50 Years of Studies on Japanese Prosody

Miyoko Sugito

Institute for Speech Communication Research, Nara Japan
msugito@nifty.ne.jp

Abstract

I will highlight this talk with some examples of my research into speech prosody over nearly the past 50 years. Starting from an acoustic, physiological, and perceptual perspective, F0 contours were extracted from several hundreds of words of both Tokyo and Osaka dialects for a study of Japanese word accent. The durations of pauses and F0 contours in formal newscasts and casual conversations were investigated. This work resulted in tools both for study and for education; for example the "CD-ROM Osaka-Tokyo Accent Speech Dictionary" and the "SUGI Speech Analyzer" were produced and made available. Three CD-ROMs for phonetic science and acoustic education were also produced recently. The Speech Database collected as part of the national research project "Japanese Prosody and its Education" (funded by the Ministry of Education and Science, 1989-1993) included speech data from all Japanese dialectal regions, resulted in a corpus of thirty-six DVD are to be released in April, 2004.

1. Introduction

With its foundations in moraic rhythm and tonal accent system, the Japanese language provides a rich source for explaining the productive and perceptive mechanisms of speech prosody. Since in the early 1940s, it was widely believed that Japanese word accent could not be explained by experimental methods [1]. However, I observed and measured each waveform of hundreds of words, using a Pen-Oscillograph donated by the SONY Laboratory in 1961. The results of these experiments made clear that Japanese accent was a pitch accent and provided a solution to the problem of words having delayed accents or those with devoiced accented vowels. English word accent was also examined by the same methods and these experiments produced similar results to those realized for Japanese accent. Moreover, early in 1970 physiological studies were carried out and perception tests were conducted using synthesized words. Those studies encouraged me to continue my research into speech prosody for many years.

2. Review of Studies on Japanese Word Accent

2.1. Japanese dialectal word accent; a historical view

In spite of the great variety of word accent systems in the dialects of Japanese, they can be categorized into four main types. The Osaka dialect, together with those of Kyoto, Nara and other cities in the Kansai district, has been the normative dialect of Japanese for more than 1,300 years. The Kansai accent patterns are the only ones that can be traced back to the 11th century by referring to the dictionaries and the ancient literature, which provided accent marks beside the letters. For instance, two-mora words can be classified into five groups according to their accent patterns (HH, HL, LL, LH, L-HL) [2] where H and L represent high and low moraic pitch and /-HL/ is used to represent a falling tone on a single mora. The words that belonged to the third group (LL) underwent an accent change to HL during the 17th century. Dialectal word accent systems have been surveyed extensively by many researchers using the above five accent categories [3]. The Osaka-Kyoto accent system is fundamental in the Japanese language and can be viewed as the point of intersection of the synchronic and diachronic studies of Japanese accent [4]. However, nowadays, dialectal accent is changing as a result of influences from the mass media.

2.2. Words with devoiced accented vowels and with delayed accent

The devoiced, accented vowel has been a central topic in discussions on whether the Japanese language has pitch accent or not. Sakuma extracted F0 contours of two mora words with devoiced accented vowels of the Kyoto dialect, and compared these with words having a L-HL accent, however, he failed to explain the falling F0 contours of the second vowels, and in 1931 [5] he concluded that the words were not HL but that all were of accent type LH. He insisted that experiments on Japanese accent were of little value. Later, Hattori and Kawakami reported that devoiced vowels might be heard as accented because of the greater intensity in the devoiced accented mora [6][7]. Chiba reported that, as in many languages, amplitude of speech waves is correlated with fundamental frequency, and concluded that Japanese too has both pitch and stress accent (1935) [8]. Neustupny (1966) [9] wrote that both pitch and intensity (stress) are indispensable to the Japanese word function of delayed accent, and confirmed that both are distinctive features of accent as Chiba reported.

3. Japanese Accent is a Pitch Accent

My experiments began with the extraction of amplitude and fundamental frequencies (henceforth, F0) from speech sound waveforms of 556 words of both the Tokyo and Osaka dialects using the pen-oscillograph. Using this method, F0 movement of vowels, voiced consonants, and segmental boundaries of words could be precisely observed. The results proved conclusively that Japanese word accent was a pitch accent and did not involve stress. We found that the falling F0 contours of the following vowels caused the preceding vowels to be perceived as accented. The abrupt falling F0 contours on the following vowel play an important role in causing listeners to perceive an accent on the first devoiced vowel. Figure 1 shows examples of the word /ike/ (accent HL) 'pond' with delayed accent (1), and the word /chikai/ (HLL) 'near or allied to' with a devoiced accented vowel (2), both uttered in sentences. Their second vowels have falling F0 contours that start from the points marked with black arrows in the figure.

With respect to delayed accent, Figure 2 shows F0 contours of ten stimuli using the synthesized word /asa/. The first vowels have level tones, while the second vowels were

synthesized with the same falling tone. Stimuli 1-7 are perceived as HL, and stimuli 5-10 are perceived as LHL. Stimuli 4-7 are perceived as HL. They can be called words with delayed accent.

Figure 3 shows speech waves of words with synthesized second vowel /a/. The falling F0 contour is the same as that of Figure 2. (1) /kusa/ is a word with natural [kus] followed by synthesized /a/, and (2) shows /ksa/, where the first vowel /u /was omitted from /kusa/(1). Both of the words are perceived as /kusa/ with accent HL. Those experiments show that the falling F0 contours of the following vowels cause the preceding mora to be heard as accented even though the first mora of the second sample has no voiced vowel. Production studies have confirmed that even though the vowel is devoiced in natural speech, the cricothyroid muscle is activated as for an accent [10].

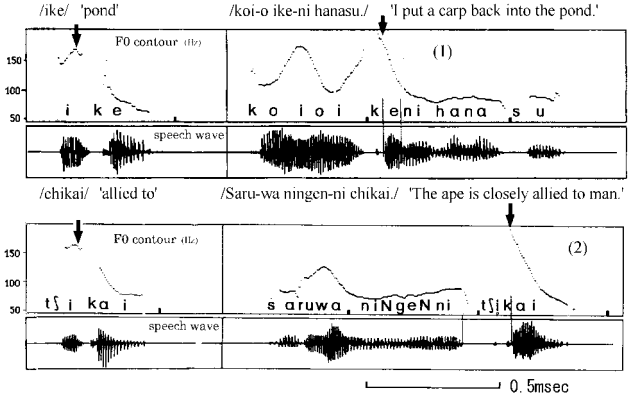

Figure 1: *F0 contours of words and sentences*
(1) An example of delayed pitch fall, /ike/ (HL) 'pond',
(2) An example of a word /chikai/ (HLL) 'allied to' with a devoiced accented vowel. (Black arrows show where pitch fall begin.)

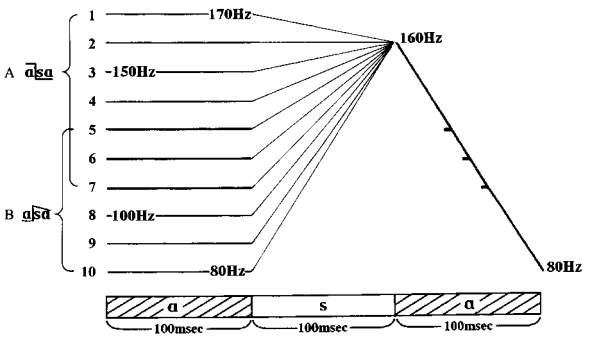

Figure 2: *Synthesized F0 patterns of stimulus /asa/ from No.1-10.*
V1: level tones, 170Hz-80Hz. difference of each stimuli: 10Hz.
V2: descending tones, 160Hz to 80Hz. (Results: stimulus No.1-7 → Type HL, No.5-10→Type L-HL)

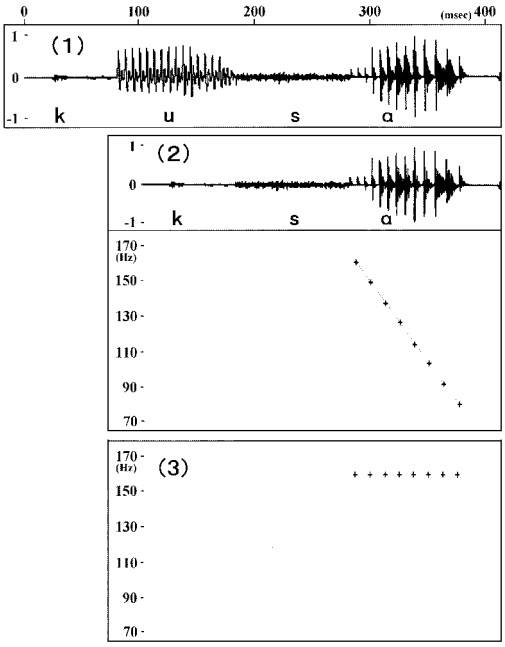

Figure 3: *Natural and synthesized F0 patterns of stimulus /kusa/ 'grass'.*
(1) /kusa/: natural [kus] + synthesized /a/ with a falling F0 contour found in Fig. 2.→perceived as accent HL.
(2) /ksa/: the first vowel [u] was eliminated from [kusa](1) → perceived as [kusa](HL) with devoiced accented first vowel.
(3) /ksa/: the second vowel /a/ was made level→[kusa](LH).

4. Physiological and Perceptual Experiments

The physiological research was carried out using electro-myography on productions of Osaka accent. The data have confirmed that certain muscular systems are active, even throughout the devoicing of vowels, to raise and lower the pitch of an utterance according to the rules of each regional variant. The results of these acoustic and physiological experiments are presented in [11][12]. Research on production and perception of Japanese word accent was also conducted using the Fujisaki-Model [13] [14] to reproduce speech of a native Osaka dialect speaker. The speaker's (laboratory-speech) F0 contours were analyzed and the starting time point of voice raising and falling (T1, T2) were extracted from the F0 contours using the model, along with the other necessary parameters. Using these parameters, words with forty different kinds of HL, L-HL, LH, and HH F0 contours were synthesized. Figure 4 shows the forty different contours synthesized for the word /ame/, whose T1 and T2 timing were changed in small incremental steps. Four kinds of perception tests were conducted. The stimuli with circled numbers in Figure 4 show the boundary between the two types of accent for each column, as obtained from the perception tests which used five Osaka dialect speakers as subjects. These results confirm the influence of timing of F0 rising or falling and segment boundaries plays important roles in perception of the Japanese accent system.

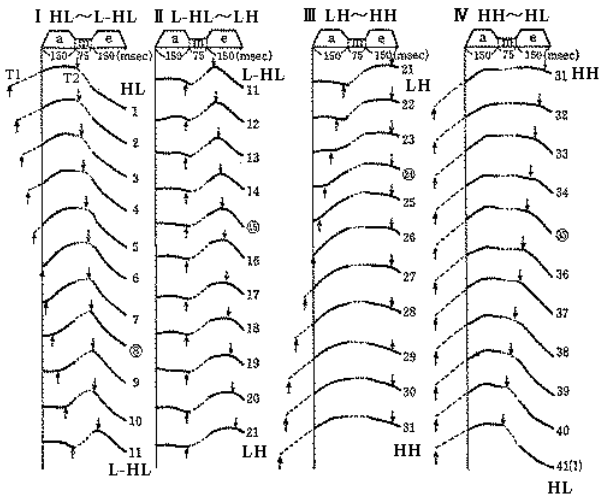

Figure 4: *Forty different F0 contours synthesized for the word /ame/. Timing of F0 rising(T1) and falling(T2) are changed in incremental steps. F0 contours with circled numbers show the boundary between two types of accent.*

5. Pause and Intonation in Conversation

Pauses and intonation of newscast speech and the casual conversations of several speakers were also analyzed. It was made clear that pauses in speech are indispensable not only for the speakers physiologically, allowing them to breathe, but also for the listeners, to provide time for processing the speech information [15][16]. Speech without pauses became impossible to comprehend in news stories, and extremely difficult in conversational speech. Speech without pauses was perceived as having a very high speaking rate, in spite of the actual segmental durations being unchanged.

6. "CD-Rom Osaka Tokyo Accent Speech Dictionary"

The CD-ROM Osaka Tokyo Accent machine-readable dictionary provides text examples that contrast the accents of Osaka and Tokyo Japanese on 65,928 words [17]. For the Osaka accent, we listed the patterns of six native speakers, three elderly and three young, male and female. These show how accent change occurs across generations. We also recorded speech with accents derived from the marks in ancient texts, reflecting the usage patterns of about 900 years ago, and these can be compared with those of the present-day speakers for 2,365 words. Moreover, recorded speech samples of words, and their equivalents in sentences, for 5,684 basic words in both the Osaka and Tokyo dialects can be heard. Their F0 contours are displayed with the corresponding speech waves as illustrated in Figure 1. Samples of dialectal conversations and read speech materials in the voices of native Osaka speakers are also included. This database can be called the first multimedia dictionary. It took me more than 30 years to collect the materials.

7. Tools For Research on Prosody and for Speech Education

The SUGI SpeechAnalyzer [18] is a suite of speech processing software that was produced as a result of my observations and measurements of the waveforms of hundreds of words starting with the Pen-Oschillograph, and was used extensively during the development of the CD-ROM and for analyzing speech data. An English-language version of this software will shortly become available. This work and the database explained in the following chapter, has also resulted in three kinds of CD-ROM for education [19][20][21] : one is for the study of dialectal variation of the accents from all over Japan, resulting from the dialectal and historical study, one provides a speech database of dialectal accent that allows interactive listening using a map of the dialectal regions of Japan, and the third was designed for the teaching and education of speech and speech communication using the data derived from my experiments.

8. Speech Database Produced by National Project, "Japanese Prosody and Education"

As many of the dialects are rapidly changing due to the influence of mass media, it was considered necessary to preserve them, taking advantage of the recent developments in digital tape recorders. In 1988 the Ministry of Education and Science decided to let starting the research project of "Integrated Studies on Prosodic Feature of Current Japanese Language with Application to spoken Language Education" (in short, "Japanese Prosody and its Education") (1989-1993, project leader, Sugito, M.) [22] uniting 281 researchers across Japan. The teams of researchers covered the country to collect data for national research into Japanese prosody, generously financed by the Japanese Government. (Figure 5 and Figure 6 show the survey points.)

These recordings provided the basis of a speech database of dialects from all over Japan which has become the focus of co-studies uniting linguistic, dialectological, and engineering-based approaches for the study of speech prosody in Japan. As a result, 18 CD and 3 CD-ROM were initially produced and have been distributed. Previous research did not require the preservation of the speech samples utilized, and we have no possibility now of knowing the precise nature of their content or speaking styles. In particular, recording of the endangered Ryukyu dialect was considered necessary as many of the local speakers no longer use the dialect in their speech. The speech of the few remaining native Ainu speakers who have preserved their Ainu language will no doubt prove invaluable

Now, all of the speech data are to be put together (by cooperation of Nagoya University and National Institute of Japanese Language) in 36 DVDs, for use of researchers by April this year. These DVDs will illustrate the great variety of word accent among the dialects of Japanese, and we hope that they will encourage further research on this subject worldwide.

9. Summary

In my studies of fifty years on Japanese word accent, moraic rhythm, segments, intonation and pauses, I have described experiments on accent and pauses, CD-ROM as tools for research and education, and speech database of Japanese dialect collected by the national project. Our prosody

education is desirable now based not just on theoretical models, but also on concrete data and computer-based analysis, using tools that are widely available and currently in use throughout the country.

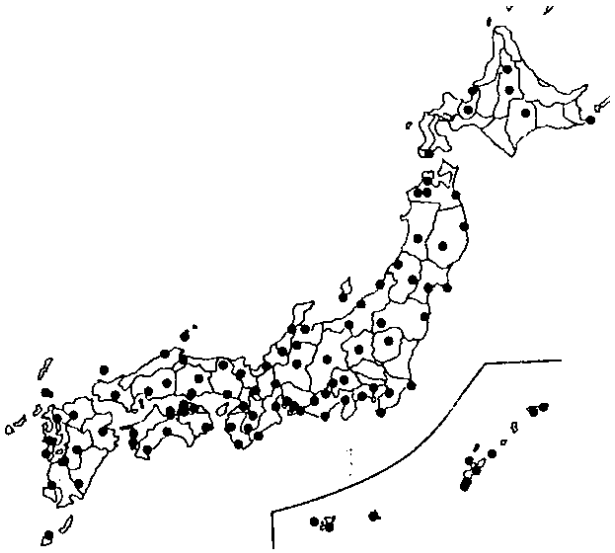

Figure 5: *The 105 points across Japan, where the speech of elderly dialect speakers at each point was recorded as input for the speech database.*

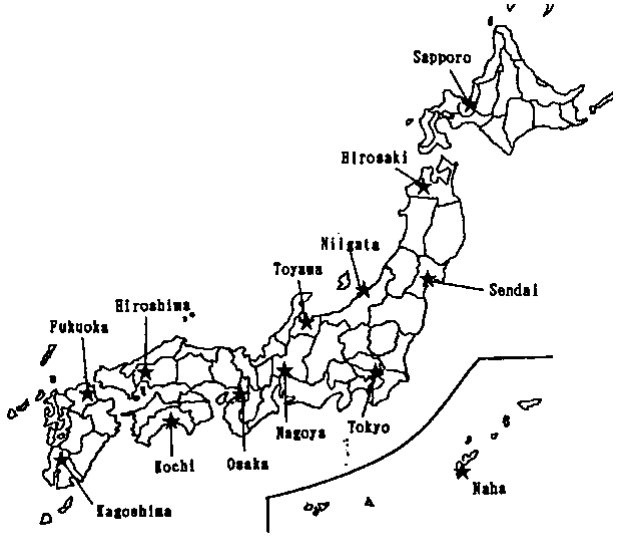

Figure 6: *The 13 major cities across the nation: Sapporo, Hirosaki, Niigata, Sendai, Tokyo, Nagoya, Toyama, Osaka, Hiroshima, Kochi, Fukukoka, Kagoshima, and Naha, where speech was recorded from up to 100 informants in ranging age from the elderly to elementary school children, divided evenly between the sexes.*

10. References

[1] Sakuma, K., "Jituyooteki Nihongo Onseigaku (*Practical Japanese Phonetics*)", Doobun Shoin, Tokyo, 1941.

[2] Kindaichi, H., "Gendai shohoogen no hikaku kara mita Heian-choo akusento (Heian accent compared with dialectal accents)," Hoogen 1-7, Syunyoodoo, Tokyo, pp. 1-43, 1937.

[3] Hirayama, T., Zen-nihon Akusento no Syosoo (Aspects of Accents in Japan), Ikueisyoinm Tokyo, 1940.

[4] Hattori, S., A General View on Accent of Japanese dialects, *Hoogen* 1-1+3+4, 2-1+4, 3-6, 1931-1933.

[5] Sakuma, K., "Kyoto-go ni okeru akusento (Accent in Kyoto dialect)," *Onsei no Kenkyuu* (Study of Sounds) IV, Phonetic Society of Japan, pp.61-77. 1931.

[6] Hattori, S., *Gengogaku no Hoohoo (The Way of Linguistics)*, Iwanami Shoten, Tokyo, 1960.

[7] Kawakami, S., "Museihaku no tsuyosa to akusento kaku (Intensity of devoiced accented mora)," *Kokugakuindai kokugo kenkyu* 27, Univ. Kokugakuin, Tokyo, 1969.

[8] Chiba, T., *Study on Accent from Viewpoint of Experimental Phonetics, Fuzanboo*, Tokyo, 1935.

[9] Neustupny, J.V., "Is the Japanese accent a pitch accent?," *The Bulletin* 121, The Phonetic Society of Japan, Tokyo, pp.1-7, 1966.

[10] Sugito, M. and Hirose, H., "Production and perception of accented devoiced vowels in Japanese," *Annu. Bull. Res. Inst. Logop. Phoniat.* 22, Univ. Tokyo, pp. 19-37, 1988.

[11] Sugito, M., "Timing relationships between prosodic and segmental control in Osaka Japanese word accent, *Phonetica 60*, Karger, Switzerland, pp. 1-16, 2003.

[12] Sugito, M., "The effect of speech rate on devoiced accented vowels in Osaka Japanese," J. van de Weijer et al. (eds.),Voicing in Japanese, to appear.

[13] Fujisaki, H. and Sugito, M., "Acoustic and perceptual analysis of two-mora word accent types in the Osaka dialect," *Annu. Bull. Res. Inst. Logop. Phoniat.* 10, Univ. Tokyo, pp. 157-171, 1976.

[14] Sugito, M., Nihongo *Akusento no Kenkyuu (Studies on Japanese Accent)*, Sanseidoo, Tokyo, 1982.

[15] Sugito, M., "Pauses and intonation," *Aspects of text and context*, The National Language Research Institute (Kokuritsu Kokugo Kenkyuusyo), pp. 107-138, 1987.

[16] Sugito, M., "On the role of pauses in production and perception of discourse," *The proceedings of the 1990 ICSLP*, Kobe, pp. 513-516, 1990.

[17] Sugito, M., *CD-ROM Osaka Tokyo Accent Speech Dictionary*, Maruzen, Tokyo, 1996.

[18] Sugito, M., SUGI SpeechAnalyzer, ANIMO, Yokohama, 2000.

[19] Sugito, M. et al., *Variation and Changes of Japanese Accent (CD-ROM)* (in Japanese), National Institute of Multimedia Education, Ministry of Education and Science, 2002.

[20] Sugito, M. et al., *Speech Database of Japanese Dialectal Accent (CD-ROM)* (in Japanese), National Institute of Multimedia Education, Ministry of Education and Science, 2002.

[21] Sugito, M., *KOENO MANDARA (Speech and Communication)* (CD-ROM), National Institute of Multimedia Education, Ministry of Education and Science, 2003.

Sugito, M.(ed.), *Nihongo Onsei (Japanese Prosody) [1][2]*, Sanseidoo, Tokyo, 1997.