Prosodic diversity according to relationship among participants in everyday Japanese conversation

メタデータ	言語: eng 出版者: 公開日: 2020-10-30 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00003052

Prosodic diversity according to relationship among participants in everyday Japanese conversation

Yuichi Ishimoto[1], *Koiso Hanae*[2]

[1]Center for Corpus Development, National Institute for Japanese Language and Linguistics
[2] Spoken Language Division, National Institute for Japanese Language and Linguistics
yishi@ninjal.ac.jp, koiso@ninjal.ac.jp

Abstract

A new corpus project of the National Institute for Japanese Language and Linguistics is building a large-scale corpus of everyday Japanese conversation, the *Corpus of Everyday Japanese Conversation*, CEJC, which contains various kinds of conversations in a balanced manner. We plan to publish the corpus with 200 hours of conversations in 2022, and have already published and released 50 hours of the conversations to the public on a trial basis in December 2018. In this paper, we investigate statistically prosodic diversity in everyday conversation by analyzing the characteristics of the fundamental frequency (F0) of utterance in the CEJC. We first examined the differences of the F0s by the relationships between the speaker and the addressee. The results indicated that speech to the family is generally produced with low F0s, and politeness to the addressee produces higher F0s in utterances. Next, we analyzed the characteristics of the F0s in situations with various types of participants to show the effect of participants on diversity. The results indicated that although the degree of the F0 change differs with each speaker, they speak in relatively low tones in situations including family members and utter with high F0 in customer service.

1. Introduction

Spontaneous speech in everyday conversation is the most basic form of human communication. In order to understand our diverse and situated interactional behavior, it is necessary to collect and analyze various kinds of conversations in our daily lives. In Japan, although several corpora of spontaneous speech have been developed, most of them are biased in terms of situations. For example, the *Corpus of Spontaneous Japanese* (CSJ) [1] contains monologues consisting of academic presentations, simulated public speech, and dialogues. This corpus is useful for research on speech recognition, natural language processing, prosodics, linguistics, and the paralinguistics of spontaneous speech. However, the CSJ is mostly limited to presentations; therefore, it is possible that the data are not representative of natural spontaneous speech in our daily life. Another example is the *Chiba Three-party Conversation Corpus* [2], which is a collection of casual and friendly conversations among three people of the same gender. The conversations were recorded in situations where some specific topics were imposed on them. Therefore, it is likewise insufficient to describe real conversation in daily life.

Due to the lack of a large amount of everyday conversation data, it appears that the diversity of spontaneous speech had been investigated mainly qualitatively, but not quantitatively. However, there has been progress toward building a new large-scale corpus of everyday conversation, which contains various kinds of conversations in a balanced manner [3].

In this paper, we investigate statistically prosodic diversity in everyday conversation by analyzing the characteristics of the fundamental frequency (F0) of utterance. We first show the differences of the F0s by the relationships between the speaker and the addressee. Next, we analyze the characteristics of the F0s in situations with various types of participants to show the effect of participants on the diversity.

2. Everyday conversation data

2.1. Corpus

A new corpus project of the National Institute for Japanese Language and Linguistics is building a large-scale corpus of everyday Japanese conversation, the *Corpus of Everyday Japanese Conversation*, CEJC [4]. We plan to publish the corpus with 200 hours of conversations for about 600 different participants in 2022. We have already published and released 50 hours of the conversations to the public on a trial basis in December 2018. The CEJC targets conversations embedded in naturally occurring activities in daily life, without the exogenous intervention of researchers imposing topics or displacing the context of action [5].

Informants for the corpus were recruited for balance with sex and age. They recorded around 15 hours of conversations from their real lives, without any interference from researchers, by using provided portable recording devices (compact action cameras and IC recorders). Then the relevant conversations were selected for the CEJC by taking into account the balance of conversation variations, quality of recorded data, and legal and ethical issues. Figure 1 shows video images from the corpus. Each participant wears an IC recorder on his breast, which allows the device to record his speech sound individually.

Table 1 shows attributes of the 20 informants included in the trial version of the CEJC. As a result of the multifarious informants, the CEJC includes a wide range of regular conversations in a balanced manner.

Although the corpus contains a total of 390 participants, in this analysis, we utilize only the speech sound data of the informants shown in Table 1 because the informants must appear in the conversations recorded by themselves. Furthermore, we excluded conversations in which relationships between the speaker and the addressee cannot be uniquely determined. In consequence, the speech data of 17 informants were analyzed.

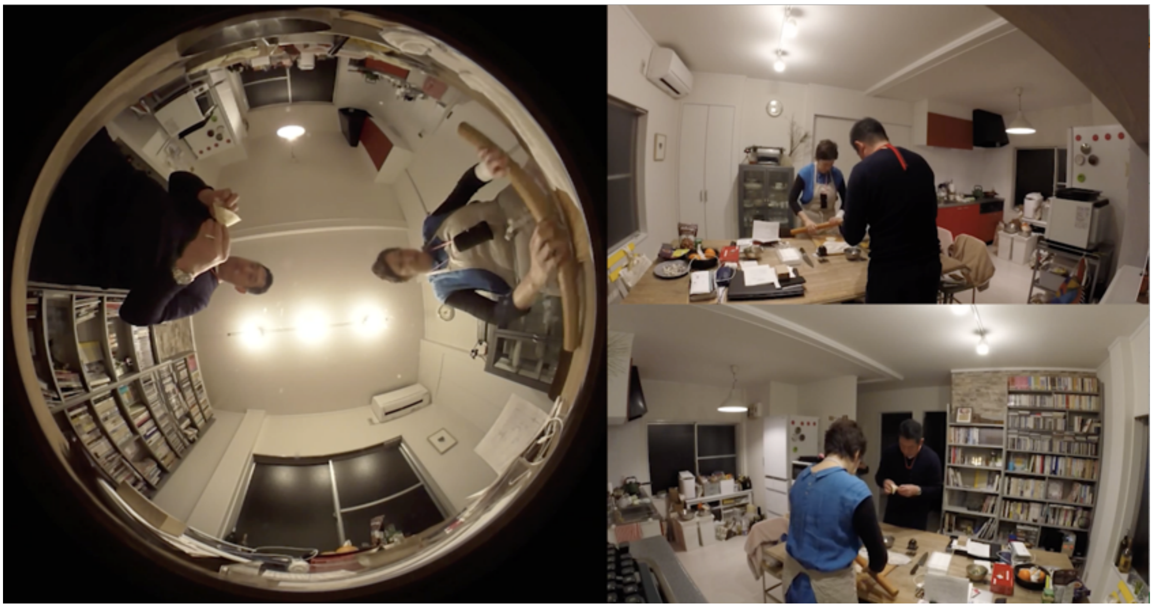

Figure 1: *Video images of a conversation between husband and wife while cooking at home in the CEJC. The left image was recorded using a 360-degree camera located on the table, while the top- and bottom-right images were recorded by two GoPro cameras placed facing each other on the bookshelf and the sideboard. Their speech was recorded using the IC recorders strapped onto their chest.*

Table 1: *Attributes of informants. The informants were almost balanced in terms of sex and age.*

ID	Age	Sex	Occupation
T010	20-24	Male	Student
T006	25-29	Male	Student
K003	20-24	Female	Student
T009	20-24	Female	Student
T001	35-39	Male	Self-employed
T005	35-39	Male	Office worker
K001	35-39	Female	Office worker
T003	35-39	Female	Housewife
T002	40-44	Male	Self-employed
T016	40-44	Male	Self-employed
C001	40-44	Female	Office worker
K004	40-44	Female	Part-time
T011	40-44	Female	Part-time
C002	55-59	Female	Office worker
K002	50-54	Female	Self-employed
S001	50-54	Male	Office worker
T015	50-54	Male	Office worker
T004	60-64	Female	Housewife
T007	70-74	Male	Volunteer
T013	65-69	Male	Office worker

2.2. Fundamental frequency estimation

As a prosodic feature, we estimated F0s at 1-ms intervals from the audio data of the CEJC by using a source information analysis function of STRAIGHT vocoder system proposed by Kawahara et al. [6][7]. The estimated F0s were applied to the voiced/unvoiced detection of the system for extraction of the F0s in voiced section. In addition, we selected the F0 values between the top 10 percent and the bottom 10 percent for each informant to eliminate estimation error such as half pitch error and double pitch error. To avoid influences from gender and individual differences, the logarithmic F0s were converted to z-scores for each informant. Finally, the mean values of the F0s for each utterance-unit were calculated. Note that because the utterance-unit segmentation is based on a *long utterance-unit* [8] defined by syntactic and pragmatic disjuncture, turn-taking does not always occur at the end of the utterance-unit.

3. Prosodic diversity in everyday conversations

In this section, we explore the diversity of the F0s in everyday conversations, focusing on types of addressees and other participants.

3.1. Difference in relationship with addressee

We analyzed whether the mean F0 of utterance differs due to relationships between the speaker and the addressee in everyday conversation. As mentioned above, conversations in which the addressee's relation to the informant cannot be uniquely determined were excluded from the analysis. For instance, when an informant is a father, a conversation with his two children were included in the analysis, but a conversation with his children and friend were excluded.

Figure 2 shows F0 distributions of the utterances of the 17 informants (i.e., speakers) for each type of addressee, namely: child, spouse, spouse, parent, sibling, colleague, client, and customer. A one-way ANOVA was conducted to compare the mean F0s of the utterances on the relationships. There was a significant difference between the relationships [F(9, 26688)=127.15, p<0.001]. Figure 2 also shows groups of the types of addressee derived from the results of post hoc comparisons using the Tukey HSD test. The groups in Figure 2 indicates the following tendencies:

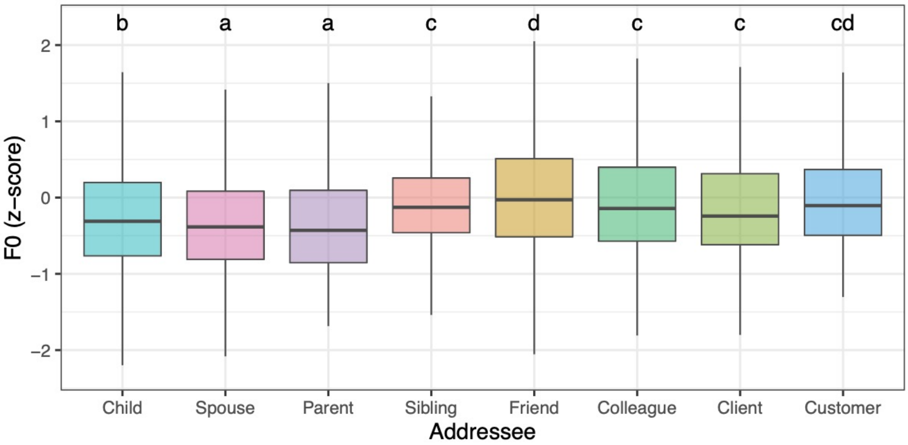

Figure 2: *Distributions of the mean F0 of utterance for relationships between speaker and addressee. As for the letters above the plots, the same letter indicates that there is no significant difference between the groups in the Tukey HSD test (p<0.05).*

- Speech to close relatives such as a child, a spouse, and a parent is uttered with lower F0s than others.
- The F0s for siblings are not as low as compared to other relatives.
- Speech to a friend has the highest F0s.
- For a colleague, customer, and client, the F0s are located somewhere between the close relatives and the friend.

It goes to show that the F0s of utterances in everyday conversations certainly differ depending on the situation. It appears to be connected to the speaker's relationships with the addressees.

3.2. Influence of participants

In the above section, we presented the effects of types of the addressee for the F0s of utterances through the overall analysis for the informants. In this section, we select some of the informants and analyze the F0s for these individuals in various situations. In the analysis, conversations in which the addressee's relation to the speaker cannot be uniquely determined were included so as to examine whether the F0 change appears in a conversation which various types of participants attend to.

The informant K002 recorded the conversations in situations including chats and meetings with friends, a meeting with a client, and a chat with her family. Figure 3 shows distributions of the mean F0s of the utterances in each situation. There are significant differences in the Tukey HSD test, except between the chats with the family and with the friends. The results indicate that the F0s were low in the cases of chats with familiar persons, namely, the family and the friend. The business talk with the client showed higher F0s than that of the chats. Speech in the meeting with the friends was also higher than that of the chat. Thus, the more the familiarity increases, the lower the F0s become. What the speech to the friend has low F0s is the opposite of the result obtained in Section 3.1.

The informant T001 recorded the conversations in situations including chats with his wife, with his wife and mother-in-law, with his wife and friends, and with friends. Figure 4 shows distributions of the mean F0s of the utterances in each situation. There are significant differences in the Tukey HSD test, except between the chats with the wife and the mother-in-law and with the wife and the friends. Speech in the situation with just the wife was uttered with lower F0s than that in the situation involving the mother-in-law. For the chats with only the friends, the F0s were higher than that in the situation with the wife and the friends. Thus, it shows that the situations involving the family bring about lower F0s.

The informant T003 recorded the conversations in situations including chats with her children, with the children and her husband, with the husband and his family, with friends, and a meeting with friends. Figure 5 shows distributions of the mean F0s of the utterances in each situation. In the Tukey HSD test, there are no significant differences between the chats with only the children and the children and the husband, and between the chats with the children and the husband and with the husband and his family. In the case of this informant, speech to the family has shown the lowest F0s. The result accords with that in Section 3.1.

The informant T007 recorded the conversations in situations including chats with his younger brother and sister, and with his wife, his daughter and daughter's husband, and a meeting with friends. Figure 6 shows distributions of the mean F0s of the utterances in each situation. There are significant differences in the Tukey HSD test, except between the chats with the siblings and with the wife, the daughter and the daughter's husband. Speech in situations with the friends had slightly higher F0s than that with the others. However, it seems that this informant does not remarkably change F0s according to the types of participants.

The informant T009 recorded the conversations in situations including chats with her family and with a friend, a meeting with her friends, and a customer service with a colleague. Figure 7 shows distributions of the mean F0s of the utterances in each situation. There are significant differences in the Tukey HSD test, except between the chat with the family and the meeting with the friends. For this informant, both the chat with the friend and the speech to the colleague and the customer yielded high F0s. However, in the meeting,

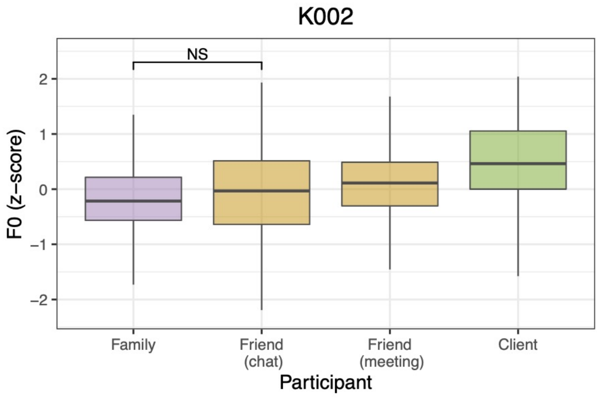

Figure 3: *Distributions of the mean F0 for types of participant in conversations of the informant K002. The letter NS between the plots indicates no significant difference in the Tukey HSD test. There are significant differences in a case without NS.*

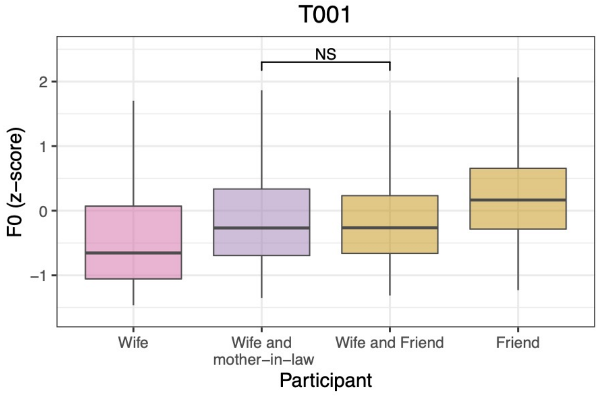

Figure 4: *Distributions of the mean F0 for types of participant in conversations of the informant T001.*

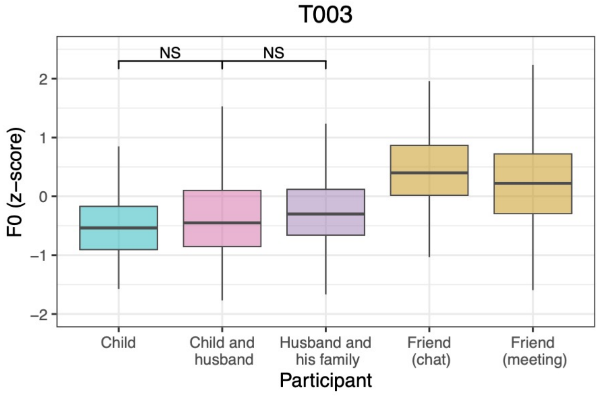

Figure 5: *Distributions of the mean F0 for types of participant in conversations of the informant T003.*

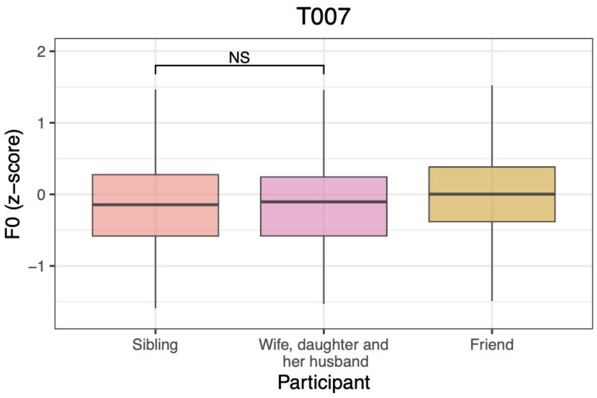

Figure 6: *Distributions of the mean F0 for types of participant in conversations of the informant T007.*

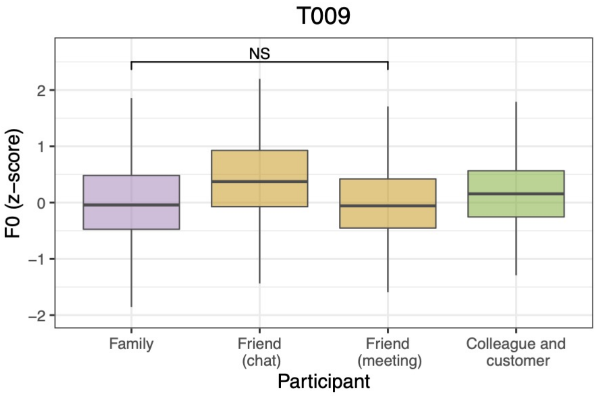

Figure 7: *Distributions of the mean F0 for types of participant in conversations of the informant T009.*

the presence of the friends did not induce high F0s because of the relatively formal situation.

4. Discussions

In section 3.1, it was presented that the F0 of utterances in everyday conversations becomes lower for close family members, such as children and spouses. On the contrary, speech to people outside of the family tends to be uttered with high F0. This may be because the politeness to the addressee enforces the high F0s on the speaker unconsciously in a social norm. In terms of familiarity, it is interesting that the family and the friend have the opposite effect though both are similar in familiarity.

In section 3.2, it was implied that not only the type of addressee but also the other participants influence the F0s of utterance. For instance, the informant T001 uttered with low F0s in the situation with only his wife, while he uttered with higher F0s in the situation including the wife and her mother, as shown in Figure 4. It suggests that even though the mother-in-law is a member of his family, he treated her distinctly from the wife because the mother-in-law in fact lives apart from him and the wife. He also uttered with high F0s in situations with only the friends, while he uttered with low F0s in situations with the wife and the friends.

On the other hand, the informants T002, T003, and T009 uttered with different F0s between the chats and the meetings in spite of situations with only the friends. In addition, focusing on customer service observed in the conversations of the informants K002 and T009, speech to the client and the customers has higher F0s than that with the friends.

In summary, although the degree of the F0 change differs with each participant and informant, what the informants showed in common are as follows:

- They speak in relatively low tones to family members.
- In situations including clients and customers, they utter with high F0.

The differences of the F0s of utterances in real life occur from complex relationships with participants and situations; therefore, the prosodic diversity of utterance appears notably in everyday conversations. Needless to say, there are other factors which account for the differences of F0 in conversations. In future works, we will focus on places where the conversations occur, as well as other detailed topics.

5. Conclusions

In this paper, we analyzed the F0s of utterances by using the *Corpus of Everyday Japanese Conversation* to observe the prosodic diversity of spontaneous speech in real life, focusing on types of the addressee. The results showed that speech to the family generally utters in low tones, and politeness to the addressee produces higher F0s in utterances. We also analyzed the F0s considering other participants in the conversations. The results indicated that the degree of the F0 change affected by the participants differs with individuals, and the prosodic diversity occurs from complex relationships with participants in conversations.

6. Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 16H03421, 17H00914, 18K11514 and 19H01252, and also supported by the NINJAL collaborative research project “A Multifaceted Study of Spoken Language Using a Large-scale Corpus of Everyday Japanese Conversation.”

7. References

[1] K. Maekawa, “Design, compilation, and some preliminary analyses of the Corpus of Spontaneous Japanese,” in K. Yoneyama and K. Maekawa, editors, *Spontaneous speech: Data and analysis*, pp. 87-108. The National Institute for Japanese Language and Linguistics, Tokyo, 2004.

[2] Y. Den and M. Enomoto, “A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation,” in T. Nishida, editor, *Conversational informatics: An engineering approach*, pp. 307-330. John Wiley & Sons, Hoboken, NJ, 2007.

[3] H. Koiso *et al.*, “Design and Preliminary Analysis of the Corpus of Everyday Japanese Conversation,” *Proc. LB-ILR2018 and MMC2018 Joint Workshop*, 2018.

[4] H. Koiso *et al.*, “Construction of the Corpus of Everyday Japanese Conversation: An Interim Report,” *Proc. LREC 2018*, pp. 4259-4264, 2018.

[5] L. Mondada, “The conversation analytic approach to data collection,” in J. Sidnell and T. Stivers, editors, *The handbook of conversation analysis*, pp. 32-56. Wiley-Blackwell, Hoboken, NJ, 2012.

[6] H. Kawahara et al., “Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity,” Proc. EUROSPEECH'99, pp 2781-2784, 1999.

[7] https://github.com/HidekiKawahara/YANGstraight_source

[8] Y. Den *et al.*, “Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme,” *Proc. LREC 2010*, pp. 2103-2110, 2010.