Manuscript for 12th ICPHS Proceedings
(University of Provence, France)

THE ACOUSTIC CHARACTERISTICS OF BOUNDARIES USED IN UTTERING TELEPHONE NUMBERS IN MANDARIN CHINESE, JAPANESE AND ENGLISH

Y. Tsukuma and J. Azuma

Ritsumeikan University, Kyoto, Japan
Kenmei Women's Junior College, Hyogo, Japan.

ABSTRACT

Through the acoustic analyses on boundaries in telephone number utterances (hereafter, T.N.U.) in Mandarin Chinese, Osaka Japanese and American English (eg. 18-8333, 188-333, where the hyphens denote boundaries), a study was made on some prosodic features which appeared at the boundaries in T.N.U. An observation was also made to determine which prosodic parameters serve as the perceptual factors for the boundaries in T.N.U. by manipulating the F0 parameter and the duration parameter including pauses both of which are considered primary prosodic features. From the experimental results, it is concluded that the speakers of each language employ both the general and language-specific prosodic features to mark the required boundaries in T.N.U. This seems to be based on the unique prosodic characteristics of each language.

1. INTRODUCTION

Prosodic features indicating a major syntactic boundary seem to be language-specific [1][2][3].

In Japanese, the most important prosodic feature to mark a syntactic boundary is the F0 contour resetting at the boundary. Whereas in Mandarin Chinese (also known as Modern Standard Chinese, the term "Chinese" is employed in this paper) and English, pause and preboundary lengthening of a syllable before the syntactic boundary.

The main objective of this study is to investigate if the same prosodic features marking the syntactic boundary are also employed to mark the boundaries found in a sequence of numbers such as T.N.U., which has stable experimental conditions without semantic and syntactic influences.

2. TEST GROUPINGS OF NUMBERS

TABLE 1. Two Groupings of Numbers Used in the Experiments

Language Examined	A	B
Mandarin Chinese	18-8333	188-333
Osaka Japanese	47-5333	475-333
American English	89-8333	898-333

Chinese monosyllabic words have four lexical tones and Japanese two-mora words have four pitch patterns in their phonological inventory. Both the numbers "yao" for 1 and "ba" for 8 in Chinese carry a high-level tone. In Japanese both the numbers "yon" for 4 and "nana" for 7 carry a high-low pitch pattern, and "goo" for 5 and "san" for 3 carry a high-high pitch pattern.

The grouping of two and four is called A, while that of three and three is called B for the sake of convenience. Thus, the hyphens denote boundaries in the sequence of numbers.

3. EXPERIMENTAL PROCEDURES

The test numbers written in Arabic numerals were uttered by the native speakers of their respective languages in declarative intonation.

TABLE 2. Details of the Subjects

Language Examined	Sex	Age	Birth Place
Mandarin Chinese	♀	37	Beijing
Osaka Japanese	♂	36	Osaka
American English	♂	48	New York

Five tokens were selected from ten repetitions of A and B of each grouping and analysed with the acoustic analysis system using NEC-PC9801RX2. The typical acoustic traces of T.N.U. are shown next.

4. AUDIO SIGNAL AND F0 CONTOURS

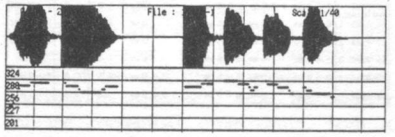

FIG. 1. Grouping A in Chinese

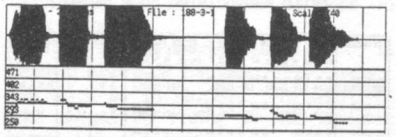

FIG. 2. Grouping B in Chinese

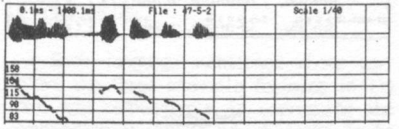

FIG. 3. Grouping A in Japanese

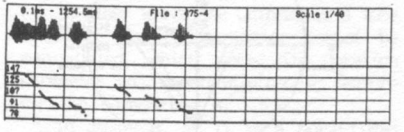

FIG. 4. Grouping B in Japanese

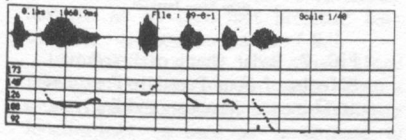

FIG. 5. Grouping A in English

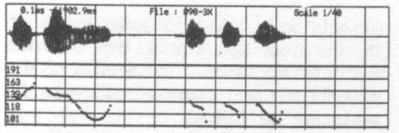

FIG. 6. Grouping B in English

5. PARCOR-SYNTHESIZED STIMULI WITH PAUSE MANIPULATION

In order to investigate to what extent the pause, the stretched duration and the F0 contour resetting at the boundary can contribute to the correct location of the boundaries in T.N.U., a perceptual experiment was carried out.

Using the editing function of the acoustic analysis system, a pause found in groupings A was comletely removed in every language. The stretched vowel duration before the boundary as a result of preboundary lengthening was also cut by 100ms and 200ms in Chinese and English. On the other hand, pauses of 200ms and 400ms were inserted between the numbers where there was no pause originally in every language.

6. PARCOR-SYNTHESIZED STIMULI WITH F0 AND PAUSE MANIPULATION

It is known as a physiological fact that resetting the phrase component takes place after a pause or syntactic boundary. Thus, the F0 parameter was manipulated in every language studied. The F0 values after the boundaries in the groupings A were lowered.

The combination of manipulated F0 and pause (including preboundary lengthening in Chinese and English) was made to examine an interrelationship among these prosodic features.

*PARCOR-synthesis (frame: 25.6ms - 256 sample / trapezoid window / 12 bit - 10kHz sampling / LPF 4.5kHz) for this study was conducted by the Speech Processing System developed by Prof. Miyoko Sugito of Osaka Shoin Women's College

7. PERCEPTUAL EXPERIMENT

The PARCOR-synthesized stimuli were randomized and presented over a cassette tape recorder to native speakers of each language for the perceptual tests. The

subjects were asked to judge which of the two groupings, A or B, each stimulus was intended to be. The details of the subjects who participated in the experiments are as follows:

TABLE 3. Details of the Subjects

Language	Sex	Age	Total	Birth Place
Chinese	♂♀	20~50	75	Beijing
Japanese	♀	18~19	122	Hyogo
English	♂♀	25~55	34	Oregon

8. RESULTS OF PERCEPTUAL EXPERIMENT

Experimental data were analyzed by a microcomputer and a Binominal Test was conducted on the difference between the judgements of each stimulus grouping. The results of the analyses are shown in TABLE 4～9.

TABLE 4. Results of Perceptual Experiment in Chinese (with Manipulated Pause)

STIMULUS	18/	PAUSE	8333	PAUSE INSERTED BEFORE 333	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
18-8333 (ORIGINAL)	770ms	432ms	1102ms	0ms	74	1	P<0.001
18-8333-1	770ms	0ms	1102ms	0ms	61	14	P<0.001
18-8333-2	670ms	0ms	1102ms	0ms	49	26	N.S.
18-8333-3	570ms	0ms	1102ms	0ms	34	41	N.S.
18-8333-4	770ms	0ms	1102ms	200ms	35	40	N.S.
18-8333-5	770ms	0ms	1102ms	400ms	15	60	P<0.001

TABLE 5. Results of Perceptual Experiment in Chinese (with Manipulated F0 & Pause)

STIMULUS	18/ (F0 of 18.	PAUSE	8333 F0 of 8333)	PAUSE INSERTED BEFORE 333	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
18-8333 (ORIGINAL)	770ms (315-306Hz.	432ms	1102ms 306-288Hz)	0ms	74	1	P<0.001
18-8333-6	770ms (315-306Hz.	0ms	1102ms 206-188Hz)	0ms	55	20	P<0.01
18-8333-7	770ms (315-306Hz.	0ms	1102ms 206-188Hz)	200ms	40	35	N.S.
18-8333-8	770ms (315-306Hz.	0ms	1102ms 206-188Hz)	400ms	6	69	P<0.001

TABLE 6. Results of Perceptual Experiment in Japanese (with Manipulated Pause)

STIMULUS	47/	PAUSE	5333	PAUSE INSERTED BETWEEN 5&3	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
47-5333 (ORIGINAL)	482ms	120ms	868	0ms	122	0	P<0.001
47-5333-1	482ms	0ms	868	0ms	121	1	P<0.001
47-5333-2	482ms	0ms	868	200ms	44	78	P<0.05
47-5333-3	482ms	0ms	868	400ms	35	87	P(0.001

TABLE 7. Results of Perceptual Experiment in Japanese (with Manipulated F0 & Pause)

STIMULUS	47/ (F0 of 47.	PAUSE	5333 F0 of 5333)	PAUSE INSERTED BETWEEN 5&3	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
47-5333 (ORIGINAL)	428ms (144-87Hz.	120ms	868ms 135-126Hz)	0ms	122	0	P<0.001
47-5333-4	428ms (144-87Hz.	120ms	868ms 95-85Hz)	0ms	113	9	P<0.001
47-5333-5	428ms (144-87Hz.	0ms	868ms 95-85Hz)	0ms	8	114	P<0.001
47-5333-6	428ms (144-87Hz.	0ms	868ms 95-85Hz)	200ms	0	122	P<0.001
47-5333-7	428ms (144-87Hz.	0ms	868ms 95-85Hz)	400ms	1	121	P<0.001

TABLE 8. Results of Perceptual Experiment in English (with Manipulated Pause)

STIMULUS	89/	PAUSE	8333	PAUSE INSERTED BEFORE 333	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
89-8333 (ORIGINAL)	654ms	192ms	996ms	0ms	33	0	P<0.001
89-8333-1	654ms	0ms	996ms	0ms	32	1	P<0.001
89-8333-2	554ms	0ms	996ms	0ms	31	2	P<0.001
89-8333-3	454ms	0ms	996ms	0ms	19	14	N.S.
89-8333-4	654ms	0ms	996ms	200ms	25	8	P<0.05
89-8333-5	654ms	0ms	996ms	400ms	9	24	N.S.

TABLE 9. Results of Perceptual Experiment in English (with Manipulated F0 & Pause)

STIMULUS	89/ (Fo of 89.	PAUSE	8333 Fo of 8333.	PAUSE INSERTED BEFORE 333 Fo of 8333)	NUMBER OF A	NUMBER OF B	SIGNIFICANCE
89-8333 (ORIGINAL)	654ms (144-119-133Hz.	192ms	996ms 138-158Hz.	0ms 144-126Hz)	33	1	P<0.001
89-8333-6	654ms (212-187Hz.	0ms	996ms 187-147-177Hz.	0ms 197-180Hz)	21	12	N.S.
89-8333-7	654ms (212-187Hz.	0ms	996ms 187-147-177Hz.	200ms 197-180Hz)	2	31	P<0.001
89-8333-8	654ms (212-187Hz.	0ms	996ms 187-147-177Hz.	400ms 197-180Hz)	0	33	P<0.001
89-8333-9	554ms (212-187Hz.	0ms	996ms 187-147-177Hz.	0ms 197-180Hz)	24	9	N.S.
89-8333-10	454ms (212-187Hz.	0ms	996ms 187-147-177Hz.	0ms 197-180Hz)	11	22	N.S.

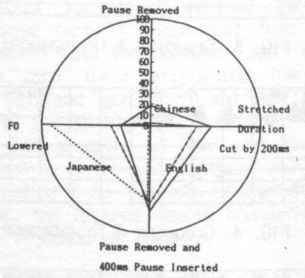

FIG. 7. Radar Chart of Perceptual Effects of Four Prosodic Features in Chinese, Japanese and English.

From the above TABLES and FIGURES, the following observations are made for each language.

Chinese: the boundaries in T.N.U. in Chinese show the following two characteristics; placing a pause at the boundary, lengthening the syllabic vowel of the last digit before the pause as an effect of preboundary lengthening. The perceptual experiments prove that these two prosodic features are found significant, whereas the Fo parameter does not account for the presence of the boundaries in T.N.U. for listeners of this language in comparison with Japanese.

Japanese: the boundaries in T.N.U. in Japanese are made with the same two characteristics as in Chinese. Moreover, resetting the Fo contour after the boundary is also prominent in this language. However, the lengthened syllabic vowel of the last digit before the pause is rare in Japanese. From the results of the perceptual experiments, both a pause insertion and an Fo contour resetting account for the presence of the boundaries in T.N.U. for listeners in this language.

English: the boundaries in T.N.U. in English also show the same three characteristics as in Chinese. In addition, the unique Fo rising of the syllabic vowel of the last digit before the boundary is observed in this language.

9. CONCLUSION

Each language mentioned above employs the three prosodic features in its own way respectively in expressing the boundaries in T.N.U., which seems to be based on the unique prosodic characteristics of each language.

In other words, as Chinese is a tone language, the Fo parameter cannot be manipulated, which makes its duration parameter an important prosodic feature for the boundaries in T.N.U.

In Japanese, as it is a mora-timed language, its duration parameter cannot be manipulated. Thus, an Fo parameter function as an important prosodic feature for the boundaries in T.N.U.

In English, as it is a stress-timed language, its duration parameter (both a pause insertion and the preboundary lengthening) is used as in Chinese to mark the boundaries in T.N.U. However, as the Fo parameter in English is less exhaustively employed to make semantic differences than in Chinese, both the duration parameter and the Fo parameter are used for the boundaries in T.N.U.

However, it should also be noted that for listeners of every language we studied, all of these three prosodic features as an interactive effect could account for the presence of the boundaries in T.N.U.

These experiments also seem to prove that the prosodic characteristics for the boundaries in T.N.U. coincide with those for ordinary sentences in each language[1][2][3].

REFERENCES

[1] AZUMA,J. and TSUKUMA,Y.(1990),"Prosodic Features Determining the Comprehension of a Syntactically Ambiguous Japanese Sentence: In the Case of the Kinki Dialect," *Syntactic Structure and Prosodic Features*, 24-33. (Report on "Prosodic Features of the Japanese Language" – Grant-in-aid for Scientific Research on Priority Areas, the Ministry of Education, Science and Culture of Japan)

[2] LEHISTE,I. (1975). "Role of Duration in Disambiguating Syntactically Ambiguous Sentences," *JASA*, 60,5,1199-1202.

[3] TSUKUMA,Y. and AZUMA,J. (1990),"Prosodic Features Determining the Comprehension of Syntactically Ambiguous Sentences in Mandarin Chinese," *Proceedings of ICSLP 90*, 505-508.

*This work was supported by the Ministry of Education, Science and Culture of Japan (Grant-in-aid for Scientific Research on Priority Areas "Prosodic Features of the Japanese Language" represented by Prof. Miyoko Sugito: Grant Nos. 01642504 and 02224204).