fragment detection and related techniques have now been used with a range of languages varying in rhythmic structure, these techniques can potentially be used as a diagnostic of rhythmic structure; diagnosis involves the establishment of similarity across result patterns from different languages. As we saw above, a syllabic pattern, for example, has emerged not only with French but also with Spanish and Catalan, and in each case in experiments with the fragment detection task. This source of perceptual evidence is thus a valuable contribution in the program of testing the RCH.

Murty, Otake, and Cutler (2007) started the evidence collection in this project by examining similarities involving the smallest of the rhythmic units: the mora. Motivated by an extensive literature claiming phonological similarity between the languages of the Dravidian family and Japanese, they conducted fragment detection experiments in Telugu (the largest Dravidian language). The experiments were directly analogous to those which had already been done in Japanese. The two language families are totally unrelated. Murty et al.'s results showed consistent parallel performance between Japanese and Telugu listeners; moreover, in both languages the response pattern when the non-native input was presented was like that of native listeners. That is, Telugu listeners responded to Japanese input like Japanese listeners did, and Japanese listeners responded to Telugu input like Telugu listeners did.

Thus Murty et al. took the language for which mora segmentation has been most robustly established (Japanese) and compared it with a completely unrelated and typologically different language, chosen on the basis of phonological claims concerning rhythmic similarity (Caldwell, 1856/1976; Shiba, 1973). In the present series of experiments we apply the same approach in examination of another rhythmic unit: the syllable. The language for which syllabic segmentation has been most robustly established is French, and we therefore undertook to compare French with an unrelated and typologically quite distinct language. Our choice in this case fell on Korean. Again we chose this language on the basis of claims in the existing literature.

Our literature search revealed that no direct analogue of the RT studies described above has previously been undertaken for Korean, but several papers report evidence from measurements or other tasks of a more metalinguistic nature. We describe below the evidence which has motivated claims that Korean has syllabic rhythm. First, however, we acknowledge that we also found three other papers making different claims about Korean. Lee and Seong (1996) demonstrated phrase-final lengthening effects in Korean, and proposed that extensive phrase-final lengthening may qualify a language for the classification stress-timed. Since it has been argued that phrase-final lengthening is in fact language-universal (Vaissière, 1983), this proposal would make all languages stress-timed. Cassandro, Collet, Duarte, Galves, and Garcia (2003) used automatic sonority detection, without segmental labeling, to classifying sentences from 15 rhythmically different languages. The value they found for Korean fell quite close to that for Japanese. However, their results in fact showed no clear groupings among the languages at all, and assigned markedly different rankings to some languages elsewhere reported to be similar. Cho (2004) proposed Korean to be mora-timed on the basis of a perceptual discrimination study analogous to one conducted by Ramus et al. (1999). Korean listeners were presented with resynthesized sentences in Korean versus Japanese (mora-timed), English (stress-timed) or Italian (syllable-timed). The resynthesis, as in Ramus et al.'s study, destroyed segmental identity by replacing all consonants with /s/ and all vowels with /a/. The task was AAX discrimination; the first two (different) sentences were in one language, and the listener's task was to judge whether or not the third sentence was from the same language. Cho's listeners were significantly better at discriminating Korean from English and Italian than from Japanese. On this basis, Cho deemed Korean to be mora-timed. It is noticeable, however, that the word length of the utterances used in Cho's study was less well controlled than in the original Ramus et al. study, and in particular the word length (in phonemes or syllables) of the Korean and Japanese sentences was more similar than either was to the English and Italian sentences.

These three isolated papers do not seem to us to constitute central contributions in the line of cross-linguistic research on rhythm classification. In contrast to these studies, reports from two separate groups working extensively on cross-linguistic studies of language rhythm have proposed Korean to have syllabic rhythm. Stockmal, Moates, and Bond (2000; see also Bond & Stockmal, 2002) grouped Korean with French, Akan, Latvian, Swahili, and Tagalog in a group they defined as syllable-timed; in language identity judgment experiments, the assigned rhythmic category membership was indeed one of the factors determining clustering of listeners' judgments. Yoon and Derwing (1995) discovered that Korean listeners' judgments of similarity between pairs of words were overwhelmingly based on match versus mismatch in syllabic units, irrespective of the number of phonemes involved in any mismatch. This is very different from the results obtained with the same task in English (Derwing & Nearey, 1994; Nelson & Nelson, 1970) or Japanese (Derwing & Wiebe, 1994). The same authors (Yoon & Derwing, 2001) replicated this finding in similarity judgment tasks, as well as in reduplication and recall tasks. They further established that the syllable in Korean listeners' representations could not be decomposed into an onset-rhyme structure as is the case for English listeners (Treiman, 1985); instead, the only unit below the level of the whole syllable which played a significant role in listeners' success rate in these tasks was the syllable body (onset plus vowel, irrespective of vowel length; Taft, 1992). Since the body of a syllable is, by definition, most of the syllable starting from the beginning, a body effect in auditory perception could arise from continuous activation in the same way that onset-overlapping words are simultaneously activated (e.g., McQueen, Norris, & Cutler, 1994; Zwitserlood, 1989). Where the body also corresponds to the initial mora (e.g., when the onset is simple and the vowel short), then such an effect could also be termed moraic. However, moraic effects should then also be observable in moraic codas, which was not the case in Yoon and Derwing's studies (see also Derwing & Nearey, 1994). The most significant finding, therefore, and indeed the finding stressed in their papers, is the absence of any effect of traditional sub-syllabic components, either onset-rhyme or individual segments. In their studies, Korean listeners treated syllables as units.

One final important factor concerns poetic structure. In languages known as stress-timed, poetic forms overwhelmingly refer to stress beats (for instance, the limerick in English is defined as a form with three stress beats in the first, second and fifth line, and two in the third and fourth; poetic rhythms in English always refer to number and placement of beats—iambic pentameter, etc.). In moraic languages, the poetic forms are determined by mora structure (so the Japanese haiku must have five, seven and five morae in its three successive lines; Telugu poetic forms likewise refer to mora structure). And in languages with syllabic rhythm, poetic forms refer to syllables. This is the case in French (De Cornulier, 1982), and, crucially for the present project, it is also the case in Korean, in which the basic structure of forms such as the *sijo* (the most classic Korean poetic form) is defined in terms of a count of whole syllables. The classic *sijo* form has three lines of 14–16 syllables; the first and second line generally match in number of syllables. For further detail, including the constraints of syntactic and semantic structure on the phonological form, see Rutt (1958) or McCann (1976).

Of course, similarity in rhythmic structure does not necessarily come paired with similarity in all aspects of language phonology. In fact the phonological structures of French and Korean differ in numerous respects. French allows complex onsets and a variety of coda types, and has a mid- to large-size phoneme inventory with many vowels. Korean also has a mid-size vowel inventory; but it allows no onset clusters, few coda clusters, and also constrains which consonants can be codas. Onsets may be consonants or null, medials are vowels or diphthongs, and codas may be one or two consonants or null. A syllable can be (C1) (S) V (C3), where parentheses signal optionality, C1 is any consonant other than /ŋ/, S is one of the semi-vowels (/w, j/), and C3 is (an unreleased variant) of any of /p/, /t/, /k/, or one of the three nasals /m/, /n/, /ŋ/ or the liquid /l/ or /s/ only before a word or suffix beginning /s, ś/ (Kim & Davis, 2006). Different as these phonologies are, the results of studies such as those of Yoon and Derwing (1995, 2001) certainly suggest that syllabic segmentation is as effective in Korean as it has been shown to be in French.

If this is indeed so, a syllabic segmentation pattern should be observable in experiments tapping into the time course of speech perception in Korean. In Experiment 1, modeled directly on the original French experiment in which syllabic segmentation was established for that language (Mehler et al., 1981), we present Korean listeners with Korean materials, and ask them to detect specified word-initial target fragments. We predict that, as in the French experiments, fragment detection will be easier where the target corresponds exactly to a syllable of the input.

2 Experiment 1 (Korean words)

2.1 Method

2.1.1 Participants

Twenty-five undergraduate students at Kyungjoo University, Korea, aged 19 to 22, participated in the experiment in return for course credit. All were native Korean speakers and none reported any hearing problems.

2.1.2 Materials

Twenty-four Korean words, comprising 12 pairs, were selected as experimental stimuli: *camay*, *campok*; *cemyeng*, *cemswu*; *chimyeng*, *chimswu*; *inam*, *insayng*; *imo*, *imsi*; *kamwun*, *kamsa*; *kumwul*, *kumcwu*; *manwula*, *mantala*; *pema*, *pemcwu*; *pomwul*, *pomchel*; *sanai*, *sanswuhwa*; *samyeng*, *samkak*. Note that /wu/ is [u] (no glide), and /ye, ya, wa/, etc., count as unitary diphthongs in Korean. Hence, neither /kamwun/ nor /camyeng/, for example, has two consonants in the onset of the second syllable.

Also, /ch/ each count as a single consonant and do not make the preceding syllable heavy. The words within each pair were chosen such that they all began with the same three segments (e.g., *pem*-) but differed in that one member of the pair had CVC/CV- structure (e.g., *pem/cwu*), while the other had CV/CV- structure (e.g., *pe/ma*). The members of a pair were unrelated in meaning.

A further 214 words were selected and arranged together with the 12 pairs of experimental words into 92 sequences varying in length from two to six words: six practice sequences, and two experimental sets (A and B) of 43 sequences. Each set of 43 consisted of 24 sequences containing one of the chosen experimental stimulus words, in second, third, fourth or fifth position, and 19 filler sequences which did not contain any of the experimental stimuli. In the filler sequences, the specified target did not occur, or occurred early or last. All sequences were preceded by a spoken target syllable; there were 46 different such syllables, all either CV (e.g., *pe*-) or CVC (e.g., *pem*-).

Each experimental stimulus word occurred twice, once in Set A and once in Set B. Across sets A and B, the target assignment for the experimental words was counterbalanced. Thus if *pemcwu* was assigned target *pem*- in Set A, it was assigned target *pe*- in Set B; the target assignments for its pair, *pema*, would be the reverse (*pe*- in A, *pem*- in B). The targets for the practice and filler sequences were also CV or CVC syllables. All target syllables and words were recorded to digital audio tape in a sound-attenuated cubicle by a female native speaker of Korean.

2.1.3
Procedure

Participants were tested individually in a quiet room. They heard the target specifications and sequences over high-quality closed headphones from the computer. Twelve heard the item sets in AB and 13 in BA order. They were told that they would be listening to Korean words and that they should listen for a word beginning with the specified target, and respond as fast as possible by pressing a button as soon as they had detected the target fragment.

The DMDX program (Forster & Forster, 2003) was used for the presentation of the materials and collection of response times and missed responses. The sequences were presented at a normal speaking rate, with approximately one word every two seconds, and with longer pauses between sequences. A brief auditory signal preceded each target specification and thus alerted the listener that a new trial was beginning. Response times were measured from the onset of each target item, and timing was terminated by the subject's button press. A timeout was registered if the participant did not respond within 1.5 seconds. Mean RT in ms and mean number of missed responses for each subject and item were calculated, and analyses of variance conducted for each measure across participants (F1) and across items (F2).

2.2
Results

Mean response times (RTs) and mean missed responses per condition are shown in Figure 1 and Figure 2 respectively. The mean RT was 694 ms and the average miss rate was 4.6%. The RT curves were essentially flat, with responses to CVCCV- words

Figure 1

Mean response times as a function of target type (CV, CVC) and word type (CVCV-, CVCCV-), in Experiment 1 (Korean listeners, Korean materials)

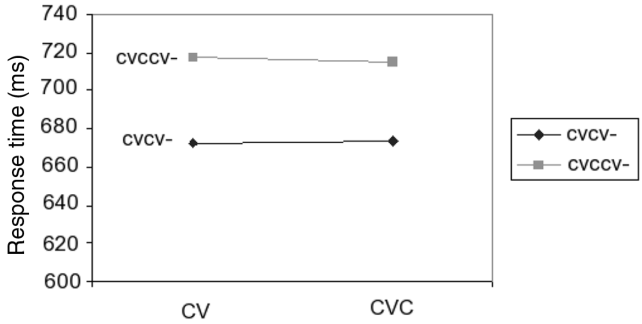
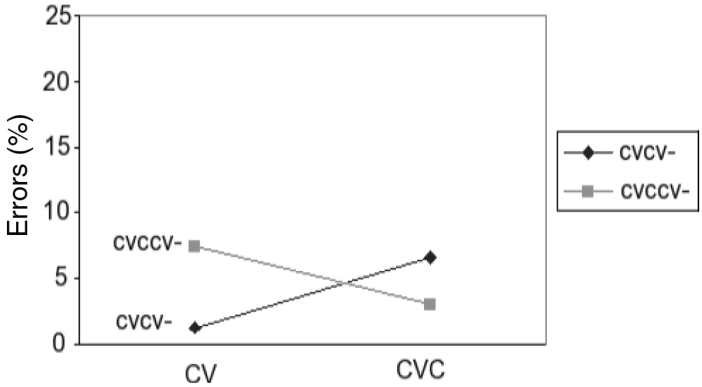


Figure 2

Mean miss rates as a function of target type (CV, CVC) and word type (CVCV-, CVCCV-), in Experiment 1 (Korean listeners, Korean materials)



slower than to CVCV- words; this presumably reflects the fact that CVCCV- words were longer than CVCV-, and the very long RTs are consistent with response delay until item completion. Analyses of variance were conducted across participants and items for each measure. Analyses for the RTs indicated that there was an effect of word type (CVCCV- words slower) ($F1[1, 23] = 54.58$, $p < .005$, $F2[1, 44] = 6.45$, $p < .05$). However, there was no difference in speed of response to the different types of target (both Fs < 1). Likewise, the analyses showed no interaction between word type and target type (both Fs < 1).

The pattern for missed responses shown in Figure 2 shows a clear interaction between word and target type. Analyses confirmed that this interaction was significant ($F1[1, 23] = 9.79$, $p < .01$, $F2[1, 44] = 23.23$, $p < .005$). This interaction was further explored in separate t-tests for each word type. In CVCV- words, there were significantly fewer misses for CV targets (1.3%) than for CVC targets (6.7%; $t[24] = 2.78$, $p < .05$). The difference between the miss rates in CVCCV- words for CVC targets (3.0%) versus CV targets (7.3%), however, narrowly failed to reach significance ($t[24] = 1.83$, $p = .08$). The main effects for word and target type were not significant (all Fs < 1).

It has been demonstrated, as described above, that French listeners apply their native syllabic segmentation to non-native input (in English, Japanese, Spanish, etc.). However, the French stimuli when presented to listeners from other rhythmic classes (e.g., English, Japanese) do not typically attract the syllabic style of response (Cutler et al., 1986; Otake, Hatano, & Yoneyama, 1996a). The results from Experiment 1 suggest that Korean belongs to the same rhythmic class as French. Given this, Korean listeners should show the French syllabic segmentation pattern if they are presented with input in French. This was investigated in Experiment 2.

3 Experiment 2 (French words)

3.1 Method

3.1.1 Participants

Thirty-six undergraduate students at Sejong University, Korea, aged 19 to 22, participated in the experiment in return for course credit. All were native Korean speakers, none reported hearing problems, and none had experience of learning French.

3.1.2 Materials

The materials were those originally tested on French listeners by Mehler et al. (1981), and also used by Cutler et al. (1983, 1986, 1992) with French, English, and French–English bilingual listeners, by Otake et al. (1996a) with Japanese listeners, and by Cutler (1997) with Dutch listeners. The critical target-bearing items were 10 French words, comprising five pairs: *balance*, *balcon*; *carotte*, *carton*; *garage*, *gardien*; *palace*, *palmier*; *tarif*, *tartine*. As in Experiment 1, words within each pair all began with the same three segments (e.g., *bal*-) but differed in that one word had CVC/CV- structure (e.g., *bal/con*), while the other had CV/CV- structure (e.g., *ba/lance*). A further 10 words served as filler target-bearing items, along with five serving as practice target-bearing items. A further 124 words filled out the 40 sequences. There were two experimental sets (A and B) of 20 (10 experimental and 10 filler) sequences varying in length from two to six words. The experimental materials were repeated such that if *bal*- was an initial target for *balance* in file A, then when it was repeated the target *ba*- was used. The opposite item allocation was used in file B. These repeated experimental materials were presented along with 10 additional filler sequences. Thus each set of 40 sequences included 20 experimental target items, 10 filler sequences with target and 10 filler items without target. Targets appeared in second, third, fourth, or fifth position.

The recording was by a female native speaker of French, with the target specifications that preceded each sequence spoken separately by another speaker with similar voice quality. Other preparation of the material was similar to that of Experiment 1.

3.1.3
Procedure
The procedure was as in Experiment 1 except that participants were told that they would hear French words. Seventeen heard the item sets in AB and 19 in BA order.

3.2
Results
Mean RTs and mean missed responses per condition are shown in Figure 3 and Figure 4. The mean RT was 649 ms and the mean miss rate was 14.0%. The pattern of the RT data was similar to that shown in Experiment 1 with the Korean materials. That is, responses made to CVCCV- words were slower than those made to CVCV- words, although this effect was only secure in the participant data ($F1[1, 34] = 9.06$, $p < .005$, $F2[1, 16] = 1.61$, $p > .05$). As with the Korean materials, there was no difference in RTs to the different types of target ($F1[1, 34] = 1.06$, $p > .05$; $F2 < 1$). Likewise, analyses showed no interaction between word type and target type (both *F*s < 1).

Figure 3

Mean response times as a function of target type (CV, CVC) and word type (CVCV-, CVCCV-), in Experiment 2 (Korean listeners, French materials)

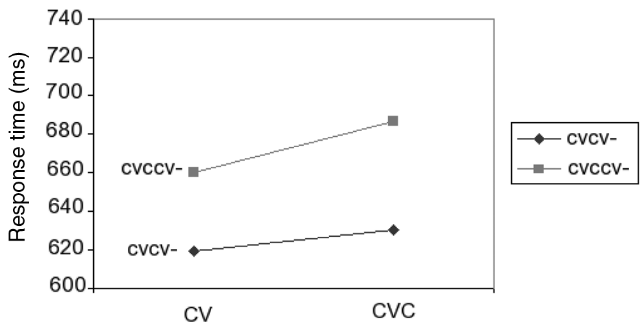
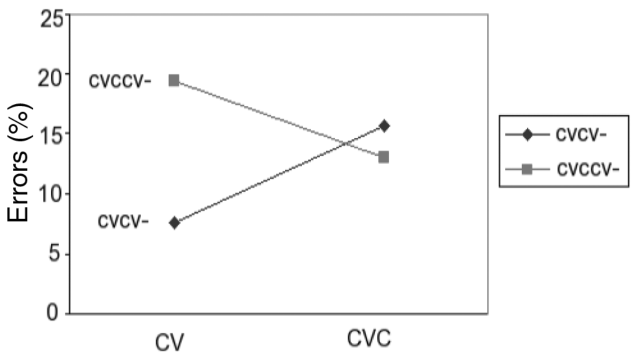


Figure 4

Mean miss rates as a function of target type (CV, CVC) and word type (CVCV-, CVCCV-), in Experiment 2 (Korean listeners, French materials)



The pattern of the miss rates (Figure 4) was also similar to that found for the Korean materials; that is, there was here a significant interaction between word and target type ($F1[1, 34] = 12.23$, $p < .001$; $F2[1, 16] = 4.6$, $p < .05$). Separate *t*-tests were again conducted for the average miss rates for each target type in the two word types. In CVCV- words, there were significantly fewer misses for CV targets (7.6%) than for CVC targets (15.7%; $t[35] = 2.64$, $p < .05$). In CVCCV- words, the difference between the miss rates for CVC targets (13.1%) and CV targets (19.4%) again narrowly failed to reach significance ($t[35] = 1.72$, $p = .09$). There was no difference in the rate of misses for the different target types (both *F*s < 1). There were fewer misses for targets in CVCV- words (11.7%) compared to CVCCV- words (16.3%); this difference was secure in the participant analysis but not in the item analysis ($F1[1, 34] = 5.44$, $p < .05$; $F2 > 1$).

4 General discussion

The results of our two experiments offer clear support to the hypothesis that the rhythmic similarity which has been observed between French and Korean is paralleled by similarity in the perceptual processing of speech by French and by Korean listeners. The error patterns in each of our experiments were exactly as predicted for syllabic segmentation: targets which corresponded to the initial syllable of the word (e.g., *pe*- in *pema*, *pem*- in *pemcwu*, in the Korean materials of Experiment 1) were detected more accurately than targets which mismatched with the syllabic boundaries (e.g., *pem*- in *pema*, *pe*- in *pemcwu*). This pattern appeared not only with the native Korean materials that consisted of words familiar to the listeners, but also with the French materials of Experiment 2, which were completely unfamiliar to the listeners. Their responses here again showed the syllabic effect: *ba*- in *balance* and *bal*- in *balcon* were detected more accurately than *bal*- in *balance* and *ba*- in *balcon*; exactly this effect also appeared when native French listeners heard the same input.

We did not find significant syllabic effects in the RT analyses, but this is not unusual with listeners unpracticed in such tasks (for instance, in the Japanese study of Otake et al., 1993, and in the Sesotho study of Cutler, Demuth, & McQueen, 2002, the predicted effects were statistically reliable in miss rates only). Consistent with an unfamiliar-task account, note that the RTs in our two experiments were very long by comparison with other fragment detection studies (for example, the Experiment 2 RTs were about twice as long as those of the original 1981 study with native listeners: Mehler et al., 1981). In such cases, the double view of performance afforded by the fragment detection task proves its worth.

Thus just as Telugu belongs with Japanese in the class of mora-based languages, so it appears that Korean belongs with French in the class of syllable-based languages. This had already been claimed on the basis of phonological analysis, comparison of poetic forms, and metalinguistic judgments. Our study tapping into the process of speech segmentation has produced clear converging evidence in support. The body of available evidence gives grounds for accepting the proposal that languages group into rhythmic classes, defined purely by phonological structure and without necessary reference to etymological relationship.

Besides the tendency for rhythmic structure to determine the poetic forms of a language, it is also frequently associated with effects in the orthography. Thus, as Murty et al. (2007) report, the special symbol which the Japanese kana orthographies reserve for moraic nasal codas is paralleled in Telugu orthography as well by a special symbol for the nasal codas. In Korean, the role of the syllable in the orthography is especially obvious. The Korean Hangul orthography directly codes syllabic structure; each syllable is written as a separate unit, centred on the vocalic nucleus, with onset and coda affixed. Ambiguity of syllabic boundary assignment, easily possible in a stress language like English, does not occur in Korean phonology, and the orthography reflects this exactly.

What implications does membership of the same rhythmic class as French (and other languages with syllabic rhythm) have for Korean listeners? Obviously, our findings suggest that Koreans listening to their native language draw on speech segmentation strategies resembling those applied by French listeners to speech in their own language. We suggest that there are also further implications, however, beyond the native language case. Consider that one of the strongest effects of rhythmically based segmentation is its language-specificity. Exploiting the rhythm of the native language facilitates segmentation of the native language, but, as we described in the introduction, not all languages have the same rhythm; thus the segmentation strategies appropriate for the native language will be utterly inappropriate for use with speech in many other languages. Listening to natural continuous speech in a foreign language is often unexpectedly difficult, and one reason for this may be misapplication of native segmentation procedures which are inappropriate for the foreign input. Indeed, there is empirical evidence that listeners do misapply native segmentation procedures in this way; French listeners apply syllabic segmentation to English (Cutler et al., 1986) and Japanese (Otake et al., 1993), Japanese listeners apply moraic segmentation to English (Cutler & Otake, 1994) and French and Spanish (Otake et al., 1996a).

Taking the rhythmic mismatches into account is thus potentially helpful in understanding just what makes listening to foreign languages difficult. The RCH suggests, however, that inappropriate segmentation will only happen across different rhythmic classes. A quite different situation applies when the foreign speech input comes from a language which is another member of the same rhythmic class as the listener’s native language. In that case, the native procedures will indeed be appropriate for application to the input in the other language, with the potential consequence that listening—at least, insofar as segmentation is involved—will be relatively unimpaired. Irrespective of how closely related the languages are, rhythmic class membership should determine how appropriate or inappropriate the segmentation procedures are. That is, the RCH suggests that non-native listening need not always imply segmentation difficulty. In the present case, it suggests that French listeners should have relatively little difficulty segmenting continuous speech in Korean, and Korean listeners should have relatively little difficulty segmenting continuous speech in French. Although this may have little relevance for listeners who have never attempted to acquire the other language in question, it does suggest that acquisition of the other language as an L2 would be facilitated (again, in all aspects concerning speech segmentation). Thus, learning to listen to a second language acquired in adulthood is not necessarily difficult in all respects; at least the segmentation problems will not arise if one learns a language from the same rhythmic class as the native tongue, even if the two languages are otherwise unrelated.

The RCH in essence proposes that there may be only a limited number of ways in which languages offer rhythmic support to make listening easier. But important though this support is, and however useful the practical effects it may have for L2 learning, the extent of the RCH's effects in language phonology is in some ways limited. For instance, the measurement studies with the nPVI and related indices have not succeeded in fitting every language neatly into a rhythmic class. Many languages exhibit mixed patterns or fail to show any clear class membership at all. Further, even with languages which do allow clear assignment to a rhythmic class, the effects of the rhythmic class membership in listening are also not always clear-cut. Morais, Kolinsky, and Nakamura (1996), for example, have found that Japanese listeners can manipulate syllabic structure as well as moraic structure in certain tasks, and Otake, Davis, and Cutler (1995) reported that English listeners prefer syllabic segmentation irrespective of stress pattern in explicit segmentation tasks.

Moreover, perceptual effects of rhythmic class are not necessarily predictable; listeners who misapply their native rhythmic strategy to one language may avoid misapplying it to another language. Consider the extent to which Japanese listeners misapply their native moraic segmentation procedure, for example. They show evidence of applying it to French and Spanish (Otake et al., 1996a), as described above, and both of these languages exhibit syllabic rhythm. Therefore one might be inclined to predict that Japanese listeners would misapply moraic segmentation to any language with syllabic rhythm—such as, for example, Korean. But this appears not to be the case. Cutler, Kim, and Otake (2006) presented the Korean materials used in the present study to Japanese listeners and found that, while the listeners were in fact very good at detecting the specified targets, they did not do so in a way that indicated application of moraic segmentation.

For natural reasons, published results are more likely to be positive than negative findings; but the literature on segmentation does reveal further cases in which there was no misapplication of segmentation. For instance, Murty et al. (2007) report that Telugu materials which included phonological patterns impossible in Japanese—such as coda clusters—elicited no moraic segmentation, although such segmentation was applied to Telugu materials when codas were restricted to the types also permissible in Japanese. Again, whereas the Japanese listeners of Cutler and Otake (1994) detected both vowels and consonants faster in moraic rather than non-moraic position in Japanese, in English the Japanese listeners only detected consonantal targets faster in "moraic" position (/n/ in *candy* versus *canopy*); with English vowel targets they showed no response difference for "moraic" (/o/ in *kiosk*) versus "non-moraic" vowels (/o/ in *abolish*). In words like *kiosk* the Japanese listeners apparently did not hear successive vocalic elements (which might have triggered a moraic response), but rather a vowel-glide-vowel.

In Murty et al.'s (2007) and Cutler and Otake's (1994) studies, that is, the listeners arguably did not perceive all types of input as allowing a moraic analysis. Cutler et al. (2006) proposed that the explanation of these failures was to be found in the degree to which the listeners could analyze the input as conforming to requirements for application of a moraic segmentation procedure. The phonemic sequences in the Korean input were putatively too far from Japanese phonological legality for a match to native-language expectations to be possible. If this is the correct explanation, Cutler et al. argued, then it may be that listeners will only draw on their native segmentation heuristics when encouraged to do so by the structure of the speech input they are presented with. Experience with the native language has encouraged listeners to develop these heuristics because speech recognition thereby becomes more efficient. Where the phonological structure of a non-native input affords a match to structural expectations from the native language, then, the same experience suggests that listening to this language should benefit from application of the familiar heuristics. This may prove to be the case; but if the apparent match is in fact a spurious one, interference will result instead of benefit. In either case, it is the presence in the non-native input of specific features, matching the listener's expectations for the native language, which has triggered the use of the native segmentation procedures.

This proposal obviously motivates further research. What are the specific features which trigger a given segmentation procedure? Are all segmentation procedures subject to such trigger features? And are trigger features at all mutable? In the latter case, it could be that with increasing experience in the course of acquiring a second language, features which were initially assimilated to native categories might come to be perceived as distinct from those categories (so that the native segmentation procedure would no longer be triggered). Conversely, of course, features of the second language which were not initially perceived as similar to the native language might turn out, on better acquaintance, to share regularities of phonological patterning with those features (and thereby come to trigger the relevant native segmentation procedure after all).

It is clear, therefore, that the research program concerned with rhythmic classes has not yet been completed. Nonetheless, the existence of rhythmic classes which include languages unrelated in origin has received further support from the present data, and this perceptual approach has again proved valuable in providing converging evidence to phonological and phonetic analyses.

References

ABERCROMBIE, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

BARRY, W. J., & RUSSO, M. (2003). Measuring rhythm: Is it separable from speech rate? In A. Mettouchi & G. Ferré (Eds.), *Actes de Interfaces Prosodiques* (pp.15–20). Nantes: Université Nantes.

BLOCH, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, **26**, 86–125.

BOND, Z. S., & STOCKMAL, V. (2002). Distinguishing samples of spoken Korean from rhythmic and regional competitors. *Language Sciences*, **24**, 175–185.

CALDWELL, R. (1856/1976). *A comparative grammar of the Dravidian or South-Indian family of languages*. London/Madras: University of Madras Press.

CASSANDRO, M., COLLET, P., DUARTE, D., GALVES, A., & GARCIA, J. (2003). *A universal linear relation among acoustic correlates of rhythm*. Unpublished MS. Retrieved 29 October 2008, from http://www.ime.usp.br/~tycho/participants/a_galves/linearity.pdf

CHO, M. (2004). Rhythm typology of Korean speech. *Cognitive Process*, **5**, 249–253.

CLASSE, A. (1939). *The rhythm of English prose*. Oxford: Blackwell.
CONDON, W. S., & SANDER, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, **45**, 456–462.
CUTLER, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes*, **12**, 839–845.
CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218–236.
CUTLER, A., DEMUTH, K., & McQUEEN, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science*, **13**, 258–262.
CUTLER, A., KIM, J., & OTAKE, T. (2006). On the limits of L1 influence on non-L1 listening: Evidence from Japanese perception of Korean. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. Auckland, New Zealand (CD-ROM).
CUTLER, A., MEHLER, J., NORRIS, D., & SEGUÍ, J. (1983). A language-specific comprehension strategy. *Nature*, **304**, 159–160.
CUTLER, A., MEHLER, J., NORRIS, D., & SEGUÍ, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385–400.
CUTLER, A., MEHLER, J., NORRIS, D., & SEGUÍ, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381–410.
CUTLER, A., & NORRIS, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113–121.
CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824–844.
DE CORNULIER, B. (1982). *Théorie du vers: Rimbaud, Verlaine, Mallarmé*. Paris: Seuil.
DEMANY, L., McKENZIE, B., & VURPILLOT, E. (1977). Rhythm perception in early infancy. *Nature*, **266**, 718–719.
DERWING, B. L., & NEAREY, T. M. (1994). Sound similarity judgments and segment prominence: A cross-linguistic study. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP 94)*, (vol. 1, pp.351–354).Yokohama, Japan. Retrieved 29 October 2008, from http://www.isca-speech.org/archive/icslp_1994/i94_0351.pdf
DERWING, B. L., & WIEBE, G. (1994). Syllable, mora or segment? Evidence from global sound similarity judgements in Japanese. In P. Koskinen (Ed.), *Proceedings of the 1994 Annual Conference of the Canadian Linguistic Association* (pp.155–163). Toronto: Toronto Working Papers in Linguistics, Linguistic Graduate Course Union.
DETERDING, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, **29**, 217–230.
FORSTER, K. I., & FORSTER, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavioral Research: Methods, Instruments & Computers*, **35**, 116–124.
FRAUENFELDER, U. H., & KEARNS, R. K. (1996). Sequence monitoring. *Language and Cognitive Processes*, **11**, 665–673.
GOW, D. W., & GORDON, P. C. (1993). Coming to terms with stress: Effects of stress location in sentence processing. *Journal of Psycholinguistic Research*, **22**, 545–578.
GRABE, E. (2002). Variation adds to prosodic typology. In B. Bel & I. Marlin (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp.127–132). Aix-en-Provence: Laboratoire Parole et Langage.
GRABE, E., & LOW, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp.515–546). Berlin: Mouton.
HOEQUIST, C. E. (1983). Durational correlates of linguistic rhythm categories. *Phonetica*, **40**, 19–31.

KEANE, E. (2006). Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech*, **49**.

KIM, J., & DAVIS, C. (2006). Literacy acquisition in Korean Hangul. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp.377–387). Hillsdale, NJ: Erlbaum.

KOLINSKY, R., MORAIS, J., & CLUYTENS, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language*, **34**, 19–40.

LADEFOGED, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

LEE, H., & SEONG, C. (1996). Experimental phonetic study of the syllable duration of Korean with respect to the positional effect. *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96)*, (vol. 2, pp.1193–1196). Philadelphia. Retrieved 29 October 2008, from http://www.asel.udel.edu/icslp/cdrom/vol2/355/a355.pdf

LOW, E. L., GRABE, E., & NOLAN, F. (2000). Quantitative characterization of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, **43**, 377–401.

McCANN, D. R. (1976). The structure of the Korean Sijo. *Harvard Journal of Asiatic Studies*, **36**, 114–134.

McQUEEN, J. M., NORRIS, D., & CUTLER, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621–638.

MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U., & SEGUÍ, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298–305.

MORAIS, J., KOLINSKY, R., & NAKAMURA, M. (1996). The psychological reality of speech units in Japanese. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp.203–216). Berlin: Mouton.

MURTY, L., OTAKE, T., & CUTLER, A. (2007). Perceptual tests of rhythmic similarity. I: Mora rhythm. *Language and Speech*, **50**, 77–99.

NAZZI, T., BERTONCINI, J., & MEHLER, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 756–766.

NAZZI, T., & RAMUS, F. (2003). Perception and acquisition of linguistic rhythm by infants, *Speech Communication*, **41**, 233–243.

NELSON, D. L., & NELSON, L. D. (1970). Rated acoustic (articulatory) similarity for word pairs varying in number and ordinal position of common letters. *Psychonomic Science*, **19**, 81–82.

OTAKE, T., DAVIS, S., & CUTLER, A. (1995). Listeners' representations of within-word structure: A cross-linguistic and cross-dialectal investigation. *Proceedings of EUROSPEECH 95* (vol. 3, pp.1703–1706). Retrieved 29 October, 2008 from http://www.isca-speech.org/archive/eurospeech_1995/e95_1703.pdf

OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258–278.

OTAKE, T., HATANO, G., & YONEYAMA, K. (1996a). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp.183–201). Berlin: Mouton.

OTAKE, T., YONEYAMA, K., CUTLER, A., & Van der LUGT, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, **100**, 3831–3842.

PARKER JONES, O. (2006). Durational variability and stress-timing in Hawaiian. In P. Warren, & C. I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. Auckland, New Zealand (CD-ROM).

PIKE, K. L. (1946). *The intonation of American English*. Ann Arbor: University of Michigan Press.

RAMUS, F. (1999). *Rythme des langues et acquisition du langage* [doctoral dissertation]. Paris: EHESS.
RAMUS, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In B. Bel & I. Marlin (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp.115–120). Aix-en-Provence: Laboratoire Parole et Langage.
RAMUS, F., NESPOR, M., & MEHLER, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.
RUTT, R. (1958). An introduction to Sijo: A form of short Korean poem. *Transactions of the Korea Branch of the Royal Asiatic Society*, **34**, 7–11.
SEBASTIÁN-GALLÉS, M., DUPOUX, E., SEGUÍ, J., & MEHLER, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31**, 18–32.
SHIBA, S. (1973). Daravida-go to Nihongo (Dravidian languages and Japanese). *Jinbun Ronsoo*, **22**, 19–53.
STOCKMAL, V., MOATES, D. R., & BOND, Z. S. (2000). Same talker, different language. *Applied Psycholinguistics*, **21**, 383–393.
TAFT, M. (1992). The body of the BOSS: Subsyllabic units in the lexical processing of polysyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 1004–1014.
TREIMAN, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology*, **39**, 161–181.
VAISSIÈRE, J. (1983). Language-independent prosodic features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp.53–66). Berlin: Springer.
VROOMEN, J., Van ZON, M., & De GELDER, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24**, 744–755.
WHITE, L., MATTYS, S. L., SERIES, L., & GAGE, S. (2007). Rhythm metrics predict rhythmic discrimination. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp.1009–1012). Saarbrücken: Saarland University.
YOON, Y. B., & DERWING, B. L. (1995). Syllable saliency in the perception of Korean words. In K. Elenius (Ed.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (vol. 2, pp. 602–605). Stockholm: KTH & Stockholm University.
YOON, Y. B., & DERWING, B. L. (2001). A language without a rhyme: Syllable structure experiments in Korean. *Canadian Journal of Linguistics*, **46**, 187–237.
ZWITSERLOOD, P. (1989). The locus of the effects of sentential-semantic context on spoken-word processing. *Cognition*, **32**, 25–64.