

The perception of rhythm in spoken and written language

Anne Cutler

Max-Planck-Institut für Psycholinguistik, Wundtlaan 1, 6525 XD Nijmegen, Netherlands
MRC Applied Psychology Unit, Cambridge, UK

Part I: Introduction

Rhythm is perceptually salient to the listener. This claim is central to the research project briefly described below: a large-scale investigation of listening, in which the principal issue was how listeners segment continuous speech into words. Listeners must recognize spoken utterances as a sequence of individual words, because the utterances may never previously have been heard. Indeed, listeners' subjective impression of spoken language is that it may be effortlessly perceived as a sequence of words. Yet speech signals are continuous: there are only rarely reliable and robust cues to where one word ends and the next begins.

Research on this issue in a number of languages prompted apparently differing proposals. In English, experimental evidence (Cutler & Norris, 1988; Cutler & Butterfield, 1992; McQueen, Norris & Cutler, 1994) suggested that lexical segmentation could be efficiently achieved via a procedure based on exploitation of stress rhythm: listeners assumed that each strong syllable was likely to be the beginning of a new lexical word, and hence segmented speech input at strong syllable onsets. Experimental evidence from French, in contrast (Mehler, Dommergues, Frauenfelder & Segui, 1981; Segui, Frauenfelder & Mehler, 1981) motivated a syllable-by-syllable segmentation procedure. Similar to the French findings were results from Catalan, and, under certain conditions, Spanish (Sebastian, Dupoux, Segui & Mehler, 1992). Cross-linguistic studies with French and English listeners (Cutler, Mehler, Norris & Segui, 1986, 1992) established that these contrasting results reflected differences between listeners rather than being effects of the input itself; non-native listeners did not use the segmentation procedures used by native listeners, but instead could apply their native procedure—sometimes inappropriately—to non-native input.

Language rhythm offered a framework within which the English and French results could be interpreted as specific realizations of a single universal procedure. The rhythm of English is stress-based, while French is said to have a syllable-based rhythm. Thus in both these cases the segmentation procedure preferred by listeners could be viewed as exploitation of the characteristic rhythmic structure of the language. To test this hypothesis, experiments similar to those conducted in French and English were carried out in Japanese, a language with a rhythmic structure based on a subsyllabic unit, the mora (Otake, Hatano,

Cutler & Mehler, 1993; Cutler & Otake, 1994). These showed indeed that Japanese listeners could effectively use moraic structure to segment spoken input, thus supporting the hypothesis that listeners solve the segmentation problem in speech recognition by exploiting language rhythm—whatever form the rhythmic structure of their native language may take. The segmentation problem, however, is specific to listening. One of the many differences between listening and reading is that in most orthographies the segmentation problem does not arise—words are clearly demarcated in most printed texts. Therefore it is at least arguable that the reader, who is confronted with no segmentation problem, has no need of language rhythm. In other words: Rhythm will *not* be perceptually salient to the reader.

The following section of this chapter was conceived as a (self-referential) test of this suggestion. The reader is invited to read the piece as it originally appeared (Part II). For those still unconvinced, Part III may serve as a useful comparison to Part II.

Part II: The Perception of Rhythm in Language

1. *The segmentation problem*

The orthography of English has a very simple basis for establishing where words in written texts begin and end: both before and also after every word are empty spaces and this demarcation surely helps the reader comprehend. In a spoken text, however, as presented to a hearer, such explicit segmentation cues are rarely to be found; little pauses after every single word might make things clearer, but the input is continuous, a running stream of sound. This implies that part of listening involves an operation whereby input is segmented, to be processed word by word, for we cannot hold in memory each total collocation, as most sentences we come across are previously unheard. Yet we listeners experience no sense of some dramatic act of separating input into pieces that are known; as we listen to an utterance it seems unproblematic—words in sentences seem just as clear as words that stand alone. Just how listeners accomplish such an effortless division is a question that psychologists have now begun to solve, and this paper will describe (although with minimal precision) some experimental studies showing what it might involve. The findings, as this summary explains, at once can vindicate the order of the problem and the hearer's sense of ease, for though speech *must* be segmented, yet the data plainly indicate that rhythm in the input makes segmenting speech a breeze.

2. *The language-specificity of rhythmic structure*

Now linguistic rhythmic structures have a noticeable feature in that language-universal they are definitely not. This fact is all too obvious to any hapless teacher who has tried to coax the prosody of French from, say, a Scot. Thus while English rhythmic structure features alternating stresses in which syllables contrast by being either strong or weak, this particular endowment is not one which French possesses, having rather one where syllables are equal, so to speak. These distinctions were expressed within traditional phonetics as uniquely based on timing (stress or syllable), though now we admit of more complexity in rhythmic exegesis and of other types of patterning that languages allow; thus in Japanese the mora is the (subsyllabic) unit which provides the root of rhythm, as phonolo-

gists maintain. An important source of evidence, and few would dare impugn it, can be found in verse and poetry: the metrical domain. So compare the English limerick, a form which thousands take up, with the haiku, a poetic form of note in Japanese; there are five lines in a limerick, and stress defines their makeup: the third and fourth are two-stress lines, the others all are threes; and analogously haiku have their composition reckoned by the mora computation, in a manner iron-cast: while the longest line in morae, having seven, is the second, there are five and only five in both the first line and the last.

3. *The use of rhythm in listening*

Just those rhythms found in poetry are also those which function in perception, as the work referred to earlier suggests; thus for English there is evidence¹ involving a conjunction of spontaneous performance and experimental tests, which together show that listeners use stress in segmentation, by hypothesizing boundaries when syllables are strong. Since the lexicon has far more words with strong pronunciation of the word-initial syllable, this method can't go wrong. In comparison with English we should surely not ignore a set of studies² run quite recently on hearers in Japan, which produced results consistent with the story that the mora is the unit that these listeners segment by when they can; while those studies³ that initiated all this lengthy series were performed on native listeners of French some years ago, and they demonstrated well beyond the range of any queries that these listeners used syllables for parsing speech *en mots*. More experiments were subsequently carried out in Spanish, and in Catalan and Portuguese and Quebecois and Dutch, which in spite of minor variance did nothing that would banish the conclusion that for hearers rhythm matters very much. So the picture that emerges is that rhythm as exhibited in verse forms of a language can effectively predict those procedures which, assuming that their use is not inhibited, allow us to declare the segmentation problem licked.

4. *The non-use of rhythm in reading*

Unexpected complications to this neat account, however, are observed when we consider rhythms found in written text. Some preliminary findings, which this section will endeavor to elucidate, at first left their discoverer perplexed. For if rhythm is so integral a part of our audition, then it ought to be the case that it is hard to overlook; but the most pronounced of rhythms can escape our recognition when they're reproduced in printing in an article or book. Late in 1989 the present author wrote a letter, in which verse (or rather, doggerel) pretended to be prose, to at least a hundred friends, from whom responses showed the better part had not perceived the rhymes at all, wherever they arose. In a follow-up, a colleague⁴ gave this ready-made material to subjects to read out, and his results were even worse: of the readers who produced the text, in strict progression serial, not one perceived the letter as a rhyming piece of verse. But the selfsame text, however, may be printed as a ballad (thus, with lines which end in rhymes), and any reader can descry where the rhythm is, which renders this interpretation valid: written rhythm's only noticed when it clearly hits

¹Cutler & Norris (1988); Cutler & Butterfield (1992).

²Otake et al. (1993); Cutler & Otake (1994).

³Mehler et al. (1981); Segoi et al. (1981).

⁴Many thanks to Aki Fukushima and Bob Ladd for conducting this study and permitting me to describe it.

the eye. But perhaps the readers' lack of use of rhythm, as conceded, if judiciously considered has a lesson it can teach: it arises just *because* no segmentation step is needed. Thus the role of language rhythm is in understanding *speech*.

Part III: The Perception of Rhythm in Language

1. The segmentation problem

The orthography of English has a very simple basis for establishing where words in written texts begin and end: both before and also after every word are empty spaces and this demarcation surely helps the reader comprehend. In a spoken text, however, as presented to a hearer, such explicit segmentation cues are rarely to be found; little pauses after every single word might make things clearer, but the input is continuous, a running stream of sound. This implies that part of listening involves an operation whereby input is segmented, to be processed word by word, for we cannot hold in memory each total collocation, as most sentences we come across are previously unheard. Yet we listeners experience no sense of some dramatic act of separating input into pieces that are known; as we listen to an utterance it seems unproblematic— words in sentences seem just as clear as words that stand alone. Just how listeners accomplish such an effortless division is a question that psychologists have now begun to solve, and this paper will describe (although with minimal precision) some experimental studies showing what it might involve. The findings, as this summary explains, at once can vindicate the order of the problem and the hearer's sense of ease, for though speech *MUST* be segmented, yet the data plainly indicate that rhythm in the input makes segmenting speech a breeze.

2. The language-specificity of rhythmic structure

Now linguistic rhythmic structures have a noticeable feature in that language-universal they are definitely not. This fact is all too obvious to any hapless teacher who has tried to coax the prosody of French from, say, a Scot. Thus while English rhythmic structure features alternating stresses in which syllables contrast by being either strong or weak, this particular endowment is not one which French possesses, having rather one where syllables are equal, so to speak. These distinctions were expressed within traditional phonetics as uniquely based on timing (stress or syllable), though now we admit of more complexity in rhythmic exegetics

The perception of rhythm in spoken and written language

and of other types of patterning that languages allow;
thus in Japanese the mora is the (subsyllabic) unit
which provides the root of rhythm, as phonologists maintain.
An important source of evidence, and few would dare impugn it,
can be found in verse and poetry: the metrical domain.
So compare the English limerick, a form which thousands take up,
with the haiku, a poetic form of note in Japanese;
there are five lines in a limerick, and stress defines their makeup:
the third and fourth are two-stress lines, the others all are threes;
and analogously haiku have their composition reckoned
by the mora computation, in a manner iron-cast:
while the longest line in morae, having seven, is the second,
there are five and only five in both the first line and the last.

3. *The use of rhythm in listening*

Just those rhythms found in poetry are also those which function
in perception, as the work referred to earlier suggests;
thus for English there is evidence involving a conjunction
of spontaneous performance and experimental tests,
which together show that listeners use stress in segmentation,
by hypothesizing boundaries when syllables are strong.
Since the lexicon has far more words with strong pronunciation
of the word-initial syllable, this method can't go wrong.
In comparison with English we should surely not ignore a
set of studies run quite recently on hearers in Japan,
which produced results consistent with the story that the mora
is the unit that these listeners segment by when they can;
while those studies that initiated all this lengthy series
were performed on native listeners of French some years ago,
and they demonstrated well beyond the range of any queries
that these listeners used syllables for parsing speech *en mots*.
More experiments were subsequently carried out in Spanish,
and in Catalan and Portuguese and Quebecois and Dutch,
which in spite of minor variance did nothing that would banish
the conclusion that for hearers rhythm matters very much.
So the picture that emerges is that rhythm as exhibited
in verse forms of a language can effectively predict
those procedures which, assuming that their use is not inhibited,
allow us to declare the segmentation problem licked.

4. *The non-use of rhythm in reading*

Unexpected complications to this neat account, however,
are observed when we consider rhythms found in written text.

Some preliminary findings, which this section will endeavor to elucidate, at first left their discoverer perplexed. For if rhythm is so integral a part of our audition, then it ought to be the case that it is hard to overlook; but the most pronounced of rhythms can escape our recognition when they're reproduced in printing in an article or book. Late in 1989 the present author wrote a letter, in which verse (or rather, doggerel) pretended to be prose, to at least a hundred friends, from whom responses showed the better part had not perceived the rhymes at all, wherever they arose. In a follow-up, a colleague gave this ready-made material to subjects to read out, and his results were even worse: of the readers who produced the text, in strict progression serial, not one perceived the letter as a rhyming piece of verse. But the selfsame text, however, may be printed as a ballad (thus, with lines which end in rhymes), and any reader can descry where the rhythm is, which renders this interpretation valid: written rhythm's only noticed when it clearly hits the eye. But perhaps the readers* lack of use of rhythm, as conceded, if judiciously considered has a lesson it can teach: it arises just *because* no segmentation step is needed. Thus the role of language rhythm is in understanding *speech*.

References

- 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.G. and Segui, 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. and Segui, 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. and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824-844.
- McQueen, J.M., Norris, D.G. and 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. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, 20, 298-305.
- Otake, T., Haiano, G., Cutler, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 358-378.
- Sebastian-Galles, N., Dupoux, E., Segui, J. & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory & Language*, 31, 18-32.
- Segui, J., Frauenfelder, U.H. & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, 72, 471-477.