PROSODIC FEATURES OF UTTERANCES IN THE CORPUS OF SPONTANEOUS JAPANESE: INTONATION PHRASE–BASED APPROACH

Yuichi Ishimoto$^1$ and Hanae Koiso$^2$

$^1$National Institute of Informatics, Tokyo, Japan,
$^2$National Institute for Japanese Language and Linguistics, Tokyo, Japan

ABSTRACT

It is unclear how the prosodic features in long spontaneous speech change. We focused on the F0 declination of an utterance to investigate the F0 transition by taking into consideration the length and syntactic structure of a Japanese utterance. In the case of an utterance without a strong syntactic boundary, the F0 range is fixed to a certain width, and the F0s decline toward the end of the utterance in that range. In addition, the F0s at the beginning of an utterance are slightly higher for a longer utterance. In the case of an utterance with a strong syntactic boundary, the F0 declination is reset at the strong syntactic boundary in the utterance, and the final lowering occurs only at the end.

Index Terms— Prosody, spontaneous speech, intonation phrase, fundamental frequency, syntactic boundary

1. INTRODUCTION

The aim of this study is to investigate the prosodic behavior of spontaneous Japanese utterances. Pierrehumbert & Beckman defined that the unit utterance is put above intonation phrases (IPs) and accentual phrases (APs) in the prosodic hierarchy[1]. The utterance unit is a region in which the fundamental frequencies (F0s) monotonously decline over time, which is called the F0 declination, and the F0s fall significantly at the end of the utterance unit, which is called the final lowering. For previously prepared speech such as news, the utterances correspond to the sentences divided by sentence final markers. For spontaneous speech, sentences are often attached with multiple clauses at the end, and the utterances extend further and further. Now, we wonder if the F0 declination and the final lowering occur in a long spontaneous utterance.

We examine how the length of the utterance and the syntactic structure affect the F0 transition in the utterance unit in this paper. For spontaneous Japanese speech, the utterance unit is defined by using a clause unit[2]. So, we examine the transition of the maximum and minimum values of the F0s for each IP that comprises the clause unit with a strong syntactic boundary.

2. F0 TRANSITION IN SPONTANEOUS UTTERANCE

2.1. Data

The corpus of spontaneous Japanese (CSJ)[3] is a large-scale database of Japanese speech, which consists mainly of spontaneous monologues including academic presentation speech and simulated public speech. We used the 70 academic presentation speeches (about 19 hours) and the 107 simulated public speeches (about 20 hours) of the CSJ for this study.

2.2. Clause unit

The clause unit[2] is adopted as the utterance unit in order to define what an utterance is. The clause units are designed for annotating discourse and sentence structure information, and are defined as basic processing units of spontaneous Japanese speech in the CSJ. The boundaries of the clause units are classified by structural type as follows:

- **Absolute boundary** explicit sentence final markers
- **Strong boundary** conjunctive particles expressing coordination
- **Weak boundary** other conjunctive particles followed by a discourse marker or speaker change.

Among these boundaries, absolute and strong boundaries are defined as clause unit boundaries.

We examine the relation between the length of the clauses and the F0 transition. In analysis 1, we targeted the clause units that have absolute boundaries at its beginning and end (as shown in Figure 1, Clauses 4 and 8). In this case, the data do not contain strong syntactic boundaries. In analysis 2, we focus on the utterances that contain strong syntactic boundaries, and investigate whether the F0s continue to decline over the clauses. Specifically, the data are two consecutive clauses that are delimited by the strong boundary (as shown in Figure 1, Clauses 2-3 and 6-7).
2.3. Intonation phrase

We extracted the representative F0s of the IPs in each clause for these analyses. An IP consists of one or more APs whose F0 ranges are recursively reduced. The F0 downstep of the APs is reset at the start of each new IP. The CSJ has been annotated with words and tone structures such as boundary tones and break indices (BIs), by using the X-JToBI scheme[4], and we defined the BIs indicated by level 3, D (disfluent), or F (filler) as the boundaries of the IPs. Then, the disfluent speech and the fillers were excluded from this study. After this definition, we extracted the maximum and minimum values of the F0s for each IP as shown in Figure 2 and follows:

F0 max either higher F0 at the phrase accent “H-” or the accent nucleus “A” in the first AP in an IP

F0 min an F0 at the final boundary tone “L%” in the final AP in an IP.

To avoid any influence from gender and individual differences, the F0s were converted to z-values, which are normalized by the mean value and the standard deviation of all F0 values for each discourse.

2.4. Results and discussion

2.4.1. Analysis 1

We examined the relation between the length of the clause units and the F0 transition. The number of IPs in a clause unit is used as an indicator of the length of the clause unit. The data for analysis 1 consisted of 4914 clause units and 26763 IPs, and the mean number of IPs in a clause unit was 5.4.

Figure 3 shows the transition of F0 max and F0 min of the IPs in the clause units. Both the F0 max and F0 min tend to decline toward the end of the clause unit, regardless of the number of IPs. This decline differs from the downstep between APs, because the F0s were extracted as the values representing each IP. In addition, the F0 min of the final IPs decline more significantly than that of the preceding IPs. We believe that this indicates the final lowering in the final AP, which was mentioned in Maekawa’s work[5].

Upon close examination of Figure 3, the F0 max values at the beginning of the clause units (especially for the second IPs that are not affected by conjunctions) are nearly the same, regardless of the length of the clause. Similarly, the F0 min values at the final IPs are very close in most clauses. In other words, utterances start from a certain common height and end at a certain common height, and therefore, the F0 range of an utterance is fixed to a certain width. Figure 4 shows the F0 max of the second IPs and the F0 min of the final IPs for the clause units that contain 2 to 13 IPs. There were statistically significant differences between the groups of the F0 max as determined by one-way ANOVA (F(11, 874.37)=9.462, p<0.001). As shown in this figure, the maximum F0 values at the beginning of an utterance become slightly higher as the utterance gets longer. Using the Tukey-Kramer multiple comparisons, although the adjacent groups have little or no statistical significance, the distant groups (e.g. groups
of 3 IPs and 8 IPs) have statistically significant differences. This means that speakers already control F0s at the beginning of an utterance according to the length. On the other hand, there were no statistically significant differences between the groups of the F0 min as determined by one-way ANOVA (F(11,139.11)=0.850, p=0.591). Namely, the minimum F0 values at the end of an utterance are maintained virtually constant regardless of the length of the utterance.

2.4.2. Analysis 2

In the utterances that contain at least one strong syntactic boundary, we examined the relation between the length of the clause units and the F0 transition. It should be noted here whether the F0 declination observed in the above section continues over the strong syntactic boundary.

Figure 5 shows some typical F0 transition results for two consecutive clause units. As an illustration, the results from 54 utterances that contain 4 IPs in the preceding clause and 5 IPs in following clause, and 21 utterances that contain 5 IPs in the preceding and 8 IPs in the following are described in Figure 5. As shown in this figure, the F0 declination occurs in the preceding clause units and the following clause units, while the declination is reset at the strong boundaries between the clause units. Furthermore, although the values of the F0 min at the end of the following clause units (namely just before the absolute boundaries) fall significantly as the final lowering, the values at the end of the preceding clause units (namely just before the strong boundaries) do not fall that much and the final lowering does not occur. That is to say, the F0s are reset at the strong syntactic boundary before finishing their falling to the lower limit.

This suggests that the F0 declination does not always continue in the utterance, and is readjusted at the strong syntactic boundary. Moreover, the F0 declination differs from the declination mentioned by Pierrehumbert & Beckman[1]. They reported that the F0s monotonously decline in proportion to the duration of the utterance. However, the F0 declination observed in our study is not monotonic and its slope changes during an utterance.

Considering these results, the lower limit of the F0s exists in the utterance, and the utterance is adjusted to reach the lower limit at its end. Therefore, in a long utterance, the F0s are reset at the strong syntactic boundary so that they do not reach the limit before the end of utterance. In addition, starting from a higher F0 in a long utterance may prevent a fall to the lower limit in the middle of the utterance.

3. CONCLUSION

By focusing on the F0 declination of an utterance, we were able to investigate the F0 transition while taking the length and syntactic structure of a Japanese utterance into consideration. In analysis 1 for an utterance without a strong syntactic boundary, the results were as follows: The F0 range of the utterance was fixed to a certain width, and the F0s decline toward the end of the utterance in that range. The slope of the declination depends on the length of the utterance. The F0s at
Fig. 4. F0s at beginning and end of utterance: F0 max for second IPs in clause units (upper panel), and F0 min for final IPs (lower panel).

Fig. 5. Transition of F0 max and F0 min of IPs in two consecutive clause units divided by strong boundary: preceding clause has 4 IPs and following clause has 5 IPs (left), and preceding has 5 IPs and following has 8 IPs (right). The vertical solid lines indicate the boundaries of the clauses.

the beginning of the utterance are slightly higher for a longer utterance. In analysis 2 for an utterance with the strong syntactic boundary, the results showed that the F0 declination is reset at the strong syntactic boundary in the utterance, and the final lowering occurs only at the absolute boundary, namely the end of the utterance. We believe that the end of utterance is characterized by these F0s behaviors.

4. REFERENCES


