Ann. Bull. RILP, No. 13, 91-98 (1979)

A PRELIMINARY STUDY ON THE RHYTHM PATTERNS OF REPETITIVE UTTERANCES

- The differences between the utterances pre-set by auditory and visual stimuli -

Seishi Hibi

Introduction

When we are asked to repeat successively a monosyllable such as /pa/ at certain rates which are pre-set by auditory or visual stimuli given to us during the performance, we seem to have more difficulty with visual stimuli than with auditory stimuli.

The relationships between patterns of repetitive production of monosyllables and pre-set auditory stimuli have been examined by many researchers and it is said that the repetitions are consistent up to 6 or 7 Hz for normal subjects.

The aim of the present study is to compare the degrees of inconsistency of responses in terms of the standard deviation of response frequencies together with the response times between the auditory and visual stimuli.

Résumé

Quand on produit la répétition de monosyllabe, tel que /pa/, a une certaine vitesses, qui sont pre-arrangés par les stimuli auditifs ou visuels donnés pendant la performance, on a l'impression d'avoir un peu plus de difficulté dans le cas dernier que dans le cas précédent.

La relation entre le modèle de production répétitive des monosyllabes et celui des stimuli auditifs a été examinée par beaucoup de chercheurs. On dit que les répétitions sont consistents jusqu'à 6 ou 7 Hz pour le sujet normal.

L'auteur a pour but de comparer les degrés de l'inconsistence de réponse sous forme de la déviation standarde des fréquences des réponses avec les délais des réponses entre les stimuli auditifs et visuels.

Procedures

1. Stimulus signal generator

a. Visual stimulus signal generator
As visual stimuli, the light of an LED (Light Emitted Diode) emitted by a pulse generator was used (Figure 1).
The duration of the signal pulse was 50 msec. and the frequencies of the signal were 1 to 6 Hz.

b. Auditory stimulus signal
As auditory stimuli, pre-recorded tone bursts with durations of 5 msec. were used. These tone bursts were made previously by using a sine-curve generator, a pulse generator and the above mentioned visual stimulus signal generator, whose output was used as a trigger pulse. Therefore, the frequencies of this signal were

exactly the same as those of the visual stimulus signal.

2. Subjects and Test utterances

The subjects were two normal male adults and two normal female adults, between 25 and 45 years old. The subjects were asked to produce the monosyllable /pa/ in time to the pre-set auditory or visual signals. The subjects were requested to produce about 30 utterances at each rate.

As can be seen in Figure 2, these utterances and the auditory or visual stimuli were recorded simultaneously.

In addition, subjects were requested to produce the monosyllables repetitively at their maximum rate of utterance without a stimulus signal in order to compare the representative patterns of repetitive utterances with those produced from the pre-set stimuli.

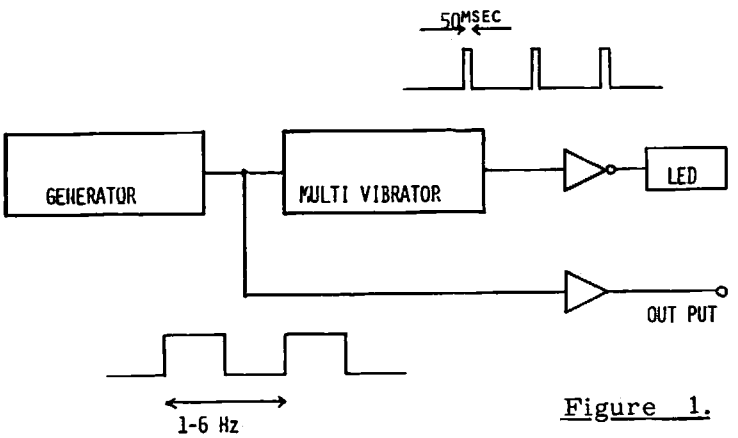


Figure 1.

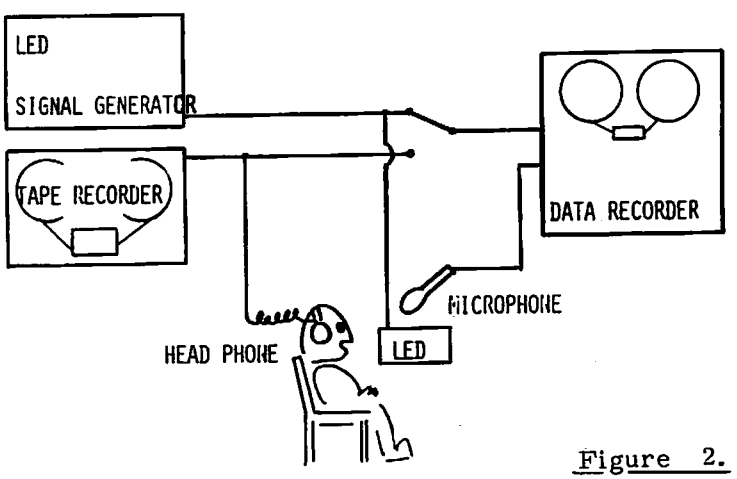


Figure 2.

Results

Table 1 shows the main part of the data obtained in the present experiment. In this table, FS indicates the ideal target frequencies between 1 to 6 Hz for those of the stimulus signals. The actual frequencies of both auditory and visual stimulus signals are slightly different from ideal ones as shown in the line of Fsig (Fsig meaning frequency of the stimulus signal) because of limitations of the tolerance of the electronics parts like resistors or capacitors.

In the line of FM-T or FM-L, the average frequencies for 21 utterances produced by the subject on listening or watching the auditory or visual stimulus signals are shown respectively. The differences from the signal frequencies are also given in parenthesis. The lines SD-T and SD-L show the standard deviation for the levelling-off of FM-T and FM-L.

It was generally observed that in all of the subjects FM-T are rather accurate (only the decimal places have differences fromFsig.)

On the contrary, inconsistency for FM-L compared with FM-T in the same subject was generally observed, i.e., the differences of FM-L from the signal frequencies were slightly bigger than those of FM-T.

In the lines of SD-T and SD-L, it can be roughly said that SD-T is smaller than SD-L for the lower frequencies such as 1, 2 and 3 Hz. This is also shown schematically in Figure 3.

There are some cases in which SD-L is smaller than SD-T, for example, 6 Hz for subject SN, 5 Hz for the subject NU etc. In these cases, FM-L differed from Fsig much more than FM-T did.

It was also found that the patterns of delay of the responses against auditory or visual stimuli differed among the individual subjects.

Table 2 shows some typical patterns of delays of the responses, and Figure 5 illustrates them schematically.

Generally speaking, there are constant delay time patterns and increasing delay time patterns. Type T-1, T-2 and L-1 are types of constant delay time patterns. Type T-1 has constant delay time whose absolute absolute values are small and observed mainly in tone-stimulated utterances as shown in Table 3. Type T-2 also has constant delay times, the absolute values of delay time are, however, fairly large. This type is also observed in tone-stimulated utterances. Type L-1 is observed in many cases of LED-stimulated utterances and has constant delay times. The absolute values of delay time are the largest among these three types.

The types of increasing delay time patterns are type L-2 and L-3. In these types, the delay time increased almost steadily as shown in Figure 4. In other words, it is revealed that these types have different frequencies from signal frequencies.

In Table 2, there are also some special cases like type T-3 and T-4. Type T-3 has constant delay time with one exception. Type T-4 has almost constant delay time values but 5 different values in the later part.

Comments

1. It is questionable whether or not undershoot for articulation occurs during repetitive utterances at rates as high as 5 to 6 Hz.

   In order to test this, the subjects were requested to produce the repetition at their maximum rates in addition to the repetition

Table I.

	F_1	F_2	F_3	F_4	F_5	F_6	Fmax	
FS	1	2	3	4	5	6		
FSig.	0.943	1.961	2.941	4.040	4.878	5.714		
FM-T	0.946(0.003)	1.959(-0.002)	2.931(-0.010)	4.096(0.056)	4.915(0.037)	5.682(-0.032)	6.033	
SD-T	0.02133	0.05955	0.1155	0.2496	0.3363	0.3779	0.3762	
SD-L	0.04517	0.08004	0.1799	0.2053	0.3099	0.2486		
FM-L	0.941(-0.002)	1.969(0.008)	2.945(0.005)	3.940(-0.100)	4.649(-0.129)	6.071(0.357)		Subject SN male
FM-T	0.948(0.005)	1.969(0.008)	2.945(0.005)	4.078(0.038)	4.906(0.028)	5.719(0.005)	7.151	
SD-T	0.03184	0.08652	0.2061	0.4806	0.5818	0.3439	0.2492	
SD-L	0.05072	0.1783	0.4468	0.6535	0.5544	0.5380		
FM-L	0.941(-0.002)	1.996(0.035)	2.952(0.011)	3.994(-0.046)	4.814(-0.064)	5.685(-0.029)		Subject HH male
FM-T	0.951(0.008)	1.952(-0.009)	2.946(0.005)	4.085(0.045)	4.937(0.059)	5.784(0.070)	7.306	
SD-T	0.05825	0.04913	0.1125	0.1957	0.4169	0.4888	0.8802	
SD-L	0.04016	0.1582	0.09859	0.1796	0.1969	0.2257		
FM-L	0.936(-0.007)	2.152(0.191)	2.889(-0.052)	3.652(-0.433)	4.395(-0.542)	4.400(-1.314)		Subject NU female
FM-T	0.943(0.000)	1.943(-0.018)	2.960(0.019)	4.097(0.057)	4.959(0.081)	5.615(-0.099)	6.728	
SD-T	0.03028	0.07082	0.2603	0.2103	0.3193	0.5015	1.063	
SD-L	0.06845	0.1370	0.2881	0.1994	0.3248	0.6559		
FM-L	0.959(0.016)	1.960(-0.001)	2.957(0.016)	3.612(-0.428)	4.740(-0.138)	6.100(0.386)		Subject KP female

FS: Freq.Standard FSig: Freq.Signal FM-T: Freq.Measured-Tone Signal SD-T: SD-Tone Signal
FM-L: Freq.Measured-LED Signal SD-L: SD-LED Signal

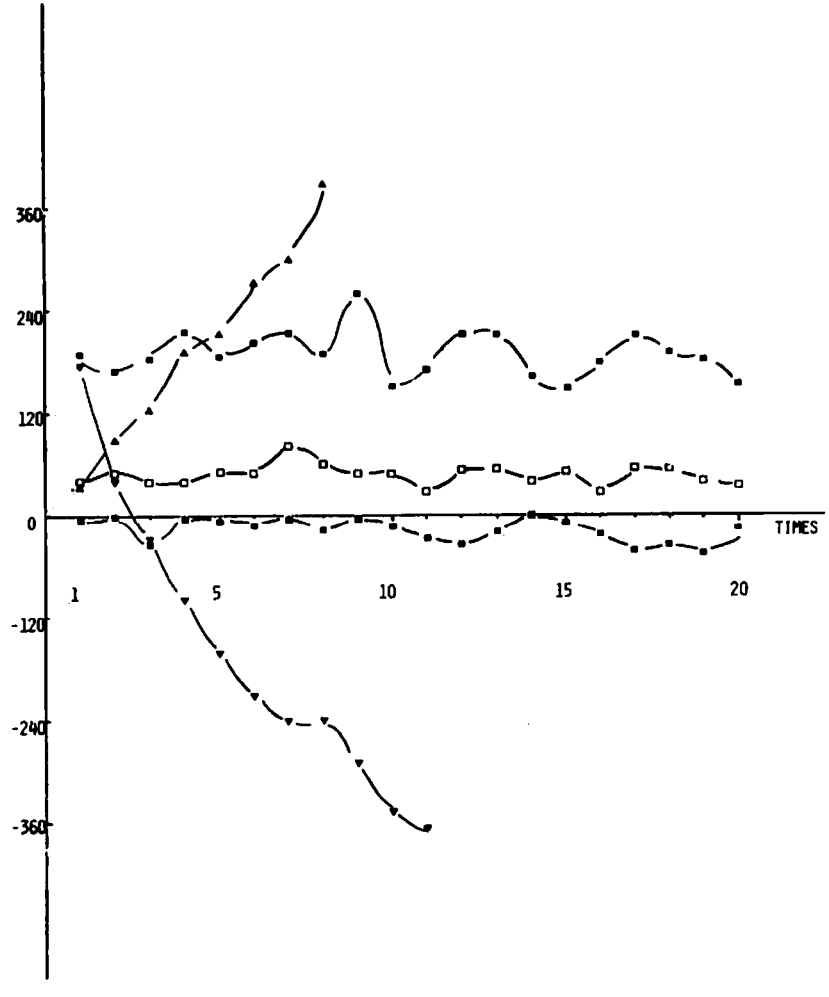


Figure 4.

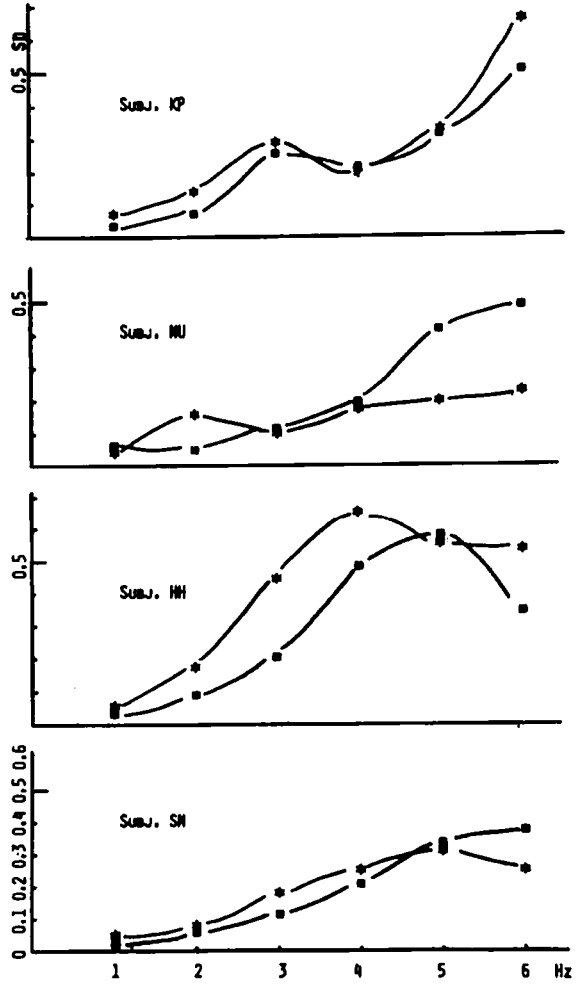


Figure 3.

TYPE T-1 (HH-2)	(HH-3)	TYPE-T-2 (HH-1)	(NU-3)	TYPE T-3 (HH-6)	TYPE L-1 (HH-1)	TYPE L-2 (NU-5)	TYPE L-3 (NU-2)	TYPE L-4 (SN-2)
-7	0	41	-42	21	193	36	177	-42.
0	-11	50	-50	21	170	88	42	-35
-35	26	39	-50	10	185	123	-29	-35
-4	21	39	-50	10	217	193	-99	-49
-7	37	50	-27	10	183	211	-162	-49
-11	0	50	-37	46	208	272	-219	-60
-4	0	80	-34	57	217	299	-240	-78
-18	-21	58	-27	36	190	386	-240	-85
-4	11	48	-32	36	248	404	-289	-92
-14	0	47	-53	51	153	442	-346	-92
-28	-11	27	-32	41	170	492	-367	-95
-35	0	54	-27	21	211	544	-402	-64
-21	11	53	-34	36	215	571	-445	-68
0	0	41	-45	82	163	571	-473	0
-7	26	48	-40	113	144	623	-480	28
-25	-5	26	-27	98	177	659	-501	28
-42	0	48	-40	72	208	729	-494	28
-35	11	49	-45	41	190	825	-480	0
-46	-26	41	-29	10	180	843	-508	-21
-14	-5	35	-50	-10	153	878	-544	-35

Table 2.

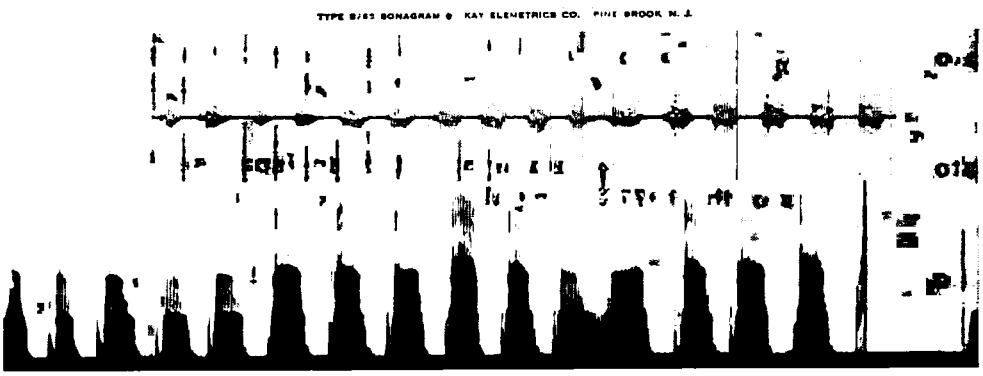
stimulated by the signals from 1 to 6 Hz.

The average of their maximum rates which were levelled-off in individuals, was higher than 6 Hz for all of the subjects.

Thus, tasks where the subjects were requested to produce the repetition pre-set by the signals of 6 Hz were considered to give no problems for articulation. During the performance at their maximum rates, there seems to exist an articulatory undershooting.

As can be seen in Sonagram 1, there are parts which show an explosion noise for /p/ and parts which show vowels /a/, and we can understand them as normal articulations of /pa/.

But there is a part which has no apparent /p/ between the vowels /a/, where the glottal vibration does not stop completely, and the timing of articulation is not accurate.



Sonagram 1.

2. Another problem to be considered is how the feedback mechanism works during the repetitive utterances. There seems to be some feedback mechanism when we produce a repetitive movement such as tapping. Also it can be predicted that there is an auditory feedback system during the production of repetitive utterances.

In order to check this, a further experiment is presently in progress by the same author. In this experiment, the subject is requested to produce the repetitive utterances in time to pre-set auditory signals which were pre-recorded together with a masking noise.

In addition, a visual feedback apparatus was constructed. The auditory wave form of an utterance /pa/ consists of a part that shows an explosion for the consonant /p/ and a part that shows the glottal vibration for the vowel /a/. The visual feedback apparatus has a LED which is emitted by a pulse wave triggered by an explosion part.

Using the above-mentioned noise masking system and the visual feedback system, the experiments will consist of the following:

a. Auditory stimulus signal without masking noise;
b. Auditory stimulus signal with masking noise;

c. Visual stimulus signal with the visual feedback system;
d. Visual stimulus signal without the visual feedback system.

My objective is to compare the degrees of inconsistency of responses in terms of standard deviation of response frequencies with the response times between the auditory and visual stimuli and also between "With Feedback" and "Without Feedback".

References

Allen, G. D. (1975), Speech rhythm: its relation to performance universals and articulatory timing. Journal of Phonetics (1975) 3, 75-86.

Nakamura, R. (1976), Disorder of rhythm formation in Parkinson's disease. Hastening phenomenon. Shinkei-Naika (in Japanese). 4:291-297, 1976.

Sato, H. and S. Hashimoto (1976), On timing of speech and phoneme duration. Reports of the 1976 Autumn Meeting of the Acoust. Soc. Japan, 111-112.

Sato, H. (1977), Analysis of Timing Structure in Japanese Words. Reports of the 1977 Spring Meeting of the Acoust. Soc. Japan, 197-198.

Sato, H. (1977), Some Properties of Phoneme Duration in Japanese Nonsense Words. Reports of the 1977 Autumn Meeting of the Acoust. Soc, Japan, 43-44 .

Sato, H. (1977), Segmental Duration and Timing Location in Speech. Reports of the 1977 Autumn Meeting of the Acoust. Soc. Japan. Transactions of Committee on Speech Research, Acoust. Soc. Japan, No. S77-31.