Tonal alignment in Tokyo Japanese

Takeshi Ishihara

Department of Theoretical and Applied Linguistics
University of Edinburgh
takeshi@ling.ed.ac.uk

Abstract

This paper reports part of the results of a large-scale study whose aim is to investigate which aspects of the accentual-phrase-initial F0 rise in Tokyo Japanese are invariant and which are variable, focusing on alignment. Acoustic data were collected with materials varying in segmental composition, accent location and within-utterance position, under different experimental conditions. I present results of normal spoken data which showed consistent alignment of the F0 peak at the beginning of the initial-accented accentual phrase, with a fixed point in the segmental string. It appears that phonological structure plays a significant role in tonal alignment in Tokyo Japanese.

1 Introduction

Pitch features such as tone and accent are bound to specific elements of the segmental string. It is generally assumed that Chinese tones are associated with syllables, English pitch accents with stressed syllables, Japanese word accents with specific syllables, and so on. This relationship is known as *association* in the autosegmental description. Recent experimental studies have shown that the phonetic realisation of such association does not simply follow this relationship, and that the temporal synchronisation (*alignment*) of F0 events with segmental events may be affected by various factors.

A well-known phenomenon of the mismatch between association and alignment is known as *peak delay*. 'Peak delay' refers to the following phenomenon: the F0 peak of a pitch accent may be aligned outside the stressed/accented syllable with which it is intuitively associated. It is widely reported across languages. Factors affecting peak delay such as adjacent prosodic boundaries and proximity of pitch features are demonstrated in several languages, though how and to what extent these factors act on the alignment of pitch features varies from language to language (Silverman and Pierrehumbert 1990; Caspers and van Heuven 1993; Prieto *et al.* 1995; Arvaniti *et al.* 1998).

More recently, further evidence has been provided concerning the temporal alignment of pitch features (Arvaniti *et al.* 1998; Ladd *et al.* 1999, 2000; Xu 1998). Overall, it reveals that when the factors affecting pitch accent alignment are properly controlled, both the F0 maxima and minima for pitch accents are consistently aligned with fixed points in the segmental string. Based on the evidence, it is claimed that the temporal alignment and F0 level of tonal targets for pitch accents are determinants for the resultant shapes of pitch accents, which is a logical consequence of the autosegmental-metrical theory of tone and intonation (e.g. Ladd 1996).

An equivalent phenomenon to peak delay in Japanese is called *ososagari* ('late fall') in which the beginning of F0 fall (i.e. the F0 peak) for a pitch accent occurs after the end of the associated mora (Sugito 1982). While ososagari has been known for some time (Neustupný 1966), there are only a few studies on factors relating to it.[1] Among the few, Sugito (1981) found that ososagari tended to occur in the initial-accented words whose second mora has a non-high vowel.

Considering the limited data about factors affecting ososagari, as well as the recent evidence for consistent tonal alignment across languages, it is still unclear what kind of factors affect the timing of pitch features in Tokyo Japanese, and whether they are invariably aligned with segmental events.

2 Method

Tokyo Japanese has an F0 rise at the beginning of the accentual phrase, and the shape of the F0 rise interacts with accent type and phonological weight (Pierrehumbert and Beckman 1988). This paper reports preliminary findings from a large-scale study whose aim is to investigate which aspects of the F0 rise are invariant and which are variable, focusing on the alignment. To this end, acoustic data were collected with materials varying in segmental make-up, accent location and within-utterance position, under different experimental conditions. The results presented in this paper are from part of normal spoken data. More specifically, they are about the alignment of the F0 peak at the beginning of words with accent on the first syllable. Further analysis of the rest of the data is in progress and will be reported in due course.

Fifteen groups of test words were prepared to vary in accent type and initial segmental composition. Each group had five words. In terms of accent type, test words were either unaccented, or (first-, second-, third- or fourth-syllable) accented. In terms of segmental composition, only unaccented and first-syllable-accented words had six different types of initial segmental make-up (Table 1), because of the limitations of collecting possible real words. Test words were systematically collected to contain only vowels and sonorants in target sequences for the F0 tracking, and to balance as much as possible the effects of intrinsic segmental properties on the F0 in each group. All the test words were put in two types of the carrier sentences: *X-ga kaitearimasu.* ('X is written.'); *Sokoni X-ga kaitearimasu.* ('X is written there.').

The data reported in this paper came from the recording of five native speakers of Tokyo Japanese (two males and three females). None of the five had any speaking/hearing difficulties. The recording was conducted in a sound-treated studio at the Phonetics Laboratory of Sophia University in Tokyo.

[1] The relationship between F0 peak location and accent perception has been thoroughly studied (Sugito 1982; Hasegawa and Hata 1992)

	Unaccented	Initial-accented
CV+CV	/ne+mimi/	/mi+nari/
CVCV	/mimi+nari/	/namida/
CVR	/mee+moku/	/nuudoru/
CVV	/mainasu/	/mairudo/
CVN	/maNneri/	/maNmosu/
CVQ	/maQ+seki/	/meQseezi/

Table 1: Six types of word-initial segmental make-up for unaccented and initial-accented words. + is a morpheme boundary. C is an onset consonant, and V is a vowel. R, N and Q are the second part of a long vowel, a moraic nasal and moraic obstruent, respectively. The second vowel of CVV is different from the first one.

Annotation and acoustic measurements were performed using Praat (Boersma and Weenink 2003). The segmentation and the identification of F0 minima and maxima were carried out manually on the basis of the visual displays of the waveform, wide-band spectrogram and F0 contour.

3 Results and discussion

There seemed to be no difference between the two word positions in an utterance (utterance-initial vs. utterance-internal). The effect of a left-hand utterance boundary on the timing of the F0 peak was not demonstrated.

Ososagari (i.e. peak delay) occurred in almost all cases of the CV+CV and CVCV sequences, showing invariant alignment of the F0 peak with the beginning of the vowel of the second syllable. It was observed across the subjects and regardless of the presence or absence of a morpheme boundary. Examples are shown in Figure 1 (all the examples presented in this section are from the data of the same speaker). As can be seen in the first two examples (/nanohana/ and /minari/), the F0 peak of the initial-accented word occurred outside of the accented syllable, and was aligned just at the beginning of the vowel of the following syllable, which corresponds to an observation in Sugito (1981) that ososagari tends to occur in initial-accented words whose second mora has a non-high vowel. However, similar alignment occurred in the opposite cases, i.e. in the initial-accented words whose second mora has a high vowel, as can be seen in the last example (/namida/) of Figure 1. In fact, more precisely, this alignment pattern was consistently observed regardless of the segmental composition of the first two syllables.

While ososagari did not happen in the cases of the CVN sequence, invariant alignment of the F0 peak was clearly observed in most cases across the subjects (see Figure 2 as examples). The F0 peak was aligned at the end of the first mora (or at the beginning of the second mora), regardless of the vowel type.

Figure 3 shows examples of the CVR and CVV sequences (‘R’ is the second part of a long vowel, and ‘VV’ of CVV is a sequence of two different vowels). In the cases of the CVR sequence, although it was impossible to locate the boundary between the two

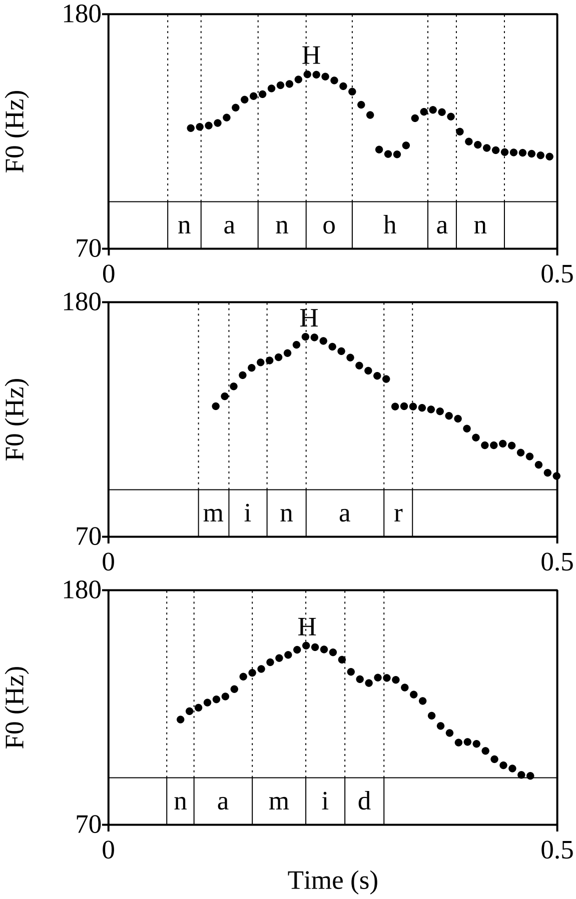

Figure 1: Examples of the F0 peak for initial-accented words which begin with the CV+CV and CVCV sequences. From the top to bottom, /na+no+hana/ ‘rape blossoms’, /mi+nari/ ‘appearance’ and /namida/ ‘tears’.

moras in the long vowel, the F0 peak occurred within the long vowel of the accented syllable. It was invariably aligned somewhere slightly after the middle of the long vowel, regardless of the vowel type, and across the subjects. As for the CVV cases, the F0 peak occurred within the vowel sequence, as in the CVR cases, though its location appeared to depend on the vowels comprising the vowel sequence. While the F0 peak occurred after the middle of the vowel sequence in almost all cases, it tended to occur later when the first vowel of the vowel sequence was a low vowel. This tendency was observed across the subjects.

Since it was difficult to establish a common measuring point for all types of the segmental make-up, statistical analysis was performed on the data of the CV+CV, CVCV and CVN sequences in which the end of the first mora could be exploited to measure

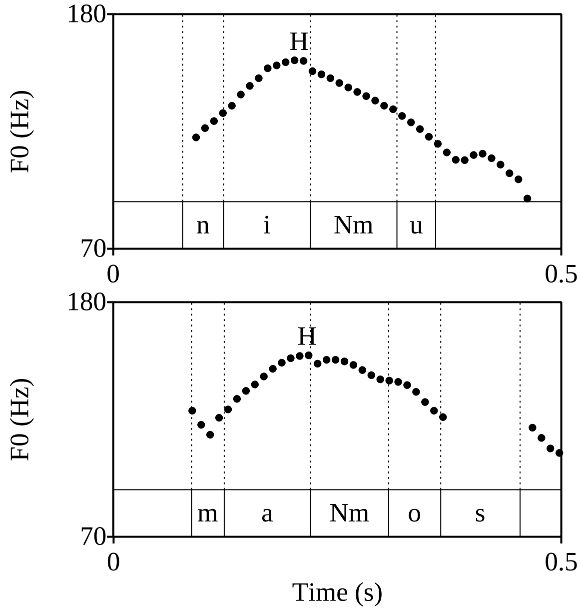

Figure 2: Examples of the F0 peak for initial-accented words which begin with the CVN sequence. /niNmu/ ‘duty’ (top) and /maNmosu/ ‘mammoth’ (bottom). N is a moraic nasal.

the alignment of the F0 peak. The data were analysed using a two-way mixed design ANOVA in which Segmental Make-up (CV+CV, CVCV and CVN) is a within-subjects factor and Speaker (TN, FI, ST, AK, and NI) is a between-subjects factor. Box plots in Figure 4 display the data of the duration from the end of the first mora to the F0 peak of the CV+CV, CVCV and CVN cases across the subjects. The horizontal line indicates the end of the first mora, so a value above the line means that the F0 peak occurs after the end of the first mora. The ANOVA revealed a significant effect of Segmental Make-up, $F(2,90) = 31.417$, $p < 0.0001$, but no speaker effect and no interaction, $F(4,45) = 0.992$ and $F(8,90) = 0.799$. A post hoc test showed a significant difference between the alignment of the F0 peak of the CVN sequence and that of the other two sequences ($p < 0.05$). These corroborate the alignment differences between the three types of the segmental sequences, detected visually.

Based on the visual inspection and the statistical analysis described above, it seems reasonable to say that the F0 peak for initial-accented words is consistently aligned with a fixed point in the segmental string, depending on the structure of the initial syllable(s). In particular the alignment patterns between the CV+CV, CVCV and CVN sequences are clearly distinct, regardless of the segmental differences in the initial syllable(s) (i.e. /m/ vs. /n/, and /i/, /e/, /a/, /o/ or /u/). Since segmental duration differs intrinsically (among other factors being equal), it seems very difficult to explain in phonetic terms these consistent alignment patterns which are due to the structure of the initial syllable(s). Moreover, in terms of the organisation of the mora and syllable in Tokyo

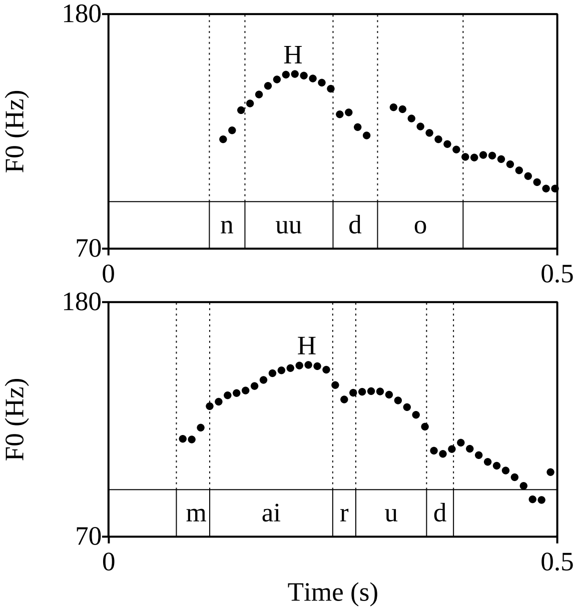

Figure 3: Examples of the F0 peak for initial-accented words which begin with the CVR and CVV sequences. ‘R’ is the second part of a long vowel, and ‘VV’ of CVV is a sequence of two different vowels. /nuudoru/ ‘noodle’ (top) and /mairudo/ ‘mild’ (bottom).

Japanese, it seems feasible to interpret these alignment patterns as a consequence of segmental anchoring. According to one of the proposals of moraic phonology (Hayes 1989), the organisation of the mora and syllable is as follows.

Given that onset consonants belong directly to the syllable, rather than to the mora, it can be claimed that the F0 peak is anchored to the beginning of the second mora across all types of the segmental sequences. This may lead to reconsideration of ososagari (i.e. peak delay in Japanese) such that ososagari is not in fact a phenomenon where the F0 peak for a pitch accent occurs after the end of the associated mora; rather it is a consequence of the anchoring of a pitch accent to the beginning of the second mora.

A possible objection to this explanation is the alignment of CVR and CVV sequences, because the F0 peak occurred after the middle of the ‘VR’ and ‘VV’ portions. However, as it has been reported that the duration of a long vowel which amounts to two moras is not twice as long as that of a short vowel which amounts to one mora

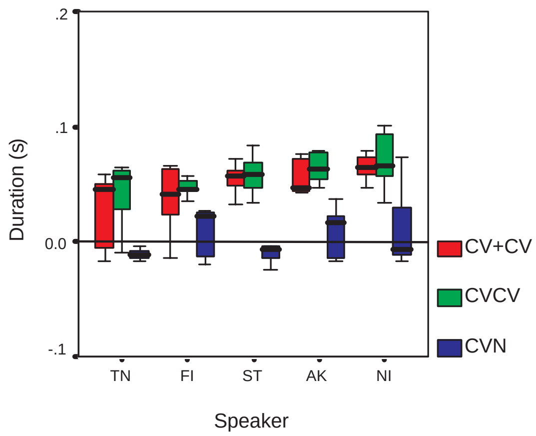

Figure 4: Boxplots of the duration from the end of the first mora to the F0 peak of the CV+CV, CVCV and CVN cases across the subjects.

(Campbell and Sagisaka 1991), the alignment of the F0 peak after the middle of these vowel sequences can be regarded as resulting from shorter duration of the latter mora of the CVR and CVV sequences.

In sum, the evidence discussed in this paper clearly shows the significance of phonological structure to the temporal alignment of pitch accent in Tokyo Japanese. It reveals that phonological conditioning plays an important role in the alignment of pitch features in at least some languages, as demonstrated in Ladd *et al.* (2000) and Xu (1998), though the details differ from language to language.

In spite of the finding of this paper, it is still unclear how the anchoring is influenced under different speaking styles such as speaking fast and raising the voice. It is also required to make clear the alignment of F0 rise of other types of the accentual phrase. Further analysis of these data will be reported in the near future.

Acknowledgments

I would like to thank Bob Ladd and Alice Turk for their advice and valuable comments. I am also grateful to Tsutomu Sugawara who gave me access to the facilities at the Phonetics Laboratory of Sophia University. Needless to say, all errors are my own.

References

Arvaniti, Amalia, D. Robert Ladd and Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25.

Boersma, Paul and David Weenink. 2003. Praat: doing phonetics by computer, version 4.1. Available on-line from `www.fon.hum.uva.nl/praat/`.

Campbell, Nick and Yoshinori Sagisaka. 1991. Moraic and syllable-level effects on speech timing. Tech. Rep. SP 90–107, Institute of Electronic, Information and Communication Engineers of Japan.

Caspers, Johanneke and Vincent J. van Heuven. 1993. Effects of time pressure on the phonetic realisation of the accent-lending pitch rise and fall. *Phonetica* 50: 161–71.

Hasegawa, Yoko and Kazue Hata. 1992. Fundamental frequency as an acoustic cue to accent perception. *Language and Speech* 35: 87–98.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253–306.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge, England: Cambridge University Press.

Ladd, D. Robert, D. Faulkner, H. Faulkner and Astrid Schepman. 1999. Constant segmental anchoring of $F0$ movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–54.

Ladd, D. Robert, Ineke Mennen and Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685–96.

Neustupný, Jiri Vaclav. 1966. Is the Japanese accent a pitch accent? *Onsei Gakkai Kaihoo* 121: 1–7. (In Japanese).

Pierrehumbert, Janet B. and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, Massachussetts: MIT Press.

Prieto, Pilar, Jan P. H. van Santen and Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23: 429–51.

Silverman, Kim and Janet B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. In Kingston, John and Mary E. Beckman, eds., *Between the Grammar and Physics of Speech*, *Papers in Laboratory Phonology*, vol. 1, 71–106. Cambridge: Cambridge University Press.

Sugito, Miyoko. 1981. The relation betwen voice lowering for accent and articulaton of vowels. *Gengo Kenkyuu* 79: 1–32. (In Japanese).

Sugito, Miyoko. 1982. *Nihongo Akusento no Kenkyuu [Studies on Japanese Accent]*. Tokyo: Sanseido.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.