Perceptual Inspection of V-V Juncture in Japanese

Kitazawa Shigeyoshi Kiriyama Shinya, Itoh Toshihiko Yukinori Toyama

Department of Computer Science
Faculty of Information, Shizuoka University
{Kitazawa, kiriyama, t-itoh, cs8064}@cs.inf.shizuoka.ac.jp

Abstract

We examined the subject of phrase boundary determined through evaluation of disjuncture in a Japanese prosodic database. In normal fluent speech, not only word boundaries but also phrase boundaries are obscured. Such phenomena are called internal open junctures, i.e. boundaries between phrases without pause, which is one of four aspects of prosody. We investigated V-V juncture through J-ToBI labeling and listening to whole phrases to estimate degree of discontinuity and to determine the exact boundary between two phrases if possible. Different levels of discontinuities were found in various levels of junctures of phrases. Appropriate boundaries were found in most cases including some overlaps. The test materials are taken from the "Japanese MULTEXT", containing read and spontaneous speech by three male speakers and three female speakers in Tokyo dialect.

1. Introduction

This paper presents results of a study concerning the "juncture" that determines the boundary between morphological units, i.e., words and phrases in a Japanese sentence. There are two types of juncture: external open and internal open [1]. External open juncture is defined by phonological features of the segmental phonemes and supra-segmental phonemes at the beginning and ending of a separated utterance. However, similar phenomena occurs inside a sentence. Here we investigate the type of internal open juncture that happens at the phrasal boundary in an utterance. The junctures we examine here are acoustic phonetic phenomena of two adjacent phrases comprising a transition between a final mora of a preceding accentual phrase and an initial mora of the succeeding accentual phrase.

Beckman points out that phrasing and pitch range are the most important features of Japanese prosody [2]. Since the phrase-internal tone is determined lexically, the phrasal intonation is poor compared to English. Therefore, juncture, the fourth component of prosody, is just as important as rhythm, intonation, and stress in Japanese.

In normal fluent speech, phrase as well as word boundaries become obscure because of fluency and become difficult to segment. This is perhaps the salient problem of speech recognition and speech synthesis. Marks such as juncture, punctuation, focus, prominence in a stream of speech sound are crucial for effective listening comprehension.

J-ToBI, a prosody annotation scheme, defines vaguely the juncture as BI label with 5 different degrees as perceived disjuncture [3]. We tried to measure this ambiguous disjuncture quantitatively through a series of perceptual experiments.

2. Prosody data base

Phonetic prosodic labeling is performed on voice data collected for Japanese prosody database.

2.1. Japanese MULTEXT prosody corpus [4]

The Japanese version of MULTEXT (multi-language prosody corpus) is created by the specification of EUROM1 [5]. It aims at recording same-content speech consisting of 40 small paragraphs, then the extraction of prosody parameter, and the prosody notation of five languages.

Speakers are between 20 to 40 years, in a total of six persons (three men and three women), all native speakers of the Tokyo dialect. A text is given for a reading and to evoke a simulated spontaneous utterance. A speaker practices and takes natural pauses fluently. Incorrect utterances and accent errors are corrected and re-recorded at the beginning of each paragraph.

2.2. Labeling

After automatic F0 extraction, F0 contour was edited by hand. Phoneme segmentation by hand-eye is good, but still it is difficult even for the expert to segment tough problems such as those arising when the same two vowels connect but do not compose a long vowel. Those difficult cases were conventionally treated at the mid point concerning equality of morae duration, in which correctness is not necessary [6].

J-ToBI labeling is applied for prosodic annotation according to [3]. Although, the X-JToBI extended the J-ToBI to spontaneities of speech, e.g. descriptions of fillers and disfluencies, it does not enhanced in description of VV junctures. So the J-ToBI is sufficient for our prepared speech.

3. Method of perceptual test

The prosodic phrases to be labeled were segmented with reference to the speech waveform and the spectrogram of wide-band and narrow-band, and then checked by listening to the speech segment.

3.1. Segmental analysis of phrase juncture

The juncture we treat is a boundary between adjacent accentual phrases in Japanese. We are going to decide a juncture boundary on a segmental sound level. When a boundary consists of VCV, in Japanese, many cases can determine obviously as a boundary of V and CV. However, in vowel continuation boundaries such as CVVC, it is difficult to determine clearly the boundary between two adjacent vowels. Furthermore, if this VV juncture consists of the same kind of vowels without any pause, boundary decision becomes much more difficult.

Actual speech data were taken from the Japanese MULTEXT prosodic corpus specifically spoken by a female speaker fhk. The examined phrases consist of the following 5 phrases taken on vowel junctures of /a/-/a/, /i/-/i/, /u/-/u/, /e/-/e/, /o/-/o/. There is no gap between these two vowels.

3.2. Preparation of speech materials

In order to investigate deviations of VV segment boundary junctures, the following short speech waveforms are prepared. Concerning the boundaries in reference to the hand labeled segment boundary as a fixed point, a front phrase and a rear phrase are separated and excised for speech materials in a perceptual experiment. The separation points are moved forward and backward from the fixed point with a step width of one vocal cord vibration period up to 5 periods. As a result, it amounted to 11 speech sounds of each side for a total of 22 speech sounds per juncture.

3.3. Phrase listening

Speech sounds are presented in random order for each subject. Each subject was asked to judge the naturalness of each phrase sound, paying special attention to the ending and beginning. Responses were scored on a scale from 5 to 0, with 5 points awarded for natural speech, and 0 for utterances appearing completely unnatural. Each answer is scored from +2, +1, 0, -1, -2 accordingly. Subjects' answers are summed and averaged for individual speech materials. The listeners participating in the perceptual experiments were 6 male students and 2 female students.

4. Results of perceptual judgment of V-V juncture

Mary E. Beckman and Janet B. Pierrehnmbert set three levels of prosodic phrasing marked by f0 features [2]. They call these three types of phrases the *accentual phrase,* the *intermediate phrase*, and the *utterance.* We put more details in lower levels. The lowest level, the accentual phrase, is a phrasal unit containing at most one accent. This unit may be a single word. However, when words are combined into sentences, it is quite usual for some to lose their status as separate accentual phrases. Noun-noun compounds typically form a single accentual phrase, as do adjective-noun sequences.

4.1. Within word juncture

The boundary of two same-vowels is usually pronounced as a long vowel: *ok**aa**saN* "mother", *ocj**ii**Qte* "fall in", ***ii**masu* "say". This level corresponds to BI=0 in J-ToBI. This level was skipped since junctures are difficult to hear as a phrase.

4.2. Word and postposition and particles juncture

A Japanese word accompanies a postposition to compose a minor phrase in a sentence. A word boundary is marked as BI=1 in J-ToBI. Concerning these boundaries, there are cases where two same vowels continue. Native Japanese can notice disjuncture between these two vowels. It is interesting to see what acoustic features exist around this boundary. The followings are examples of this sort of junctures: *oto-o* "sound-ACC", *neko-o* "cat-ACC", *ono-o* "ax-ACC", *koto-o* "matter-ACC", *tokoro-o* "place-ACC". This level was also skipped, since the succeeding phrases of junctures are difficult to hear as a phrase.

4.3. Word and word juncture within a complex word

More than two words compose a complex word. The word boundary is weakened than two separate words. The followings are the examples of this sort of juncture: *sita ato* "done after", *dai ici* "number one", *komugi iro* "light brown".

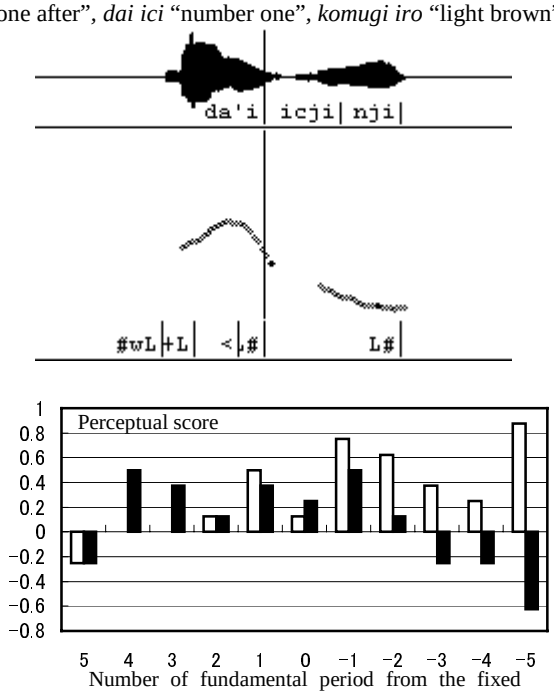

Figure 1. *Illustrated ii juncture where the fixed point is shown as a vertical line on a waveform and F0 curve with the J-ToBI tone transcription overlaid (upper) and perceptual results (lower); white bars are front phrase and black bars are rear phrase.*

Figure 1 shows words on a waveform (top), an F0 plot overlaid with J-ToBI tone transcription and perceptual results in a bar graph for a phrase *da'iicjinji*: this phrase consists of three words *dai* (number), *icji* (one), *nji* (–LOC). They normally compose two accentual phrases: *da'i icji'nji*, however, in fluent speech, these change into one phrase. Perceptual test showed maximum disjunction of front phrase at –1 with score 0.75 (real maximum is 0.88 at –5 but this is not really appropriate since the rear phrase is degraded as low as –0.6.) and also at –1 in rear phrase with score 0.5.

Figures 2 through 5 show the same structure with the identical scale. The range of F0 plots is 100 Hz at the bottom and 400 Hz at the top line.

4.4. Accentual phrase juncture within an intermediate phrase

This is a juncture between two accentual phrases in an intermediate phrase where phrase ending vowel and phrase initial vowel are the same. There is acoustic change around the boundary such as F0 lowering, glottalization and reduced co-articulation.

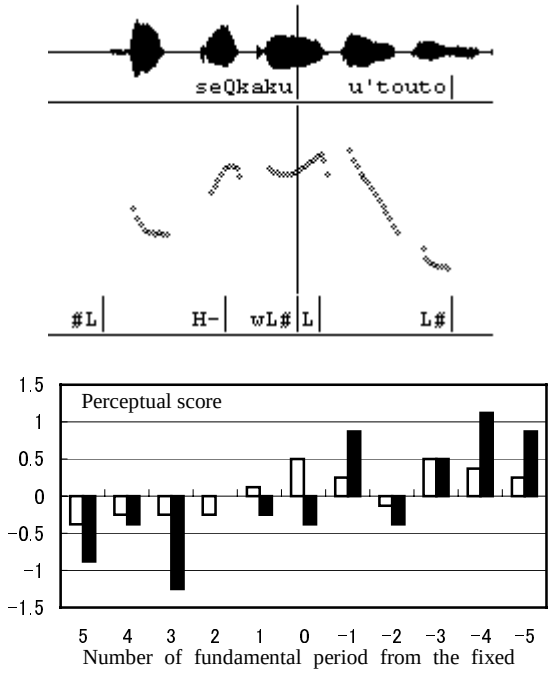

Figure 2. *An uu juncture. Scales are the same as Figure 1.*

Figure 2 shows two accentual phrases are connected together along a strong intonation curve to emphasize the intermediate phrase *seQkaku-u'touto* "precious doze off (was broken)". Perceptual test showed maximum disjunction of front phrase at 0 with score 0.5, and at –1 in rear phrase with score 0.88, however, disjunction is weakly marked as BI=2-.

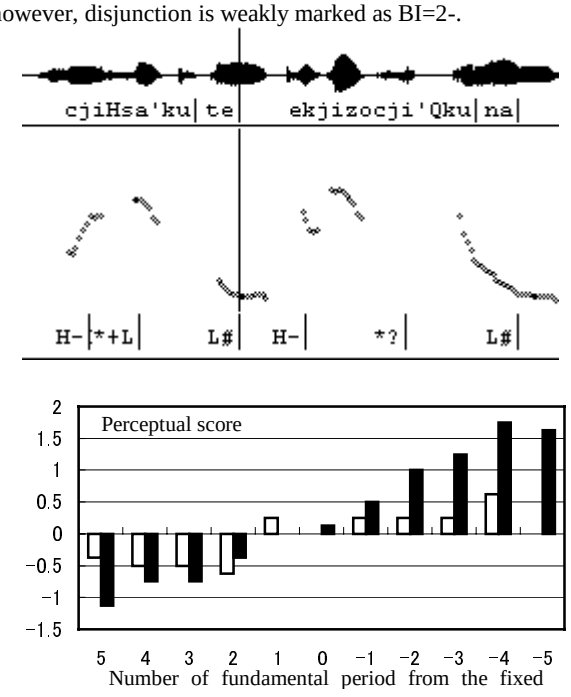

Figure 3. *An ee juncture. Scales are the same as Figure 1.*

Figure 3 shows two accented adjective phrases *cjiHsa'kute* "small" *ekjizocji'Qkuna* "exotic". The best point is the fourth period after the fixed point at score 0.624 for the front phrase, 1.75 for the rear phrase in both phrases achieving the maximum of the goodness scores. This shows clear lengthening of phrasal ending vowel and shortening of phrasal beginning vowel. In this case, a "pitch reset" is observed.

4.5. Intermediate phrase juncture

An intermediate phrase is composed as a chanking of several (only rarely more than three) accentual phrases. An intermediate phrase boundary is often marked by a pause or *pseudo-pause* (a pre-pausal "winding down" of production speeds unaccompanied by any actual momentary cessation of production). Also, the L% boundary tone for the last accentual phrase in an intermediate phrase is markedly lower than at a medial accentual phrase boundary. F0 declining characteristic of the intermediate phrase, however, is known as *catathesis*. We sometimes call this a *pitch-reset*.

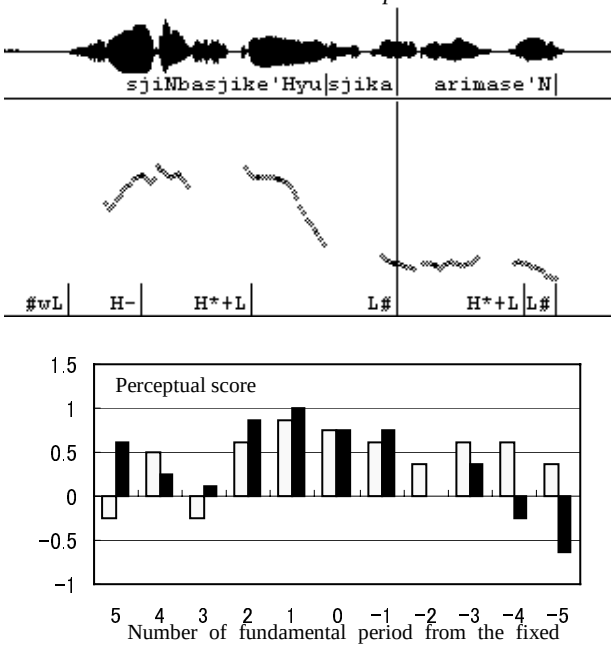

Figure 4. *An aa juncture. Scales are the same as Figure 1.*

Figure 4 shows scores concerning an intermediate phrase juncture; *sjiNbasji ke'Hyu sjika* "only via Shinbashi" and *arimase'N* "there is no other way". The former phrase composed with three accentual phrases, i.e. *sji'Nbasji keHyu sjika*, an intermediate phrase and then indicating a focus on the first accentual phrase *sji'Nbasji*. The following phrase *arimase'N* is a predicate part of an utterance ending.

The preceding phrase is best heard at the one period before the fixed point, and the succeeding phrase is also best at the one period before the fixed point. Both preceding and succeeding phrases coincided at this point. This means that the hand labeled point should be moved to left at one F0 period, then two adjacent phrases are best separated with the perceptual score 0.88 for preceding phrase and 1.0 for succeeding phrase. This disjuncture is marked as high BI as 3-.

In this case, the succeeding phrase is de-emphasized, resulting in a narrow pitch range. Therefore, no accentual peak is observed and the pitch contour is horizontal. We can point out an elbow on the F0 catathesis at this point. As shown in the figure, these two phrases are acoustically

connected into a continuous vowel /a/, but perceptually well separated. There is a prosodic boundary, i.e., a juncture.

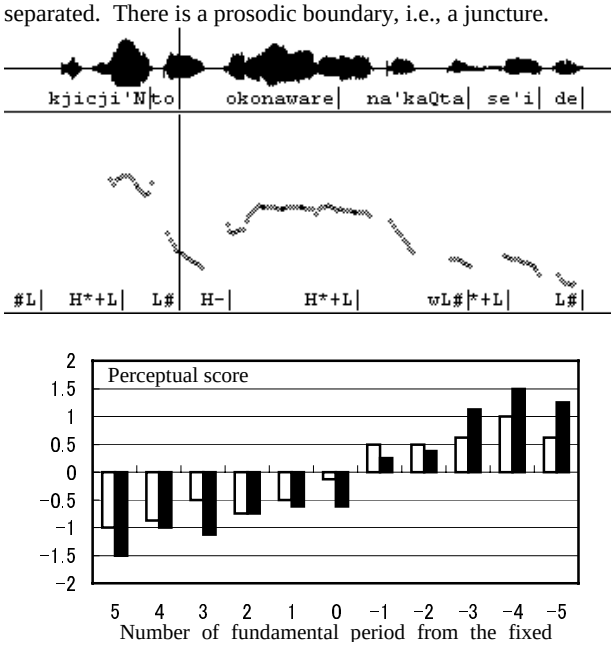

Figure 5. *An oo juncture. Scales are the same as Figure 1.*

Figure 5 shows an intermediate phrase where an accent phrase *kjicji'Nto* "accurately" is focused and emphasized, while *okonawarena'kaQta se'ide* "due to it was not carried out" is a suppressed intermediate phrase. A pitch-reset is observable in the succeeding phrase, and the pitch-range is reduced. Perceptual results show that there is a clear juncture at –4 period with scores 1.0 for the front phrase and 1.5 for the rear phrase. Lengthening at the phrase ending as well as shortening at the phrase start up was observed.

4.6. Utterance boundary

There is a pause at the end of an utterance. Therefore the boundary is clear.

5. Discussion

5.1. Levels of disjuncture

Perceptual score achieved indicate degrees of disjuncture between phrases. An ee juncture showed the best score of 1.8 in the succeeding phrase due to pitch reset. Also an aa juncture achieved good score of 1.5 in the succeeding phrase due to pitch reset. Those scoring less than 1— uu, ii, and aa – are poor disjunctures, i.e. combined or emphasized words with reduced pitch range.

5.2. Shapes of perceptual scores

As the boundary moves from the fixed point, perceptual score changes. If the score changes monotonously, the optimal boundary is the maximum score point. If the score has multiple peaks, the optimal point should be chosen by coincidence of the boundary point, with the preceding phrase and the succeeding phrase taken into account. Even if perceptual score is high for either phrase, this is not sufficient to determine the optimal boundary. Of course there are cases in which preceding phrase and succeeding phrase overlap each other or loosely connected.

5.3. Forward and backward balance

The optimal perceptual scores do not necessary balance, but rather one is high and the other is low. Our results show that preceding phrases are lower than succeeding phrases, with the exception of the ii juncture (Figure 1, this juncture resembles a strong cohesion).

5.4. BI relationship

Break indices indicate the degree of prosodic association between words and phrases. They are subjective values – perceived disjunctures between phrases. The perceptual score obtained here is directly related to the degree of disjuncture, hence the break indices.

6. Conclusion

J-ToBI labeled phrase boundaries are examined through perceptual evaluation of disjuncture. We investigated V-V juncture by listening to whole phrases to estimate degree of discontinuity and to determine exact phrase boundary. Different levels of discontinuity were found in phrases ranging between words, accentual phrases, and intermediate phrases. Disjuncture is affected from emphasis, focus, and pitch reset. Appropriate boundaries were found in most cases, including some overlaps.

7. Acknowledgements

This research is based on the domain research specific (B) (2) subject number 12132204.

8. References

[1] Lehiste, I., 1960. An acoustic-phonetic study of internal open juncture. Supplement to *Phonetica*, 5, 1-54.

[2] Beckman, M. and J. Pierrehumbert, 1986. Japanese Prosodic Phrasing and Intonation Synthesis. Proceedings of the 24th Meeting of the Association for Computational Linguistics 173-180.

[3] Venditti, Jennifer J., 2002. The J_ToBI model of Japanese intonation. In S.-A. Jun (ed.) *Prosodic Typology and Transcription: A Unified Approach*. Oxford: Oxford University Press.

[4] Kitazawa Shigeyoshi, Kitamura Tatsuya, Mochiduki Kazuya, and Itoh Toshihiko, 2001. Preliminary Study of Japanese MULTEXT: a Prosodic Corpus. International Conference on Speech Processing, Taejon, Korea, 825-828.

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

[6] Shigeyoshi Kitazawa, Itoh Toshihiko, and Kitamura Tatsuya, 2002. Juncture segmentation of Japanese prosodic unit based on the spectrographic features. Proceedings of 7th International Conference on Spoken Language Processing, Denver, USA. 1185-1188.