Acoustic and Prosodic Analysis of Japanese Vowel-Vowel Hiatus with Laryngeal Effect

Kitazawa Shigeyoshi   Kiriyama Shinya
Department of Computer Science
Faculty of Information, Shizuoka University
{Kitazawa, kiriyama }@cs.inf.shizuoka.ac.jp

Abstract
We investigated V-V hiatus 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. Appropriate boundaries were found in most cases as the maximum perceptual score. Using electroglottography (EGG) and spectrogram, the acoustic phonological feature of these V-V hiatus was phrase-initial glottalization and phrase-final nasalization observable in EGG and spectrogram, as well as phrase-final lengthening and phrase-initial shortening. The test materials are taken from the "Japanese MULTTEXT", consisting of a particle - vowel (36), adjective - vowel (5), and word - word (4).

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

This paper presents results of a study concerning the boundary between morphological units, i.e., words and phrases in a Japanese sentence. Here we investigate the phrasal boundary in an utterance comprising a transition between a final mora of a preceding accentual phrase and an initial mora of the succeeding accentual phrase consisting of the same two vowels, i.e., vowel-vowel hiatus.

J-ToBI, a prosody annotation scheme, defines vaguely the phrase structure as BI label with 5 different degrees as perceived disjuncture [1]. We tried to measure this ambiguous disjuncture quantitatively through a series of perceptual experiments. Results were also investigated using EGG analyzed data (open quotient), F0, speech waveform, and spectrogram.

2. Vowel-Vowel Hiatus in Japanese
Since Japanese consists of open syllables, if the following phrase begins with an initial vowel, a vowel-vowel (V-V) hiatus arises, the same vowel continues without pause. This vowel sequence is very common in Japanese.

2.1. Morphonology of vowel-vowel hiatus
Example hiatus taken from our corpus (3.1) are as follows.

The most common phrasal unit is a morpheme (e.g. a noun) + a particle which bears an accent to compose a particle (joshi) - vowel initial phrase. Examples are ga - aru, wa - ame,

sika - arimaseN, ni - iQtet, te - ekizo, to - omou, wo - osiite, no - otaka and some others.

The second type is an adjectives (fukushi) - vowel initial phrase. Examples are Mada - atarasisi, iQtet - itu, mosi - ikite, seQkaku - utouto, kitinNo - okonau and some others.

The third less frequent type is word - word, such as, komugi - iro, takasii - ichidai.

2.2. Phonological realization of Japanese hiatus
There are number of possible factors that helps perception of Japanese V-V hiatus.

2.2.1. Phrase-initial glottalization
Glottalization of word-initial vowel is a common phenomenon of world languages [2]. It is more strongly pronounced if the word has a stress or accent at the beginning of the word.

2.2.2. Phrase-final nasalization
Voiced velar consonant is nasalized at the non-word-initial position in Tokyo Japanese. This nasalization contrasts with the following word-initial vowel that should not be nasalized. This sort of hiatus occurs very often since a noun phrase consisting of a noun + a particle ga is very common in Japanese, and this phrase can be followed by a predicate aru for example, composing a a-n hiatus.

2.2.3. Lengthening and shortening
Phrase-initial syllable or mora is shortened, while phrase-final syllable or mora is lengthened. This mora timing is a built in rhythm of Japanese as well as other languages. Duration of the concatenated vowel might be segmented with a built in timer of the perception mechanism. The mechanism will help the human hearing to resolve the hiatus.

2.2.4. Morphonological constraints
The part of speech plays some role in realization of the hiatus. Vowel sequence at the phrase boundary often occurs in the environments stated in 2.1. This constraints help to resolve the hiatus.

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

The Japanese version of MULTTEXT (multi-language prosody corpus) is created by the specification of EUROML [4]. It aims at recording same-content of speech consisting of 40
small paragraphs, then the extraction of prosody parameter, and the prosody notation of five languages.

Speakers are native speakers of the Tokyo dialect. A text is given for a reading and to evoke a simulated spontaneous utterance. Speech was recorded with apparatus based on the specifications of EUROM1, in an anechoic chamber, using a B&K 1/2 capacitor microphone, a DAT recorder (SONY PCM2300). In addition, electroglottograph is recorded with an EGG (KAY (Co.) 4338) from which F0 and open quotient are extracted.

3.2. Phonetic and prosodic labeling

Phoneme segmentation by hand-eye is good, but still is difficult to segment when the same two vowels connect. Those cases were conventionally marked at the midpoint to achieve equality of morae duration [5].

J-ToBI labeling is applied for prosodic annotation according to the manual [1]. Although, the X-JToBI extended the J-ToBI in spontaneities of speech, e.g., descriptions of fillers and disfluencies, it does not describe V-V hiatus. J-ToBI is sufficient for our prepared speech.

4. Method of hiatus analysis

The prosodic boundary of phrases was segmented with reference to the waveform (speech and EGG) and the spectrogram of wide-band and narrow-band, and then evaluated by listening to the separated accentual phrases.

4.1. Perceptual analysis of phrase

The hiatus we treat is a V-V boundary between adjacent accentual phrases in Japanese. Samples were taken from the Japanese MULTEXT prosodic corpus spoken by a female speaker fhk. The examined phrases consist of 45 phrases producing hiatus of /ai/-/aI , /i/-/iL , /u/-/uI , /e/-/eI , /o/-/oI . There is no gap between these two vowels.

4.1.1. Preparation of speech materials

In order to investigate deviations of V-V segment boundary, the following short speech waveforms are prepared. Referring to the hand labeled boundary as a fixed point, a front phrase and a rear phrase are separated and excised for speech materials in a perceptual experiment. The exciting points are moved forward and backward from the fixed point with a step width of one vocal cord vibration period up to 5 periods (vertical lines in Figure 1). As a result, it amounted to 11 sound fragments for each side, to a total of 22 speech sounds per hiatus.

4.1.2. Phrase listening [5]

Speech sounds are presented in random order to each subject. Subjects were asked to judge the naturalness of each phrase sound, paying special attention to the ending and beginning (the phrase in either side is shown over the waveform at the top of the Figure 1). 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.

4.2. Electroglottography waveform analysis

Electroglottography waveforms were analyzed for the open quotient (as shown in the bottom tier of the Figure 1) and the fundamental frequency (the middle tier in Figure 1) computed from the each glottal cycle using the KAY CSL tool [6]. The open quotient is related to the voice quality, i.e., over 50% is harsh voice, 50% is modal voice, and 20-30% is breathy voice. The quotient changes smoothly along time, but abrupt change can be an evidence of glottalization. The fundamental frequency extracted from EGG is unlike conventional F0 that are smoothed by a filter but drops simultaneously with glottalization as well.

5. Analysis results of vowel-vowel hiatus

Hiatus resolution is possible based on glottalization in most cases and nasalization in some cases. And the resultant lengthening of phrase-final vowel and shortening phrase-initial vowel is common.

5.1. Hiatus observed in phrasal tone

An intermediate phrase is composed as a chunking of several (only rarely more than three) accentual phrases. An intermediate phrase boundary is often marked by a pause or pseudo-pause. Also, F0 declining characteristic of the intermediate phrase, however, is known as catahesis. We sometimes call this a pitch-reset.

Figure 1 shows an accent phrase kiciji’Nto “accurately” is focused and emphasized, while okonawarena’kaQta seede “due to it was not carried out” is suppressed. A pitch-reset is observable in the succeeding phrase, and the pitch-range is reduced. Perceptual results show that there is a clear hiatus at 4 period with scores 1.0 for the front phrase and 1.5 for the rear phrase. An abrupt F0 step down and a dip in open quotient indicate glottalization. Lengthening at the phrase ending as well as shortening at the phrase start up was observed.

5.2. Phrase-initial glottalization

In most cases, the phrase-initial vowel is stressed on its phonation by glottalization. This is also true in case the preceding phrase-endning vowel is the same as the following phrase-initial vowel. An example is shown in Figure 1 in a lower part, the EGG open quotient once goes up to 80% then decreases down to 45% and then goes up to 60%. Before this bottom (10.43s) is the most appropriate point to separate the phrases. Simultaneously the F0, which is computed from the EGG period, showed lowering: 235 Hz to 142 Hz then 178 Hz movements.

Similar phenomena were observed in other cases in different degree of prominence. Among 45 hiatus analyzed, 17 showed clear glottalization with dip in F0 and open quotient, and 23 showed weak glottalization accompanied with other features, and remaining 5 showed phrase-final nasalization. If the phrase-initial word is emphasized, the boundary showed prominent glottalization, while a particle ga - a-initial phrase and a particle wo - o-initial phrase depict rather vague features of glottalization if the following phrase is not emphasized. In those cases, the acoustic features are too small in the magnitude to detect visually. Syntactic condition may help to resolve hiatus and probably human perception of glottalization must be far more sensitive than present signal processing.
5.3. Phrase-final nasalization

Switching from nasalized to non-nasalized may sign with perceptual cue, and indicates a segmental boundary by spectrogram texture as relatively lower high frequency energy for nasalized speech. The nasalization contrast is observable even in a continuation of the same vowel. A phrase-final particle ga have to be nasalized in Tokyo Japanese, though the speaker fkh, as a younger generation in the drift of phonological change, tend to pronounce these ga with non-nasals or at least weakly nasalized.

Nasalization is helpful to detect hiatus since the preceding phrase-final nasalization ends and the following phrase starts without nasalization. This +/-nasal contrast as well as glottalization resolves many hiatus (in our example, +/-nasal contrast alone was 4 cases and 12 cases accompanied with glottalization as well as nasal contrast).

5.4. BI and perceptual score relationship

This is a boundary between two accentual phrases, where the phrase ending vowel and the phrase initial vowel are the same, which is usually marked as BI = 2 but deviates a bit.

Materials analyzed here are boundaries between phrases labeled in BI (break index) as 2 or 2-, 2m, 3 according to perceptual degree of disjunction and intonation curves. Indices are followed to labeler’s subjective impression; therefore they do not necessarily agree with the syntax.

![Figure 2. Break Index and mean perceptual scores at the best boundary point with the standard deviation.](image)

Break indices indicate the degree of prosodic association between words and phrases. They are subjective values – perceived disjunction between phrases. The perceptual score obtained here is also related to the degree of disjunction, hence the break indices. We have shown a BI tier and a tone tier in an example below the F0 tier as well as a phoneme tier in Figure 1. A correlation between BI and perceptual score (4.1.2) is shown in Figure 2. Although the number of samples is too small to generalize the results, there is a trend that the perceptual score increases as the break index increases toward more disjunction.

5.5. Segmental duration analysis

A phrase ending mora is lengthened to indicate the end of a phrase, while a phrase initiating mora is shortened in order to catch up with the isosyllabic mora timing. A V-V sequence of the same vowels has duration of about two morae, however, the boundary is usually searched in the right half region.

Statistics of our 45 hiatus, the ratio of duration of vowel segment in the preceding phrase-final position to the following phrase-initial position was 1.7 in average. In cases the following word is emphasized, the preceding vowel is not lengthened, but the following vowel keeps normal duration, then the ratio becomes as low as 0.76 for example.

6. Conclusion

J-ToBI labeled phrase boundaries are examined through perceptual evaluation of disjunction, i.e. tidiness or flawless perfection. We investigated V-V hiatus by listening to whole phrases. The best perceptual score was obtained in most cases as the maximum perceptual score of a single peak:

A phrase-final particle - a vowel, the most common pattern of V-V hiatus, was found to have the following acoustic features: (1) ga - a, nji - t, no - o: these show +/-nasal contrast in the spectrographic pattern, since ga is normally nasalized while the following a is not nasalized. (2) wa - a, sjika - a, te - e, to - o: phrase initial vowel is glottalized. This glottalization is observable in F0 drop and a dip in EGG open quotient. (3) wo - o: another frequent pattern, glottalization is not so distinct since EGG open quotient is not stable, but spectral change is also useful.

A phrase-final adjective - a vowel and a word ending a vowel and a word beginning a vowel, those cases are characterized with stronger glottalization than above cases.

Phrase-initial glottalization observable in EGG open quotient, F0 or period of each cycle, and phrase ending nasalization are important to explain the hiatus phenomena.

Break indices (BI) have shown to correlate with perceptual scores. Duration of vowels of hiatus depends on mutual emphasis of the phrases adjacent.

7. Acknowledgements

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

8. References

Figure 1. Illustrated oo hiatus where the fixed point is shown as a vertical line on a waveform and F0 curve with the J-ToBI transcription overlaid (upper) and perceptual results (lower); white bars are front phrase and black bars are rear phrase. This phrase consists of three phrases kicjiNto (accurately (adj.)), okonawarenaQta (was not done (predicate)), seede (owing to).