Mouthing vowels interferes with language rhythm but writing circles does not: evidence from dual-task experiments

Katsumi Nagai
Kagawa University, JAPAN

ABSTRACT

Accuracy of various rhythm patterns was tested by experiment. There were fast and slow series of stimuli, which had a sparse-dense pattern, a dense-sparse pattern, and a symmetric pattern. The subjects were each asked to reproduce the rhythms by tapping a keyboard on the one hand in the control condition. They silently articulated nonsense phrases before their tapping in the articulatory suppression condition, and they drew circles in the spatial condition. The result indicated that the mouthing of the nonsense phrases significantly affected the accuracy of the rhythm reproduction. The experiment also proved that the fast rhythm, with 250ms intervals in each segment, produced more accurate rhythm reproduction than the slower rhythm did. In the meantime, the subjects needed less response time to reproduce the slow rhythm, probably because the slow rhythm was processed simultaneously as they heard the stimuli. These results imply that language and rhythm can be processed in the same process in the Working Memory, where the resources are shared with the slave system such as the Phonological Loop.

Keywords: rhythm production, dual-task, mouthing and spatial suppressions

1. Introduction

1.1. The Working Memory Model and the Phonological Loop

Baddeley (1986, 1992) presents his working memory model in order to reanalyze the previous experimental results which look opposed to the previous dual storage model (Atkinson and Shiffrin 1971). His model suggests that phonological information is stored in the 'Phonological Loop.' Visual information, on the other hand, is stored in the 'Visuo-spatial Sketch Pad.' These two devices correspond respectively to Echoic Memory (Glucksberg & Cowan 1970) and Iconic Memory (Sperling 1960), which have different storage capacities. The Central Executive, in his model, handles processing at higher cognitive levels, and the Phonological Loop holds linguistic information phonologically. Both are separated in the working memory model. Consequently, the activity of answering true-false question sentences does not affect the performance of simultaneous numeric retrieval (Baddeley & Hitch 1974). In other words, the response time and ratio of correct responses are not affected by the answering questions because the retrieving of words and answering of the questions are tasks different in nature. However, if the numeric retrieval exceeds a maximum of the ‘magical number seven’ units (Miller 1956), the ratio of the correct responses decreases drastically. This means that the Phonological Loop has a storage limit. It needs to borrow processing power, or ‘resources,’ from the Central Executive when the amount of information exceeds the storage limit. Baddeley’s working memory model succeeds in explaining the results of the past experiments on memory retrieval by sharing resources for information processing. The resource sharing is the most marked characteristic of the model, although the structure of the Central Executive still remains unclear. More recently, Just & Carpenter’s model (1992), an ingenious improvement of the levels of the processing model (Craik & Lockhart 1972, 1973), defines the function of the Central Executive more successfully.

The retrieval of a series of phonologically similar letters, like *B, G, V, P,* and *T,* is more difficult than that of the less similar series of *Y, H, W, K,* and *R* (Conrad 1964, Conrad & Hull 1964). This type of interference is called the Phonological Similarity Effect, of which the reality has been verified by a large number of researchers. The effect is obtained regardless of the way in which the test list is presented — it functions both audibly and visually. However, if the subjects are asked to repeat, by articulation, or ‘mouthing,’ familiar words like *one, two, three* or *a, i, u, e, o* before the retrieval, the Similarity Effect is seen to fade away when the list is presented visually (Murray 1967, Murray 1968, Peterson & Johnson 1971). This fact leads to posit two distinctive modules in Phonological Loop model (Baddeley 1992). One of the components, called Phonological Store, holds the phonological information through repetitive rehearsals at the other device, the Articulatory Control Process. Visually presented stimuli are phonologically coded and input into the Phonological Store through the Articulatory Control Process. Because the Similarity Effect takes place in the Phonological Store, memory retrieval of a visually presented list is disturbed while the subjects are articulating familiar words through the experiment. Human memory has been investigated in a large number of experiments about the retrieval of alphabets or nonsense words as stated above. However, not many previous research have studied memory retrieval using rhythm patterns, which are important factors of natural languages.

1.2. Memory for rhythm reproduction

Umemoto *et. al.* (1983) are one of the few earlier studies of rhythm reproduction. They have measured the accuracy of rhythm reproduction and conclude that rhythm synchronization is easier when the rhythm has a sparse-dense (O-OOO) sound pattern than when it has a dense-sparse (OOO-O) sound pattern. They presented auditory pulses (800Hz, 50ms) to their subjects at the intervals of 250ms. They analyzed the response time, which was measured from the presentation of stimuli to the beginning of subjects’ responses synchronized with the stimuli. Their discovery is that the response time is shorter when the rhythm has a sparse-dense pattern. A similar experiment on rhythm reproduction was carried out by Saito (1994, 1997a, 1997b). He asked the subjects to reproduce the rhythm after listening to the stimuli because he did not use a synchronization task in his experiment. Before reproducing the rhythm, a mouthing-suppression group repeated *aiueo* silently, and a spatial-suppression group drew as many 2cm squares as possible. His result indicates that mouthed *aiueo* significantly impairs the correct recall of the rhythm in contrast to the data obtained from the control group.

In both experiments, the inter-onset-interval was set at 250ms because it was considered to be the shortest interval in which people can produce the responses naturally. Meanwhile, as clarified in Kono's report (1993), there exist two different kinds of human rhythm processing. One is analytic processing of rhythm at a slow tempo, and the other is holistic processing at a fast tempo. In the present paper, the subjects were asked to reproduce a variety of rhythm patterns by tapping a keyboard. The tempi are set at two levels (fast tempo with intervals of 250ms and slow tempo with intervals of 750ms). This difference in tempi is expected to cause a switch from holistic processing to analytic processing in subjects.

2. Experiment

2.1. Aim

An articulatory suppression (double-task) method is used in this experiment. The aim of this experiment is to examine the relation between tempi and suppression conditions such as mouthing and drawing before reproducing the rhythm. The inter onset intervals of stimuli in this experiment are set at 250ms (fast tempo) and at 750ms (slow tempo). Another aim is to examine the relation between tempi and rhythm patterns such as sparse-dense (O-OOO) and dense-sparse (OOO-O) patterns. To achieve these aims, null hypotheses are set as below:

H_0: Response time with articulatory (mouthing) suppression is shorter, and the number of correct responses is larger, when the tempo is slow.

H_0: Response time of a sparse-dense rhythm pattern is shorter, and the number of correct responses is larger, when the tempo is slow.

Can rhythm perception and production be explained within the frameworks of the Rhythm Processing Model and the Working Memory Model? Can the effects found in Umemoto *et. al.* (1983) and Saito (1997) be observed if the interval is set long enough to start analytic rhythm processing?

2.2. Stimuli

A sine wave at the frequency of 800Hz was generated and edited with Kay's Multi Speech 3700 on an IBM/PC compatible computer. The stimuli had the structures shown in Figure 1, and varieties of rhythm patterns were synthesized as seen in Table 1. Total duration of fast rhythm patterns was 1,750ms, and that of slow rhythm patterns was 5,250ms.

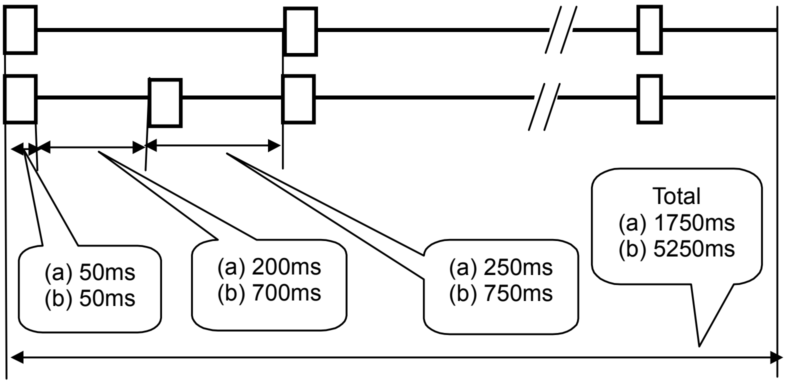

Figure **1** Structure of stimuli used in Experiment 4

Table **1** Patterns of stimuli

fast(250ma)	slow(750ms)	patterns
a1	b1	O O O O O O O
a2	b2	O - O O O O O
a3	b3	O O - O O O O
a4	b4	O O O - O O O
a5	b5	O - - O O O O
a6	b6	O O - - O O O
a7	b7	O - - - O O O
a8	b8	O - O - O O O
a9	b9	O - O - O - O
a10	b10	O O O - O - O
a11	b11	O O - - - O O
a12	b12	O O O - - - O
a13	b13	O O O - - O O
a14	b14	O O O O - - O
a15	b15	O O O O - O O
a16	b16	O O O O O - O

2.3. Subjects

This experiment was organized with fifteen native Japanese speakers. They were graduate students at Osaka University aged from 23 to 36, who all had normal hearing. An experiment in Pechmann & Mohr (1992) reports that the memory of pitch is affected only when the subjects are people who have learned music for a long time. To avoid the effect of musical experience, the subjects in this experiment were limited to those who had never learned music or musical instruments more than six months. Another experiment by Friedman, *et. al.* (1975) has examined the effect of right or left handedness and concludes that tapping with their right hand interferes with the work of right handed subjects. All of the subjects in this experiment were right handed and asked to use their right hands.

2.4. Procedure

The whole procedure was controlled by a computer program written by the author. The subjects sat each in front of a personal computer in a sound proof room at Osaka University. Synthesized pulse patterns were presented from a loudspeaker of the personal computer (Fujitsu FMV-NC13D). To enhance the time resolution in a multi-thread and multi-task environment, all possible means were tried by direct use of Windows Application Interface

(WinAPI). Statistical tests were carried out with SPSS-PC+ package software on the same computer.

The subjects were allowed to adjust the volume and screen to most comfortable levels. First, they read the instructions written on one sheet of paper in Japanese and practiced a few times with the help of the author. The responses were stored in the computer.

In the controlled condition, the subjects were asked to listen to the rhythm patterns and repeat them by tapping a keyboard of the computer. In the articulatory suppression (mouthing) condition, they were asked to repeat articulating (mouthing) *aiueo* during the time in which they heard the stimuli and before reproducing the rhythm by tapping a keyboard. In the spatial suppression condition, they were asked to draw as many circles as possible on a sheet before tapping. They were not instructed beforehand about the size or the number of the circle.

Responses were judged by two different criteria. One is an absolute criterion, with absolute zero set at the onset of the first pulse. By this absolute criterion, all tapping pulses of the subjects' response fall within plus or minus 20 percent of the absolute frame. The other criterion was a relative criterion, with a scale applied to each pulse individually. This relative criterion only required adjacent pulses to be within the limitation of plus or minus 20 percent of the set duration. For example, if the first pulse was 20 percent faster, and then the second pulse 20 percent faster again from the first pulse, this response was judged ‘incorrect’ by the absolute criterion, but was judged ‘correct’ by the relative criterion because the second and the third pulses fall within the limit of plus or minus 20 percent. Response time was also measured separately. It was the duration between the end of the stimuli and the onset of the first tap by the subjects.

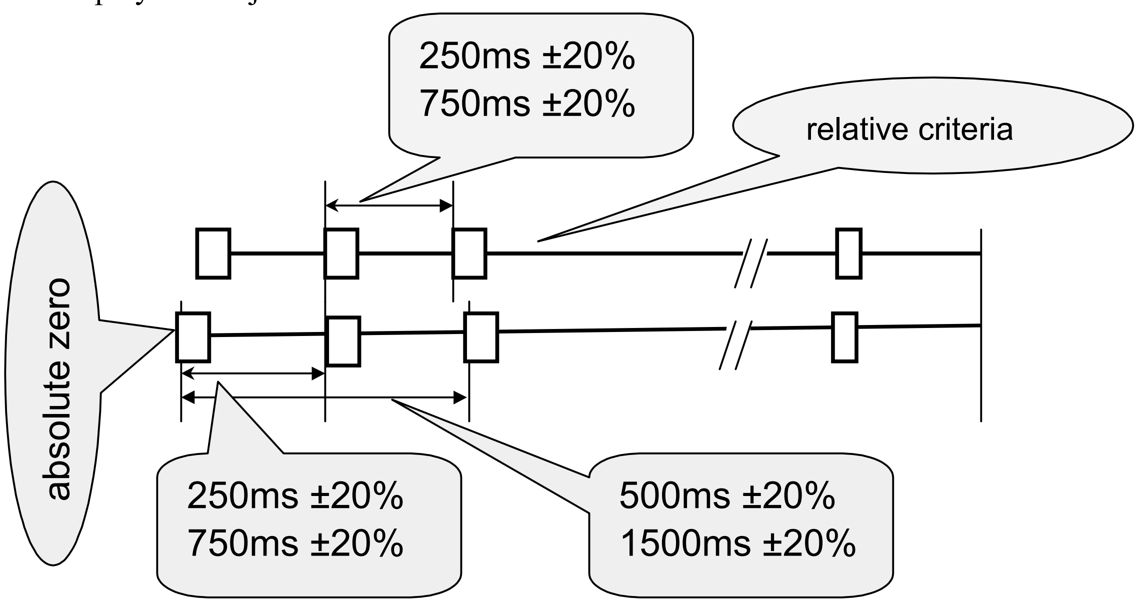

Figure **2** Measurement criteria

2.5. Results

The mean numbers of correct responses are shown together in Figure 3 and Figure 4. The mean response time is plotted in Figure 5 and Figure 6. Figure 7 shows the number of circles that the subjects drew in the experiment in the spatial condition.

Although the results of this experiment are frequency data, ANOVA is applied because the population is comparatively large (Komaki 1995:52, Umemoto 1983). The result reveals that the response to quick rhythm is better than to slow rhythm overall both by the absolute criteria ($F_{(1,94)}$=4.7415, p<0.05) and by relative criteria ($F_{(1,94)}$=6.2079, p<0.05).

When measured by the absolute criteria, the scores for quick rhythm surpass those for the slow rhythm in the control condition (t=3.50, p<0.05), but in the articulatory suppression (mouthing) condition there is no significant difference in the score between the quick and slow rhythms (t=0.34, p=0.738). Scores for quick rhythm are better in the spatial condition though the difference is not significant (t=2.06, p=0.57).

When measured by the relative criteria, quick rhythm yields significantly better tapping reproduction than slow rhythm (t=3.77, p<0.01). No difference in the tempo is seen in the articulatory condition (t=0.17, p=0.865). Scores for the quick rhythm are better than for the slow rhythm in the spatial condition by the relative scoring (t=2.45, p<0.05).

Since the three experimental conditions by absolute scoring have a significant main effect at the fast tempo ($F_{(2,45)}$=18.4759, p<0.001), multiple comparisons by Least-significant difference (LSD) method, by Tukey's method, and by Scheffe's test were made at the significant level of five percent. The results show that scores in the articulatory (mouthing) suppression condition are significantly lower than those in the control condition. At the slow tempo, no significant main effect is observed among the three experimental conditions ($F_{(2,45)}$=0.9649, p=0.3888).

As regards relative scoring, the three conditions at a fast tempo have a significant main effect ($F_{(2,45)}$=9.0107, p<0.001), and the multiple comparison shows that the scores in the articulatory suppression condition are significantly worse than those of the control and the spatial suppression conditions. No significant main effect is available at the slow tempo by relative scoring ($F_{(2,45)}$=0.2573, p=0.77438). In sum, it can be said that the fast tempo gets high scores of rhythm reproductions and that mouthing *aiueo* before tapping disturbs the rhythm reproduction at the fast tempo.

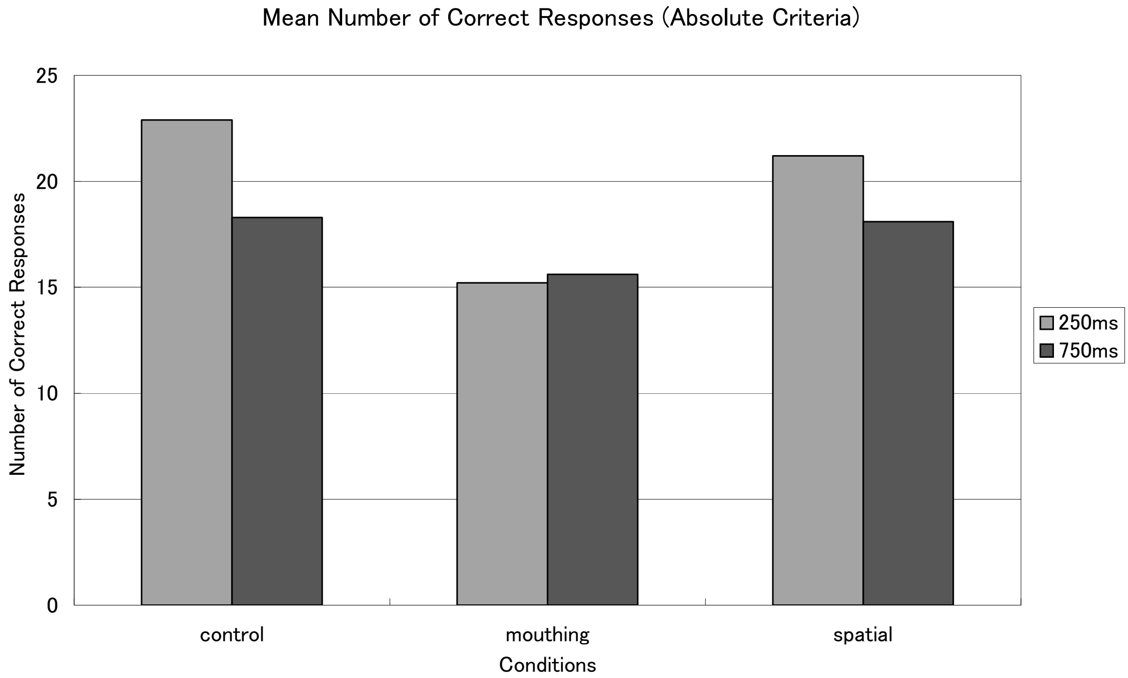

Figure **3** Mean numbers of correct responses by absolute criteria

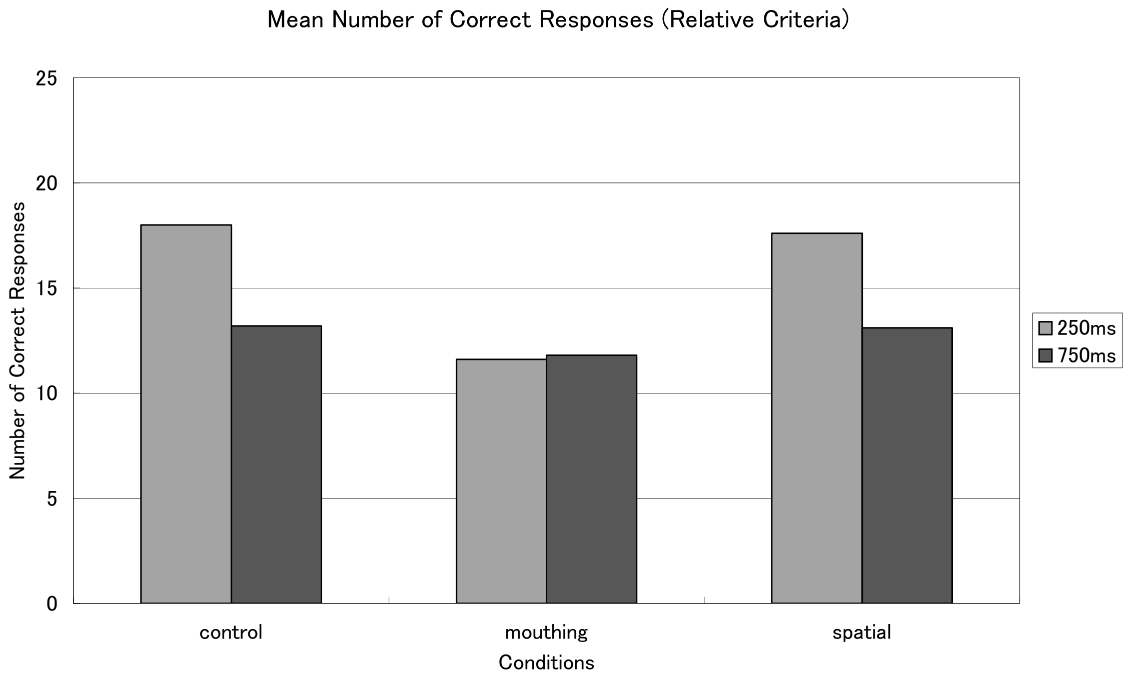

Figure **4** Mean numbers of correct responses by relative criteria

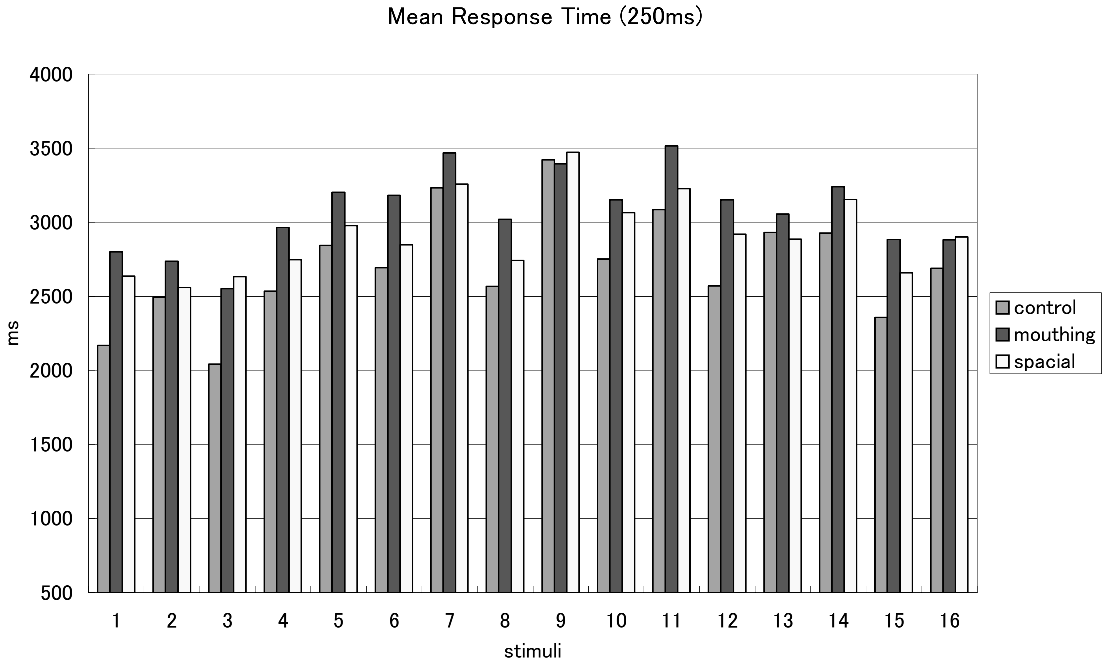

Figure **5** Mean response times (250ms)

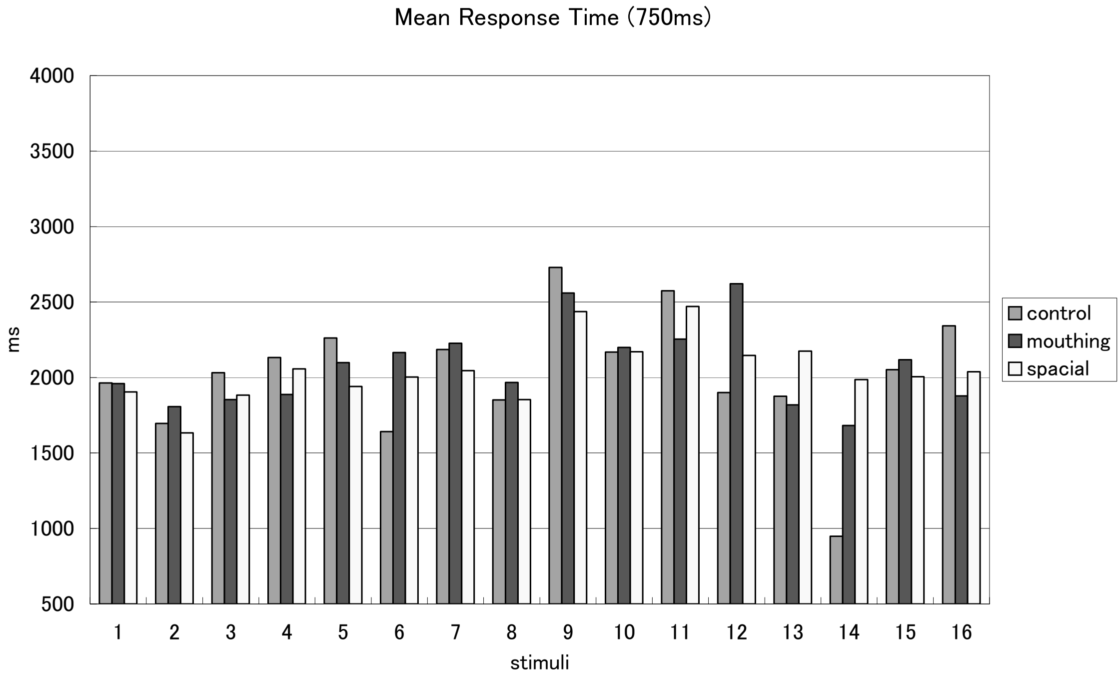

Figure 6 Mean response times (750ms)

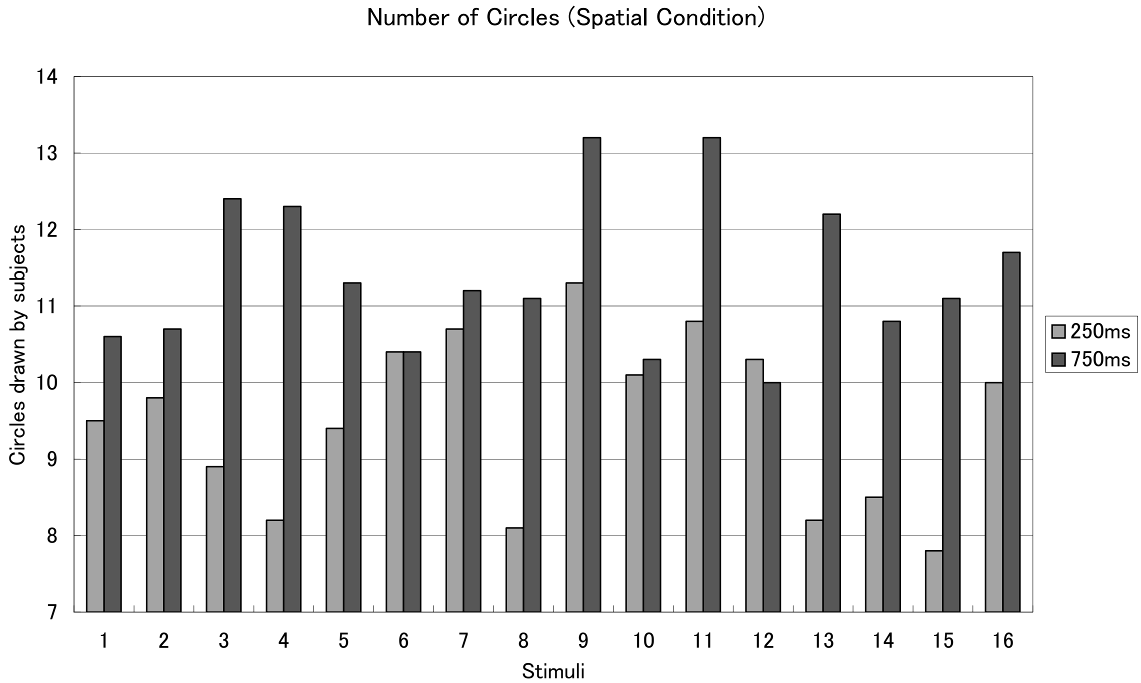

Figure **7** Numbers of circles drawn by subjects

3. Discussion

There are some types of rhythms that can be memorized and reproduced more easily than others. The numbers of correct responses and response time were analyzed to make cluster analyses charts (Figure 8 to Figure 11). By both absolute and relative criteria, the rhythms which have symmetrical structures like stimuli No. 1 and 9 make their cluster first. Another noticeable rhythm is No. 8, which reminds Japanese people of the rhythm of a cheering group's song. These patterns are also the rhythms which are more correctly reproduced. When the correct responses at the faster tempo were counted, symmetrical stimuli make long legs in the cluster analyses. However, this effect is somewhat blurred at slow tempo. The figures also indicate that the first part of the rhythm serves much more for formation of the primitive cluster than the last part. Overall, this result implies that rhythm grouping is processed linearly along the time coordinate. The next question would be whether rhythm reproduction makes parallels with syllabification process of speech sounds.

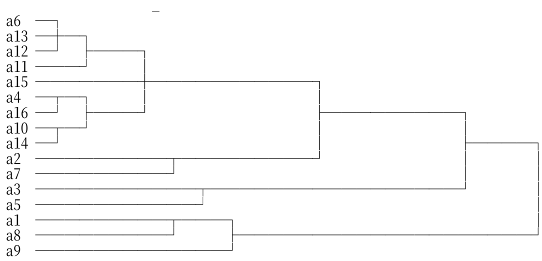

Figure **8** Cluster analysis of the number of correct responses by absolute criteria (250ms)

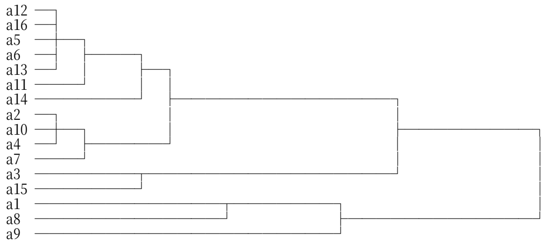

Figure **9** Cluster analysis of the number of correct responses by relative criteria (250ms)

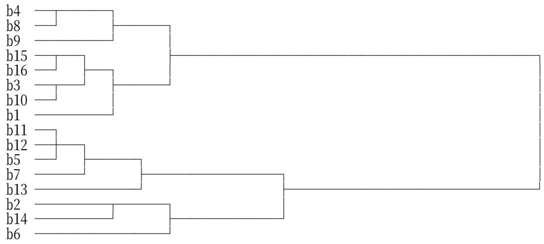

Figure **10** Cluster analysis of the number of correct responses by absolute criteria (750ms)

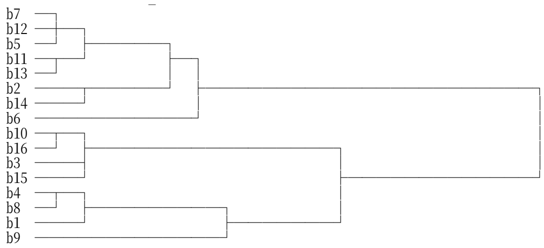

Figure **11** Cluster analysis of the number of correct responses by relative criteria (750ms)

Response time, which is defined as the time span between the offset of the stimulus and the onset of the first pulse reproduced by the subject, is also affected by the tempo of the stimuli. That is, the slow rhythm makes the reproduction significantly quicker ($F_{(1,2879)}$=528.508, p<0.001). This suggests that the slow rhythm is processed simultaneously (analytically one by one) as they hear the stimuli. Because the main effects of the three conditions are also significant ($F_{(2,2877)}$=7.9446, p<0.001), LSD and Tukey's multiple comparisons are made. The results indicate that the response time of the fast rhythm in control condition is the shortest by one percent significant level. In sum, data of the response time indicate that slow tempo and mouthing *aiueo* in repetition yield the longer response time in the experiment.

Cluster analyses with indexes of the response time show that stimuli of which the number of correct tapping reproductions is smaller tend to make their own cluster. Another conspicuous characteristic of the clusters is that the stimuli which have a long blank period in the latter half as in No. 12 and 14 make their own clusters as seen in Figure 12 to Figure 17.

Table **2** LSD analysis of response time by the difference of stimuli (* denotes pairs of groups significantly different at the 5% level)

(*) Denotes pairs of groups significantly different at the .050 level (LSD Procedure)

Mean	Group	Grp 3	Grp 2	Grp 1	Grp 8	Grp 15	Grp 4	Grp 14	Grp 16	Grp 6	Grp 13	Grp 12	Grp 5	Grp 10	Grp 7	Grp 11	Grp 9
2.1715	Grp 3																
2.2210	Grp 2																
2.2387	Grp 1																
2.3463	Grp 8																
2.3677	Grp15																
2.4241	Grp 4	*															
2.4323	Grp14	*	*														
2.4712	Grp16	*	*	*													
2.4742	Grp 6	*	*	*													
2.5602	Grp13	*	*	*	*												
2.5663	Grp12	*	*	*	*												
2.5666	Grp 5	*	*	*	*												
2.5908	Grp10	*	*	*	*	*											
2.7531	Grp 7	*	*	*	*	*	*	*	*	*							
2.8629	Grp11	*	*	*	*	*	*	*	*	*	*	*	*	*			
3.0141	Grp 9	*	*	*	*	*	*	*	*	*	*	*	*	*	*		

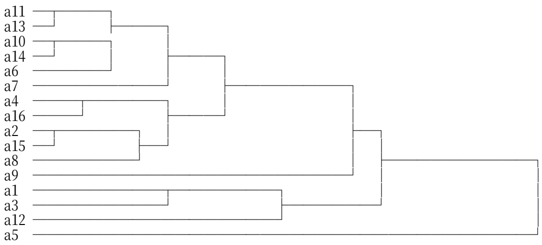

Figure **12** Cluster analysis of response times in control condition (250ms)

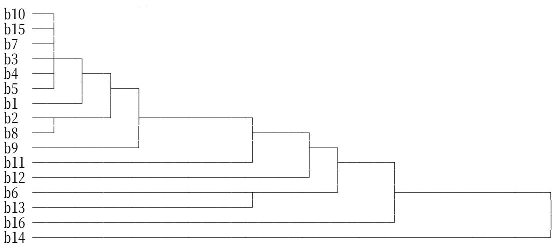

Figure **13** Cluster analysis of response times in control condition (750ms)

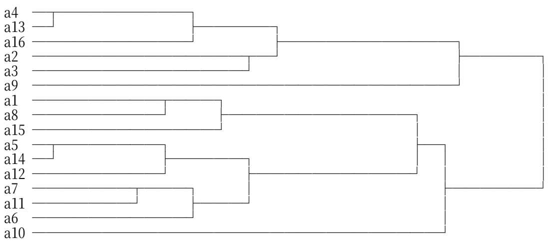

Figure **14** Cluster analysis of response times in mouthing condition (250ms)

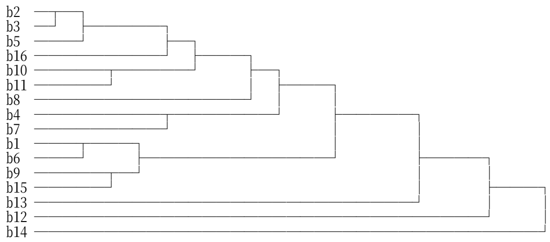

Figure **15** Cluster analysis of response times in mouthing condition (750ms)

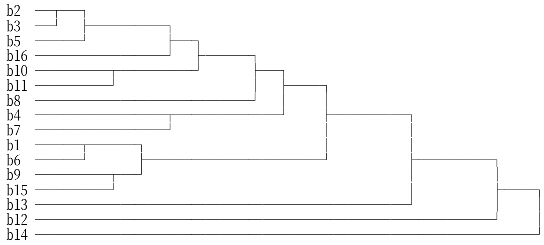

Figure **16** Cluster analysis of response times in spatial condition (250ms)

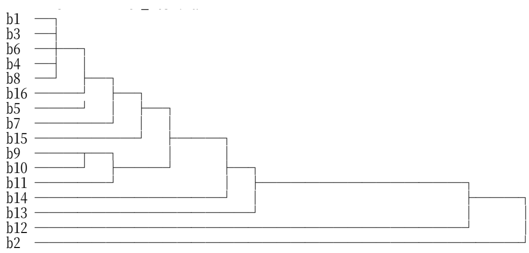

Figure **17** Cluster analysis of response times in spatial condition (750ms)

The stimuli can be classified into two groups from the viewpoint of their structures. One is the sparse-dense type, which has a silent blank in the first half (2, 5, 7, and 8). The other type is the dense-sparse type, which has a silent blank in the latter half (10, 12, 14, and 16). If the pulses in the first part of the stimuli work as a trigger for the formation of a well-formed pattern, the stimuli that have long blanks in the latter half ought to have an advantage for the better grouping. This sort of effect is not observable in the result of the quick rhythms in the mouthing condition scored by the absolute criteria as seen in Figure 18. Statistical testing is not available because of the population size, but rhythms that have a sparse-dense pattern give comparatively better results when scored by the relative criteria as seen in Figure 19.

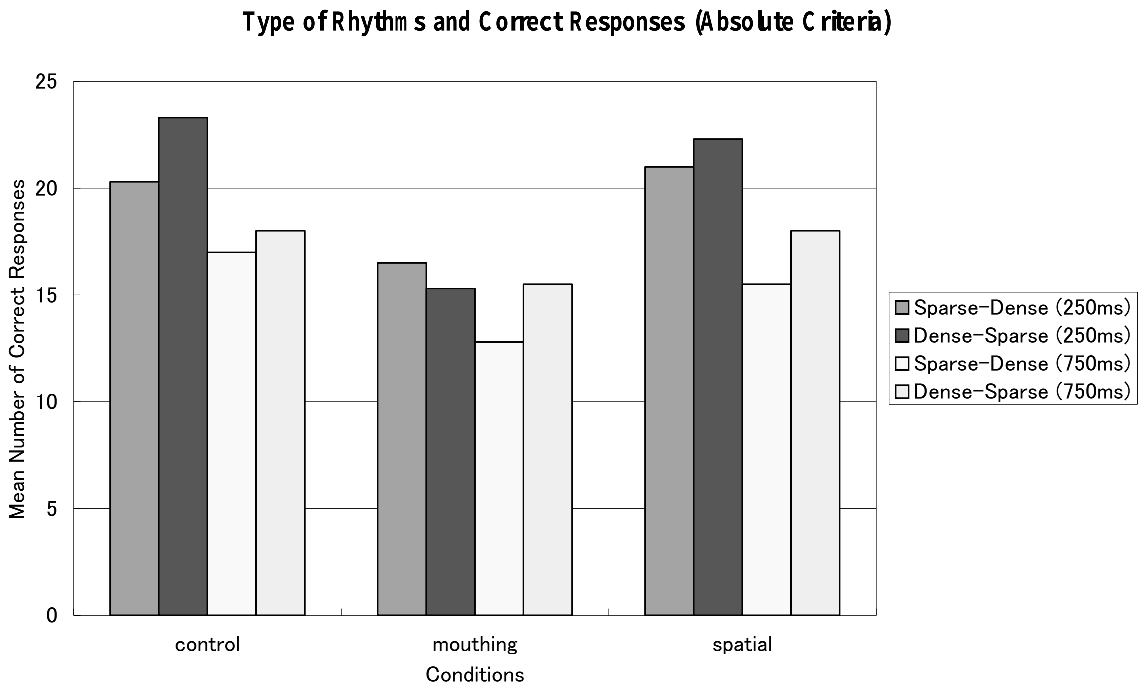

Figure **18** Effect of Sparse-Dense stimuli by absolute criteria

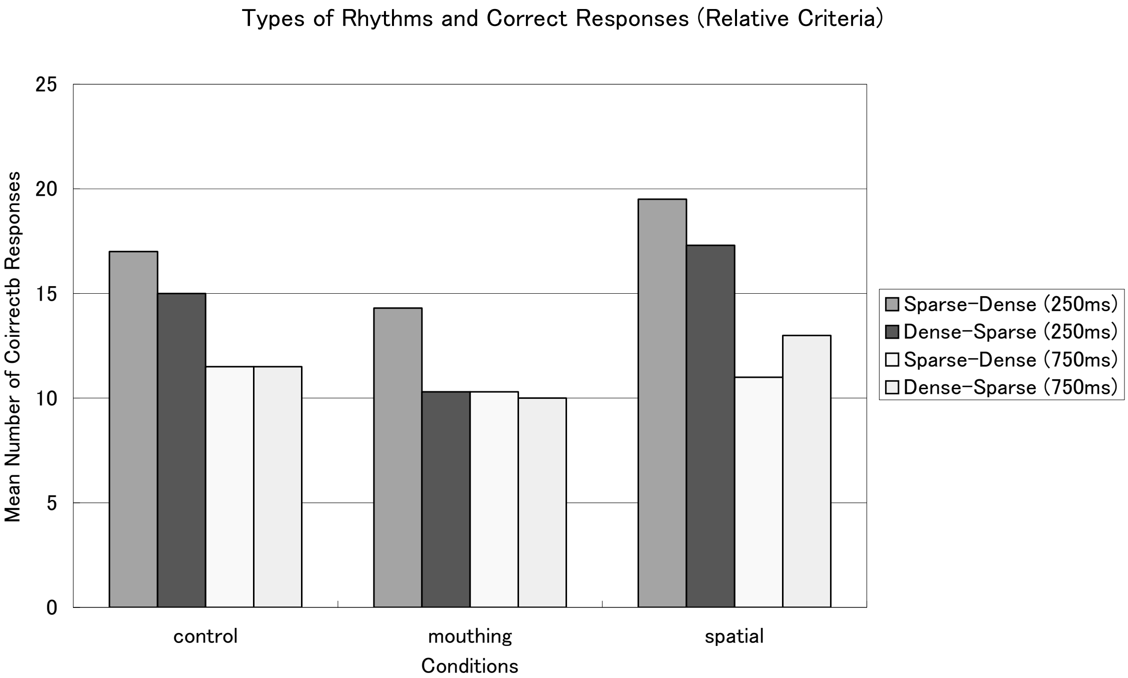

Figure **19** Effect of Sparse-Dense stimuli by relative criteria

4. Conclusion

As it was stated in this experiment, various rhythmical patterns made by editing sine waves were presented to the subjects. The subjects were asked to reproduce the patterns at fast and slow tempi by tapping a computer keyboard. The subjects articulated *aiueo* in the mouthing condition and drew circles in the spatial condition before reproducing the rhythm. The result revealed that the faster tempo gave better results. It suggests that slow tempo hinders the correct analysis of the rhythm.

Statistical tests also showed that scores in mouthing condition were constantly worse than the other two conditions when the interval were set at 250ms. This means that mouthing *aiueo* interferes with rhythm perception and production at the fast tempo. The result is in agreement with earlier studies that used 250ms intervals.

The results are fully conformable to the frameworks of the Phonological Loop and the Working Memory Model constructed by memory retrieval tests of the English alphabet. If we assume that the stimuli presented audibly in this experiment disturb rehearsing at the Articulatory Control Process in the mouthing condition, the information is not kept longer in the Phonological Store. That is, the resources in the Phonological Loop are used up by the articulation of *aiueo*. Drawing circles, on the other hand, does not affect the rhythm processing because the suppression is processed in a different device, such as the Visio-spatial Sketch Pad. It also implies that rhythm and languages are processed in the same device.

5. References

Atkinson, R. C., & Shiffrin, R. M. (1971). "The control of short term memory." *Scientific American*. 225. 82-90.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). "Working memory". *Science.* 255. 556-559.

Baddeley, A. D., & Hitch, G. (1974). "Working memory." In G. H. Bower *ed*. *The psychology of learning and motivation.* 8. 47-90. Academic Press.

Baddeley, A. D., Thomson, N. and Buchanan, M. 1975. "Word length and the structure of short term memory." *Journal of verbal Learning and Verbal Behavior.* 14. 575-589.

Bond, Z. S., & Small, L. H. (1983). "Voicing, vowel and stress mispronunciations in continuous speech." Perception & Psychophysics. 34. 470-474.

Conrad, R. (1964). "Acoustic confusions in immediate memory." *British Journal of Psychology*. 55. 75-84.

Conrad, R., & Hul A. J. (1964). "Information, acoustic confusion and memory span." *British Journal of Psychology*. 55. 429-432.

Craik, F. I. M., & Lockhart, R. S. (1972). "Levels of processing: a framework for memory research." *Journal of Verbal Learning and Verbal Behavior*. 11. 671-684.

Craik, F. I. M., & Watkins, M. J. (1973). "The role of rehearsal in short term memory." *Journal of verbal Learning and Verbal Behavior*. 12. 599-607.

Crowder, R. G. (1993). "Short-term memory: Where do we stand?" *Memory & Cognition*. 21. 142-145.

Ebbinghaus, H. (1885). "Memory: A Contribution to Experimental Psychology." Translated by Henry A. Ruger & Clara E. Bussenius (1913) published in New York by Teachers College, Columbia University.

Friedman, S. L., & Stevenson, M. B. (1975). 'Developmental changes in the understanding of implied motion in two-dimensional pictures.' *Child Development*, 46, 773-778.

Gatherecole, S. E., & Buddeley, A. D. (1993). *Working memory and language*. Lawrence Erlbaum Associates.

Glucksberg, S., & Cowan, G. N. Jr. (1970). "Memory for nonattended auditory mate4rial." *Cognitive Psychology*. 1. 149-156.

Glanzer, M., & Cunitz, A. R. (1966). "Two storage mechanisms in free recall." *Journal of Verbal Learning and Verbal Behavior.* 5: 351-360.

Hibi, S. (1983). 'Rhythm perception in repetitive sound'. *Journal of the Acoustic Society of Japan (E).* 4. 2. 83-95.

Just, M. A., & Carpenter, P. A. (1992). "A capacity theory of comprehension: Individual differences in working memory." *Pschological Review.* 99. 122-149.

Komaki, J. (1995). *Data bunsekiho jyosetu.* Kyoto: Nakanishiya.

Kono, M. (1989). *`Pause no Listening Comprehension ni ataeru eikyouni tsuite'.* IEEE Technical Report. 88. 482.

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

Miller, G. A. (1956). 'The magical number seven plus or minus two: some limits on our capacity for processing information.' *Psychological Review.* 63: 81-97.

Murray, D. J. (1967). "The role of speech responses in short-term memory." *Canadian Journal of Psychology.* 21. 263-276.

Murray, D. J. (1968). "Articulation and acoustic confusability in short-term memory." *Journal of Experimental Psychology*. 78. 679-684.

Nagai, K. (1995). "A study of rhythm perception model." *Gengo Bunka-gaku*. 5:189-201. (Reprinted in *Eigogaku Ronsetsu Shiryo* vol.30) Osaka: Osaka University.

Pechmann, T., & Mohr, G. (1992). 'Interference in memory for tonal pitch: Implications for a working-memory model.' *Memory & Cognition*. 20. 342-344.

Peterson, L. R., & Johnson, S. T. (1971). 'Some effects of minimizing articulation short-term retention.' *Journal of Verbal Learning and Verbal Behavior*. 10. 346-354.

Saito, S. (1994). 'The disappearance of the phonological similarity effect by complex rhythmic tapping.' *Psychologia*. 36. 27-33.

Saito, S. (1997a). *Oninteki sado-kiokunikansuru kenkyu*. Tokyo: Kazama shobo.

Saito, S. (1997b). '*Onin-loop kenkyuno tenkai.*' *Japanese psychological review.* 40. 188-202.

Saito, S., & Ishio, A. (1997). 'Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythm.' Japanese Psychological Research.

Sperling, G. (1960). "The information available in brieff presentations." *Psychological Monographs.* 74. 498.

Umemoto, A. (1983). Rhythm kozono rikaio chusintosuru rhythm kodono kenkyu. Monbusho Kakenhi houkokusho.

Watkins, M. J., & Peynircioglu, Z. F. (1983). "Three recency effects at the same time." *Journal of Verbal Learning and Verbal Behavior*. 22. 375-384.