Periodicity of Japanese Accent in Continuous Speech

Kitazawa Shigeyoshi, Kitamura Tatsuya, Mochizuki Kazuya, and Itoh Toshihiko

Department of Computer Science
Shizuoka University
(kitazawa; kitamura; cs7093; t-itoh)@cs.inf.shizuoka.ac.jp

Abstract

Japanese is a tonal language, and it is also mora-timed language. In this paper, we investigated the hypothesis that accent kernels appear at regular intervals. If duration of each mora is almost equal, then Japanese accent kernel is a constant time marker of continuous Japanese. The Japanese prosodic corpus we investigated comprises 3 hours and 37 minutes of speech involving six different speakers (3 males and 3 females). The recordings on which the corpus is based are passages translated from the MULTEXT prosodic corpus distributed through the European Language Resource Association (ELRA). The text was translated from five European languages and modified into Japanese, consisting of 40 different passages, with 6523 morae including 1085 accent kernels. Speakers are professional narrators and actors who read the text naturally before recording. During the recording, the standard Japanese accent was kept. Accent kernels are marked by two experienced researchers of Japanese phonology. ANOVA assured the average occurrence of accent kernel is one out of every six morae.

1. Introduction

Japanese is a tonal language, and a mora-language. The duration of each mora is equal (at least psychologically). Japanese prosody is based on this mora counting [1]. In this paper, we investigated the hypothesis that accent kernels appear at regular intervals. If the duration of each mora is almost equal, then the Japanese accent kernel is a constant time marker of continuous Japanese.

Japanese prosody is defined by “word accent (high/low)”, with the accent determined by pitch type (High or Low in fundamental frequency). English accent, on the other hand, is based on stress (Strong or Weak in speech power). This word accent is a stable and regular feature of a sentence.

And there are more global features such as sentence accent, rhythm, and intonation of Japanese. The sentence accent is so-called because stress is applied to some word in a sentence with enhanced intensity where the word bears important information of the sentence, prominence or focus.

Japanese rhythm is based on mora-timing, that is, each mora keeps an almost equal time interval, while western rhythm is stress-timed, that is intervals between stresses are kept almost equal. We have shown that Japanese rhythm is based on bimoraic units and have developed a TEMAX-gram that visualizes this rhythmic pattern, as well as speech rate [2]. Intonation of Japanese is not as evident as it is in Western languages. For example, ending pitch is raised a little in enquiry sentence, but usually spoken in falling inotonation.

Japanese prosody is characterized by stable word accent, constant mora-timed speaking rhythm, and an ambiguous intonation pattern.

This study is a part of a Japanese prosodic corpus development. At least in Japan, a database of prosody has never been developed before. Although several collections of speech samples are called “prosodic databases” and used to control pitch movement of synthetic speech, these are not sufficient for research purposes. There is no phonological study concerning a large number of accent kernels in continuous Japanese.

2. A Japanese prosodic corpus

A prosodic corpus of Japanese is going to be developed as a scheduled project by university researchers in Japan. A new project focusing on "Realization of advanced spoken language information processing from prosodic features" headed by Professor Keikichi Hirose (Professor, Department of Frontier Informatics, School of Frontier Sciences, The University of Tokyo) has been underway since October 2000. This is a project sponsored by the government, the Ministry of Education, Culture, Sports, Science and Technology, financed with a fund called the “Grants-in-Aid for Scientific Research”. The project consists of several subgroups including theoretical, phonological, pathological, interactive, discourse, as well as speech recognition and speech synthesis. Cooperating researchers are from departments of computer science, linguistics, psychology and medicine as well as some company researchers. The planned term of the project is from 2000 to 2003. The goal of this research project is the development of a prosodic corpus.

2.1. Prosodic corpus based on the MULTEXT [3,4]

The idea for a prosodic corpus based on the MULTEXT [3,4] came from the ELRA prosodic corpus MULTEXT [5]. Its intended use is mainly for the study of intonation, but it contains other useful ideas: written text contains various situations expressed in several sentences (passages) that is better than a single sentence to express prosodic and emotional attitude. Passages are translated in several different languages with small adaptations that means the passages are easily translatable into other languages, including Japanese. The MULTEXT CD data is read in an emotionally neutral tone, but can be performed with different emotional paralinguistic attitudes (semi-naturally).

We recorded three males and three females including professional narrators and actors and actresses in a sound proofed room at the Tokyu Construction Technical Research Institute. Data was recorded with the precision apparatus for sound measurement in this perfectly non-reverberant room (11.6 x 11.6 x 6.5 m). The recording equipment used included: a B&K 4190 condenser microphone, a B&K 2669 preamplifier, a NEXSUS 2690 conditioning amplifier, and a SONY PCM2300 DAT recorder. Speakers were equipped with KAY 6103 EGG electrodes and recorded simultaneously with a microphone.

The recordings are digitally transferred to computer's hard disk drive in a sampling rate of 48 kHz, 16 bits.

2.2. Text translation from MULTEXT

The text consists of 40 small paragraphs (passages) edited from translated MULTEXT consisting of various topics, situations and styles of talking such as: ordering something by telephone, a telephoned complaint, an urgent report, a telephone reference, a presentation, report, traffic information, an apology, boast, a letter, occasional thoughts, a discourse, a lecture, a novel, a monologue, etc.

Five different Japanese texts were created by translating from five MULTEXT languages (English, French, German, Spanish, and Italian), with the originals unified and modified for appropriate Japanese context. The translation into Japanese from the various languages is rather free and often constitutes an adaptation to the local culture (for proper names, food, etc.). Therefore, each sentence is a little long for oral utterance accompanied with the hardness of a translation tone. They also have the feature of written language tone rather than spoken-language tone.

2.3. Recordings procedure

The recording procedure is drawn from the MULTEXT multi-lingual prosodic corpus. We recorded all 40 passages for each speaker in two different modes: text reading mode and semi-natural performance mode. Every speaker was asked to read a whole set of the passages and to try to use as natural an intonation as possible. The recorded material was controlled during acquisition so that bad quality recordings (noisy or misread sentences) were directly cancelled and repeated.

Table 1 gives the number of passages read by speakers and duration per language, where the row Japanese RD is read speech and the row Japanese NT is performed speech.

Table 1: Duration per language.

Language	Passages per speaker	Total duration (h:m:s)	Average duration per passage (s)
Japanese RD	**40**	**01:47:52**	**26.9**
Japanese NT	**40**	**01:49:11**	**27.4**
English	15	00:43:55	17.6
French	10	00:36:30	21.9
German	20	01:13:09	21.9
Italian	15	00:54:18	21.7
Spanish	15	00:52:21	20.9

3. Accent kernel labels

Our data can be used for various prosodic studies. We started from accent kernels.

3.1. Accent kernel

The accent kernel in Japanese is a sudden drop in pitch that has been claimed to be the primary phonetic correlate of Japanese word accent.

3.2. Hand labeling

Accent kernels are hand labeled by hearing the speech and marked on the transcription in terms of mora of Japanese by two experienced Japanese. The labelers are researchers of Japanese accent and teachers of Japanese for foreign students. The accent kernels are marked where the tone height is heard falling. Accents identified for the study are those that are very weak or those that are abnormally placedl. Two labelers worked independently, then they met to cross-validate each transcription. We use one of the labeler's normal labels consistently.

3.3. Prosodic analysis of the passages

Each passage contains, on average, 163 morae and 27 accent kernels, as shown in Table 2. That is, an accent appears, on average, in every six morae. Figure 1 shows the actual distribution of intervals between adjacent two accent kernels. The most frequent interval is five morae, however there are a very few long intervals. Taking the average duration of a passage given in Table 1 into account, the average mora rate is six including pauses between phrases, which means normal speaking rate (163.1/26.9=6.06, 163.1/27.4=5.95). In fact, since the semi-natural performances include more pauses between utterances than text reading, duration of each syllable is shorter in semi-natural performances, i.e. speaking faster.

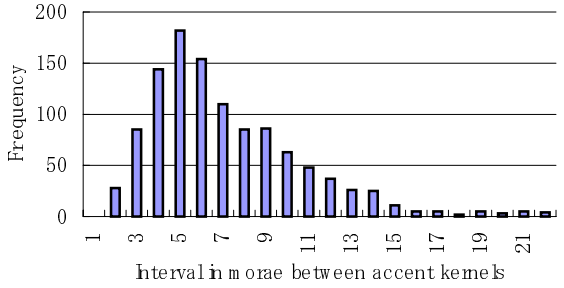

Figure 1. A histogram of intervals in morae.

Table 2: Statistics of mora interval between two accent kernels in the passages

	Morae	Accent kernels
Average	163.1	27.1
Standard deviation	5.17	0.91

3.4. Accent coincidence between speakers

Speakers pronounced the text correctly concerning accentuation, since they lived in Tokyo for a long time, moreover the recording director checked the misarticulations of accent and restarted the recordings. Therefore, as shown in Table 3, accent kernels are almost coincident between all speakers.

Table 3: Coincidence of accent between speakers.

Coincided speakers	occurrence	percent
6 (all)	830	74.9
5	116	10.4
4	52	4.7
3	43	3.9
2	17	1.5
1	50	4.5

3.5. Labeler's consistency

Two labelers are experienced professional teachers of Japanese. However, there are some differences between their identification of the accent kernels. They marked the accent kernels independently on the same speech corpus, and then they crosschecked each other and made modifications. There still remained a small number of differences, but most of the labels coincided.

3.6. Marking accent kernels

Figure 2 shows a screen we used to mark accent kernels; the software used is Wavesurfer [6]. We use a narrow-band spectrogram to find F0 movements. In the figure, accent kernels are marked along the transcription pane "acc" using a label character "@" to put it at the peaks of F0 contour.

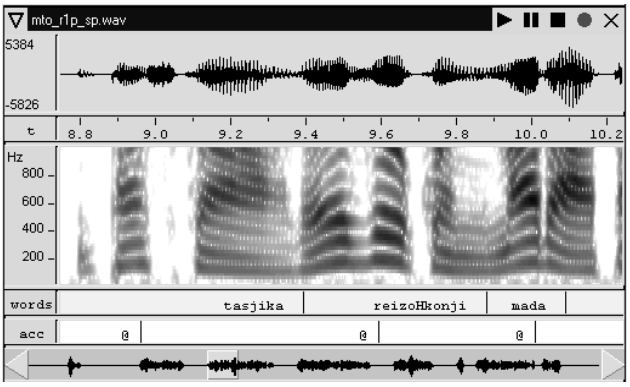

Figure 2: Accent labeling screen.

3.7. ANOVA of accent kernel intervals

3.7.1. Horizontal analysis: average per stream

Figure 3 shows the number of accent kernels (AK) in 40 passages. Each passage consists of several sentences containing about ten words. The number of accent kernels is proportional to the number of morae in a passage. Open circles in the figure show the ratio: the number of morae in the passage is divided by the number of accent kernels. The chart shows that these results are consistently distributed around six (morae/accent kernels). ANOVA of the average morae per accent kernel for each passage assures to be equal, moreover the average morae per sentence is also assured to be equal in all passages (probability with 0.61). Furthermore, average morae per a continuous phrase, separated by at least two short pauses, is also shown to be equal (probability with 0.97) in one speaker. This fact means that although the interval between AK deviate depending on the lexical and syntactic constraints, the probability of the AK occurrence is constant.

3.7.2. Vertical analysis: temporal structure

How do these accent kernels start, repeat, and finish within a sentence in terms of morae intervals between accent kernels?

3.7.2.1 Starting accent kernel intervals

The first position is defined as the first AK from the beginning of a sentence after a long pause or a phrase after a short pause. The number of morae before the first AK is 4.4 morae on average, and the most frequent occurrence (28%) is one (that is AK is placed at the initial morae). The second position is defined as the second AK from a pause. The number of morae between the first AK and the second AK is 6.0 in average, and the most frequent occurrence (15%) is five.

In a similar way, the third position shows 5.5 on average, with the most frequent occurrence (19%) being four. Short AK intervals are repeated with a periodicity of about five morae.

3.7.2.2 Ending accent kernel intervals

AK intervals before the end of a sentence or a short pause are investigated. The last position is defined as immediate morae before a pause until an AK. The number of morae is 3.0 in average, and the most frequent occurrence (28%) is one (that is AK is placed at two morae before the final pause). These correspond to typical (*/de'su/*, */ma'su/*) Japanese ending phrases. At the previous position, the average number of morae is 6.2, and the most frequent occurrence (14%) is four. At the third position, the average number of morae is 6.0; the most frequent occurrence (16%) is four.

Accordingly, AK temporal structure is symmetric. Longer intervals are observed around intermediate positions.

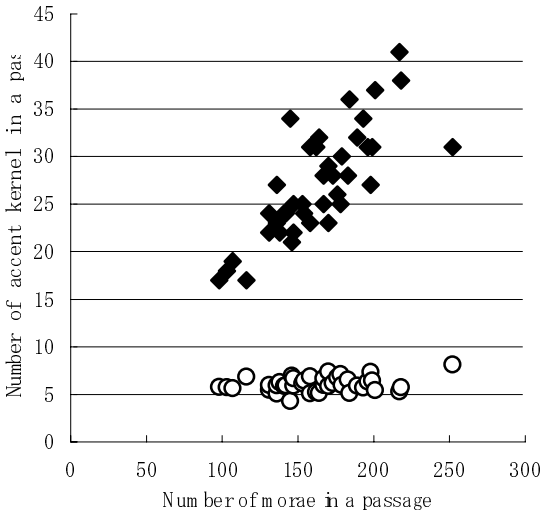

Figure 3: Accent kernel and morae in a passage.

3.8. Reduction of word accent in continuous speech

In continuous speech, and especially in fast speech, some accent kernels are skipped and intonation is flattened. Here we show an example. The lexical transcription considers accentual modification in compound words. Underlined portions are differences for the accent type x: <x>----</x>

The following example includes 5 differences: in each case the original word accents are made flattened due to 1st case a particle "/no/", 2nd case an auxiliary verb "/mitai/", 3rd and 4th cases an auxiliary verb "/masu/", and finally 5th case put between a particle and an auxiliary verb.

3.8.1. English passage: [7]

I have a problem with my water softener. The water-level is too high and the overflow keeps dripping. Could you arrange to send an engineer on Tuesday morning please? It's the only day I can manage this week. I'd be grateful if you could confirm the arrangement in writing.

3.8.2. *Lexical accent of a translated Japanese passage:*

<2>ie</2>no<3>zyoHsuikji</3>no<0>cyoHsji</0>ga<2>warui</2>desu. <0>suiacu</0>ga<0>takasugjiru</0>mitaide, <0>haisuikoH</0>kara<0>zuQto</0><0>suitekji</0>ga<1>tare</1>te<0>i</0>masu.<4>sumimaseN</4>ga, <2>kayoHbi</2>no<1>gogo</1>nji<3>naosji</3>nji<1>kji</1>te<0>morae</0>maseNka? <0>koNsyuH</0>wa<2>kayoHbi</2>sjika<0>cugoH</0>ga <2>cuke</2>rarenainodesu. <1>kji</1>te<0>kureru</0> <1>mae</1>nji, <5>neNnotame</5>nji<0>deNwa</0>o <0>sji</0>t <0>moraeru</0>to<4>arigatai</4>desu.

3.8.3. Realized accent kernel in continuous speech:

<0>ieno</0><3>zyoHsuikjino</3><0>cyoHsjiga</0><2>waruidesu</2>. <0>suiacuga</0><6>takasugjirumitaide</6>, <0>haisuikoHkara</0><0>zuQto</0><0>suitekjiga</0><1>tarete</1><2>imasu</2>. <4>sumimaseNga</4>, <2>kayoHbino</2><1>gogonji</1><3>naosjinji</3><1>kjite</1><5>moraemaseNka</5>? <0>koNsyuHwa</0><2>kayoHbisjika</2><0>cugoHga</0> <4>cukerarenainodesu</4>. <1>kjite</1><0>kureru</0> <1>maenji</1>, <5>neNnotamenji</5><0>deNwao</0> <0>sjite</0> <0>moraeruto</0> <4>arigataidesu</4>.

4. Discussion

4.1. Periodicity of accent kernel

We considered the periodicity of accent kernel intervals in terms of number or morae between accent kernels. This shows psychological periodic feeling in listening to continuous speech, if we could regard that each mora is heard to occupy constant duration. However, physical periodicity is not yet investigated in this paper. We have shown a bimoraic rhythm in Japanese with a spectrographic schema we call a TEMAX-gram [2]. Isosyllabism is not proven for Japanese. Is there isochronic structure in Japanese? Here we have shown the periodicity of accent kernels in terms of morae, a kind of isochronicity. Now we are labeling accent kernels onto our prosodic corpus. Whether accent kernels are physically periodic will be investigated in upcoming research. Of course, the word "periodic" does not mean exactly constant, but approximately constant with probabilistic deviation. The periodicity of vowels is the fundamental frequency (F0), however the exactly constant F0 vowel sounds non-human, mechanical, and unpleasant. Fluctuation in F0 frequency is essential for human-likeness. It seems that this is just as true for Japanese as it is for other languages.

4.2. Rhythmic structure of sentence

As we have discussed here, Japanese is a mora-timed language, and it may be an accent-timed language with an average interval between adjacent accent kernels of around six morae. Of course, actual intervals deviate probabilistically; therefore, the previous statement can be restated as "the expected interval between adjacent two accent kernels is 6 morae".

There may be a sort of rhythm in the deviation of intervals. We could observe some typical accent kernel patterns of starting and ending of a continuous utterance after and before a short or a long pause.

5. Conclusions

We have investigated the periodicity of accent kernels in Japanese concerning our prosodic corpus, a Japanese MULTEXT composed of 3 hours 37 minutes of speech and involving 6 different speakers. The average interval between two adjacent kernels is about six. ANOVA assured this average is equal over different speakers, different passages, and different sentences in the corpus. We could observe the accent kernel occurrence patterns of starting and ending of a continuous utterance after and before a short or a long pause.

6. Acknowledgements

We are grateful to the Ministry of Education, Culture, Sports, Science and Technology for supporting this research through the Grants-in-Aid for Scientific Research, project number 12132204, during 2000-2001. We would also like to thank Mr. Kawatsu Motoi and Ms. Yanagisawa Emi for listening to the recording and marking accent kernels, and also the students who edited digitized signals. The accent type analysis in section 3.8 was conducted by Mr. Yamagishi Toyohide.

7. References

[1] Kubozono, H., 1993. *The Organization of Japanese Prosody*, Kurosio Publishers, Tokyo.

[2] Kitazawa, S.; Ichikawa, H.; Kobayashi, S.; Nishinuma, Y., 1977. Extraction and Representation Rhythmic Components of Spontaneous Speech. *EUROSPEECH 1977*, Greece, 641-644.

[3] Kitazawa S., 2001. Preparation of a Japanese Prosodic Database. *The ELRA Newsletter*, Vol. 6 no. 2, 4-6.

[4] Kitazawa S.; Kitamura T.; Mochiduki K.; Itoh T., 2001. Preliminary Study of Japanese MULTEXT: a Prosodic Corpus. *International Conference on Speech Processing*, Taejon, Korea, 825-828.

[5] Campione, E.; & Véronis, J., 1998. A multilingual prosodic database. *5th International Conference on Spoken Language Processing (ICSLP'98).* Sidney 3163-3166.

[6] Sjölander K.; Beskow J., 2000. WAVESURFER - AN OPEN SOURCE SPEECH TOOL. *ICSLP'2000*. paper number 01557, Beijing.

[7] 1998. Passage O2. *MULTEXT Prosodic Database –English-*, Version 1.0, Disk 1 of 5, MULTEXT Project, LRE 62-050, ELRA/ELDA.