Perceptual Tests of Rhythmic Similarity: I. Mora Rhythm

Lalita Murty[1], **Takashi Otake**[1],
Anne Cutler[1,2]

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

Key words

Japanese

rhythm

segmentation

Telugu

word recognition

Abstract

Listeners rely on native-language rhythm in segmenting speech; in different languages, stress-, syllable- or mora-based rhythm is exploited. The rhythmic similarity hypothesis holds that where two languages have similar rhythm, listeners of each language should segment their own and the other language similarly. Such similarity in listening was previously observed only for related languages (English-Dutch; French-Spanish). We now report three experiments in which speakers of Telugu, a Dravidian language unrelated to Japanese but similar to it in crucial aspects of rhythmic structure, heard speech in Japanese and in their own language, and Japanese listeners heard Telugu. For the Telugu listeners, detection of target sequences in Japanese speech was harder when target boundaries mismatched mora boundaries, exactly the pattern that Japanese listeners earlier exhibited with Japanese and other languages. The same results appeared when Japanese listeners heard Telugu speech containing only codas permissible in Japanese. Telugu listeners' results with Telugu speech were mixed, but the overall pattern revealed correspondences between the response patterns of the two listener groups, as predicted by the rhythmic similarity hypothesis. Telugu and Japanese listeners appear to command similar procedures for speech segmentation, further bolstering the proposal that aspects of language phonological structure affect listeners' speech segmentation.

1 Introduction

The language-specificity of listening is well-established; two decades of research have produced a wealth of evidence that listeners process spoken language in ways

which are subtly tailored to suit the phonological structure of their native language. One aspect of language structure affecting adaptation of processing in this way is rhythm. Rhythm refers to the regularities which govern grouping of elements in a language's phonological structure. It is expressed in timing effects, but also in such factors as variability of syllable structure and the function of accent; these factors vary across languages, and thus so does rhythmic structure. Psycholinguistic research has indicated that languages with similar rhythmic structure encourage similar types of processing heuristics; languages that differ in rhythmic structure correspondingly encourage different types of processing.

The segmentation of continuous speech into its component words is an important part of our listening ability. Speech signals are truly continuous and contain no particular acoustic markers of most word boundaries. Yet the only way that a speech signal can be understood as a whole is by understanding the individual words of which it is made up; representations of all the words we know are stored in our lexical memory, but we do not store whole utterances, because (with a few exceptions) each utterance is new to us. Thus understanding is based on word recognition, and the prerequisite for word recognition is lexical segmentation of utterances. This cannot simply amount to recognition of each word in linear order, because words overwhelmingly contain other words embedded within them, in particular at their beginnings (McQueen, Cutler, Briscoe, & Norris, 1995); this means that the first word that arrives may not be the intended word at all (*star* may turn into *start, starch, startle* etc.). Spoken-word recognition involves multiple concurrent activation of words with full or partial support from the signal, and competition between them, and this process can itself support eventual segmentation of the continuous stream, in that the candidates with the most support finally win out (see Cutler, 2004, for a review). Nonetheless, listeners also exploit many other sources of information that can assist with the segmentation process, including phonotactic boundary cues (McQueen, 1998; Weber & Cutler, 2006), syllable duration (Salverda, Dahan, & McQueen, 2003) and phonological constraints on word structure, such as vowel harmony (Suomi, McQueen, & Cutler, 1997). Among these segmentation aids is linguistic rhythm.

Thus early demonstrations of syllabic segmentation of speech by French listeners but not by English listeners (Cutler, Mehler, Norris, & Seguí, 1983; 1986) were followed by demonstrations of stress-based segmentation by English listeners (Butterfield & Cutler, 1988; Cutler & Butterfield, 1992; Cutler, Mehler, Norris, & Seguí, 1989, 1992; Cutler & Norris, 1988; Smith, Cutler, Butterfield, & Nimmo-Smith, 1989). Since French and English had long been held up as the classic examples supporting the rhythmic distinction of "syllable-timed" versus "stress-timed" languages (Abercrombie, 1967; Classe, 1939; Pike, 1946), the observed differences in segmentation behavior prompted the proposal that language-specific segmentation of continuous speech was based on language-specific rhythm. Cutler et al. (1992) argued that infants, already known to attune early to the timing of the ambient language (Condon & Sander, 1974; Demany, McKenzie, & Vurpillot, 1977), use language rhythm as a handy aid to identify potential discrete units in the continuous speech stream. This enables them to build up a stock of potential word forms, as the initial step in bootstrapping a vocabulary; vocabulary initiation is otherwise a difficult process, given that over 90% of the words infants hear in the first year of life occur in continuous speech (van de Weijer, 1999). The

early experience with using rhythmic structure to segment continuous speech then translates into continued reliance on this handy aid in later life; and because rhythmic structure differs across languages, segmentation procedures based upon rhythmic structure will be language-specific too.

The rhythmic segmentation hypothesis thus proposed that segmentation in French was syllable-based because the rhythm of French was syllable-based, and segmentation in English was stress-based because the rhythm of English was stress-based. The crucial test for this hypothesis consisted in the search for a different rhythmic structure (i.e., one that was neither stress- nor syllable-based), and the demonstration that it, too, was the basis of a language segmentation procedure. The moraic rhythm of Japanese, already established in the phonetic literature as distinct from the stress rhythm of, for example, English and the syllabic rhythm of, for example, French (Hoequist, 1983; Port, Dalby, & O'Dell, 1987) stood up to this test; mora-based speech segmentation was attested in a series of studies in Japanese (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993; Otake, Hatano, & Yoneyama, 1996a; Otake, Yoneyama, Cutler, & van der Lugt, 1996b).

The consequences of language-specific segmentation procedures, it was further argued (see Cutler, 2001, for a review), included difficulty in segmenting non-native languages with rhythmic structure different from that of the native language. Certainly there was abundant experimental evidence that non-native listeners did not necessarily segment speech in the same way as native listeners. Thus not only did English listeners fail to show the French-like syllabic segmentation pattern with French speech materials (Cutler et al., 1983, 1986), but Japanese listeners (Otake et al., 1996a) and Dutch listeners (Cutler, 1997) also failed to show this pattern. French listeners presented with Japanese did not show the mora-based segmentation which was evident in native Japanese listeners' response patterns (Otake et al., 1993), and neither did English listeners (Cutler & Otake, 1994; Otake et al., 1993) or Dutch listeners (Otake et al., 1996b). The patterns of English listeners' responses with English materials were likewise replicated neither with French listeners (Cutler et al., 1983, 1986) nor with Japanese listeners (Cutler & Otake, 1994).

This series of studies largely made use of a paradigm introduced by Mehler, Dommergues, Frauenfelder, and Seguí (1981), later dubbed fragment detection (Frauenfelder & Kearns, 1996): response times and miss rates are recorded as listeners attempt to detect prespecified targets — fragments of words — in spoken input (typically isolated words, but it may also be continuous speech, or nonsense). The rationale underlying the use of this task to study lexical segmentation is that insofar as the boundaries of the target fragment in the input correspond to boundaries which the listener finds easy to detect anyway, the task will be easier to perform. The several types of effects which were repeatedly observed are schematically depicted in Figure 1.

Groups of languages that produced similar patterns of results emerged in the course of this research program. Spanish and Catalan listeners proved to respond similarly to French listeners (Bradley, Sánchez-Casas, & García-Albea, 1993; Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992), and Dutch listeners produced results which strongly resembled

Figure 1

Schematic representation of the four principal effect types observed with the fragment detection paradigm. In all experiments, detection of CV versus CVC targets is compared in polysyllabic words beginning CVCV- versus CVCC-. An English example is detection of *ba*- versus *bal*- in *balance* versus *balcony*. The dummy numbers on the vertical axis stand for a dependent response variable, which may be response time (RT) or percent missed responses. (a) "Syllabic crossover" effect: Responses are facilitated (faster RTs or fewer missed responses) when the target matches the first syllable of the carrier word compared with when it does not match. This pattern was shown by French listeners with materials in several languages (Cutler et al., 1986; Mehler et al., 1981; Otake et al., 1993). (b) "Mora" effect: Only targets which mismatch the mora boundaries of the input are difficult. This effect was shown by Japanese listeners only, and only with Japanese input (Otake et al, 1993, 1996a). In Japanese, only nasal codas are permitted, so CVC=CVN, where N stands for any nasal consonant. (c) Word type effect: Whatever the target, responses are always facilitated for one word type. This effect was shown by English listeners with materials in several languages (Cutler et al., 1986; Cutler, Norris, & Williams, 1987). (d) Target type effect: Whatever the word type, responses are always facilitated for one target type. An advantage for CV targets appeared when Japanese listeners were presented with foreign-language materials (Otake et al., 1996a), and was interpreted as a partial moraic effect; an advantage for CVC targets appeared in experiments with English and with Dutch speakers (Bradley et al., 1993; Zwitserlood et al., 1993) and was interpreted as a partial syllabic effect
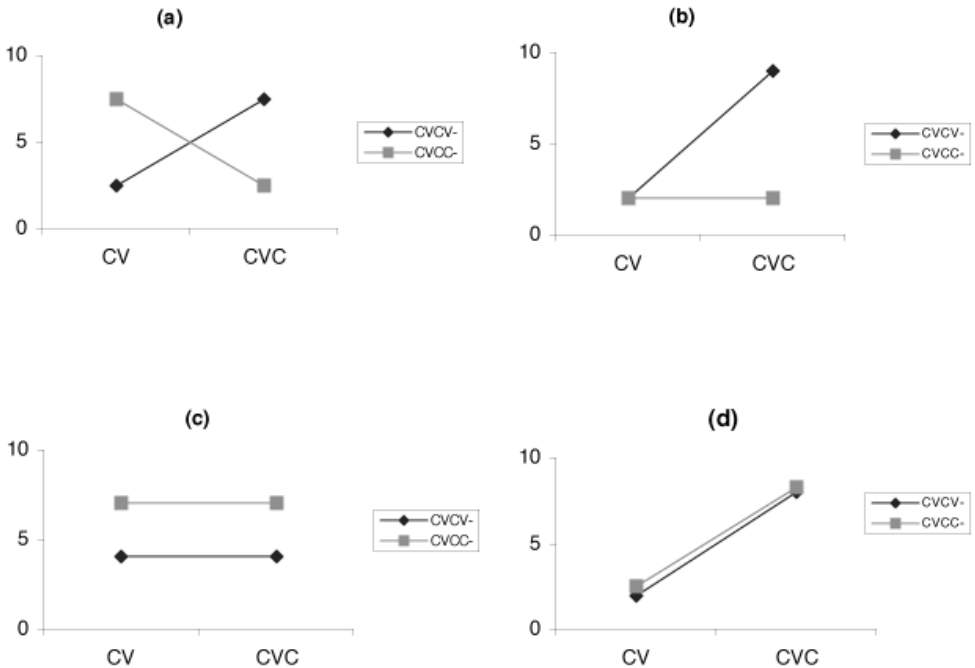
those of English listeners (Vroomen & de Gelder, 1994; Vroomen, van Zon, & de Gelder, 1996). An effect interpreted by Bradley et al. (1993) and Zwitserlood, Schriefers, Lahiri, and Donselaar (1993) as a partial syllabic effect was also observed for both

English and Dutch, although later analyses suggested that this effect was largely due to acoustic differences in the materials (Donselaar & Stoutjesdijk, 1994).

These similarity groups, of course, involve languages with a common ancestral origin. An interesting question which was raised by several groups of researchers, however, concerned the possibility of rhythmic similarity classes comprising sets of potentially unrelated languages. The early stress- versus syllable-timing proposals (see, e.g., Abercrombie, 1967) had been intended to embrace any and all languages, whereby languages were preferably exemplars of one or the other category. These proposals were based strictly on regularity of timing units (stress feet, syllables), though this proved to be too simple to do justice to real speech (see, e.g., Laver, 1994; Roach, 1982). More complex proposals based on phonological structure broadened the definition of rhythm to include factors such as the variety and nature of syllable types in a language, and the patterning of accents (e.g., Dasher & Bolinger, 1982; Dauer, 1987), and also allowed for languages possessing features of both categories (Dauer, 1987; Roach, 1982).

Two recent initiatives have attempted to provide a phonetic measure which would enable grouping of languages into rhythmic classes. One is classification on the basis of the Pairwise Variability Index (PVI; Deterding, 2001; Grabe, 2002; Grabe & Low, 2002; Low, Grabe, & Nolan, 2000), on which languages can be assigned a point in a space defined by the average variability in the normalized duration of successive vowels and the average variability in the duration of successive intervals between vowels in speech samples. The other approach similarly assigns languages to a point in a space defined by the *SD* of the intervocalic intervals and the proportion of speech time taken up by vowels (Ramus, 2002; Ramus, Nespor, & Mehler, 1999). Each of these approaches resulted in a clear separation between languages (e.g., British English; French) which were traditionally viewed as differing in rhythm. However, the two approaches vary in agreement on where other languages fall in the space they plot (see, e.g., Grabe & Low, 2002 and Ramus, 2002, for comparisons of samples on both measures). Extension of the measures to other languages also produced mixed results (Keane, 2006; Stockmal, Markus, & Bond, 2005). More seriously, the measures vary with speech rate (at faster rates, differences between languages become smaller: Barry & Russo, 2003, though distance between dialects does not: Deterding, 2001), with variety, both dialectal (Deterding, 2001; Grabe, 2002; Low, Grabe, & Nolan, 2000) and colloquial versus formal (Keane, 2006), and even across speakers of the same variety (Keane, 2006). It is possible that very large and carefully representative standardized samples from each language could produce greater convergence between the measures, as well as eventual stability of groupings of languages on each measure; we certainly hope that in the future it will be possible to apply these (currently labor-intensive) measurement tools to large samples.

In this project, however, we have taken a different route to the identification of rhythmic similarity. We start from the perceptual data and ask whether we can observe similarities in listening performance as a function of language background. Given the substantial amount of perceptual data already available, this route is a very tractable way to begin a similarity exploration. We chose to start by seeking a language which we suspected might pattern perceptually like Japanese. There were several reasons for starting at that point. First, the existing body of research summarized above had

found several languages which produced patterns of results similar to English and similar to French, respectively, but so far none similar to Japanese. Second, Japanese is a typological isolate, so that similar perceptual results would be unlikely to be due to contamination from Japanese. Third, the mora — the putative rhythmic unit of Japanese — is the smallest of the rhythmic units proposed to be relevant in the segmentation of continuous speech, and is relatively easy to define (whereas stress, e.g., is notoriously variable in its cross-linguistic manifestations).

The phonological structure of Japanese syllables is simple. Vowels may be short (giving the syllable structures V, CV, CCV) or long (VV, CVV, CCVV); a consonantal coda can only be a nasal (VN, CVN, CCVN) or a geminate with the onset of the following syllable (thus only nasals can be word-final codas). Syllables with a short vowel consist of one mora, syllables with a long vowel or nasal coda are bimoraic. Syllables do not carry context-independent tones. Trimoraic syllables do not occur. The majority of syllables in most speech samples are CV (Otake, 1990).

The world in fact contains many languages with Japanese-like syllable structure. Among them are some with huge speaker populations (making them relatively accessible to psycholinguistic experimentation) — for example, the Dravidian languages of South India, one of the world's larger language families. The principal Dravidian languages are Telugu (75 million speakers; values estimated from SIL's website *ethnologue.com* and from the 2001 Indian National Census), Tamil (63 million speakers), Kannada (50 million speakers), and Malayalam (30 million speakers), though there are many smaller languages in the family (see Steever, 1998). The similarity between Japanese and the Dravidian family was pointed out over a century ago (Caldwell, 1856), and has been noted by both Japanese linguists (e.g., Ohno, 1981; Shiba, 1973) and Indian scholars (e.g., Dutt, 1992).

Among these languages, only Tamil has been subject to phonetic analyses of the type described above. A set of analyses using Ramus' measures found values for Tamil very close to those for Japanese (Mehler & Dupoux, 2004), while another set using the same measures on a different sample found that Tamil patterned differently from Japanese (and from all 17 other languages in that study; Grabe & Low, 2002). On the PVI, the latter Tamil sample patterned similarly to Singapore English (Grabe & Low, 2002). Mehler and Dupoux (2004) claimed that Tamil should group with Japanese in a rhythmic class, while Grabe and Low (2002) concluded that Tamil was unclassifiable. On the basis of analyses of colloquial and formal Tamil speech from five speakers Keane (2006) also argued that Tamil belongs in none of the traditional rhythmic classes. Note that Tamil is both highly diglossic and the most dialectally variable of the Dravidian languages (Annamalai & Steever, 1998; Keane, 2004, 2006). Dialects of one language can differ on these phonetic indices (Low, Grabe, & Nolan, 2000). Keane's results showed large differences across the colloquial and formal varieties, suggesting that the measures can also reflect variety differences. Thus the earlier differing results may simply reflect differing dialects or varieties in the Tamil samples which were analyzed.

Our investigation involved Telugu, the native language of the first author. Since Brown (1857), phonological descriptions of Telugu have argued that Telugu syllable structure is weight-sensitive and moraic; we base our account on Krishnamurti (1998).

In native Telugu words, most syllables are CV. Vowels can be short or long; geminate and nasal codas both occur. As in Japanese, there is no coda after long vowels. The only permissible word-final coda is a nasal /m/ or semivowel /w, j/. Syllables do not carry context-independent tones. Nativized forms of borrowed words show the effects of these phonological constraints. Thus if the borrowed word has a short vowel in the final syllable, the last consonant will be geminated and an epenthetic vowel added to avoid impermissible consonants in word final coda position; the English words, *pen*, *jug*, *tulip*, for example, become *pennu*, *juggu*, and *tulippu* respectively. If the borrowed word has a long vowel, then only vowel epenthesis takes place; thus *glass*, *bar*, and *blouse* become *glassu*, *baru*, and *blousu* (Sailaja, 1999; Vasanta, 1997).

Although the constraints on word-final codas are uncontroversial, there has been controversy over whether syllabic codas other than nasal and geminate are possible (word-medially) in Telugu. Positions range from that of Mohanan (1989), who claims that the Dravidian syllable has no codas, through Reddy (1979), who claims that the nasal coda is the only true coda and all other syllables are open, to Rao (1996), who claims that word-medial clusters can result in closed syllables. Studies of language games and explicit syllabification decisions in Telugu, Malayalam, and Kannada have in fact found a surprisingly large amount of variation in the way word medial clusters are syllabified. Thus Mohanan (1989) cites studies in which, in violation of predictions from the sequencing of the Sonority Hierarchy, word medial clusters were syllabified consistently with the following vowel (e.g., *sa-kti*). Sailaja (1999) found in a production task that speakers similarly preferred not to split clusters but to produce them as syllabic onsets. The principal exception to this pattern was word-medial clusters of nasal plus obstruent; in both production and judgment tasks in the studies of Sailaja (1998, 1999), these were consistently split across the two syllables, leaving the nasal in coda position.

This suggests that the nasal consonants have a special status in the language, and certainly constitute the most acceptable syllabic coda. Note that there are several respects in which the position of the nasal coda in Telugu resembles that in Japanese. For instance, in both languages there is a default form (in Japanese /ŋ/, in Telugu /m/) which, however, in word-medial position always assimilates to the following consonant in place of articulation: Japanese *samban* /sambaɴ/ 'third', *sanju* /sandʒɯ/ '30', *sangatsu* /saŋgatsɯ̥/ 'March', Telugu *kempu* /kɛmpu/ 'coral', *panTa* /pʌɳʈa/ 'crop', *penku* /pɛŋku/ 'tile'. Likewise, in both languages the orthography represents syllable-final nasals with a single grapheme, which differs from the graphemes used for the nasals in other positions. In Japanese kana scripts, the word-medial nasal has a special symbol; in Telugu, it is always represented by a grapheme called *anuswara*, written as a circle 'o'. The anuswara is traditionally listed in the alphabet with vowels and is considered a part of the preceding syllable (Krishnamurti & Gwynn, 1985).

Finally, Telugu also passes the poetic test for moraic rhythm. The poetic structures preferred by a linguistic community correlate reliably with listeners' sensitivity to rhythmic structure in speech segmentation. Thus just as English, Japanese, and French listeners respectively use stress, moraic structure and syllabic structure in speech segmentation, so is English poetry overwhelmingly stress-based, while Japanese poetry is largely based on moraic patterning and French poetry on syllabic patterning. Telugu poetic forms crucially make reference to moraic structure; thus, for instance,

certain kinds of feet are constrained to contain two short (monomoraic) syllables or a single long (bimoraic) syllable in a given position (Brown, 1857).

Although Telugu phonology is not in all respects analogous to Japanese phonology (for instance, Telugu has many more phonemes than Japanese), there was thus sufficient motivation, in the existing literature (e.g., Dutt, 1992; Ohno, 1981) as well as in the phonological similarities listed here, for us to consider Telugu a candidate for rhythmic similarity to Japanese. We began our investigation by presenting Telugu listeners with the Japanese fragment detection materials previously used in the study of Otake et al. (1993). In Otake et al.'s experiment, Japanese listeners heard lists of words and pressed a button as soon as they detected a specified word-initial fragment. The words began either CVCV- (e.g., *tanishi*) or CVNC- (e.g., *tanshi*), and the targets were either CV (e.g., *ta-*) or CVN (e.g., *tan-*). Listeners' responses were accurate and rapid for CV targets such as *ta-* in both word types (*tanishi*, *tanshi*). They also accurately detected CVN targets (*tan-*) in the CVNC- words (*tanshi*). However, detection of *tan-* in *tanishi* was extremely difficult — the majority of targets were not detected at all. Recall that no other listener group tested with this set of materials has shown this pattern; both French listeners and English listeners produced different patterns of response (Otake et al., 1993). In Experiment 1 we examine the response of Telugu listeners to these Japanese materials.

2 Experiment 1

2.1
Method

2.1.1
Participants

Forty-one undergraduate students at Sai Sudhir College in Hyderabad, India, aged 19 to 22, participated in the experiment in return for a small payment. All subjects had Telugu as mother tongue, none had any previous experience with Japanese, and none reported any hearing problems. Twenty heard one version of the experiment and 21 the other.

2.1.2
Materials

The materials were those of Otake et al. (1993; Experiment 3): 16 meaningful words, in eight pairs: *tanishi, tanshi; monaka, monka; kanoko, kanko; sanaka, sanka; nanoka, nanka; kinori, kinri; haneda, handa; shinigao, shingao*. Within each of these Japanese pairs, words had the same initial and final mora. Thus both *tanishi* and *tanshi* have *ta* and *shi* as their initial and final moras respectively, but they differ in the medial mora, *tanishi* having a CV mora beginning with a nasal, *ni*, and *tanshi* a nasal consonant coda. The members of a pair were unrelated in meaning (in this pair, e.g., *tanishi* is a kind of snail, *tanshi* a terminal). Accent pattern was matched across pairs. 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 once in one of the last 32 sequences.

Sequence length varied from three to six words, with the target words in second, third, fourth or fifth position.

Two test versions were compiled, with type of target counterbalanced across order for each pair of items. Thus in version A subjects monitored for *ta-* in a sequence containing the target word *tanishi,* and for *tan-* in a sequence with *tanshi*. Subjects in version B monitored for *tan-* in *tanishi* and for *ta-* in *tanshi*.

The words were spoken by a male native speaker of Standard Tokyo Japanese, and the target specifications that preceded each sequence were spoken separately by the same speaker. The sequences were pronounced 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. The sequences with their target specifications were recorded onto the hard disc of a notebook computer. The sequences (also those for Experiments 2 and 3) are available at <*http://www.mpi.nl/world/persons/private/anne/ materials.html*>.

2.1.3
Procedure

Subjects were tested individually in a quiet room in the college. They heard the target specifications and sequences over high-quality closed headphones from the computer. They were told (in Telugu and English) that they were going to be listening to Japanese, and that they should monitor for a word beginning with the sounds specified as target for each sequence, and press a button as soon as they detected the target fragment. Response times were measured from a signal (inaudible to the participants) aligned with the onset of each target item, and timing was terminated by the subject's button press. Response times were collected by the notebook computer, using the experimental control software NESU (<*http://www.mpi.nl/tools/nesu.html*>). 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 subjects (F1) and across items (F2).

2.2
Results and Discussion

The mean RT was 895 ms. The miss rate in this experiment was high (19.7%), and comparable with the original study of Otake et al. (1993), where the overall miss rate was 22%. The miss rate and RT results are displayed in Figure 2. The pattern is very similar to that of the original Japanese results; the miss rate for the CVN targets in CVCV- words is double that in the other three conditions, and the responses to those targets are slower than in any of the other three conditions.

Statistical analyses (using $\alpha = .05$ as criterion for significance) confirmed the impression given by Figure 2. Analyses of variance for the RTs showed a significant interaction between word type and target type, $F_1(1, 40) = 4.9$, $p < .04$; $F_2(1, 14) = 4.76$, $p < .05$. This interaction was further explored in separate *t* tests for each word type; these revealed no significant effects of target type in CVNC- words, but significantly slower RTs to CVN than to CV targets in CVCV- words $t_1(40) = 2.99$, $p < .01$; $t_2(7) = 4.05$,

Figure 2

Mean response times and mean miss rates, across subjects and items, as a function of target type (CV, CVN) and word type (CVCV-, CVNC-), in Experiment 1 (Telugu listeners, Japanese materials)
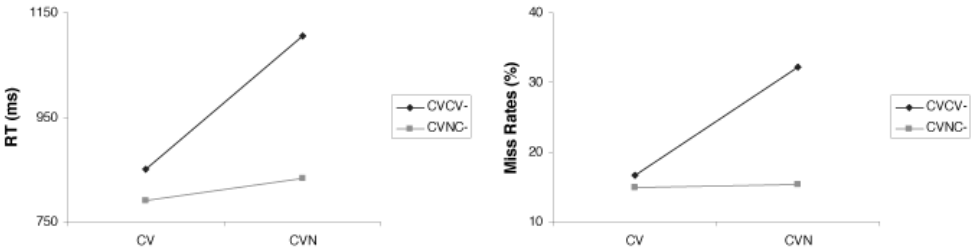
$p<.005$. Analyses of variance for the miss rates likewise showed that the interaction between word type and target type was significant, $F_1(1, 40) = 11.86$, $p<.005$; $F_2(1, 14) = 13.48$, $p<.005$; again, t tests showed no significant differences in the miss rates across target type in CVNC- words, but significantly higher miss rates for CVN than for CV targets in CVCV- words, $t_1(40) = 3.98$, $p<.001$; $t_2(7) = 5.67$, $p<.001$.

The pattern of results both in RTs and miss rates for the Telugu listeners with the Japanese material of Otake et al. (1993) is very similar to the pattern observed in the original study. Telugu listeners are the first non-native group tested with these materials to show a similar pattern to the native Japanese listeners. This result motivated us to construct analogous materials for a comparable study in the Telugu language.

Recall that there are controversies among Telugu phonologists concerning the status of syllabic codas in the language. We first desired to construct an experiment as closely as possible modeled on the materials of Experiment 1. However, it also seemed of potential interest to use the perceptual yardstick afforded by this experimental paradigm to shed light on the phonological questions at issue. We thus constructed two sets of materials with which we conducted two experiments; in Experiment 2a, the materials were closely modeled on those of Experiment 1, while in Experiment 2b the materials contained a wider sample of possible Telugu syllabic forms.

3 Experiment 2

3.1
Method

3.1.1
Participants

Forty undergraduate students at Satavahana College in Vijayawada, India, took part in Experiment 2a, and 40 undergraduate students at Sai Sudhir College in Hyderabad took part in Experiment 2b. The average age of each group was 20 years. Each participant received a small payment for taking part; all were native speakers of Telugu with normal hearing. For each experiment, 20 subjects heard each version.

3.1.2
Materials

For Experiment 2a, 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.* As in the materials of Experiment 1, these words form 16 pairs containing a word-medial mora in the form of either a Nasal+V sequence (the first word of the pair) or a single Nasal (the second word). 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 all target specifications and a further 278 words onto the hard disc of a computer. The filler words were not always structures which would also be possible in Japanese, but an attempt was made to avoid sounds and sequences which would be especially difficult for Japanese listeners. The words were then arranged into 10 practice sequences and 64 experimental sequences, varying in length from four to seven words, with the target word in third, fourth, fifth, or sixth position. As in Experiment 1, 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*, while subjects in version B monitored for *ma-* in a sequence containing *mandu,* and for *man-* in a sequence with *manishi.*

For Experiment 2b, 32 Telugu words were again selected, again in pairs beginning with the same three initial phonemes; however, the syllabic coda was in this case not always a nasal. The pairs were *benuku, benda; bharine, bharta; chamuru, champa; chemata, chembu; chinuku, chinta; godugu, godraalu; kaanuka, kaanti; kaasepu, kaasta; kaluva, kalpana; kunuku, kunda; manishi, mandu; muuredu, muurcha; paatalu, paatlu; panulu, panche; pasupu, pastu; waasana, waastu.* Filler words also varied across the phonological repertoire of Telugu. Recording and arrangement in experimental sequences was as for Experiment 2a.

3.1.3
Procedure

In both experiments, the procedure was as for Experiment 1 except that listeners were told that they would hear words of their native language. Analyses were conducted in the same manner as for Experiment 1.

3.2
## Results and Discussion

3.2.1
Experiment 2a

After the experiment had been run, it was discovered that the timing signals for two items had not triggered the computer's timers, resulting in no responses being available for these items. These words and their pairs with the same target (*mandu*/*manishi* with target *man-*; *winta*/*winati* with target *wi-*) were therefore removed from all

analyses. Mean RTs and miss rates calculated over the remaining items are displayed in Figure 3a.

Figure 3

Mean response times and mean miss rates, across subjects and items, in Experiments 2a and 2b (Telugu listeners, Telugu materials). Experiment 2a: Target type is CV, CVN, word type is CVCV-, CVNC-; Experiment 2b: Target type is CV, CVC, word type is CVCV-, CVCC-
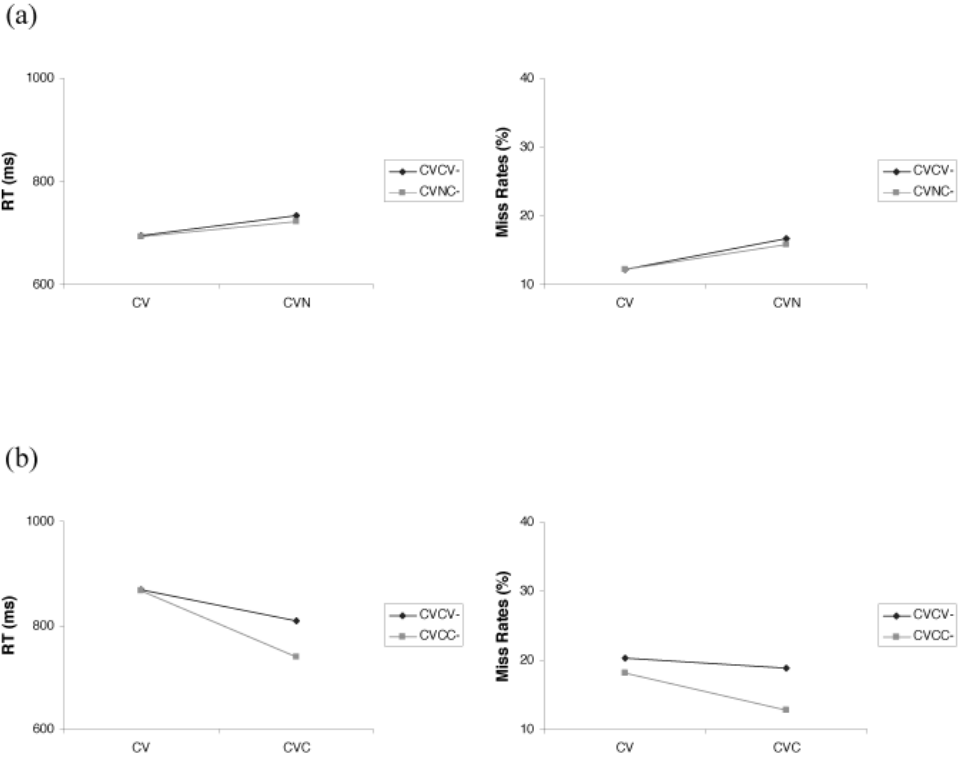
The mean RT was 711 ms and the mean error rate 14.2%. In the analysis of variance for RTs, no effects reached our .05 criterion of significance. The analysis of variance for miss rates showed that fewer missed responses occurred for CV than for CVN targets, although this effect was only significant across subjects, $F_1(1, 39) = 4.85$, $p < .05$; $F_2(1, 26) = 2.89$, $n.s.$; the effect of word type and the interaction between the two main effects were insignificant.

3.2.2
Experiment 2b

All items were analyzed for this experiment. The mean RTs and miss rates are displayed in Figure 3b. The overall mean RT was 821 ms and miss rate 17.5%. Again the effect of word type and the interaction were insignificant in all analyses; the effect of target

type was significant but in the opposite direction for Experiment 2a, in that CVC targets produced faster RTs, $F_1(1, 39) = 8.39$, $p < .01$; $F_2(1, 30) = 3.69$, $p < .065$, and fewer missed responses, $F_1(1, 39) = 5.62$, $p < .025$; $F_2(1, 30) = 2.41$, *n.s.*, although, again, this effect was in each case statistically robust only across subjects.

The results of both these experiments are unlike the results of equivalent experiments with Japanese listeners and Japanese materials. Although we in fact observed no effects which were statistically robust across both subjects and items, the pattern in Experiment 2a, a general advantage for CV targets, resembles most strongly the pattern which Otake et al. (1996a) found for Japanese listeners hearing English, French, and Spanish materials, and which those authors interpreted as mora-based. The pattern in Experiment 2b, a general advantage for CVC targets, resembles the finding interpreted as a syllabic effect by Bradley et al. (1993) and Zwitserlood et al. (1993).

In Experiment 3, finally, we presented the Telugu materials to Japanese listeners. The rhythmic similarity hypothesis predicts that Telugu is similar enough to Japanese for a moraic response pattern to emerge with listeners who are sensitive to moraic structure. In Experiment 1 we observed that Telugu listeners indeed displayed such a pattern with Japanese speech; that experiment, involving materials to which the listeners could assign no meaning, in fact may have allowed a particularly unconfounded view of Telugu listeners' phonological sensitivities. This would also hold for Japanese listeners hearing Telugu input, so we predict that Experiment 3 will allow moraic sensitivity to emerge.

4 Experiment 3

4.1
Method

4.1.1
Participants

Thirty-eight undergraduate students, aged 18 to 21, from Dokkyo University participated in Experiment 3a and 40 students from the same population in Experiment 3b, always in return for course credit. All subjects had normal hearing and none had any experience with the Telugu language. Nineteen subjects heard each test version of Experiment 3a and 20 each version of Experiment 3b.

4.1.2
Materials and Procedure

The materials were those of Experiment 2a and 2b respectively. Subjects were tested individually in a quiet room. They were instructed to listen first for the target specification, and then to press a response button as soon as they heard a word in the following sequence of words beginning with the target sound. Analyses were conducted in the same manner as for Experiments 1 and 2.

4.2
Results and Discussion

4.2.1
Experiment 3a

The same two target pairs which were omitted from Experiment 2a were omitted here also. The results are displayed in Figure 4a. The overall mean RT was 835 ms and the mean miss rate 18.5%. Analysis of variance across RTs showed a significant interaction between word type and target type, $F_1(1, 37) = 17.22$, $p < .001$; $F_2(1, 26) = 18.52$, $p < .001$; t tests on the components of this interaction revealed no target type effects for CVNC- words but significantly slower RTs for CVN than for CV targets in CVCV- words, $t_1(37) = 3.42$, $p < .005$; $t_2(13) = 5.27$, $p < .001$. Analysis of miss rates revealed the same pattern: a significant interaction between word and target type, $F_1(1, 37) = 12.46$, $p < .005$; $F_2(1, 26) = 13.16$, $p < .005$, and t tests showing no significant effects of target type in CVNC- words but significantly more missed responses for CVN than for CV targets in CVCV- words, $t_1(37) = 3.09$, $p < .005$; $t_2(13) = 5.84$, $p < .001$.

Figure 4

Mean response times and mean miss rates, across subjects and items, in Experiments 3a and 3b (Japanese listeners, Telugu materials). Experiment 3a: Target type is CV, CVN, word type is CVCV-, CVNC-; Experiment 3b: Target type is CV, CVC, word type is CVCV-, CVCC-
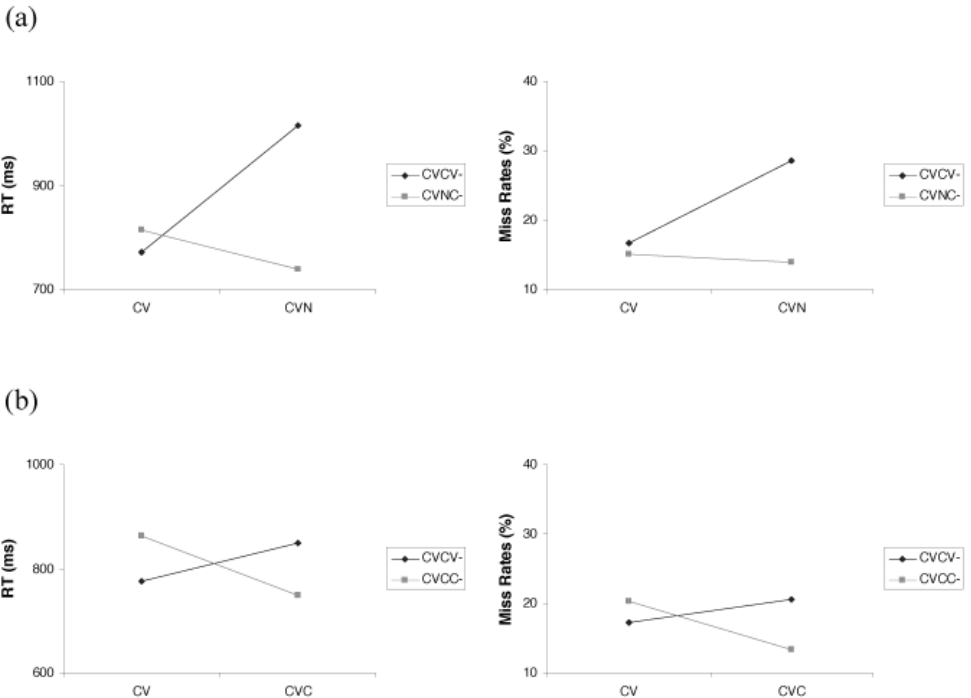
4.2.2
Experiment 3b

The results are displayed in Figure 4b. The overall mean RT was 809ms and miss rate 17.9%. The main effects of word type and target type were insignificant in all analyses. However, a significant interaction emerged, though robust only across subjects, in both RTs, $F_1(1, 39) = 7.33$, $p < .01$; $F_2(1, 30) = 3.71$, $p < .065$, and miss rates, $F_1(1, 39) = 6.53$, $p < .02$; $F_2(1, 30) = 3.97$, $p < .056$. Separate t-tests again explored this interaction; no target type effects were observed in any analysis for CVCV- words, but for CVCC- words, across subjects only, CVC targets enjoyed an advantage over CV targets in both RTs, $t_1(39) = 2.28$, $p < .03$, and miss rates, $t_1(39) = 2.37$, $p < .03$.

The response pattern for RTs and for miss rates was once again consistent in each experiment. It is clear that with stimulus material resembling that of the original Japanese experiment (Experiment 3a), Japanese listeners produce a pattern of responses very similar to what they show with native speech materials. With the more complex phonological structures which Telugu allows but Japanese does not (Experiment 3b), the Japanese listeners do not show a clearly mora-based response pattern; if anything, their responses most resemble the syllabic segmentation effect characteristic of French listeners with native (Mehler et al., 1981) and non-native materials (Otake et al., 1993; Cutler et al., 1986), although it may be unwise to assign much weight to this result given the lower levels of statistical significance. The most robust component of the results of Experiment 3b was an advantage for CVC targets, also shown by the native Telugu listeners with the same materials (Experiment 2b).

Thus despite the absence of a clear and consistent pattern of results for the Telugu listeners presented with their native language in Experiment 2, Experiment 3 has confirmed the suggestion, based on the observed resemblances of phonological structure and the results of the present study's Experiment 1, that there is indeed similarity in the patterns of response provided by Japanese and by Telugu listeners. We consider the implications of this similarity further below.

5 General discussion

The results of our three experiments provide important new insights for the rhythmic similarity debate. First, we observed that Telugu-speaking listeners, when presented with spoken materials in Japanese, a language quite unfamiliar to them, show a response pattern highly similar to that of Japanese native listeners. Since these materials had previously been presented to numerous other groups, none of which had produced responses resembling those of the original Japanese listener group, these results demonstrate that the Japanese response pattern is not unique to Japanese listeners. When Japanese is heard by another group of listeners whose language exhibits phonological similarities to Japanese, a "Japanese-like" response pattern appears. This finding is as predicted by the rhythmic similarity hypothesis; similarity in language phonological structure will be reflected in similarity in response patterns in listening experiments.

Second, however, our experiments with Telugu speakers presented with materials in their own language did not lead to a response pattern matching that of listeners

from the same population with the Japanese materials. The results of the experiments were statistically too weak to motivate any major conclusions concerning a unitary classification of Telugu rhythmic preferences, but it would appear that Telugu speakers are not dependent upon moraic segmentation to the extent that Japanese listeners, to judge by the results of earlier experiments, appear to be. Despite the absence of a single clear pattern in the Telugu native listening results, however, the findings provide a potentially very informative first data set on speech segmentation in this language. We consider below the implications of our findings for the interpretation of Telugu phonological structure and how Telugu listeners exploit it.

Third, however, when Japanese listeners were presented with the same Telugu materials the resulting pattern of responses once again indicated support for the rhythmic similarity hypothesis. Previous experiments in which Japanese listeners had been presented with non-native materials had produced results which were consistent (in that an advantage for CV targets recurred) but different from those which Japanese listeners demonstrate with input in their native language. With Telugu, however, when the target words were chosen so as not to violate Japanese coda constraints (Experiment 3a), the pattern which Japanese listeners showed was just like the pattern they show in Japanese. With more varied Telugu materials (Experiment 3b) their responses were less obviously mora-based; in fact, in this case the statistical analyses of their response pattern, which showed a weak syllabic effect, most strongly evident for CVC targets, resembled the analyses of the pattern produced by the Telugu native listeners.

Experiments such as our Telugu studies are, it should be noted, immensely difficult to undertake. This is not due to any difficulty in access to listener populations—as described above, Telugu is a major world language, and the undergraduate groups we tested, and the conditions of testing, did not noticeably differ from those encountered in undertaking the previous studies in this long-term research program. Likewise, the fragment detection task is easy for listeners in all languages, and appears to be a natural way of examining the boundary detection involved in lexical segmentation in speech comprehension. The difficulty arises, rather, in controlling the linguistic materials when constructing experiments. Electronic dictionaries, and corpora from which frequency counts can be computed, exist for only a tiny minority of the world's languages. This minority includes English, French, Japanese, Dutch, and Spanish—the principal languages studied in rhythmic experiments so far—but it does not as yet include Telugu. Without such resources, it is virtually impossible to control the cohort of potential words activated by a given input, so that we cannot be sure, for example, how strong lexical effects in our Telugu experiments with Telugu listeners may have been, nor can we exclude possible differences along this dimension between the present study and previous experiments with Japanese and other populations.

One way of excluding lexical effects, of course, is to conduct experiments with nonwords—though in the absence of lexical databases, even constructing nonwords in Telugu is difficult. Interestingly, words in an unknown foreign language are effectively nonwords. Thus Experiments 1 and 3 in that sense constitute the cleanest case for a test of our hypothesis. And, as we have seen, Experiments 1 and 3a showed a clear moraic effect. In other words, when Japanese and Telugu listeners are attending simply to the sounds of the input, and neither group is in a position to consider the meaning of the incoming words, each set of listeners is equally able to make use of moraic structure.

What do our experiments tell us, next, about the Telugu language itself, and how Telugu speakers listen to it? With respect to the controversies about Telugu syllable structure, summarized in the Introduction, our results first offer support for the proposal (Reddy, 1979; Sailaja, 1998) that nasal codas have a special status. With Japanese materials in which only nasals occurred as codas, and with only the Telugu materials having the same structure, Telugu listeners showed evidence of exploiting moraic structure in performing the fragment detection task. Second, our results do not support the most radical proposal (Mohanan, 1989), that Telugu syllables have no codas, so that word-medial clusters may be interpreted by listeners as complex onsets. This proposal would have predicted an overall advantage for CV targets; a weak CV advantage was observed in Experiment 2a, but in Experiment 2b, in which complex onsets would have been most in evidence, the reverse effect was observed: an advantage for CVC targets, indicating that these listeners preferred to assign the first element of a word-medial cluster to a coda. As described above, this claim was also put forward in the phonological literature on Telugu by Rao (1996), and this aspect of our results can be seen as supporting Rao's position.

With respect to lexical segmentation in Telugu speech, our findings suggest that Telugu listeners certainly command a moraic segmentation strategy, but they also suggest that this listener population does not make as much use of the strategy as Japanese listeners do. Ito (1988) has argued for Japanese that nasal segments are treated as differing in status as a function of position in the syllable (onset vs. coda), and experiments with Japanese listeners have provided support for this position (Otake & Yoneyama, 1995). Similar experiments with Telugu have not been conducted as yet but could be potentially revealing. Neither in Japanese nor in Telugu did our experiments exploit geminate consonants in coda position — it is not necessarily possible to find the requisite minimal pairs among real words of the language — but it may be that in both languages such materials with geminate consonants would, like the materials with nasal codas, encourage use of moraic structure in segmentation.

Experiments with child listeners performing a simple game of breaking words into pieces show early sensitivity to moraic structure in Japanese (Inagaki, Hatano, & Otake, 2000), although there is evidence that moraic structure is less salient than syllabic structure for very young children, becoming more salient as children grow older (Otake & Imai, 2001). Again, we have no comparable child listening data from Telugu, but such data likewise would be very useful in establishing how easily available moraic structure is to listeners in that language. The orthographic marking of the moraic nasal as a special case in each language underlines the availability of this structure to adult listeners; for Japanese, however, the sensitivity of preliterate children to moraic structure is evidence that adult sensitivity is not itself entirely determined by the orthographic marking. Thus it would also be valuable to conduct child listening experiments, of the same kind conducted by Inagaki et al. (2000), in Telugu.

The syllable and the mora are phonological constructs in terms of which speech in any language may be described. This is not true of all aspects of phonological description, of course — lexical tone distinctions, for example, are absent from many languages, and likewise, many languages fail to exploit stress distinctions. But the syllable and the mora are simply language-general descriptive terms. Thus it is not the case with these structural features that some languages offer contrasts which

other languages simply do not present. Exploitation of moraic structure or syllabic structure in speech segmentation in principle is available to every listener in the world. The listeners who use either one of these, therefore, obviously use it because their language makes it helpful for them to do so, and the listeners who fail to use it ignore it because in their language it is of little use to them. Everything that decades of psycholinguistic experimentation on speech segmentation have revealed confirms that listeners are extremely efficient in using every aspect of their language's structure which facilitates performance of the segmentation task (and correspondingly ignoring structure which is irrelevant in this respect).

In some languages, then, moraic structure usefully aids segmentation of speech into words, and in some languages syllabic structure performs the same function. In yet other languages neither of these levels of structure proves very useful in spoken word recognition (English is one such language). In extension of this account, it is not necessarily the case that only one such level of structure would be always useful. There is no reason in principle why listeners should not be able to call upon the different strategies which may be more or less useful given the nature of the input. Indeed, evidence from experiments using the migration task has shown that Japanese listeners can manipulate both moraic and syllabic structure (Morais, Kolinsky, & Nakamura, 1996). This is not surprising; as Kubozono (1998) has argued, both morae and syllables play a role in accent assignment in Japanese, and Tanaka (1999) has likewise argued that Japanese poetic structures can involve syllabic as well as moraic constituents. As we observed above, preferred poetic structures in a linguistic community offer a reliable guide to listeners' knowledge about the relevant phonological structure of their language — whereas English poetry is mostly stress-based, Japanese poetry mostly mora-based and French poetry mostly syllabic, so are the *most preferred* English segmentation procedures stress-based, French procedures syllable-based, and Japanese procedures mora-based. Telugu poetic structures are likewise not exclusively moraic but involve reference to both moraic and syllabic components (Brown, 1857).

If listeners in a given language may call upon procedures exploiting more than one type of structure, the extent to which they make use of one procedure rather than another in segmenting speech into words may depend on several factors: the nature of the input, the nature of the segmentation task which they are performing at the time, or even listener-specific factors such as individual vocabulary size (in particular the availability of low-frequency loan words from other languages) or perhaps individual differences in sensitivity to rhythmic structure in speech. Some tasks encourage even the use of levels of structure which effectively play no role in on-line recognition or poetry — for instance, tasks in which listeners are asked explicitly to segment words into their component parts produce a syllabic preference in English irrespective of stress pattern (Otake, Davis, & Cutler, 1995).

In Japanese, nearly all tasks that have been used in studies of speech segmentation produce a moraic advantage (the migration task being one of the few exceptions), which suggests that the mora is more efficient than any other structural level in listening to this language. In French, nearly all such tasks produce a syllabic advantage, similarly suggesting that the syllable is overwhelmingly efficient for segmenting French. But in many languages there may be no such overwhelming advantage for a single structural

unit. We suggest that Telugu is one such language. We suspect that there are in fact many such languages. It is possible that these languages may be identified by use of the Grabe and Ramus measures discussed in the introduction; on their measures the languages in question are likely to produce mixed or unclassifiable results.

This suggests, in turn, that the connection between the phonological structure tapped by these measures and the phonological structure which plays a role in listeners' segmentation procedures may not be as close as we might perhaps have hoped. Many languages may support no compellingly efficient primary strategy, but may still allow listeners considerable benefit from multiple alternative segmentation strategies. Thus even if the language does not lend itself to a single strategy, listeners may still be able to make use of phonologically based segmentation procedures, possibly switching between alternative procedures according to the dictates of efficiency. Accordingly, a failure to classify a given language on the measures does not imply that listening experiments in the same language will produce null results for any of the known effects (of the type summarized in Fig. 1). It is probably never going to be possible to assign all the languages of the world to mutually exclusive rhythmic classes; however, there may be several dimensions of rhythmic structure on which given language pairs display more versus less similarity, with corresponding implications for similarities in the performance of perceptual tasks.

Use of multiple strategies in native language also has implications, we note, for second-language listening. The rhythmic similarity hypothesis itself implies that learning to listen to certain non-native languages acquired in adulthood should not prove hugely difficult — at least the segmentation problems will be minimized if one learns a second language with the same rhythmic structure as the native language, even if the two languages in question are typologically unrelated. If speakers of a given language indeed command multiple segmentation strategies, use of these strategies may generalize to other languages which encourage the same procedures.

In conclusion, we argue that the perceptual approach has produced clear grounds for accepting the proposal that languages group along lines of rhythmic similarity. Phonetic measurements will always also be useful, but given that such measurements with current techniques are so labor-intensive that effectively they always yield too little data, it would be very useful to have alternative, more easily implemented approaches to the issue as well. We predict that the perceptual approach should be able to isolate groups of languages which pattern similarly in several respects; our results suggest that listening to non-native materials, and possibly also listening to "native" nonsense materials, may offer the easiest route to informative results. Methods such as the fragment detection task we have used, where the listening results are based on purely acoustic-phonetic decisions and do not depend on knowledge of the language and its lexicon, allow such tests. Where previous results from a given language are based on such methods, the statistical patterns in the past findings can serve as a basis of comparison for results from future rhythmic similarity investigations in other languages. We hope that many more such investigations will follow, and others' investigations may be facilitated in that, as noted above, the present materials may be downloaded from our website.

References

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

ANNAMALAI, E., & STEEVER, S. B. (1998). Modern Tamil. In S. B. Steever (Ed.), *The Dravidian languages* (pp. 100–128). London: Routledge.

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.

BRADLEY, D. C., SÁNCHEZ-CASAS, R. M., & GARCÍA-ALBEA, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes*, **8**, 197–233.

BROWN, C. P. (1857, reprinted 1981). *A grammar of the Telugu language*. Madras: Asian Educational Services.

BUTTERFIELD, S., & CUTLER, A. (1988). Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. *Proceedings of SPEECH '88, Seventh Symposium of the Federation of Acoustic Societies of Europe,* 827–833. Edinburgh.

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

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. (2001). Listening to a second language through the ears of a first. *Interpreting*, **5**, 1–23.

CUTLER, A. (2004). Segmentation of spoken language by normal adult listeners. In R. D. Kent (Ed.), *MIT encyclopedia of communication sciences and disorders* (pp. 392–395). Cambridge, MA: MIT Press.

CUTLER, A., & BUTTERFIELD, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218–236.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUÍ, J. (1983). A language-specific comprehension strategy. *Nature*, **304**, 159–160.

CUTLER, A., MEHLER, J., NORRIS, D. G., & 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. (1989). Limits on bilingualism. *Nature*, **340**, 229–230.

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., NORRIS, D., & WILLIAMS, J. N. (1987). A note on the role of phonological expectations in speech segmentation. *Journal of Memory and Language*, **26**, 480–487.

CUTLER, A., & OTAKE, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824–844.

DASHER, R., & BOLINGER, D. (1982). On pre-accentual lengthening. *Journal of the International Phonetic Association*, **12**, 58–71.

DAUER, R. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the XIth International Congress of Phonetic Sciences* (ICPhS), **5**, 447–450. Tallinn.

DEMANY, L., McKENZIE, B., & VURPILLOT, E. (1977). Rhythm perception in early infancy. *Nature*, **266**, 718–719.

DETERDING, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, **29**, 217–230.

Van DONSELAAR, W., & STOUTJESDIJK, O. (1994). The acoustic/phonetic factors behind the "syllable effect." *OTS Yearboook, 1993*, 25–43.

DUTT, N. V. B. S. (1992). *Japanese and Telugu easy learner.* Vishakapatnam.

FRAUENFELDER, U. H., & KEARNS, R. K. (1996). Sequence monitoring. *Language and Cognitive Processes*, **11**, 665–674.

GRABE, E. (2002). Variation adds to prosodic typology. In B. Bel & I. Marlin (Eds.), *Proceedings of Speech Prosody 2002* (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.), *Laboratory Phonology 7* (pp. 515–546). Berlin: Mouton.

HOEQUIST, C. E. (1983). Durational correlates of linguistic rhythm catgories. *Phonetica*, **40**, 19–31.

INAGAKI, K., HATANO, G., & OTAKE, T. (2000). The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology*, **75**, 70–91.

ITO, J. (1988). S*yllable theory in prosodic phonology*. New York: Garland Publisher.

KEANE, E. (2004). Illustrations of the IPA: Tamil. *Journal of the International Phonetic Association*, **34**, 111–116.

KEANE, E. (2006). Rhythmic characteristics of colloquial and formal Tamil. *Language and Speech*, **49**, 299–332.

KRISHNAMURTI, Bh. (1998). Telugu. In S. B. Steever (Ed.), *The Dravidian languages* (pp. 202–240). London: Routledge.

KRISHNAMURTI, Bh., & GWYNN, J. P. L. (1985). *A grammar of modern Telugu*. Oxford: Oxford University Press.

KUBOZONO, H. (1998). On the universality of mora and syllable. *Journal of the Phonetic Society of Japan*, **2**, 5–15.

LAVER, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

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

McQUEEN, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.

McQUEEN, J. M., CUTLER, A., BRISCOE, T., & NORRIS, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, **10**, 309–331.

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.

MEHLER, J., & DUPOUX, E. (2004). Combien de syllabes dans "psycholinguistique"? Discussions autour de Juan Seguí. In L. Ferrand & J. Grainger (Eds.), *Psycholinguistique Cognitive: Essais en l'honneur de Juan Seguí* (pp. 35–56). Brussels: De Boeck.

MOHANAN, T. (1989). Syllable structure in Malayalam. *Linguistic Inquiry*, **20**, 589–625.

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.

OHNO, S. (1981). *Nihongo to Tamiirugo*. (Japanese and Tamil). Tokyo: Shinchoosya.

OTAKE, T. (1990). Rhythmic structure of Japanese and syllable structure. *IEICE Technical Report*, **89**, 55–61.

OTAKE, T., DAVIS, S., & CUTLER, A. (1995). Listeners' representations of within-word structure: A cross-linguistic and cross-dialectal investigation. *Proceedings of EUROSPEECH 95*, **3**, 1703–1706. Madrid.

OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 358–378.

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., & IMAI, Y. (2001). Awareness of subsyllabic structure by Japanese preschool children and its universality. *Phonological Studies*, **4**, 81–88.

OTAKE, T., & YONEYAMA, K. (1995). Moraic status and syllable structure in speech perception. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (ICPhS), **2**, 686–689. Stockholm: KTH and Stockholm University.

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

PALLIER, C., SEBASTIÁN-GALLÉS, N., FELGUERA, T., CHRISTOPHE, A., & MEHLER, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, **32**, 373–389.

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

PORT, R. F., DALBY, J., & O'DELL, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574–1585.

RAMUS, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of Speech Prosody 2002*, 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.

RAO, G. U. M. (1996). A nonlinear analysis of syllable structure and vowel harmony in Telugu. *PILC Journal of Dravidic Studies*, **6**, 55–84.

REDDY, K. N. (1979). Problems of syllable-division in Telugu. *Work in Progress*, **12**. Edinburgh: Department of Linguistics.

ROACH, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Edward Arnold.

SAILAJA, P. (1998). Orthography and phonological awareness: Phoneme and syllable manipulation abilities of Telugu-English biliterates. *Paper presented at SALA XIX*. York: University of York.

SAILAJA, P. (1999). Syllable structure of Telugu. *Proceedings of the XIVth International Congress of Phonetic Sciences* (ICPhS), 743–746. San Francisco.

SALVERDA, A. P., DAHAN, D., & McQUEEN, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, **90**, 51–89.

SEBASTIÁN-GALLÉS, N., 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, Part 1). *Jinbun Ronsoo*, **22**, 19–53.

SMITH, M. R., CUTLER, A., BUTTERFIELD, S., & NIMMO-SMITH, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research*, **32**, 912–920.

STEEVER, S. B. (1998). *The Dravidian languages*. London: Routledge.

STOCKMAL, V., MARKUS, D., & BOND, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech*, **48**, 55–63.

SUOMI, K., McQUEEN, J. M., & CUTLER, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422–444.

TANAKA, S. (1999). Nihongo no onsetsu to yonhaku no tenpureeto — Senryuu to puroyakyuu seien ni okeru Ziamari no bunseki, [Syllables and four-mora templates in Japanese: Analysis of "Zi-amari" in Senryu poems and baseball cheering], in Spoken Langauge Working Group (Ed.) *Bunpou to Onsei II [Speech and Grammar 2]* (pp. 261–290). Tokyo: Kurosio Shuppan.

VASANTA, D. (1997). Coarticulation in the temporal domain: Evidence from Telugu-speaking prelingually deaf children. *Asia Pacific Journal of Speech, Language, and Hearing*, **2**, 139–147.

Van de WEIJER, J. (1999). *Language input for word discovery*. PhD thesis, University of Nijmegen. MPI Series in Psycholinguistics, **9**.

VROOMEN, J., & de GELDER, B. (1994). Speech segmentation in Dutch: No role for the syllable. *Proceedings of the Third International Congress on Spoken Language Processing, Yokohama*, 1135–1138.

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.

WEBER, A., & CUTLER, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, **119**, 597–607.

ZWITSERLOOD, P., SCHRIEFERS, H., LAHIRI, A., & van DONSELAAR, W. (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **19**, 260–271.