INTER-SPEAKER INTERACTION OF F0 IN DIALOGS

Kuniko Kakita

Department of Liberal Arts and Sciences
Toyama Prefectural University
Kosugi-machi, Imizu-gun, Toyama 939-03, Japan
Email: kakita@pu-toyama.ac.jp

ABSTRACT

This paper presents some preliminary results of a study on inter-speaker interaction of fundamental frequency of voice (F0) during dialogs. Simple question-answer type dialogs (in Japanese) were recorded and analyzed. The results revealed all three possible patterns of F0 interaction. In some cases, the two speakers' F0 converged as the dialog progressed, while in other cases they diverged; there were also cases in which they maintained a parallel relation to each other throughout the dialog. The results also indicated that dialog-initial difference in the two speakers' F0 values was closely related to consequent overall pattern of F0 interaction during dialogs. Partner-dependent reorganization of dialog-initial F0 was observed for some of the speakers.

1. INTRODUCTION

People are constantly reorganizing their speech in relation to their communication partner(s) [2-6,8,9]. Sometimes they do so intentionally [e.g. 3]; at other times, they do so without apparent intention [e.g. 5]. The present study focuses on the latter type of speech reorganization. It aims to examine how two speakers' speech characteristics interact in a context that involves minimum intentionality for speech reorganization. For this purpose, dialogs consisting of simple questions and answers based on a list of numbers were recorded and analyzed.

Previous reports by the present author [5,6] have shown that in a 'relay' reading of sentences, the speech rate and pause duration of the 'following' speaker assimilate to those of the 'preceding' speaker. Based on this finding, it was expected that, in a dialog, the speech characteristics of the two speakers, differing at the start of the dialog, would assimilate to those of the other speaker, and become increasingly similar as the dialog progressed. The primary goal of the present study was to examine if this was indeed the case with F0, and to clarify the nature of the inter-speaker interaction of F0 in dialogs.

2. METHOD

2.1. Dialog Procedure

A dialog proceeded in the following manner. Each of the two speakers were given a sheet of paper with a random list of 20 five-digit numbers. The numbers were labeled with consecutive ID numbers, 1A-1 to 1A-20. The first speaker (Sp1) selected an ID number, e.g. 1A-2, and asked the second speaker, "1A no 2 wa nandesuka? (What number is 1A-2?)" The second speaker (Sp2) searched for the number identified as 1A-2 and answered, e.g., "15765 desu. (It is 15765.)" Sp2 then continued the dialog by asking, e.g., "1A no 5 wa nandesuka?" Thus, the dialog progressed as follows: Sp1 questions - Sp2 answers - Sp2 questions - Sp1 answers - and so on. The speakers were told to continue the dialog until the experimenter asked them to stop. About 60 Q&A utterances were recorded for each dialog, out of which the first 50 utterances, 25 belonging to each speaker, were analyzed in the present study.

Eighteen dialogs were recorded. The speakers were fourteen Japanese male university students (age: 21-25). Most of the speakers took part in more than one dialog. The speaker combination was different for sixteen out of eighteen dialogs. For the two remaining dialogs, the speaker combination was the same but the order of utterance was different, i.e., Sp1 in one dialog became Sp2 in the other. The participants were told (a) to speak naturally, (b) not to worry about making mistakes, and (c) that there was no need to hurry unnecessarily in searching the list.

2.2. Recording

The dialogs were recorded in a sound-proof recording booth (Rion, AT-80) using a digital audio taperecorder (Sony , TCD-D10) and headset microphones (Sennheiser, HMD25-1). The speech signals were sampled at 22kHz, digitized with a 16bit AD-converter, and stored in a personal computer (Apple, Macintosh 8100/AV). The speech signals were analyzed and measured by use of a speech analysis software (G. W. INSTRUMENTS, SoundScope/16).

2.3. Measurements

F0 was measured at a selected location in each question utterance, i.e. the vowel portion of /ichi<u>e:</u>/ ("1A"). This was because (a) "1A" portion was common to all the question utterances, (b) it had "high" (as opposed to "low") pitch level and was pronounced clearly, (c) it did not bear emphatic stress, which, when placed, was on the one/two digit number that followed "1A". In this study, the F0 value obtained for this

particular portion of the utterance was considered the representative F0 value of that utterance. For each dialog, 50 F0 values were obtained, 25 for each speaker.

3. RESULTS AND DISCUSSION

3.1. The Pattern of Two Speakers' F0 Change Over the Course of a Dialog.

Figure 1 shows the three typical patterns of F0 interaction throughout the course of dialogs. In each plot, the vertical axis gives the F0 values (Hz) and the horizontal axis gives the utterance numbers (1-50). The filled circles represent the first speaker in the dialog, while the unfilled circles represent the second speaker. Straight lines are fitted through the data points by using linear regression.

Out of 18 dialogs analyzed in the present study, 6 dialogs showed the (a)-pattern, the two speakers' F0 converging as the dialog proceeded. In contrast, 8 dialogs showed the (c)-pattern, the two speakers' F0 diverging as the dialog proceeded. The degree of convergence/divergence differed across dialogs. The 4 remaining dialogs showed the (b)-pattern, the difference between the two speakers' F0 remaining unchanged throughout the dialog. (In the present study, an F0 difference below 3 Hz was regarded as "no change" on the basis that jnd in frequency is generally reported to be 2-3 Hz below 1000 Hz [1,7].)

As was mentioned in the Introduction, it was expected on the basis of previous experimental results [5,6] that the two speakers' F0 would most likely converge over the course of a dialog. However, this was not the only pattern observed; there were cases where two speakers' F0 diverged, as well as cases where they maintained a parallel relation throughout the dialog.

3.2. What Determines the F0 Interaction Patterns in Dialogs?

The finding that F0 interaction was manifested in varying ways led to the second question: what determines the pattern of F0 interaction in a dialog? The simplest assumption is that the initial state determines subsequent interaction.

In **Figure 2**, the difference between the dialog-initial F0 difference of the two speakers and the dialog-final F0 difference of the two speakers [dF0(ini) - dF0(fin), vertical axis] is plotted against the dialog-initial F0 difference between the two speakers [dF0(ini), horizontal axis]. F0(ini) and F0(fin) are derived from the results of linear regression that yielded regression lines. The positive F0 difference along the vertical axis indicates that the two speakers' F0 converged, while the negative values indicate that they diverged. Zero indicates that the difference between the two speakers' F0 remained unchanged during the dialog, or , in other words, showed a parallel change.

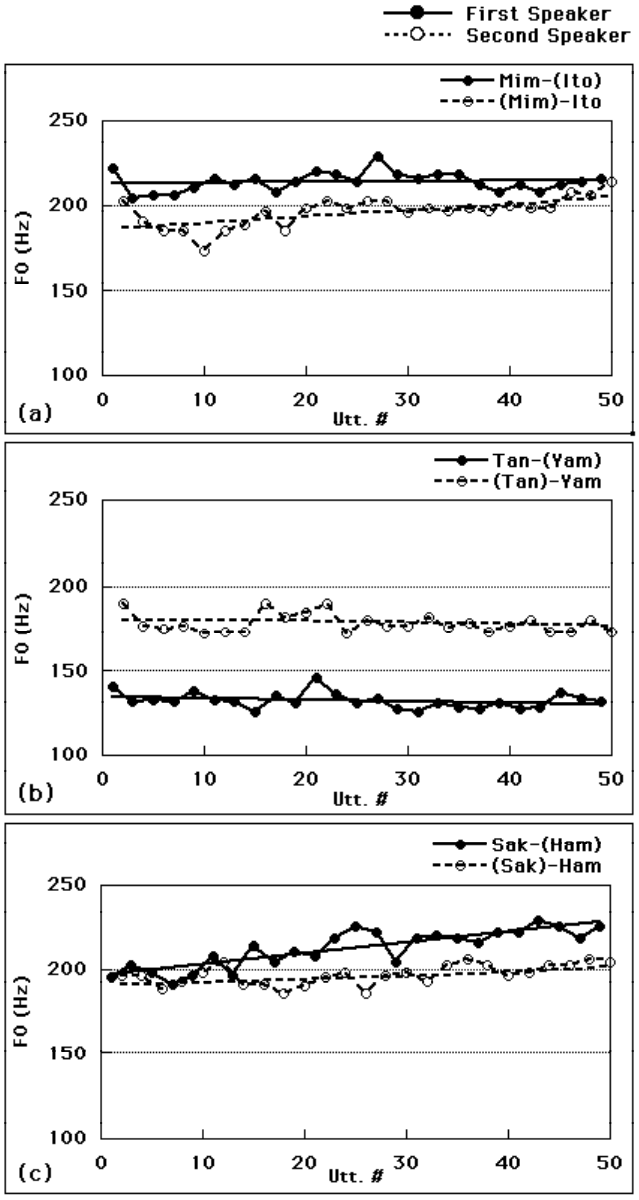

Figure 1: Three typical patterns of F0 change of the two speakers during dialogs (a: convergence, b: parallel change, c: divergence). In each plot, the vertical axis gives the F0 values (Hz) and the horizontal axis gives the utterance numbers (1-50). The filled circles represent the first speaker in the dialog, while the unfilled circles represent the second speaker. Straight lines are fitted through the data points by using linear regression.

The following are the major observations made.

- When dF0(ini) is small, i.e. roughly below 5 Hz, the subsequent F0 interaction is realized mainly as divergence. The degree of divergence seems to be negatively correlated to the degree of initial difference, i.e. the smaller the initial difference, the greater the divergence and vice versa.

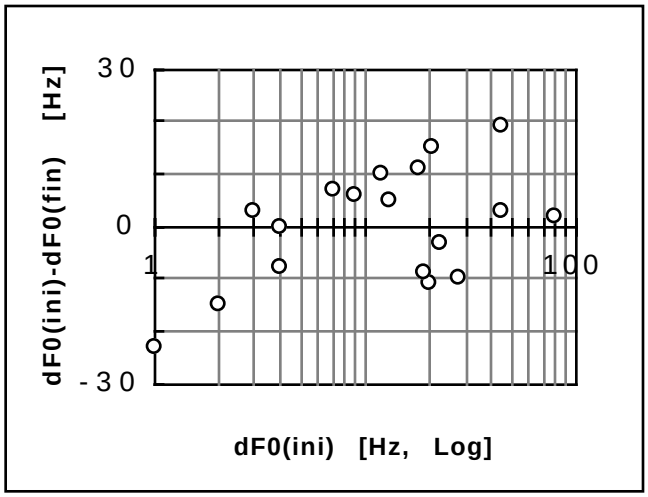

Figure 2: The difference between the dialog initial F0 difference of the two speakers and the dialog final F0 difference of the two speakers [dF0(ini) - dF0(fin), vertical axis] plotted against the dialog-initial F0 difference between the two speakers [dF0(ini), horizontal axis].

- When dF0(ini) exceeds a certain amount, i.e. roughly 5 Hz, the cases of convergence begin to prevail. Generally speaking, the degree of convergence is positively correlated to the degree of dF0(ini), i.e. the greater the dF0(ini), the greater the convergence and vice versa.
- Divergence is observed for 5 Hz < dF0(ini) region, too, but within a limited range, i.e. around 20 Hz - 30 Hz. The amount of divergence in this region is, on the whole, smaller compared with those in the dF0(ini) < 5 Hz region.
- In the 5 Hz < dF0(ini) < 20 Hz region, convergence is the only pattern observed. There is no instance of divergence or parallel change in this region.
- The cases in which the difference between the two speakers' F0 changes little or does not change at all during the dialog -- see data points close to y = 0 line [y = dF0(ini) - dF0(fin)] -- are observed both in the dF0(ini) < 5 Hz and in 20 Hz < dF0(ini) regions but not in between.

The above results may be interpreted in the following way.

When dF0(ini) is very small, there is little room for the two speakers' F0 to get closer. In principle, then, the possible F0 interaction would be either parallel change or divergence. The present results confirm this point, and suggest that the region to which this rule applies is dF0(ini) < 5Hz.

When, on the other hand, dF0(ini) is large enough, the two speakers' F0 should theoretically be able to converge, diverge, or remain parallel to each other. The present results are in general agreement with this prediction, and all three patterns of F0 interaction are observed for the 5Hz < F0(ini) region.

However, a closer look reveals that convergence cases are missing in 5Hz < dF0(ini) < 20 Hz and 30 Hz < dF0(ini) regions.

The lack of divergence data in the 30 Hz < dF0(ini) region may indicate that when the two speakers' F0 are too far apart to start with, they will not become any farther. Two out of three data points in this region are found close to y = 0 line suggesting that when two speakers' F0 are too far apart at the start of the dialog, the two speakers' F0 difference is likely to remain more or less unchanged throughout the dialog.

The lack of divergence data in the 5 Hz < dF0(ini) < 20 Hz region is both intriguing and interesting. There seems to be no plausible reason for this. If this is not due to an accidental lack of data, it may be possible to hypothesize that there is a preferred range of dF0(ini) that uniquely induces convergence, and that, in the case of F0 interaction, this preferred range is roughly between 5 Hz and 20 Hz. This aspect of F0 interaction will have to be examined in more detail with a larger body of data.

3.3 Partner-Dependent Variation in Dialog-Initial F0

In the present study, 12 out of 14 speakers participated in more than two dialogs. **Figure 3** shows how one speaker's F0 varied in relation to his dialog partners' F0. The F0 data plots in Figure 3 show the median of the first 5 F0 values in each dialog [F0(1-5)]. In each plot, the F0 values of the speaker himself [F0(Self), vertical axis] are plotted against the F0 values of his partners' [F0(Partner), horizontal axis] in different dialogs. The maximum(250 Hz)/minimum(100 Hz) values of the vertical and horizontal axes are common to all 12 plots.

Since the number of dialogs per speaker is not large (n=2-5), the results can only be considered preliminary. The plots show that some of the speakers, e.g. Figure 3(a)-(e), did vary their F0(1-5) according to their partners' F0(1-5). The amount of variation in F0 ranged between 20 Hz and 30 Hz. Other speakers, e.g. Figure 3(i)-(l), showed compartively little partner-dependent F0(1-5) variation in spite of the fact there was a considerable amount of F0(1-5) variation across partners.

4. CONCLUSIONS

In order to examine the manner in which two speakers' F0 interacted, controlled dialogs of a simple question-answer type were recorded, and the changes in F0 over the course of dialogs were analyzed. The major results of the present study may be summarized as follows. (1) All three possible patterns of F0 interaction were observed, i.e., the two speakers' F0 either converged, diverged, or maintained a parallel relation to each other over the course of dialogs. (2) Dialog-initial difference in the two speakers' F0 values was closely related to the

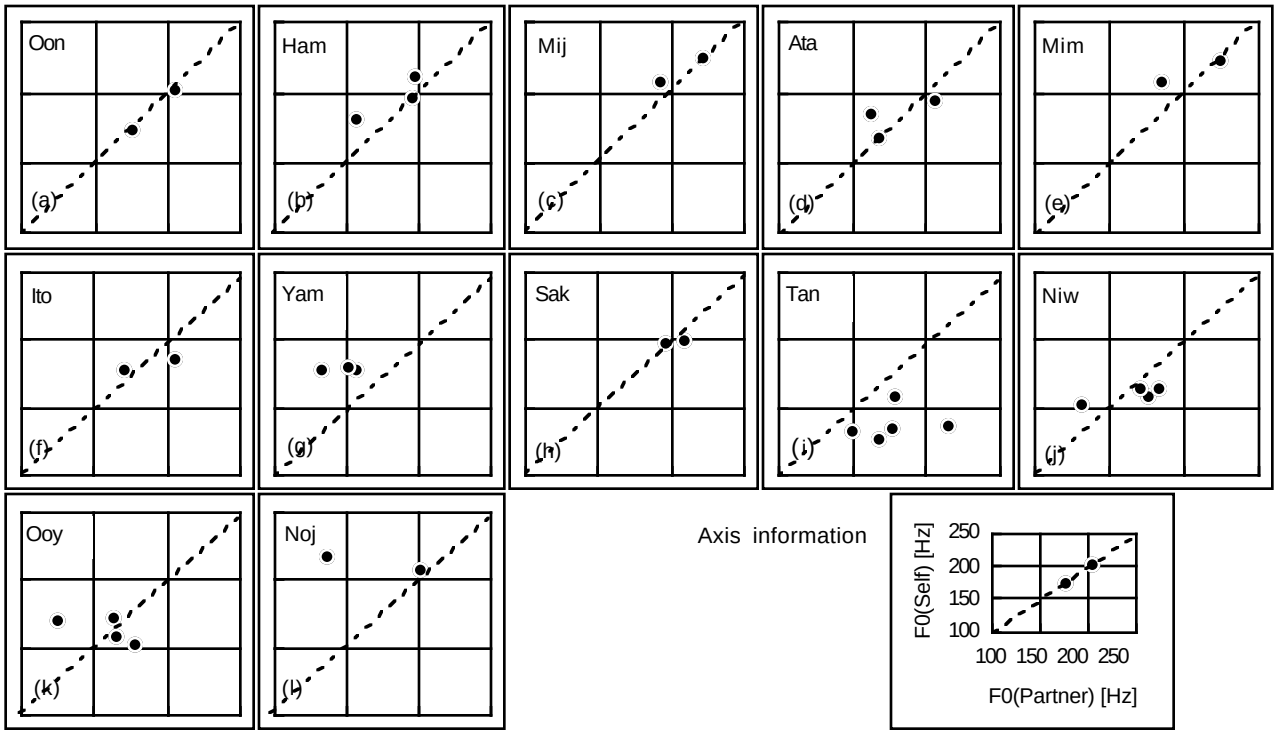

Figure 3: The F0 values (the median of the first five F0 values in each dialog) of the speaker himself [F0(Self), vertical axis] plotted against the F0 values of his partners [F0(Partner), horizontal axis] in different dialogs. The maximum (250Hz) and minimum (100Hz) values of the vertical and horizontal axes are common to all 12 plots.

consequent overall pattern of F0 interaction during dialogs. (3)Some of the speakers reorganized their dialog-initial F0 [F0(ini)] according to their partner's F0(ini), while others were more or less unaffected by their partners' F0(ini).

5. ACKNOWLEDGMENTS

This work was partly supported by the Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science and Culture, Japan (No. 06232213). I am grateful to Professor Yuki Kakita for allowing me the use of the sound-proof recording booth at Human Intelligence Laboratory, Kanazawa Institute of Technology.

6. REFERENCES

1. Denes, P. B. and Pinson, E. N., *The Speech Chain: The Physics and Biology of Spoken Language* Anchor Press, New York, 1963.
2. Feldstein, S. and Welkowitz, J., "A chronography of conversation: In defense of an objective approach" In *Nonverbal behavior and communication*, edited by A. W. Siegman and S. Feldstein, 329-378, Erlbaum, Hillsdale, NJ, 1978.
3. Imaizumi, S., Hayashi, A. and Deguchi, T., "Listener adaptive characteristics of vowel devoicing in Japanese dialogue", *J. Acoust. Soc. Am. Vol. 98*, 768-778, 1995.
4. Jaffe, J. and Feldstein, S., *Rhythms of dialogue*, Academic Press, New York, 1970.
5. Kakita, K., "Inter-Speaker Interaction of the Duration of Sentences and Intersentence Intervals", *Proceedings of the Fifth Australian International Conference on Speech Science and Technology, Vol. 1*, 34-38, 1994.
6. Kakita, K., "Inter-Speaker Interaction in Speech Rhythm: Some Durational Properties of Sentences and Intersentence Intervals", *Proceedings of the ICSLP 94, Vol. 1*, 131-134, 1994.
7. Ladefoged, P., *Elements of Acoustic Phonetics, 2nd Ed.*, Univ. of Chicago Press, Chicago, 1996.
8. Matarazzo, J. D., Weitman, M., Saslow, G. and Wiens, A. N., "Interviewer influence on durations of interviewee speech", *J. Verbal Learn. Verbal Behav., Vol. 1*, 451-458, 1963.
9. Webb, J. T., "Interview Synchrony. An Investigation of Two Speech Rate Measures in an Automated Standardized Interview", In *Studies in Dyadic Communication: Proceedings of a Research Conference on the Interview*, edited by A. W. Siegman and B. Pope, 115-133, Pergamon Press, New York, 1970.