Title	A Study of a Rhythm Perception Model
Author(s)	Nagai, Katsumi
Citation	大阪大学言語文化学. 1996, 5, p. 189-201
Version Type	VoR
URL	https://hdl.handle.net/11094/78122
rights	
Note	

A Study of a Rhythm Perception Model*

Katsumi NAGAI**

神宮(1989)の時間知覚モデルに基づき、リズム知覚モデルを提案する。知覚された時間のグループ化、チャンク化がリズム知覚の鍵となるが、その処理は (1) 全体的知覚と分析的知覚の区別がある (Hibi 1983), (2) その区別は母語により変わることはない，という仮説の下に，日本語話者と英語話者によるリズム産出実験を行った．

1　Introduction

Time flies whether people are conscious of it or not. Since people do not have a sensory organ unique for time and rhythm perception, subjective time and rhythm cannot be directly perceived. When people say "Taro speaks faster," it means the number of phonemes in Taro's utterance is larger than some standard, and when we say "Hanako reads slower," the number of words she reads is smaller than some criteria. The continuous subjective time can be felt only by this discrete information [1]. Such kind of information for discreteness is called *temporal information* by Fraisse (1981). Jingu (1989) explains that being conscious about time is an unconscious and automatic procedure of forming continuous subjective time from discrete inner events. As long as natural language processing can not be free from its rhythm perception, time and rhythm perception must be the basis of language perception.

Tapping on a table is a basic experiment of rhythm production to examine subjective time. It's drawback is that the produced intervals by tapping is a mixture of delays of subjective time and delays of output behaviours. Wing

---

[1] Neurons of sensory organs follow an all-or-none law. It is a beginning of changing continuous stimulus into discrete information.

and Kristofferson (1973a,b) did a continuation task [2] after a synchronization task, and tried to distinguish the subjective time and the delays of produced intervals. When subjective time is C_j , Delay of response is D_j , and reproduced intervals are I_j (its average is μ_I, its variance is $\delta^2{}_I$), as Figure 1 shows,

$$Ij = Cj - Dj - 1 + Dj$$

Correlation with lug one ($\rho_I(1)$) can be calculated

$$\rho_I(1) = \gamma_I(1)/\gamma_I(0)$$

With the following equation:

$$
\begin{aligned}
\gamma_I(1) &= E[(I_j - \mu_I)/(I_{j-1} - \mu_I)] & (16.1)\\
\gamma_I(0) &= E[(I_j - \mu_I)/(I_j - \mu_I)] & (16.2)\\
&= E[(C_j - D_{j-1} + D_j - \mu_I)/(C_j - D_{j-1} + D_j - \mu_I)] & (16.3)
\end{aligned}
$$

$\{\delta^2\}_C + 2\{\delta^2\}_D$ can be calculated [3].

Figure 1 (Wing et al. 1973a)

It means that variances of subjective time can be calculated from the results of laboratory experiments (I). They found that variants of subjective time are larger than reproduced intervals between 180ms and 350ms. It proved the significance of rhythm production experiments to examine subjective time.

When tapping on the table with a pencil listening to the metronome, most people find themselves adjusting the timing when the rhythm is slow. If they think the first stroke is earlier, they can lengthen the following interval to

---

[2] Continuation task is an experiment of tapping that subjects continue tapping after hearing the last stimuli. Synchronization task is, on the other hand, an experiment of tapping to the rhythm of stimuli.

[3] As long as C and D are independent of each other, the covariance is 0.

correct the timing. If again the second stroke is out of the rhythm of the metronome, next they hit the third stroke after a shorter interval. During this procedure they recognize clearly what they are doing as Kono (1993) reviews. This kind of procedure is called *analytic rhythm*.

When the rhythm is faster than some threshold, subjects can no longer follow the strokes one by one. All they can do is to grasp the whole pattern of strokes. This kind of procedure is *holistic rhythm*. Hibi's paper (1983) is a pioneering work to distinguish the two processes mentioned above. He concludes that the switching duration of the two procedures is about 330ms.

In the field of psychophysics, time shown by a physical clock is physical continuum, while time felt by human beings is psychological continuum. Both times change continuously in quantity, but the latter time needs consciousness, and is called *subjective time*. Time and rhythm perception shares its characteristics with language processing because both are independent of perceptual modes such as vision and speech.

2   Rhythm perception model

Jingu (1989) made an inner procedural model of time perception (Figure 2). A new rhythm perception model in this paper is based on his time perception model. (Figure 3).

2.1   The pacemaker

It is quite natural to assume the existence of an equivalent for a quartz oscillator of clocks in a human brain. This device is an internal clock. The internal clock and its pulse counter has been considered to be used directly to evaluate temporal information and generate subjective time. A large number of researchers tried to find the relation between the standard of the internal clock and pulses of a heart, rate of respiration, and temperature of the body, but no sufficient data has been reported so far (Jingu 1994).

When physical clock time and its subjective time has the smallest difference in experiments, the physical clock time is called indifference time. It is about
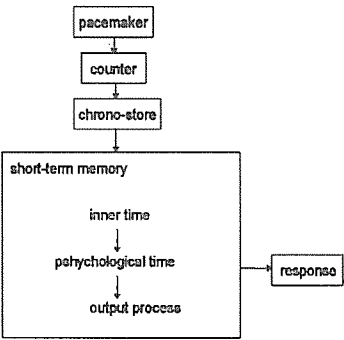
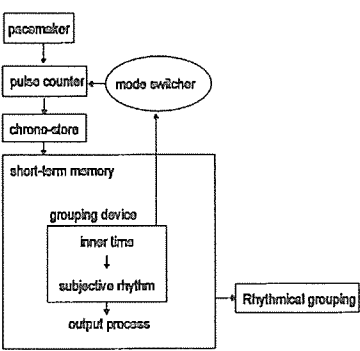Figure 2 Time perception model   Figure 3 Rhythm perception model

600ms. In other words, physical clock time generates the shorter subjective time when the physical clock time is more than 600ms. Woodrow (1951) named this effect Vierodt law and a large number of researchers set this 600ms as their standard time.

The threshold physical time to detect the two successive pulses has been also used as a standard time unit. This minimum physical time which can generate subjective time is 100ms according to Stroud (1956). Kristofferson (1967)'s time quantum model adopted a half of this time (50ms) as his basic unit. Jingu's model posits a substantial pacemaker based on the reverberating circuit in a brain (Jingu 1989). It is known that brain waves have innate cycles such as 10Hz of an α wave. The pacemaker of this rhythm perception model is a black-box, but the same cycle of pulses as Jingu's model (4ms) is postulated as a standard.

2.2   The Pulse counter

The *Pulse counter* is a device to count pulses generated by the pacemaker. This counter usually counts every 4ms pulses (*mode H*), but when the mode switcher changes the mode into *mode A*, the counter counts every two or three pulses to make longer intervals (8ms or 12ms).

2.3   The Chorono-store

The *Chrono-store* in the rhythm perception model is an equivalent of *icon* in vision and *echoic-memory* in auditory perception. It is a temporary storage space for pulses from its pacemaker. According to Jingu (1989), the capacity of this chrono-store is generally 4±2. In the rhythm perception model, the limit is set 3, which enables 12ms pulses in storage on its default mode (*mode H*). The stored information is forwarded to to the short term memory. Suppose rhythm is a gestalt of time, this chrono-store is the birth place of the unconscious rhythm.

2.4   The Mode Switcher

Here a mechanism named *Mode Switcher* is assumed for the rhythm perception model. It receives information which is given as feedback from the grouping device in short term memory. When subjective rhythm is quick and is made of less than about 330ms intervals, pulse counter counts every pulse which is produced by the pacemaker (*mode H*). When subject rhythm is slow of which the intervals are more than 330ms, the mode switcher changes the mode of its pulse counter to an analytic mode (*mode A*). The pulse counter counts every other or three pulses in its analytic mode. Around 330ms is a threshold found in Hibi (1983).

2.5   The Short-term memory

The *Short term memory* is said to be able to hold information about 15s (Jingu 1994). Continuous rehearsals can hold the contents of the memory [4]. The span of the immediate memory corresponds to the capacity of descrete time information, and Miller's magic number 7 ± 2 units are said to be held there. If converted into minimum time units, the number of chunks stored in the short term memory is 16 in Kristofferson's experiment (1980) and 23 in Jingu's paper(1989). These numbers exceeds the limit of 7 ± 2. It leads

---

[4]Long term memory, on the other hand, plays an important role in time estimation.

to a chunking in the short term memory to reduce the burden of short term memory. Time units sent from the chrono-store are again grouped together to make other larger units. They bring about the sense of subjective time at the same time.

2.6   The Grouping device

The *Grouping device* is a mechanism which controls grouping in short term memory. It receives information from the chrono-store and re-groups it to make larger basic subjective time units. If four times grouping is done recursively, maximum $24 = 16$ units are stored in the short term memory [5]. The basic time from the chrono-store is 12.5ms, which was originally 4ms in the pacemaker and was grouped in the chrono-store. After frouping four-times, the maximum time unit of default setting (*mode H*) can be 200ms. If the basic time from the chrono-store is about 50ms, which is generated by every three pulses of the pacemaker, the maximum time unit of *mode A* amounts to be 800ms. These groupings are the sources of subjective time perception in short term memory (Table 1). The behavior of this grouping device is always given as feedback to its mode switcher.

Table 1

Chunking and base times (ms; PM:pace maker)

PM	chrono store base time	mode	grouping for time units				chunking for subjective time		
4	(1/1)12.5	*Holistic*	25	50	100	200	400	800	1.6s
4	(1/2)≒25	?	50	100	200	400	800	1.6s	3.2s
4	(1/4)≒25	?	100	200	400	800	1.6s	3.2s	6.4s

---

[5] Grouping of three items is also possible (Jingu 1989), but the rhythm perception model adopt binary branching system.

3   Experiment

3.1   Aim

The aim of this experiment is to examine the existence of two different modes of perceiving rhythm. The two different modes are the holistic mode and the analytic mode.

In this experiment subjects are asked to listen to a sequence of pulses and reproduce the temporal sequence at the same time by saying monosyllabic 'ta'. They try to produce the rhythm with exact synchronization with the pulses they hear, though some deviations are inevitable. The mean Inter-Onset-Intervals (IOIs hereafter) are expected to be synchronized with the pulses of a metronome. But each IOIs cannot be free from some deviations. The duration between the adjacent onsets are measured visually on a computer.

Based on the experiment by Kristofferson (1973b), a function of autocorrelation $R_{(j)}$ is defined here to study the correlations of the deviation and adjusting. It is a correlation of one variance $C_{(i)}$ and $C(i + j)$. The covariance of the IOIs can be calculated as:

$$C_{(j)} = \frac{1}{n-j} \sum_{i=1}^{n-j} (T_i - \bar{T})(T_{i+j} - \bar{T})$$

Autocorrelation $R_{(j)}$ can be calculated from the covariances:

$$R_{(j)} = \frac{\sum_{i=1}^{n-j}(T_i - \bar{T})(T_{i+j} - \bar{T})}{\sum_{i=1}^{n}(T_i - \bar{T})^2}$$

Autocorrelations between adjacent intervals, which occurred in the utterance when the rhythm was faster than some threshold should be positive because people cannot adjust the intervals one by one. Calculating autocorrelations is an effective way to examine two kinds of rhythm perception.

Another question is whether English speakers display a significant difference from Japanese speakers when they produce temporal intervals. If both English and Japanese speakers share the same autocorrelations, it is plausible that the time and rhythm perception do not vary among speakers of different languages.

It can be a good motivation for learners of second languages who have a difficulty in learning their unfamiliar rhythms.

3.2   Stimuli

A pulse of quartz metronome (SEIKO SQ-77) was put into the analog-digital converter (GW Instruments Mac ADIOS II) on a personal computer (Macintosh II). Figure 4 shows a waveform of the original pulse.
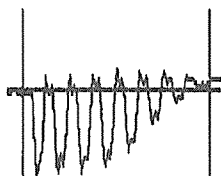
Figure 4 Original Pulse

The pulse has 8ms in duration, and it was copied to make uniformly spaced temporal sequences by Mac Speech Lab II (GW Instruments). Intervals of the IOIs are set 200ms, 500ms, and 800ms. The pulses were recorded on a digital tape recorder (SONY TCD-D10) with a 48KHz sampling frequency. It was presented to subjects as stimuli through a headphone (SONY MDR-AV7).

3.3   Subjects

Four Japanese speakers and four English speakers participated in this experiments. All of them were native speakers aged from 23 to 34 and all Japanese subjects were graduate students at Osaka University, who speak the Osaka dialect. In this experiment age and dialects were not taken into consideration. All English speakers were engaged in language teaching in Osaka prefecture. They all had normal hearing. Each recording took about 40 minutes.

3.4   Procedure

The subjects were asked to listen to the pulses on digital tape and to repeat the monosyllable /ta/ sixteen times synchronously with the pulses. The space between the mouth of the speaker and the microphone (SONY ECM-959A) was set 30cm to catch stable aspiration. The reproduced sound of subjects were recorded on a 2CH digital audio tape recorder (SONY TCD-D7) with

stimuli and later stored on a personal computer through an analog-digital converter (GW Instruments Mac ADIOS II) and LPF (4.54KHz). The sampling frequency was 48KHz, and it was linearly digitized at 16 bits with 6db/oct pre-emphasization. All the waveform and spectrogram of the utterances and pulses were displayed on a screen display of Mac Speech Lab II in a Mac IICX. The intervals between onsets were measured visually. Figure 5 shows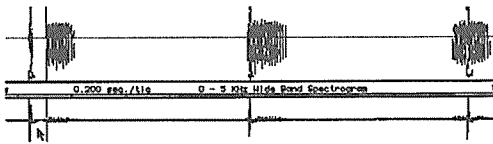 an example of the utterances and pulses. To save errors committed by inexperience or tiredness, the first and last two trials were omitted.

Figure 5 Adjustment of timing

3.5 Results

Table 2 shows the mean time interval and its standard derivation reproduced by subjects. This table shows there is a close relation between stimuli and the reproduced intervals but no significant difference can be seen statistically between English and Japanese speakers.

Table 3 shows examples of the autocorrelations calculated as defined above. At every 200ms and 500ms interval, the first second or third autocorrelations of the responses differ largely from each other because the subject cannot grasp the pulses at the beginning. After some pulses, the subject comes to follow the interval and the autocorrelations become smaller [6].

Table 2

Reproduced Intervals (ms)

---

[6] See Box & Jenkins (1976) about standard errors.

Subjects	Interval-200ms		Interval-500ms		Interval-800ms	
	m	SD	m	SD	m	SD
J-1	194	16	492	39	784	80
J-2	196	14	498	38	773	112
J-3	192	23	497	27	804	42
J-4	198	12	500	28	796	62
E-1	197	13	496	28	782	79
E-2	199	17	490	36	792	46
E-3	199	14	492	35	800	41
E-4	197	11	497	26	797	48

Table 3 Autocorrelations

```
200ms
###############################################################
     Auto-  Stand.
Lag  Cor(*) Err(.)    -.75 -.5 -.25  0  .25  .5
                      +---+---+---+---+---+---+
 1  -.011  .234              .      *        .
 2  -.415  .226          .********|           .
 3   .198  .217              .     |****      .
 4   .124  .208              .     |**        .
 5   .051  .198              .     |*         .
 6  -.050  .188              .    *|          .
 7  -.067  .177              .    *|          .
 8  -.042  .166              .    *|          .
 9  -.097  .153              .   **|          .
10  -.036  .140              .    *|          .
11  -.075  .125              .    *|          .
12  -.062  .108              .    *|          .
13  -.013  .089              .     *          .
###############################################################
```

```
500ms
###############################################################
     Auto-  Stand.
Lag  Cor(*) Err(.)    -.75 -.5 -.25  0  .25  .5
                      +---+---+---+---+---+---+
 1  -.680  .234     *****.********|           .
 2   .385  .226              .     |********.
 3  -.356  .217          . ********|         .
 4   .338  .208              .     |*******.
 5  -.299  .198          . ******|          .
 6   .258  .188              .     |*****   .
 7  -.284  .177          .******|           .
 8   .262  .166              .     |*****   .
 9  -.275  .153          ******|           .
10   .210  .140              .     |****  .
11   .011  .125              .    *          .
12  -.093  .108          .  **|            .
13  -.001  .089              .    *          .
###############################################################
```

3.6   Discussion

No significant difference can be seen between native English and Japanese speakers. It means there is no difference in rhythm production among different language speakers. If there is a close relationship between speech timing control and control of other temporal behaviors as Allen (1975) suggested, what

causes difficulty to the learners of foreign language rhythm?

The interesting thing is the fact that most of the autocorrelations of 500ms and 800ms intervals show oscillation patterns. On the other hand, the auto-correlations of 200ms intervasl have somewhat damped waves. This trend can be observed in most of the data. This means that autocorrelations of 200ms intervals are less correlative than 500ms and 800ms intervals. The result of this experiment suggests that there are two kinds of processing in production of temporal sequences.

One is an analytic processing of timing. Its autocorrelations show an os-cillation pattern in its graph. It is seen when the intervals are 500ms and 800ms, which are comparatively slow rhythms. Subjects can predict the next stimulus and check their articulatory timing with the predicted stimuli. If the prediction is wrong, subjects then adjust their next prediction. This results in a negative correlation among neighboring intervals, and the autocorrelations intervals. This ongoing way of processing temporal sequences can be called the analytic processing.

Another kind of timing processing is the holistic processing of rhythm. When the interval is 200ms, the re-adjustment of the discrepancy does not work one by one. At such a higher rate of intervals, subjects set their equally spaced patterns and compare them with the pattern of stimuli. They cannot compare each stimulus with their own production. Its graph shows no oscillation, and the adjacent autocorrelations are lower and both negative or both positive. This kind of *configuration* of a basic rhythm pattern and its *matching* is nothing but a holistic processing.

The comparatively flat line in the graph means holistic adjustment is at work, and the oscillating line means analytic, or one by one adjustment, is at work. Similar distinctions among Japanese subjects were clarified in Kono (1993). These distinctions are the very things posited in the rhythm perception model. The two kinds of rhythm are based on its subjective time and given as feedback to its mode switcher.

4   Conclusion

This short paper presented a rhythm perception model based on Jingu(1989)'s time perception model and examined two different rhythm perceptions. The model's remarkable device is its mode switcher. It changes the rate of pulse-counting (*modes*) in order to adjust the rhythm of perception. One of the modes is *mode H*, a default setting with which people grasp the holistic pattern of rhythm. The other mode is *mode A*, an analytic perception of rhythm with which people can adjust the timing consciously. The experiment of rhythm production demonstrates the existence of these two modes. It also showed that English speakers' rhythm production is the same as Japanese speakers'.

References

Allen, J. (1975). 'Some basic concepts in linguistics. Vol. 4. Speech and writing'. *Papers in Applied Linguistics*, Vol. 2, 26–28. Oxford U.P..

Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis*. Holden-Day.

Halpern, A. R. & Darwin, C. J. (1982). 'Duration discrimination in a series of rhythmic events'. *Perception and Psychophysics*, **31**, 86–89.

Hibi Seishi. (1983). 'Rhythm perception in repetitive sound'. *Journal of Acoustic Society of Japan (E)*, 4.2, 83–95. Japan.

Hoequist, Charles. (1983). 'Durational correlates of linguistic rhythm categories'. *Phonetica,* **40**, 19–31. Japan.

Jingu Hideo. (1989). *Jikan chikakuno naiteki kateino kenkyu*. Kazama Shobo. Japan.

Jingu Hideo. (1994). 'Jikan chikaku' in *Shinpen kankaku chikaku shinrigaku Handbook*. Seishin Shobo. Japan.

Kono Morio. (1993). 'Perceptual sense unit and echoic memory'. *International Journal of Psycholinguistics*, 9-1.

Kristofferson, A. B. (1967). 'Attention and psychophysical time'. *Acta Psychologica*, **27**, 93–100.

Kristofferson, A. B. (1980). 'A quantal step function in duration discrimination'. *Perception and Psychophysics*, **27**, 300–306.

Nagai Katsumi. (1994). 'Japanese geminate consonant'. Unpublished dissertation to the Graduate School of Language and Culture of Osaka University. Japan.

Nakajima Hideyoshi. (1987). 'A model of empty duration perception'. *Perception*, **16**, 485–520.

Povel, D. J. (1981). 'Internal representation of simple temporal patterns'. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 3–18.

Stroud, J. M. (1956). 'The fine structure of psychological time'. *Information Theory in Psychology*. Free Press. N. Y..

Wing, A. M. & Kristofferson, A. B. (1973a). 'The timing of interresponse intervals.' *Perception and Psychophisics*, **13**, 455–460.

Wing, A. M. & Kristofferson, A. B. (1973b). 'Response delays and the timing of discrete motor responses'. *Perception and Psychophysics*, **14**, 5–12.

Woodrow, H. (1951). 'Time perception'. *Handbook of Experimental Psychology*. New York Press.