Prosodic correlates of attitudinally-varied back channels in Japanese

Yasuko Nagano-Madsen[1] and Takako Ayusawa[2]
[1]Department of Oriental and African Languages, Göteborg University, Göteborg
[2]Department of Japan Studies, Akita International University, Akita

Abstract

Attitudinally-varied back channel utterances were simulated by six professional voice actors for Japanese. Contrary to the general assumption that a pitch accent language like Japanese cannot vary the tonal configuration for attitudinal variation as in a stress/intonation language, all the speakers differentiated two kinds of tonal configurations. Further variation was achieved by phrasing utterances differently on pitch and timing dimensions, and by adding a rising or non-rising terminal contour.

Introduction

In many stress-accent languages that have been traditionally classified as intonational languages, attitudinal meaning is expressed by means of finely defined tonal contours. In contrast, pitch-accent languages such as Japanese are assumed not to be able to choose contour types for attitudinal meaning or emotion, because the languages use lexically fixed accent shapes (Mozziconacci 2000). Thus, apart from the variation in terminal contour, the dimensions where intonation can vary are pitch range and phrasing (Beckman and Pierrehumbert 1986).

Contrary to traditional belief and assumptions, one of the findings of the present paper is the systematic variation in utterance internal tonal configuration in order to express attitudinal meaning in back channels.

Back channels in Japanese

Japanese is known to be language that uses back channels extensively and this has been studied from various perspectives. For phonological forms used for back channels in Japanese, Nagano-Madsen & Sugito (1999) presents extensive analysis and classification for both Tokyo and Osaka Japanese. The phonetic properties of back channels have only been partially studied for restricted types of back channels (Sugito, Nagano-Madsen and Kitamura 2000, Katagiri, Sugito, and Nagano-Madsen 1999, 2001). This is because these studies deal with recordings of real-life communication and the samples were therefore not systematically varied or distributed, and nor were they easily analysable, due to overlap of utterances.

In order to overcome the difficulties mentioned above and to balance the phonetic data on back channels in Japanese, we present a study of another kind: well controlled simulated utterances recorded in a good acoustic environment. The kinds of back channels presented in this study are of the 'unrepeatable back channel' type, following Nagano-Madsen and Sugito's classification based on the phonological form. Unrepeatable back channels look more like a proper utterance, whereas repeatable back channels are of /so:so:/, /haihai/ 'yes, yes' type.

The first back channel dealt with in the present study is /a-soo-desu-ka/ 'Is that so? I see..', which was the second common back channel after /N:/ 'yes' (Nagano-Madsen and Sugito 1999). Phonologically, it contains the H*L accent in the /soo/. The second type /yamada-san-desu-ka/ 'Is it Mr Yamada? / I see, it is Mr Yamada…' is classified as echo back channel where a keyword in the previous utterance is repeated as back channel (in this case Mr Yamada). This type of back channel shows a deeper concern from the listener and is frequently used where a stream of conversation becomes lively, with quick turn taking (Sugito etc.). Phonologically, it contains unaccented word /yamada/. In addition /are-desu-ka/ 'it is that? It is that…', which is similar to /yamadasandesuka/ but shorter, is also included.

Material

Our speech material consists of high quality recordings of 6 professional voice actors (3 males and 3 females) who were in their 30s or 40s at the time of recording. Each of them produced 3 back channel utterances with neutral

(NEU), joyful (JOY), disappointed (DIS), and suspicious (SUS) attitudes. In addition, the same utterance was produced as a question (Q) rather than a back channel. The second author judged the appropriateness of each attitude type at the time of recording and the best sample for each attitude for each utterance was used for analysis. The recorded material was part of the self-learning CD for Japanese accent and intonation for learners of Japanese (Ayusawa 2001). 'SUGI Speech Analyser' software installed on PC was used to do acoustic analyses. The three back channel utterances are as follows:

(1) /a-soo-desu-ka/ ‘ Is that so? I see …’

(2) /yamada-san-desu-ka/ ‘It is Mr Yamada? / It is Mr Yamada...’

(3) /are-desu-ka/ ‘It is that? It is that..’

Results

Auditory and acoustic analyses revealed that speakers modified several parameters in order to produce attitudinally-varied back channels in Japanese. This included variations in tempo, tonal configuration, pitch range, vowel quality, voice quality, and clarity of articulation. Of these, the most notable systematicity was observed for tonal configuration and phrasing in the pitch and time dimensions. The rest of the paper will focus on these aspects.

Tonal configuration

Contrary to the general assumption that tonal configuration cannot vary in a pitch accent language like Japanese, all six speakers were found to use two tonal configurations. These contours are further differentiated in phrasing in the pitch and time dimensions when expressing various attitudes.

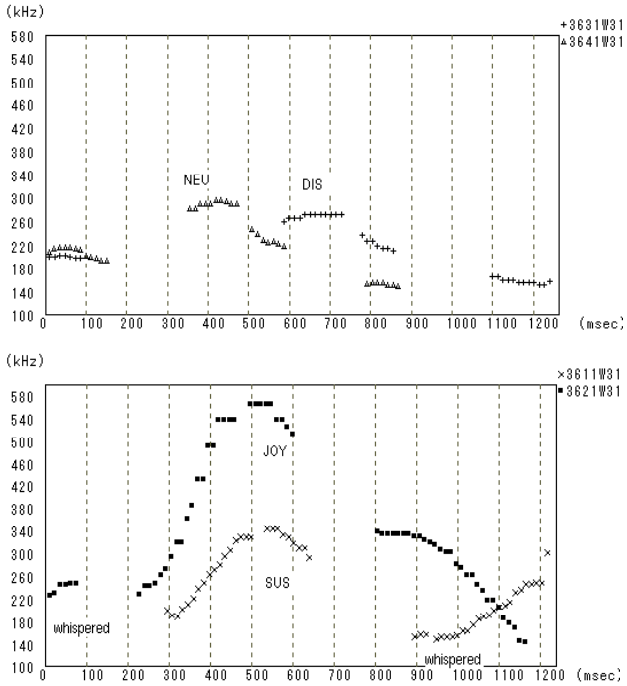
For the utterance /asoodesuka/ ‘is that so? / I see’, where /soo/ is associated with the lexical H*L accent, it is interesting to note that the expected F0 fall was largely missing. Only in one occasion (speaker Z for NEU) was there a very slight F0 fall, and all other cases were produced with either a level H or a rising LH contour. Maekawa (2004), whose data included basically the same utterance /soodesuka/ (without the initial interjection /a/) for his study of paralinguistic information in Japanese, noted this change from the H*L to the LH pattern in some of the utterances.

Table 1 shows the distribution of two contour types, H or LH, for the six speakers for varying attitudes as well as for a question Q.

*Table 1. The choice of tone for different attitudes by the six speakers for /soo/ H*L in /asoodesuka/. U,V,W are female speakers.*

	NEU	Q	JOY	SUS	DIS
U	H	H	LH	LH	LH
V	H	H	LH	LH	LH
W	H	H	LH	LH	H
X	H	H	H	LH	H
Y	H	H	H	LH	LH
Z	H(L)	H	LH	LH	H

It can be seen that the NEU, including Q, is expressed predominantly by a H tone, SUS by a LH tone. The choice of the contour was speaker dependent for JOY and DIS. SUS and Q always have a rising terminal contour while others have a falling or a level contour. There was further variation in the tonal configuration with regard to the particular mora on which the F0 peak was reached. Figures 1a and 1b show two kinds of tonal configuration produced by speaker U. She used a level H tone for NEU and DIS, and a rising (LH) tone for JOY and SUS. Note that for each pair, basically the same shape of tone are placed at different time scale (NEU and DIS) or at different pitch range (JOY and SUS). DIS and SUS are further differentiated by falling vs. rising terminal contour.



Figures 1a,b. The H contour for NEU and DIS and the LH contour for JOY and SUS (speaker U).

The utterance /yamadasandesuka/ contains an unaccented word /yamada/ and starts with a phrasal L% followed by the H*L accent on /de/ of the copula /desu/.

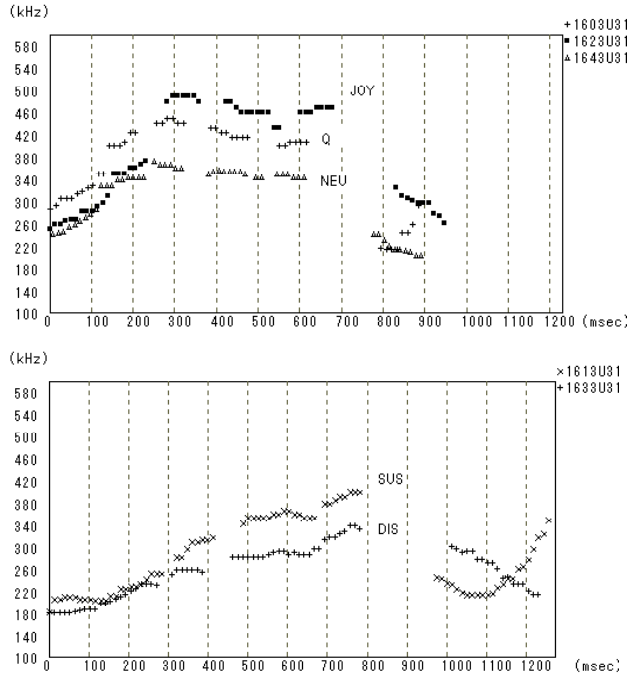


Figure 2a,b. F0 contours for JOY, Q, and NEU (above) and SUS and DIS (below) for speaker U.

As for the previous case, this utterance also had two kinds of tonal configuration that were further differentiated in phrasing and by terminal contour to express attitudinal meaning. Figures 2a and 2b show F0 contours for /yamadasandesuka/ produced by speaker U. The tonal configuration used for NEU (including Q) and JOY has a steeper initial F0 rise than that used for SUS and DIS. The former type of contour has a plateau phase. In contrast, the latter type of contour has a slow and gradual F0 rise that reaches to its peak only before the accentual fall. It does not contain any plateau phases. In Figure 2 (above) the F0 contours are placed at three equidistant pitch range intervals, Q being differentiated by a rising terminal contour. Likewise, the contours for DIS and SUS differ both in pitch range and in terminal contour.

Note that the difference in the way F0 rises initially, as described above, does not affect the phonological structure of the utterance, since /yamadasan/ is an unaccented phrase. It is the phrasal feature that varies, i.e. the H- phrasal tone which is expected to be on the second mora is greatly delayed. Interestingly enough, not only one out of the 30 tokens produced by the six speakers had the accentual F0 peak on /ma/. Instead, it was either on the third mora /da/ (10 tokens), the fourth mora /sa/ (7 tokens) and /de/ (13 tokens). None of the tokens for DIS and SUS had F0 peak on /da/. All the foregoing observation indicate that attitudinal meaning for /yamadasandesuka/ is clearly conveyed by systematic variation in tonal configuration through different phonetic implementation of the accentural-phrasal tones %L and H- (cf. X-JtoBI, Maekawa et al. 2004).

Phrasing

There was fairly good agreement among the speakers as to how the two tonal configurations were phrased in the pitch and time dimensions. With one exception, the attitude JOY always had the highest F0 peak, while the lowest peak was typically found for DIS. SUS and DIS were spoken more slowly than other types of attitudes and therefore had a longer duration. Figure 3 shows a typical example of phrasing for four attitudes by speaker Y.

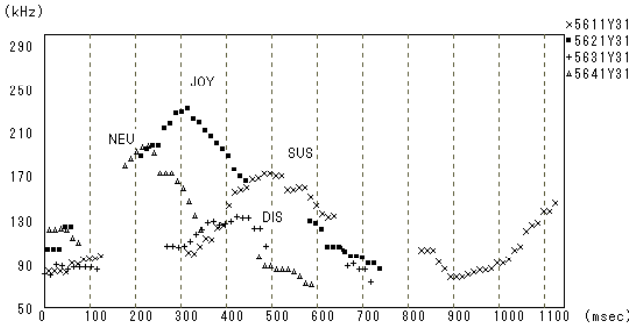


Figure 3. A typical phrasing of F0 contours for NEU, JOY, DIS, and SUS (speaker Y).

Pitch range and duration

Figures 4 and 5 show the peak F0 values and total durations for /asoodesuka/ spoken by the six speakers with varying attitudes.

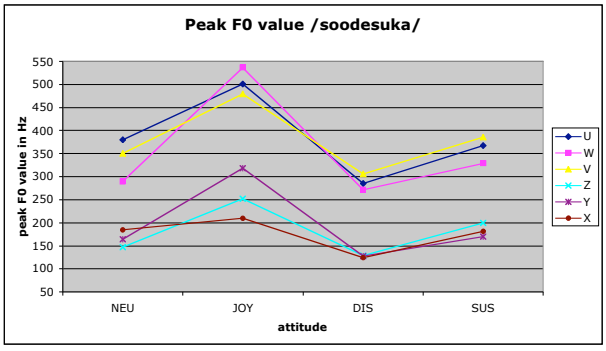


Figure 4. Peak F0 value for /asoodesuka/ produced by six speakers.

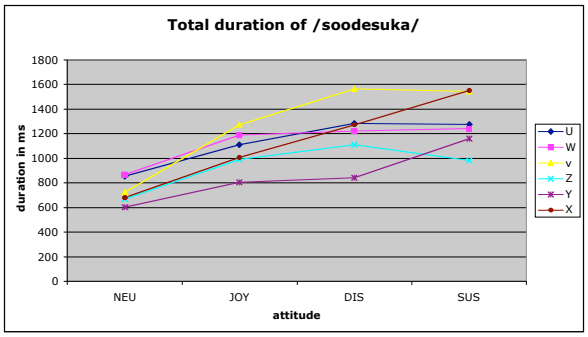


Figure 5. Total utterance duration for /asoodesuka/ produced by six speakers.

Discussion

Some of the findings reported here agree with those reported in Maekawa's (2002) study on paralinguistic information in Japanese. The present study revealed more systematic details in the way tonal configuration was varied in conveying attitudinally varied utterances.

Although the choice of tonal configuration is limited to two basic types, Japanese speakers were found to vary phrase internal F0 contours systematically in order to express attitudinally-varied back channels. The test material contained both accented (H*L) and unaccented words. In the case of the accented word /soo/ in /asoodesuka/, the lexical accent was altered to either H or LH, the former being typically used for NEU. In unaccented words, variation in tonal configuration was achieved by modifying the rate of initial F0 rise. What is consistent in both cases is that in the contours used for unmarked attitude NEU, F0 reaches its peak earlier than in marked attitude such as SUS. According to the X-JToBI (Maekawa et al 2004), Japanese ToBi labelling, 'H-' is introduced to indicate the onset of F0 plateau. It can be analysed that both accented and unaccented word can be expressed by delayed H- for marked attitudes. Exactly on which mora the initial and final F0 maxima are placed varies between speakers. It should be also reminded that the general assumption that the phrasal H- is on the second mora was not attested in the present data even for utterances with NEU and Q type. This phenomenon needs further investigation.

Apart from the phrase internal tonal variations described above, there are differences in phrasing and terminal contours. These prosodic characteristics are further modified by vowel quality, voice quality, and clarity of articulation. Identification of the more important parameters for conveying attitudinally-differentiated back channels in Japanese will require perceptual experiments.

References

Ayusawa T. (editorial supervision). (2001) Accent and intonation in Tokyo Japanese. CALL Sub-teaching material series, Japanese prosody Vol.1. National Institute of Multimedia Education.

Beckman ME and Pierrehumbert JB. (1986) Japanese prosodic phrasing and intonation synthesis. Proceedings of the 24th Meeting of the Association for Computational Linguistics

Katagiri, Y., M. Sugito, and Y. Nagano-Madsen (1999) The forms and prosody of back channels in Tokyo and Osaka Japanese. In the Proceedings of The XIIIIth International Conference on Phonetic Sciences, San Francisco.(CD).

Katagiri, Y., M. Sugito, & Y. Nagano-Madsen (2001) An analysis of forms and prosodic characteristics of Japanese 'Aiduti' in dialogue (in Japanese). In Bunpoo to Onsei (=Speech and Grammar) III, 263-274. Tokyo: Kuroshio Publication.

Maekawa K. (2002) Production and perception of 'paralinguistic' information. Proceedings of International Conference:Speech Prosody 2004, Nara, 367-374.

Maekawa K, Igarashi Y, Kikuchi E, and Yoneyama S. (2004) Intonation labelling for 'Spoken Japanese Corpus', version 1.0 (in Japanese). Electronical document for The corpus of Spontaneous Japanese.

Mozziconacci, S. (2000) Prosody and Emotions. In Proceedings Online: ISCA Workshop on Speech and Emotion.: A conceptual framework for research.

Nagano-Madsen, Y. and M. Sugito. (1999) Analysis of back channel items in Tokyo and Osaka Japanese (in Japanese). In Japanese Linguistics 5, 26-45. Tokyo: National Language Research Institute.

Sugito, M., Y. Nagano-Madsen, and M.,Kitamura.(1999) The pattern and timing of the repeat-back channels in natural dialogue in Japanese (in Japanese). In Bunpoo to Onsei (=Speech and Grammar) II, 3-18. Tokyo: Kuroshio Publication.