Rhythmic patterns in languages and psychology of speech perception

MORIO KOHNO
Kobe Stella Maris Women's College, Japan

1.0. AIM OF THIS PAPER

Phoneticians have so far classified the rhythmic types of spoken languages according to Pike (1945), and Abercrombie (1967), that is, 'stress-timed', and 'syllable-timed'. In addition to these, Hockett (1955) proposed 'mora-timed' as a sub-category of 'syllable-timed'.

The stress-timed language, English, for example, has more or less isochronous intervals among sentence stresses and the syllable-timed language such as Spanish, manifests isochronism in syllables. In Japanese, long syllables are twice as long as short ones, and therefore the short syllable is the basic metrical unit which can show isochoronism, and this unit is called the'mora'.

The aim of this paper is to show that the above mentioned dichotomy (or trichotomy) is never a fixed classification, but a superficial manifestation of the human innate systems of rhythm processing, and to propose a new, reasonable classification, as well as the psycholinguistic mechanism of spoken language timing regulation.

2.0. THE HUMAN' INNATE SYSTEM OF RHYTHM PROCESSING

The author has already discussed this theme in the paper entitled «A psycholinguistic study of rhythm processing» (Kohno-A, henceforth) which is included in these *Proceedings*. Here is the summary of the paper:

1) Human beings have two systems of processing repetitive sound sequences – holistic, out (all-at-a-time, time-independent, Gestaltic) processing, and analytic, (one-by-one, ongoing, prediction-testing) processing. The holistic system copes with rapid tempo whose inter-beat intervals are less than 330ms, and the analytic system corresponds to slow tempos whose inter-beat intervals are more than 410ms. The two processings are neuropsychologically different from each other, and may be innately set in every human being.

2) The two systems are automatically activated at the sound sequences with rapid or slow tempo respectively. That is to say, syllable sequences in which each syllable is connected with less than 330ms interval (IVI) evokes holistic processing. The ones with 410ms or longer IVIs, analytic processing. The coverage of the holistic processing, according to Miller (1955), is 7±2 syllables at its maximum, which corresponds to the span of immediate memory, or the phonological short-term store. It also offers the basement of 'perceptual sense unit' (PSU) which is a given meaning unit In the phonological short-term store. The PSU is usually designated by an intonation contour and pauses at the ends of PSUs. The analytic system pursues the semantic and grammatical relations among PSUs, dealing with them one by one. The processing scope of the analytic system is therefore infinite.

3) The two systems function in hierarchical, module-modulator relations (See Kohno & Tsushima in this *Proceedings)*

3.0. WHY DOES THE UNIT OF MORA EXIST IN JAPANESE?

3.1. Syllable structure and speech rhythm

The minimum speech unit which has the physical reality as well as psychological reality is said to be a syllable. The smaller unit such as phoneme is an abstract unit in which we cannot find any of its constant acoustic features.

Syllable duration (Inter-syllable intervals) in any spoken languages never exceed 330ms, which evokes human's holistic perception. Table 1 verifies it.

Table 1 – ISIs in spoken language

	X	SD
English	244	85.7
Spanish	201	74.0
Japanese	145	27.8

Some languages, however, can easily manifest their rhythm by the holistic processing system, and the other language cannot, depending on their syllable construction. In languages like English, the types of syllable structure are full of variety so that it is very difficult to utter each of them in the same amount of time, but in Spanish, for example, the realization of isochronism in syllables is somewhat easy, because the syllable structure is more or less unified. Let us describe the syllable structures of these languages.

English(stress-timed): V(a), CV(to), VC(up), CVC(book), CCV(play), CCVC(stop), CCCV(splay), CCCVC(straight), VCC(elm), VCCC(asks), CVCC(pact), CCVCC(breathed), CVCCC(camps), CCVCCC(trumped), CVCCCC(sixths), CCVCCC(glimpsed), CCCVCCCC (strengths)

Spanish (syllable-timed): V(y), CV(la), CVC(pan), CCV(fre-'freza'), VC(es), CCVC(flor), VCC(obs-'obstar'), CVCC(cons-'constipar'), CCVCC(plect-'plecto')

[note] Dauer (1983) points out that, in Spoken Spanish, the isochronism among syllables is manifested by using simple syllables far more often than complicated ones, for example, the rate of CV syllable use is 58%, the rate of CVC 22% and that of CCV 6%. And consonant deduction occurs very much.

Japanese (mora-timed): V (i (?) = stomach), CV(ka (?) = mosquito), C([s]in[s:ki] ??= I love) (n (?) = this negative particle is often cited as an example of C, but it is usually classified an nasal vowel.

Table 1 shows that, English, whose syllables have great variety, has the longest syllable length on the average, and the S.D. value is also the greatest, while Japanese, whose syllable structure is the simplest, has the shortest syllable length and the smallest diversity among syllables. Spanish is middle in every aspect. In other words, English cannot manifest isochronism by equal length of syllables, but Spanish can to some extent.

3.2. Psycholinguistic mechanism of mora-timed rhythm

In Japanese, owing to its simple structure of syllables, the time necessary to pronounce one syllable is very short – 145ms on the average (Table 1) – and therefore if the length of syllable is extended twice, the total length is still within 330ms, the sphere of human holistic processing. In other words, the lengthened syllable (2 morae) is psycholinguistically equal to the short one (one morae), because both can evoke the holistic processing and therefore both can constitute one element of the PSU. The syllables of Spanish and of English, however, are about 201 and 244ms in length respectively (Table 1), and if they were lengthened twice, their durations are in most cases over 330ms and therefore they can never evoke the holistic processing. In other words, overly long syllables can never constitute the holistically perceived PSU with other syllables. Generally speaking, listening comprehension is a process of pursuing the grammatical and meaning relations among PSUs by the analytic system, but if the PSUs are never formed, it is impossible for the system to work smoothly (Kohno, 1993). This is the reason why neither Spanish nor English can produce the unit of mora. Only Japanese can do so.

Morphologically, the mora system is indispensable for Japanese, which lacks many types of syllable variety. Japanese has one mora and two morae words per almost every CV combination, distinguishing their meanings at each pair. For example /i/ =stomach, /ii/ =good; /o/ =tail, /oo/ =king; /ki/ =yellow or tree, /kii/ =strange or key. Japanese can even say Oooooo='Let's chase the king', by distinguishing two morae words from one mora words in this sequence. (oo=king, o=particle to show objective case, oo=chase, o=an auxiliary word meaning 'let's'). It is not hard to imagine that, but for the mora unit, owing to the paucity of syllable variety, Japanese would produce too many homonyms to transfer information,

4. PSYCHOLINGUISTIC MECHANISM OF STRESS-TIMED RHYTHM

As already said, English, whose syllables are of so many varieties that we cannot pronounce them allotting the same amount of time, is obliged to express its rhythm by the use of the analytic system. The analytic system activates itself, reacting the primary stresses which are placed at least once per one PSU, whose length is usually more than 420ms. The PSUs themselves are the products of the holistic processing system consisting of 7±2 or less syllables. If the PSU is too short, the speakers tend to place pauses between them, and if they are too long, speakers tend to speak rapidly , to make isochronism.

Phoneticians in the past expected a rigid isochronism among stresses in English. Shen & Peterson, however, instrumentally measured the lengths among primary stresses, with the results of 380~3610ms (including the intervals across terminal junctures) (from Lehiste, 1977). Bolinger (1965) reported that 13 inter-stress intervals (ISIs) out of 53 spread double from short intervals to long ones. Lea (1974) asked eight English speakers to read aloud 31 sentences and measured ISIs. He reported that , although the mean of the ISIs was 532ms, the S.D. value was hopelessly large, 230. Being confronted with these disappointing facts, many phoneticians gave up the idea that English has isochronism among stresses. In spite of the above results, on the other hand, some phoneticians like Uldall (1971, 1972), Halliday (1967) and Lehiste (1977) still assert the isochronism, reviewing these data psychologically and phonologically. The author can also assert the psychological isochronism in English from another viewpoint. As already pointed out, as the PSU is a product of the holistic processing system, it is perceived all-at-once, irrespective of its actual length: which theoretically extends to 9◊244=2196 or

9◊330=2970 (9 = 7+2, 244= average length of English syllables. (Table 1), 330=maximum length of holistic perception). The set of meaning attached to the PSU, would more accelerate the feeling of all-at-a-time perception.

5. PERCEPTUAL SENSE UNIT OF JAPANESE

The PSU is made on the base of the human beings' innate ability of holistic, all-at-a-time perception, whose scope is 7±2 syllables. 7±2 syllables designates the basic span of the immediate memory or phonological short term store, but the span naturally expands when some meaning unit binds these syllable clusters closely.

Based on the above fact, the following experiment was held to show the specific nature of Japanese PSUs.

Experiment 1

Four nonsense words, each of which consisted of 14 syllables, were prepared and read by a Japanese university student (male: a Kansai dialect speaker) setting two words within 330ms IBIs produced by SEIKO Rhythm Trainer SQM-348. These utterances were recorded onto tape. On the other hand, we prepared another 14- syllable sequence which consisted of 7 two-syllable(morae) words, such as tane(=seed), tora(=tiger), kao(=face), tera(=temple). We made 7 lists of this kind in total. And all these lists were read by the same Kansai dialect speaker in the same way as the nonsense word, setting one word (two syllables) per (330ms) interval . Four nonsense words and four 7-word lists were chosen after being instrumentally examined as to whether or not they were clearly recorded, allowing two syllables placed in each 330ms rhythm beat intervals, and then they were presented to university students, all of whom were also Kansai dialect speakers, using a language laboratory. The university students were divided into two homogeneous groups. One group was requested to do multiplication of two digit numbers, and the other group to draw thirty circles in the space provided. Thirty seconds were allotted each to the two works. Both groups were then asked to reproduce the stimuli by writing them on an answer sheet with Kana (Japanese alphabet). In order to compare the amount of memory, the nonsense words composed of 7 syllables (inter syllable intervals were 330ms) were also given to the same university students under the same conditions such as doing multiplication and drawing circles after having listened to the stimuli. The results are shown in Table 2.

Table 2 – Perceptual sense unit and meaning

stimuli	post-presention work	n	x	S.D.
(A) nonsense word	drawing circles	44	7.05	4.19
composed of 14 syllables	calculation	44	6.95	3.13
(B) 7 two-syllable	drawing circles	27	14.26	5.36
word list	calculation	27	12.59	4.73

A (drawing circles) ? A (multiplication) t = 0.10 N.S.
B (drawing circles) ? B (multiplication) t = 1.44 N.S.
B (drawing circles) ? A (drawing circles) t = 6.93 p<0.01
B (multiplication) ? A (multiplication) t = 5.46 p<0.01
(cf.) 7-syllable nonsense word (drawing circles): x=14.20 S.D.=3.50
(multiplication): x=13.18 S.D.=3.89. Full marks are 21 per each stimulus.

Judging from the fact that the two stimuli were presented in rapid tempo with ISI of less than 330ms, it may be a natural result that no significant difference was detected between 'drawing circles' and 'doing multiplication' (Kohno-A).

The interesting thing, however, is the fact that the retention of the nonsense words was far worse than that of the two-syllable word list, in spite of the fact that they were both composed of 14 syllables ($p<0.01$). The fourteen-syllable nonsense words were still farther worse than that of 7-syllable nonsense words, which showed more or less the same retention as two-syllable word list. It is obvious that the length of the 14-syllable nonsense word would exceed the span which should be dealt with by the holistic system. If two syllables are united into one semantic unit, however, the same length of 14 syllables recovers its retention to that of 7-syllable nonsense words (Miller ,1956). This is also the very reason why Japanese can have the'mora' system.

We should notice, however, that the length of the PSU has its natural limitation. Even in Japanese, it is impossible, generally speaking, to say three or more syllables (morae) within 330ms. Three syllables would take 435ms, four syllables 580ms, which is over the scope of holistic processing, and they would automatically activate the analytic processing. It means that these long syllables would never constitute an element of the PSU. In order to verify this hypothesis, we made a list of seven words, each having three or four syllables, such as 'tokei' (=clock), 'enpitsu' (=pencil). They were read by a Kansai dialect speaker at normal speed (two syllables (morae) in 330ms and presented to Kansai-dialect subjects in just the same procedure as that of the above experiment. The result were very bad, only half as much score ($x=7$, full mark 28) as the retention of the seven-syllable nonsense word. This means the PSU will never extend unboundedly, because it is basically formed by the innate human ability to process holistically.

6.0. CONCLUSION

The above-mentioned study shows that the mora-timed rhythm is not a sub-category of the syllable-timed, but rather it is the most fundamental rhythm type. The difference between syllable-timed and stress-timed is only a relative difference. The former can more or less manifest its isochronism in syllable lengths, although it is far from perfect. Owing to this imperfection, it must also show their rhythms by sentence accent as Dauer pointed out.

The traditional dichotomy (trichotomy) classification should thus be reexamined on the base of the 'two neural clock' theory.

REFERENCES

Pike, K. L. 1945. *The Intonation of American English,* Univ. of Michigan Press.
Abercrombie, D. 1967. *Elements of General Phonetics,* Edinburgh Univ. Press.
Kohno, M. 1993. Perceptual sense unit and echoic memory, *International J. of Psycholinguistics,* 9-1, 13-31.
Hockett, C. F. 1955. A manual of phonology, *International J. of American Linguistics* 21-4, 1-246.
Lehiste, I. 1977. Isochrony reconsidered, *J. of Phonetics* 5, 253-263.