

Influence of Personal Characteristics on Nonverbal Information for Estimating Communication Smoothness

Yumi Wakita^(✉), Yuta Yoshida, and Mayu Nakamura

Osaka Institute of Technology, Osaka, Japan
yumi.wakita@oit.ac.com

Abstract. To realize a system that can provide a new topic of discussion for improving lively and smooth human-to-human communication, a method to estimate conversation smoothness is necessary. To develop a process for estimating conversation smoothness, we confirmed the effectiveness of using fundamental frequency (F0). The analytic results of free dyadic conversation using the F0 of laughter utterances in conversation are strongly dependent on personal characteristics. Moreover, F0s without laughter utterances are effective in estimating conversation smoothness.

Both the average value and the standard deviation (SD) value of F0s in smooth conversation tend to be higher than in non-smooth conversation. The differences between the SD of the “smooth” and “non-smooth” segments are shown to be significant when using a t-test, where the confidence level is 95 %.

Keywords: Conversation smoothness · Nonverbal information · Fundamental frequency

1 Introduction

Recently, with more senior citizens living alone and reclusive, many communities, companies, or schools think human-to-human communication is very important. They are interested in a system to aid human-to-human communication. The purpose of this study is to explore a system which provides a topic of discussion for carrying on lively and smooth human-to-human communication.

Several studies have been proposed which use systems to guide smooth communications by introducing some appropriate topics [1, 2]. However the communication atmosphere changes over time. The topic, timing, and method of communicating should be changed when considering the communication atmosphere.

We are developing a system which provides information about suitable timing by considering the communication atmosphere. The system can understand whether or not the human-to-human communication is proceeding smoothly. When sensing there has been little progress during the conversation, the system attempts to provide a topic for leading a smoother discussion. For practical use, it is necessary to find factors which are effective for measuring the communication and developing a process of recognizing the communication atmosphere.

2 Conventional Approaches to Estimate the Communication Atmosphere

Several papers illustrated the relationship between communication atmosphere and nonverbal communication [3, 4]. These papers suggest it is possible to estimate communication atmosphere using nonverbal communication. In addition, several methods which can measure the liveliness in human-to-human conversations using nonverbal information have been proposed [5–7]. These papers explain that the system measures liveliness based on whether many people speak at the same time or everyone speaks in turn.

However, more detailed analysis of nonverbal communication is necessary to improve the estimate's accuracy. When thinking deeply or listening attentively, we sometimes keep quiet in free conversation. Even if all member keep quiet, the conversation isn't always deadlocked. Even if only one person continues to speak for a long time, the listeners are not always bored. The conventional methods cannot estimate these communication atmospheres correctly. It is necessary to estimate not only "liveliness," but also "smoothness" to provide a topic of conversation at a suitable time.

On the other hand, in spite of nonverbal communications including several useful factors to understand the communication atmosphere, they have not been used effectively. Several papers suggest that because nonverbal behavior greatly depends on people, it cannot be used. If the system cannot understand whether a nonverbal behavior depends on the person or the communication atmosphere, it would be difficult to estimate the communication smoothness.

We would like to develop a method to estimate conversation smoothness using nonverbal information. To achieve highly accurate estimates, it is necessary to remove personal characteristics from nonverbal information before the estimation begins. For nonverbal information, we selected "fundamental frequency" (F0) and analyzed the relationship between F0 and communication smoothness by considering the factors dependent on personal characteristics.

In this paper, we report the results of our analysis and suggest the most effective factors for estimating conversation smoothness.

3 Ability of Nonverbal Information to Estimate Conversation Smoothness

We made several video recordings of free dyadic conversations and confirmed the probability of the communication smoothness estimate using nonverbal information in the recorded database.

3.1 Conversation Database

We recorded ten sets of three-minute long, free dyadic conversation (between two persons). The conditions are shown in Table 1. All members were not meeting for the first time, but we could make pairs with those who have never spoken before.

Table 1. Conditions of conversation

Number of speakers	7 males, 3 females
Number of conversations	10 (four pairs of males, two pairs of females, four pairs with one female and one male)
Conversation periods	Three minutes
Conversation condition	Free dyadic conversation

We observed video data of two people conversing in two different scenarios. The two scenarios were as follows:

- Smooth conversation (S): the topic had not been decided yet. Speakers searched for a topic which interested both of them.
- Non-smooth conversation (NS): the topic for both of the speakers was already chosen and they spoke smoothly or eagerly.

The results of pairing two people were very similar. The 96 % parts paired both of the speakers in each scenario. Table 2 shows the ratio of “smooth” and “non-smooth” scenarios for all observations.

Table 2. Configuration ratios of each part–“smooth (S)” and “non-smooth (NS)”

Pair	A	B	C	D	E	F	G	H	I	J
S	0.64	0.63	0.82	0.95	0.96	0.96	1.00	1.00	1.00	1.00
NS	0.36	0.37	0.18	0.05	0.04	0.04	0.00	0.00	0.00	0.00

Most “non-smooth” parts were at the beginning part of each conversation. In the case of the pairs A–F, they found a suitable topic in the middle of the conversation, and subsequently, the conversation atmosphere became smoother. In the case of G–J, the speakers explored several topics before initiating their conversations and the conversations were considered “non-smooth” throughout.

We analyzed the A–F conversations by comparing the “smooth” and “non-smooth” scenarios. At first, we noted the “silent interval length (no speech)” and “length of one utterance.”

Table 3. Ratio of silent interval length for each scenario

Pair	A	B	C	D	E	F
S	0.14	0.38	0.23	0.13	0.20	0.25
NS	0.20	0.33	0.08	0.33	0.16	0.29

Table 4. Average length of one utterance [Sec.]

Pair	A	B	C	D	E	F
S	2.44	0.89	1.69	2.83	1.72	1.34
NS	1.90	1.04	3.11	1.16	1.63	1.23

In general, it seems that when a conversation progresses smoothly, the silent interval lengths are shorter and utterance lengths are longer. However, the results of Tables 3 and 4 show that both are unrelated to smoothness. The results suggest that other factors should be determined to estimate conversation smoothness.

Examples of silent interval in smoothness are:

- The communication progresses using gestures (hand movements, nodding, etc.)
- Carefully assessing their answers.

Examples of non-smooth conversation intervals are:

- Searching for a conversation topic.
- An ingratiating smile while speaking.

3.2 The Probability of Estimating Conversation Smoothness Using Nonverbal Information

We processed three types of communication data from the original recorded video. The three types of data are as follows:

1. Only images without speech
2. Only speech without images
3. Only nonverbal communication (deleting language information from 2.)

To delete language information, the speech data was processed through low-pass-filter with a cut-off frequency of 300 Hz.

We asked three people to watch or listen to the original video and the three processed conversations, and to select the scenes from each where they could provide a new topic. We also asked that they do not interrupt when they felt the conversation progress was smooth. All the scenes extracted by these three people were included in “non-smooth” parts.

We compared the extracted scenes by using the original video and the processed data. The scenes extracted from the original video were regarded as “correct” and calculated the recall and precision rates of other processed data.

We used only the A–C conversations in Table 2 because the other conversations had very few “non-smooth” segments and we couldn’t select the multiple scenes.

The number of correct scenes from the A–C conversations are three from A, four from B, and five from C. We calculated the recall and precision rates for each processed dataset compared to the original data. Each rate is shown in Figs. 1 and 2. The recall and precision rates are the average of three people.

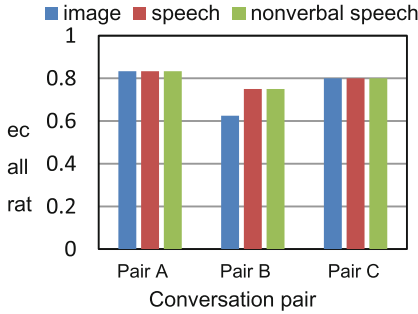


Fig. 1. Recall of each conversation

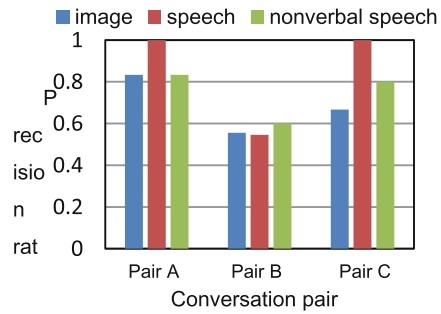


Fig. 2. Precision of each conversation

These figures illustrate the following:

- All of the recall rates are neither perfect nor low. These are all over 60 %.
- The differences of the processes among the pairs are very small.
- The performance of the precision rates depends on the pairs.
- The processed “speech” data has the highest precision rate. However, the differences between using the verbal information and nonverbal information are small for all pairs.
- Comparing nonverbal information, recall and precision rates were higher in “non-verbal speech” than in “image only.”

These results indicate the difference between verbal and nonverbal is small. Both recall and precision are high. The nonverbal speech information can be useful for estimating smoothness. However, the estimation accuracy using only one factor of nonverbal information has limitations. Many factors would be necessary to estimate conversation smoothness perfectly.

4 A Study of Estimating Conversation Smoothness Using Fundamental Frequency (F0)

4.1 The Relationship Between the F0 of Each Utterance and Conversation Smoothness

We analyzed the relationship between the F0 and conversation smoothness. We calculated the averages of F0 of each utterance (Ave-F0), and their standard deviations of F0 (SD-F0) for six speakers from the A-C conversations. Figure 3 shows the average of the Ave-F0 and Fig. 4 shows the average of the SD-F0.

These results show that F0 values do not depend on smoothness. The differences between “smooth” and “non-smooth” are small for all six speakers. The standard deviations of F0, which shows the dynamics of F0 for an utterance, tended to illustrate

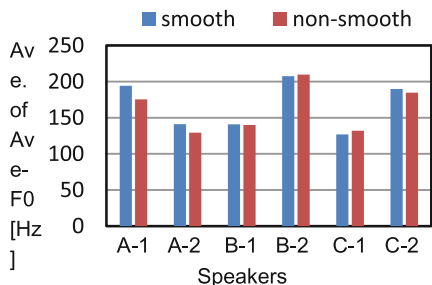


Fig. 3. Average of Ave-F0 for each speaker

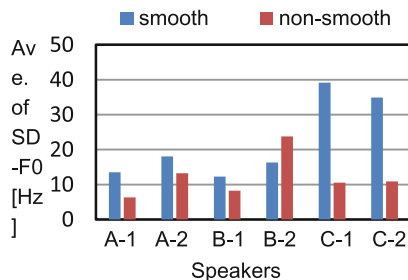


Fig. 4. Average of SD-F0 for each speaker

that the values of non-smooth are smaller than those for smooth. However, this tendency depends on the speakers. By using a t-test, we can confirm that the differences between smooth and non-smooth are not significant.

4.2 The Influence of F0 on Laughter Utterance for Estimating Smoothness

The conversation data include several utterances of laughter. Figure 5 shows the ratio of laughter utterances in “smooth” or “non-smooth” scenarios to all utterances for each conversation.

The total laughter utterance ratio of both “smooth” and “non-smooth” are about 25 %. The characteristics of laughter utterances should be clear to estimate smoothness correctly.

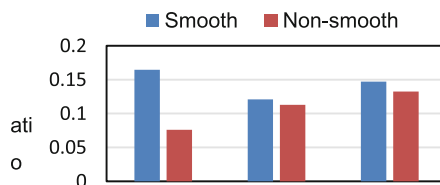


Fig. 5. Ratio of laughter utterances to all utterances

4.2.1 Classification of Laughter in the Data

Nishio and Koyama [8] explained that laughter utterances can be classified in general by “pleasantness” or “sociability.” We classified the laughter into two types—basically, “pleasantness” and “sociability.” However, several occurrences of laughter include speech. We added two more types, “pleasantness with speech” and “sociability with speech.” We asked two people to add a type tag to each utterance of laughter in the data. Before tagging, we explained the meaning of “pleasantness” and “sociability” using Nishio’s paper, and confirmed that they understood. The tagging results by both persons are very similar. Table 5 shows the number of laughter utterances for each type. The numbers in Table 5 are only laughter utterances which were regarded as the same type by both taggers.

Table 5. Number of laugh utterances for each class

Total	Pleasantness		Pleasantness with speech		Sociability		Sociability with speech	
	NS	S	NS	S	NS	S	NS	S
A-1	1	1	1	0	2	7	0	0
A-2	0	0	2	2	0	0	0	3
B-1	6	2	1	2	2	3	0	2
B-2	2	3	0	2	3	1	0	0
C-1	0	5	0	1	5	0	0	0
C-2	0	4	0	0	4	0	0	0

Table 5 shows that “pleasantness” tends to occur in smooth conversation. However “pleasantness” laughs do not always occur in smooth conversation and “sociability” laughs do not always occur in non-smooth conversation. The relationship between the laugh types and smoothness depends on the speakers.

4.2.2 Analysis of Laughter Utterances

The number of laughs for each speaker is not large. But two speakers, 2-A and 2-B, had pleasant laughs in both periods of “smooth” and “non-smooth” conversation. We compared the F0 of the pleasant laughs during “smooth” and “non-smooth” conversation. These results are shown in Figs. 6 and 7.

With both speakers, the F0 of the pleasant laughs in smooth conversation is higher than in a non-smooth one. However the tendency of the standard deviations depends on the speakers.

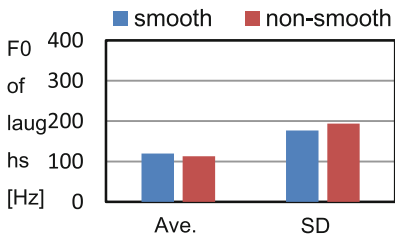


Fig. 6. Comparing the F0 of laughs during “smooth” and “non-smooth” conversation by Speaker B-1

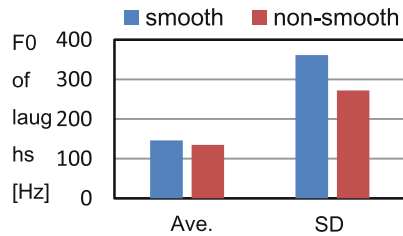


Fig. 7. Comparing the F0 of laughs during “smooth” and “non-smooth” conversation by Speaker B-2

Figure 8 shows the distribution of Ave-F0 and SD-F0 values of two speakers. The results of a t-test show the difference between “smooth” and “non-smooth” is not significant, but the difference between the two speakers is significant when the confidence level is 95 %. These results suggest that the laughter utterances are not useful for estimating smoothness. We should remove these from utterance observations before estimating smoothness.

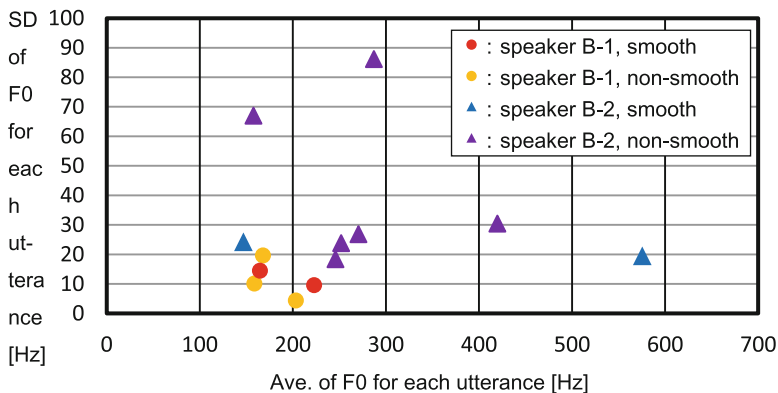


Fig. 8. Distribution of the Ave. and the SD for F0 of pleasant laughs for two speakers

4.3 Relationship Between the F0 of Utterances After Removing Laughter Observations and Conversation Smoothness

As a result of analyzing laughter utterances, we removed the laughter segments from the data and re-analyzed the relationship between the F0 and the conversation smoothness. The results are shown in Figs. 9 and 10.

With all speakers, the average F0s in smooth conversation are higher than in non-smooth conversation. The differences of four speakers are more significant than those of two speakers based on the results of a t-test which has a confidence level of 95 %.

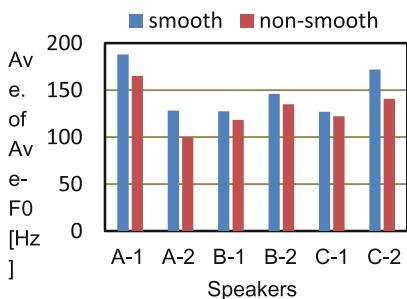


Fig. 9. Average of Ave-F0 for each speaker after removing the laughter segments

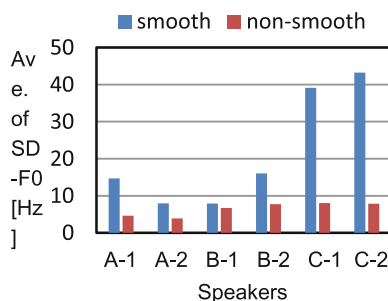


Fig. 10. Average of SD-F0 for each speaker after removing the laughter segments

The standard deviations of F0 in smooth conversation tend to be bigger than those of non-smooth conversations. The differences between “smooth” and “non-smooth” are significant as a result of a t-test which has a confidence level of 95 %.

Figure 11 shows the distribution of Ave-F0 and SD-F0 values of utterances after removing laughter segments for six speakers. The results clearly show that the SD of F0 in smooth conversation is increased more as compared with non-smooth conversation. These results suggest that the SD values of utterances are useful in estimating conversation smoothness. However, it is necessary to remove laughter utterances from the data before making the estimate.

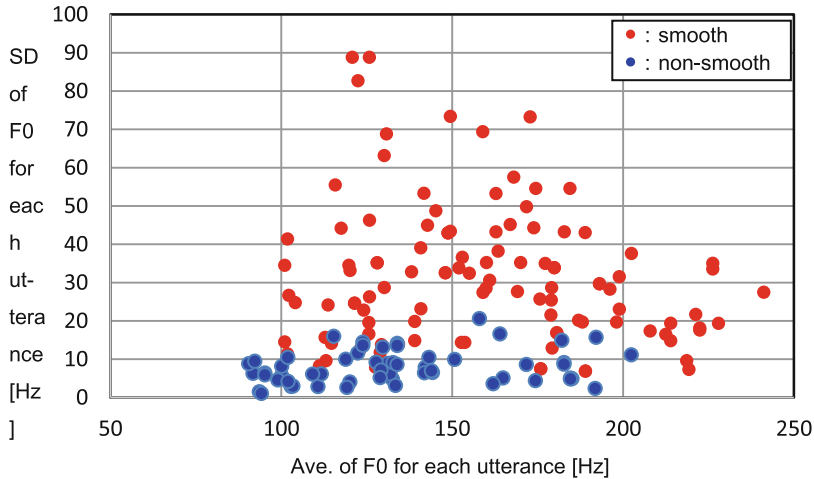


Fig. 11. Distribution between the Ave. and the SD of each utterance after removing laugh parts for six speakers.

5 Discussion

We confirmed the following through analysis of free dyadic conversation data:

- The length of a silent interval and the length of one utterance are not useful for estimating conversation smoothness.
- One may estimate the conversation smoothness using nonverbal information, images of conversation, or nonverbal information with speech removed.

These are our goals for using nonverbal information to estimate conversation smoothness.

Nonverbal information in free conversations includes two kinds of characteristics. One is a factor which depends on the communication atmosphere, and the other depends on personal characteristics. When we use nonverbal information for estimating the communication atmosphere, the personal characteristics should be removed from nonverbal information.

Our analysis results show:

- Laugh utterances depend on speakers more than conversation smoothness
- Standard deviation of the F0 for each utterance is useful for estimating the conversation smoothness.

- It is necessary to remove laughter utterance segments and use only speech utterances for the estimation.

However, there is possibly not enough data to make a clear estimate, especially if we use only two speakers for laugh segment analysis. In the future, the characteristics of laughter observations should be decided more clearly using additional data.

6 Conclusion

We analyzed the relationship between the F0 in free dyadic conversations and conversation smoothness to confirm that the F0 is an effective factor for estimating conversation smoothness. As a result, we confirmed not only that the standard deviation of the F0 for each utterance is useful to estimate conversation smoothness, but that it is also necessary to remove personal, independent utterances such as laughter before making an estimation.

In the future, we will confirm the reliability of our results using a larger quantity of data. In addition, other factors of nonverbal communication such as gestures will be analyzed to obtain a more accurate estimate.

References

1. DiMicco, J.M., Pandolfo, A., Bender, W.: Influencing group participation with a shared display. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, CSCW 2004, pp. 614–623. ACM Press, New York, NY, USA (2004)
2. Kim, T., Chang, A., Holland, L., Pentland, A.S.: Meeting mediator: enhancing group collaboration using sociometric feedback. In: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, pp. 457–466 (2008)
3. Krauss, R.M., Chen, Y., Gottesman, R.F.: Lexical gestures and lexical access: a process model. In: Language and Gesture, pp. 261–283. Cambridge University Press, Cambridge (2000)
4. Wrede, B., Shiriberg, E.: Spotting “hot spots” in meetings: human judgments and prosodic cues. In: Proceeding of INTERSPEECH, pp. 2805–2808 (2003)
5. Gatica-Perez, D., McCowan, I.A., Zhang, D., Bengio, S.: Detecting group interest-level in meetings. IDIAP Research report 04-51, IDIAP, pp. 1–8 (2004)
6. Jayagopi, D.B., Ba, S., Odobez, J.M., Gatica-Perez, D.: Predicting two facets of social verticality in meetings from five-minute time slices and nonverbal cues. In: Proceedings of the 10th International Conference on Multimodal Interfaces, pp. 45–52 (2008)
7. Toyoda, K., Miyakoshi, Y., Yamanishi, R., Kato, S.: Estimation of dialogue moods using the utterance intervals features. In: Watanabe, T., Watada, J., Takahashi, N., Howlett, R.J., Jain, L.C. (eds.) Intelligent Interactive Multimedia: Systems and Services - IIMSS 2012. Smart Innovation Systems and Technologies, vol. 14, pp. 245–254. Springer, Heidelberg (2012)
8. Nishio, S., Koyama, K.: A criterion for facial expression of laugh based on temporal difference of eye and mouth movement. IEICE **80**(8), 1316–1318 (1997)