Communication Smoothness Estimation Using F0 Information

Yumi Wakita^{a,*}, Shunpei Matsumoto^a

^aOsaka Institute of Technology, 5-1-16 Oomiya Asahi-ku, Osaka, Japan

*Corresponding Author: yumi.wakita@oit.ac.jp

Abstract

The purpose of this study is to explore a system that provides a topic of discussion for carrying on lively and smooth human-to-human communication. When the system senses a slight progress in the conversation, it attempts to provide a topic for leading a smoother discussion.

To develop a process for estimating conversation smoothness, we confirmed the effectiveness of using fundamental frequency (F0). For each utterance, we calculated F0's average values (Ave_F0) and standard deviation values (SD_F0). Then, we compared these values between "smooth" and "non-smooth" utterances. The differences are shown to be significant when using a t-test, where the confidence level is 95%. Linear discriminant analysis (LDA) was applied to the distribution between Ave_F0 and SD_F0 for classifying all utterances into "smooth" and "non-smooth". The classification rates, that is, "the percentage of correctly classified" are over 70% for four speakers.

If laughter sounds are included in the utterances, the estimation rates decreased drastically. For high performance, excluding the laughter parts from utterances is necessary. We confirmed Ave_F0 and SD_F0 values to be effective in detecting the laughter sounds. The t-test, where the confidence level is 95%, was applied on the distribution of Ave_F0 and SD_F0 for each speaker. As a result of the t-test, the difference between the laughter parts and speech parts are significant for all four speakers.

Keywords: communication smoothness estimation, fundamental frequency.

1. Introduction

Recently, with more senior citizens living alone and reclusive, many communities, companies, or schools are interested in a system to help human-to-human communication. The purpose of this study is to explore a system that provides a topic of discussion for carrying on lively and smooth human-to-human communication. For example, a topic providing terminal setting on a table support a discussion around the table.

When sensing slight progress during the conversation, the system attempts to provide a topic for leading a smoother discussion. For practical use, finding factors effective for measuring communication and developing a recognition process for the communication atmosphere is necessary.

We define a communication in which both speakers have already found a topic of common interest, and they speak smoothly or eagerly as "Smooth (S)". However, an atmosphere in which the topic has not been decided yet and speakers search for a topic that interests them both is "Non-Smooth (NS)". The system decides whether the atmosphere is "S" or "NS" by using nonverbal factors of speech such as intonations, frequency characteristics, and lengths of utterance.

We already confirmed that F0 is able to estimate smoothness. SD of F0 for each utterance in "S" conversation is greater than that of "NS" utterance. The difference between "S" and "NS" is shown to be significant when using a t-test, where the confidence level is 95%. When the utterance database includes laughter sounds, the difference between "smoothness" and "non-smoothness" is not significant. Excluding laughter parts from utterances for smoothness degree estimation is necessary ⁽¹⁾.

In this paper, we describe the method to exclude laughter parts from utterances and the method of estimating the conversation smoothness degree by using F0's information of speech parts without laughter sounds.

2. The Characteristics of laughter sounds

2.1 Classification of laughter parts

Nishio et al.⁽²⁾ explained that laughter utterances can be classified in general by “pleasantness” or “sociability.” We classified laughter into two types, “pleasantness” and “sociability.” However, several occurrences of laughter include words. We added two more types, “pleasantness with speech” and “sociability with speech.”

We extracted the laughter parts from a free dyadic conversation between two persons and classified laughter parts into the four abovementioned types by listening to each part. Table 1 shows the conditions of our free dyadic conversation database. Table 2 shows the number of laughter utterances for each class in all conversations of six speakers out of 10.

Table 1. Condition of our free dyadic 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

Table 2. Number of laughter utterances for each class

Total	Pleasantness		Pleasantness with speech		Sociability		Sociability with speech	
	NS	S	NS	S	NS	S	NS	S
Speaker								
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

2.2 Difference between speech parts and laughter parts

Ito et al.⁽³⁾ explained the method to detect a smile and laughter using smile facial expressions and laughter sound

using a Gaussian Mixture Model (GMM) model. According to his report, the recall and precision rate of laughter sound detection is approximately 70%. The evaluation was done using the same speakers, both learning GMM and evaluation.

Table.2 shows that we often talk while laughing. The rate at which laughter parts occurred with speech was approximately 25% of all laughter parts. In these cases, the frequency characteristics would depend on the words more than the laughter’s characteristics. Selecting independent factors from words for detecting laughter parts is necessary.

We tried to use F0 as a suitable factor for laughter parts detection. We analyzed the average and SD values of F0s by each utterance using the four speakers’ database.

We analyzed Ave_F0 and SD_F0 for each utterance and calculated the averages of Ave_F0s and SD_F0s for each speaker. Figure1 shows the averages of Ave_F0s, and Figure2 shows the average of SD_F0s. These figures illustrate the following:

- Ave_F0s of laughter parts tend to be higher compared to those of speech parts.
- SD_F0s of laughter parts are different from those of speech parts. However, the tendency depends on the speakers.

Fig.1 The Average of F0 for each utterance

Fig. 2 The SD of F0 for each utterance

The tendency of the difference of the SD values between laughter parts and speech parts depends on the speakers. Figures 3-1 and 3-2 show the distribution between Ave. and SD of each utterance of two speakers Male 1 and Male 3. In the case of Male1, the SD values of the laughter parts are lower than those of the speech parts. However, in the case of Male 3, the reverse holds true. The boundary between the laughter parts and speech parts would be clear for both the distribution figures.

Fig.3-1 Distribution of the Ave_F0 and the SD_F0 for Male 1.

Fig.3-2 Distribution of the Ave_F0 and the SD_F0 for Male 3

The t-test was applied on the distributions of Ave_F0 and SD_F0 for each speaker. As a result of the t-test, which has a confidence level of 95%, the difference between the laughter parts and speech parts is significant for all four speakers.

From Section 2, we confirm the following:

- When the laughter parts are classified into the

four classes shown in Table 2, the classes are not independent on conversation smoothness degree

- Comparing the distribution of Ave_F0 and SD_F0 between laughter parts and speech parts, the difference between the laughter parts and speech parts is significant for all four speakers.
- This tendency of the difference between laughter parts and speech parts depends on speakers

3. Conversation Smoothness Estimation

3.1 Comparison of F0 characters between smooth and non-smooth conversations

From Section 2, we see that the characteristics of laughter parts are different from those of speech parts, but this depends on the speakers not on the degree of conversation smoothness. This result suggests that when estimating the conversation smoothness, laughter parts and speech parts should be separated.

Fig.4 shows the distribution between Ave.F0 and SD of four speakers for each speech part.. Fig. 5 shows the distribution of each laughter part for the four speakers, same as in Fig.4.

Fig.4 shows clearly that SD of F0 in smooth conversation is greater than that in non-smooth conversation. The average of F0 in smooth conversation is slightly greater than that in non-smooth conversation. The t-test was applied on the distribution of Ave_F0 and SD_F0 for each speaker. As a result of the t-test, which has a confidence level of 95%, the difference between the non-smooth and smooth is significant for all four speakers. However, in the case of laughter parts shown in Fig. 5, the SD values of F0 in a smooth conversation are higher than those in the non-smooth conversation, but the difference between the non-smooth and smooth is not clear.

Fig.4 Distribution between the Ave.F0 and the SD of each speech parts.

Fig.5 Distribution between the Ave.F0 and the SD of each laughter parts

The results show that for the estimation of conversation smoothness, the laughter parts are not useful. After removing the laughter parts from the dialogue database, only the speech parts should be used for conversation smoothness estimation.

3.2 Conversation smoothness estimation using SD of F0

Linear discriminant analysis (LDA) was applied to the distribution between Ave.F0 and SD for each speech part for classifying all utterances to “smooth” or “non-smooth”. LDA class was developed with three speakers and evaluated the performance of the other speaker. The performance was

evaluated four times by changing the speaker for developing the LDA class. Table 3 shows the results of the classification rate of non-smooth and smooth. The classification rates denote the percentage of correctly classified speakers. All the classification rates are over 70%.

Table 3. Classification rate of non-smooth and smooth.

Speaker for evaluation	The classification rate (%)		
	Smooth	Non-smooth	Overall
Female 1	75	100	87.5
Male 1	71.4	100	85.7
Male 2	85.7	88.3	87
Male 3	86.6	88.9	87.6

4. Conclusions

To develop a process for estimating conversation smoothness, we analyzed the relationship between F0 in free dyadic conversations and conversation smoothness. We confirmed that F0s of the speech parts, without laughter sounds, are effective for estimating conversation smoothness. For detecting laughter parts, F0 is also effective.

In 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 estimation.

Acknowledgment

This work is supported by JSPS KAKENHI Grant Number 16K01293.

References

- (1) Yumi Wakita, Yuta Yoshida, Mayu Nakamura : “Influence of Personal Characteristics on Nonverbal Information for Estimating Communication Smoothness”, The Proceeding of HCI International 2016
- (2) Shuichi Nishio, Kenji Koyama : “A Criterion for Facial Expression of Laugh Based on Temporal Difference of Eye and Mouth Movement”. IEICE Vol.J80-ANo.8, 1316-1318,1997
- (3) Akinori Ito, Xinyue Wang, Motoyuki Suzuki, and Shozo Makino : “Smile and laughter recognition using speech processing and face recognition from conversation video”, International Conference on Cyberworlds 2005, pp.41-46, 2005

|