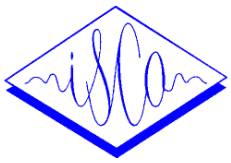


First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

PERCEPTION OF RHYTHM : A COMPARISON BETWEEN AMERICANS AND JAPANESE

Noriko Uosaki and Morio Kohno

1-402 2-9-5 Nakajosanjima-cho Tokushima, 770 Japan
/ Kobe City University of Foreign Studies Nishi-ku,
Kobe, 673 Japan

ABSTRACT

Three kinds of experiments were held to find out some answers to the following questions:(1)Do speakers of different language perceive rhythmic sequences differently? If so, is it due to the language difference? (2) Is the grouping effect consistent in different language groups? (3) Is there any difference in rhythmic perception between sounds with and without linguistic information? The results showed that (1) statistically significant difference was found only in pitch, but not in intensity nor duration, (2)difference in grouping effect among sequences of different frequency, pitch and duration existed only in Japanese subjects, but not in American subjects, (3)the types of stimuli with some linguistic information and without it did not produce any difference in rhythmic perception, (4)perception of rhythm by American subjects is deeply concerned with their native language but no particular relation was found in Japanese subjects.

I. INTRODUCTION

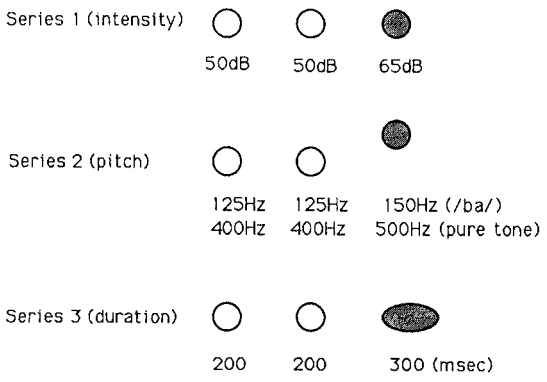
Rhythm, in the psychological sense, means the perception of a series of stimuli as a series of groups of stimuli. In listening to a series of sounds, some of which were louder than others, people showed a strong tendency to hear them in rhythmical groups[1]. Jakobson et al. reported that there is a discrepancy in the perception of identical rhythmic sequences due to the listeners' language difference[2]. When a Czech hears the sounds at even intervals, with every third louder, the pause is claimed to fall before the louder sound,while a French hears the pause after the louder, and a Polish hears the pause one after the louder. The different perception corresponds exactly to the position of the word stress in the language stress in the language involved. Unfortunately Jakobson et al. did not provide any formal experimental data in support of their claim. Thus the question regarding language effect upon the perception of rhythm still remains open. They dealt with intensity contrasted sounds. There are two other possible rhythmic sequences; pitch contrasted series of sounds and duration contrasted. Then (1) do speakers of different language perceive these three types of rhythmic sequences differently ? If so, is it, as Jakobson et al. reported, due to the language difference? Woodrow(1911), by testing American subjects, contended that intensity had a group-initial effect, and duration, a group-final effect, and that pitch had no significant grouping effect[3]. (2) Is this consistent in different language groups? Moreover, it is reported that perception of duration and pitch was affected by whether the stimulus sounds had linguistic information or not [4][5]. Then, (3)is there any difference in rhythmic perception between sounds with and without linguistic information? The purpose of this study is to find out some answers to the above hypothetical questions by testing two language groups, namely American speakers of English and Japanese speakers.

II.EXPERIMENT

Subjects and stimuli

Twelve Americans (4 males and 8 females, age:23-47 $\bar{X}$=30) and 12 Japanese (7 males and 5 females, age 28- 47 $\bar{X}$=36) were tested. At the time of testing, all of the Japanese subjects had been in the United States less than one year. They spoke Japanese at home and their exposure to English was 30-60 minutes a day on average and two hours a day at most. All subjects reported normal hearing in both ears.

Two types of basic synthetic stimuli were generated on the Klatt .formant synthesizer installed in the IBM AT computer in the Phonetics Laboratory at the University of Illinois. They consisted of (1) the CV syllable /ba/; and (2) an approximation of a pure tone- -i.e., a sound which was acoustically simple and which had no linguistic information. In order to present each of the stimulus sounds in three different conditions: with the succession of sounds occurring at equal intervals with every third (marked) sound having (1) a greater amplitude, (2) a higher fundamental frequency, or (3) a longer duration than the two preceding (unmarked) sounds, six kinds of test tapes were generated using Sound Waveform Analyzing Program developed by Professor C.C.Cheng (cf. Figure 1). Each



Fiegure 1. Main Features of the Series.

stimulus sound was 200 msec long except in condition (3) in which marked sound was 300 msec long. Pause between stimuli was 300 msec long. Thus the interstress interval - stimulus plus pause - was 500 msec long. This rate was chosen because natural preferred rates have been found to range around an average of 500 msec of interstress interval which is called central rhythmic area[6]. Each test tape, which contains 120 stimulus sounds, was one minute long. The subjects were tested individually in a sound-attenuated room in the Phonetics Laboratory at the University of Illinois. They were instructed to indicate(by drawing a vertical line at the end of each group on answer sheets) how they perceived the succession of sounds as a group.of three sounds. Each test tape began with unmarked sound followed by marked sound, and unmarked sound, the next.

III. RESULTS AND DISCUSSION

The results are shown Figures 3 a - f. If the marked sound came first in the group, we called it Pattern A, second in the group, Pattern B and third, Pattern C as shown in Figure 2.

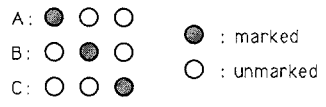


Figure 2. Pattern type

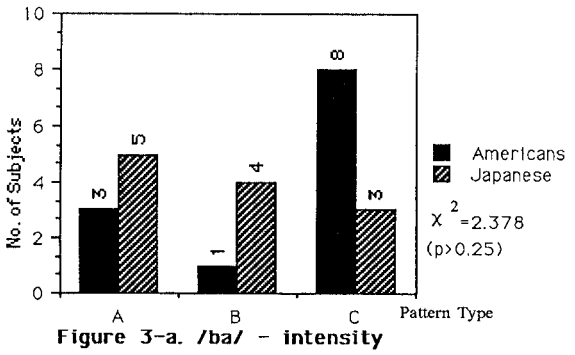


Figure 3-a. /ba/ - intensity

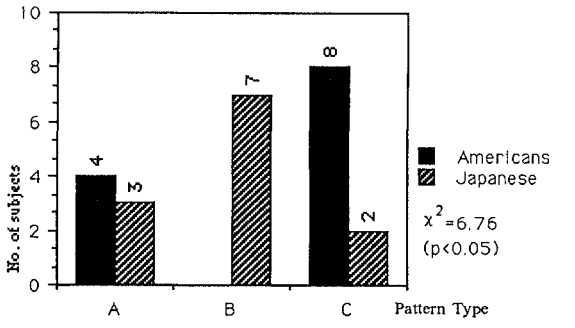


Figure 3-b. /ba/ - pitch

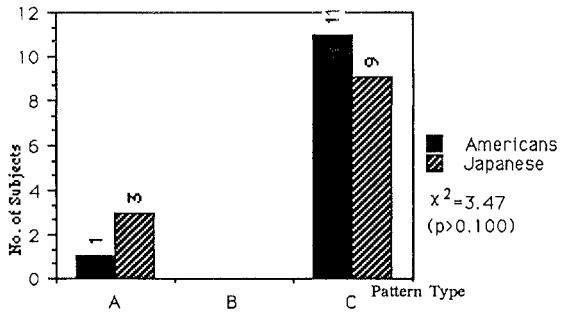


Figure 3-c. /ba/ - length

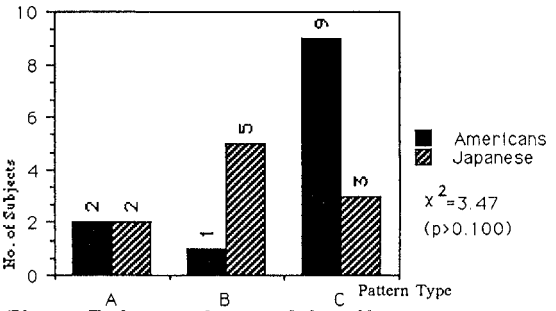


Figure 3-d. pure tone - intensity

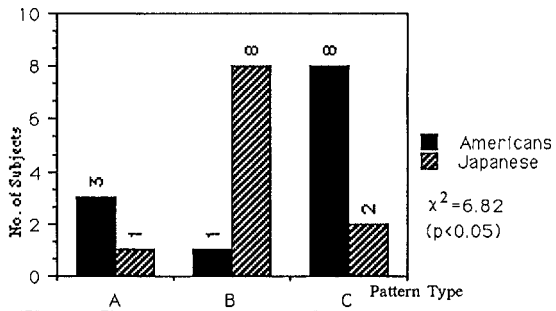


Figure 3-e. pure tone - pitch

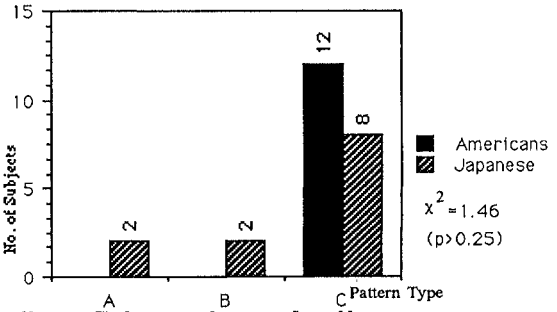


Figure 3-f. pure tone - length

As for intensity, there was no statistically significant difference between the two language groups (cf.Figures 3-a,d). Even though Woodrow(1911) reported that intensity

had a group-initial effect, our American subjects tended to hear the strong sound at the end of the group and there was no particular tendency in perception of Japanese subjects. As for duration, most of the listeners of the both languages heard the long sound at the end of the group (cf. Figures 3-c,f). This result agrees with Woodrow's contention. Thus we might be able to say that group-final effect of long duration is not a phenomenon of a particular language group, but a universal phenomenon. The outstanding difference was found in pitch contrasted stimuli for both speech and non-speech sounds: /ba/- pitch (χ^2 =6.76 $p<0.05^2$ cf. Figure b) and pure tone - pitch (χ^2=6.82 $p<0.05$ cf. Figure e). Most Japanese heard them as pattern B, while most Americans heard them as Pattern C. Why did many Japanese subjects hear it as low-high-low pattern which American subjects rarely heard? One possibility is that, as Woodrow (1909) pointed out that the patterns were likely to change according to the first sound listeners hear, the subjects felt the first sound of the stimuli, the first sound of the group. The stimuli began with a low-pitched sound followed by a high-pitched sound,and low pitched sound, the next. It made the subject form low-high-low group.

EXPERIMENT 2

Did the first Sound really affect the Japanese listeners? In order to verify this, 26 Japanese were asked to listen to pitch contrasted pure tone stimuli which begin with high-low-low(pattern A) and low-low-high(pattern C) sounds. Thirteen of them (7 males and 6 females, age:21-37 $\bar{X}$=30) listened to the pattern A tape and the rest of them (6 males and 7 females, age:27-43 $\bar{X}$= 34) listened to the pattern C tape.The results were shown in Figure 4.

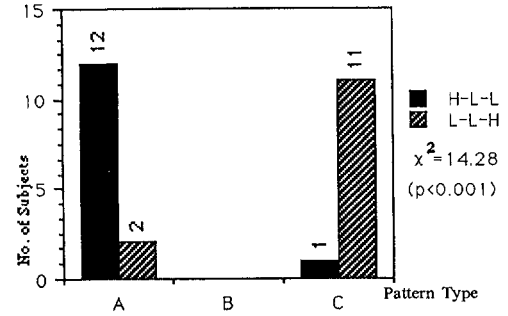


Figure 4. H-L-L(Pattern A) vs. L-L-H(Pattern C)

As shown in Figures 4, most of the pattern A tape listeners heard pattern A, most of the pattern C tape listeners heard pattern C(χ^2=14.28 $p<0.001$). It is clear that the perception of the Japanese subjects.were affected by the first sound of the stimuli rather than by their native language. This result automatically supports Woodrow(1911)'s contention that pitch has no particular grouping effect. Then why did most American listeners, who were not affected by the first sound, hear pattern C(cf. Figures b and e)? This result means that pitch has grouping final effect in the case of the American subjects. In fact American listeners tended to hear the marked sound at the end of the group, whichever the type of stimuli is, while Japanese listeners tended to hear differently according to the types of stimuli (χ^2=25.30 $p<0.001$.). Why do many Americans heard the marked sound at the end of the group ? Is this tendency connected with the characteristics of their native language? In order to answer this question, we held another experiment.

EXPERIMENT 3

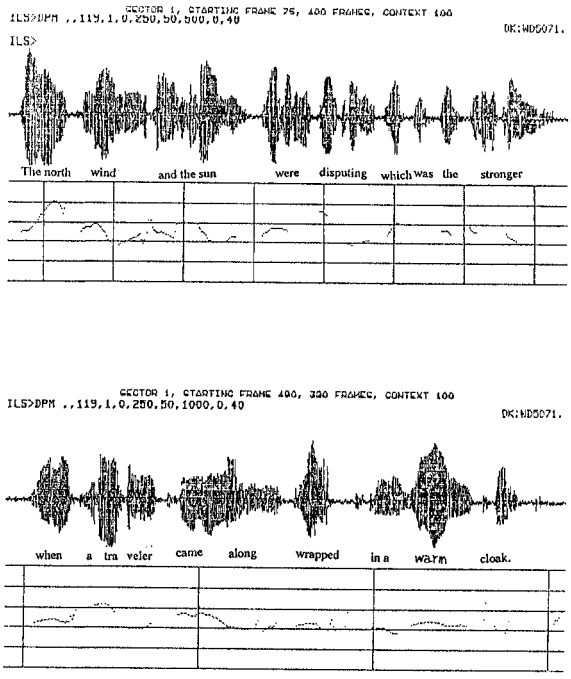
In order to see if there is any emphasis in the utterance of phrase final position in English, the story written by Aesop, "The north wind and the sun." was read by a native speaker of English(American, male, age 31). In speech perception, human beings feel duration of syllables by hearing the length between voice onset points of syllables or interbeat intervals (IBIs) rather than real syllable duration. So IBIs among syllables in the story were measured by the use of the ILS(DEC Micro Computer PDP11/73 connected with DAS BOX-12 and NF filter P-84). The pitch contours were also detected by the same apparatus. The following is part of it.

(/ = the place of a pause, ,/ = phrase boundary).

The north wind// and the sun// were dis - put -ing
IBIs 116 381 100 207 100 272 194

// which was the stron - ger// when a tra - vel - er
IBIs 393 186 263 232 85 255 136 81

came a - long// wrapped in a warm cloak.
IBIs 227 132 129 367 89 200 418 (msec)



These examples show the tendency that IBI was lengthened at the end of phrase. .We can also see that pitch contour raises at the end of phrase. We compared the phrase final IBI(d_1) with non phrase final IBI(d_2) and found that there was a statistically significant difference ($\bar{X}_{d1}$=240.88msec SD= 79.83, $\bar{X}_{d2}$=190.96msec SD=83.92. t=2.60 $p<0.01$). We then analyzed the tape of a Japanese short story (82 words) read by a native speaker of Japanese (23 year old female), and found that there was no statistically significant difference

between phrase final syllable duration (d_3) and non phrase final syllable duration(d_4)($\bar{X}_{d3}$=148.75msec SD=60.82, $\bar{X}_{d4}$=122.54msec SD=31.95, t=1.46 p>0.10). The preceding studies support our result that phrase final syllables were lengthened in English[7] - [11]. why then does English have such long syllables at the end of phrases? Concerning this point, Kohno(1990) referred to the structure of English syllables [12]. English syllables can consist of various number of consonants and one vowel such as V(a), CV(to), VC(up), CVC(book), CCV(play), CCVC(stop), CCCV(splay), CCCVC(straight), VCC(elm), VCCC(ants), CVCC(pact), CCVCC(breathed), CVCCC (camps), CCVCCC(trumped), CVCCCC(tempts), CCVCCCC (glimpsed) and CCCVCCCC(strengths), while Japanese syllables mostly consist of consonant and vowel one-to-one combinations Naturally, the amount of time to utter English syllables will have much variety, compared with the case of Japanese syllables. Kohno compared the retainability of non-sense syllable succession whose interbeat intervals (IBIs) are full of variety and with the retainability of ones where the IBIs are fixed and showed that the non-sense words where syllables were arranged in irregular IBIs are very little retained in memory. In order to compensate the difficulty in perception, English needs some marks to designate the phrase boundaries by lengthening the syllable duration of the last words of the phrases and perhaps by raising their pitch levels. We, therefore, may be able to say that this characteristic of English contributes to the American subjects' judgment of rhythmic group as a unmarked-unmarked-marked pattern.

ACKNOWLEDGMENTS

We would like to thank Dr.Molly Mack for comments and advice, Dr. C.C.Cheng for help with the stimulus making, Pin Mim Kuo for help with computer operation, and Sandra Bott for help with test tape dubbing. Any errors in fact or interpretation are the sole responsibility of the authors.

References

[1] Meumann, E. (1894(. Beitrage sur psychologie des zeitsinn. *Philos. Studien.* 9, pp.264-306.
[2] Jakobson, R., Fant,C., Gunnar, M., and Halle, M.(1951) *Preliminaries to Speech Analysis.* Cambridge, Massachusetts: MIT Press.
[3] Woodrow, H.(1911). The role of pitch in rhythm. *Psychological Review.* 18, pp.54-77.
[4] Lehiste, I.(1970) *Suprasegmentals.* Cambridge, Massachusetts: MIT Press
[5] Mack, M. & Gold,B. (1986). The effect of linguistic content upon the discrimination of pitch in monotone stimuli. *Journal of Phonetics* 14, pp.333-337.
[6] Allen, G.D.(1975) Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics* 3, pp.75-86.
[7] Oller, D.K.(1973) The effect of position n utterance on speech segment duration in English. *J.Acoust. Soc. Am.* 54, pp.1235-1247.
[8] Klatt, D. K. (1976) Linguistic uses of segmental duration in English; Acoustic and perceptual evidence *J. Acoust. Soc. Am.* 59, pp.1208-1221.
[9] Cooper, W.E. and J. Paccia-Cooper (1980) *Syntax and Speech* Harvard University Press, Cambridge.
[10] Cooper, W. E. and M. Danly. (1981) Segmental and temporal aspects of utterance-final lengthening. Phonetica 38, pp.106-115.
[11] Delattre, P. (1966) A Comparison of syllable length conditioning among languages. *International Review of Applied Linguistics.* 4, pp.183-198.
[12] Kohno, M.(1990) The nature of timing control in language. *Abstracts of the papers, PSJ 1990 annual convention at Chiba University*, Chiba, Japan.(in printng).