Automatic Detection of the Chewing Side Using Two-channel Recordings under the Ear

Akihiro Nakamura,
Hiroshi Mineno, Masafumi Nishimura
Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu, Shizuoka, 432-8011, Japan
cs16067@s.inf.shizuoka.ac.jp
mineno@inf.shizuoka.ac.jp
nisimura@inf.shizuoka.ac.jp

Takato Saito, Daizo Ikeda, Ken Ohta
Research Laboratories, NTT DOCOMO, Inc.
3-5 Hikari-no-oka, Yokosuka, Kanagawa, 239-8536, Japan
takato.saitou.bu@nttdocomo.com
ikedad@nttdocomo.com
ootaken@nttdocomo.com

Abstract—**Eating behavior is an important parameter of the state of health. A previous study confirmed that the method of recording eating sounds under the ear along with the long short-term memory-connectionist temporal classification (LSTM-CTC) were effective in detecting chewing events. This study examined the possibility of identifying the left and right sides of chewing to improve the analytical ability of eating behavior. More accurate detection was achieved through the utilization of the two-channel recordings and their cross-correlation as a new feature than through the conventional mel-frequency cepstral coefficients (MFCC) features.**

Keywords— ***Cross-correlation, LSTM-CTC, Event Detection, Chewing***

I. Introduction

A decrease in the quality of eating behavior may adversely affect health. The reduction of meal quantity is associated with the lowering of a person's immunity and overeating that causes conditions such as obesity and lifestyle illness. Chewing behavior is particularly important, and people who chew less frequently and eat quickly are more prone to obesity. Therefore, the study of a system that can monitor a series of eating behaviors in detail is vital from the viewpoint of health maintenance.

Many studies have examined the monitoring of eating behavior. Ando et al. [1] classified dietary behavior using the Gaussian Mixture Model, but the model used a small amount of data and did not show high performance. The correct label is necessary for each frame of the input in conventional machine-learning, and the cost of labeling is a problem. Graves et al. [2] proposed a model that can learn input and output sequences of different lengths using long short-term memory (LSTM) with connectionist temporal classification (CTC) as the loss function. This method does not require an accurate label for each frame. Billah et al. [3] confirmed that the combination of LSTM-CTC and masticatory sounds recorded through a single-channel condenser microphone under the ear was effective in detecting chewing behavior.

However, no effective method could be devised to automatically detect the balance of the chewing. Partial chewing can cause tooth loss, facial distortion, etc. Shogetsu et al. [4] investigated the possibility of detecting the chewing side using a myoelectric potential sensor and a microphone and confirmed the existence of a difference in sounds between left and right chewing. To analyze the quality of eating behavior in terms of the balance of the chewing, this paper proposes a method that can detect the chewing side automatically and examines its performance.

II. Proposed Method

The collected eating sounds were utilized to classify each event into the following three divisions: left chewing, right chewing, and others. The last category included front chewing and swallowing.

Fig. 1 shows the flow of the automatic detection of left and right chewing proposed in this paper.

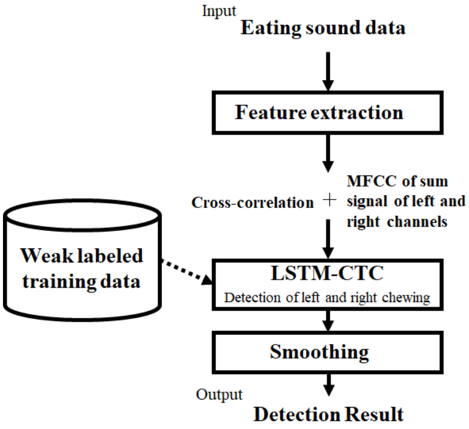
Input
Eating sound data
Feature extraction
Cross-correlation + MFCC of sum signal of left and right channels
Weak labeled training data
LSTM-CTC
Detection of left and right chewing
Smoothing
Output
Detection Result

Fig. 1. Flow of automatic detection of left and right chewing

First, the eating sound was recorded using a two-channel condenser microphone under the ear (16 bit, 22 KHz sampling). The installation example is shown in Fig. 2. In addition, the LSTM-CTC was trained with weak labels for left and right chewing. An online application was developed to reduce the cost of labeling. The subjects generated an event log by pressing keys on the keyboard corresponding to left/right chewing and swallowing at the same time as eating, and the log was used to denote the weak labels. The weak labels did not carry the correct time information.

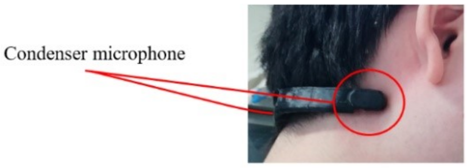

Fig. 2. Installation of the two-channel condenser microphone under the ear

Next, the feature extraction was accomplished through the use of two-channel signals recorded under the ear to detect the chewing side. The MFCC of the sum signal of the left and right channels was estimated as an enhancement processing of the observation signal. The MFCC was estimated through the window width of 80 ms and the frame shift of 40 ms, and it was combined with the cross-correlation of two-channel signals.

The features were then inputted into the LSTM-CTC, an automatic detector, and the left and right chewing sequences were estimated. Since the length of the input/output sequence was different, learning through the weak labels without the time series information became possible in the LSTM-CTC.

Finally, the LSTM-CTC output was smoothed. If there was a change of chewing side in less than one second, it was modified to record chewing only on one side and frequent changes in chewing sides within short spans were avoided.

III. Evaluation

A. Experimental Condition

The eating sounds of chewing gum, crackers (Ritz), and cabbage (shredded) were recorded from 18 subjects in their 20s. The number of collected weak labels was 17691 for chewing (left 8871 times, right 7718 times, front 1102 times) and 1982 for swallowing.

The proposed method was compared with the MFCC of the left signal, concatenation of the two MFCCs of two-channel signals, and MFCC of the sum of the two-channel signals. The MFCC consisted of 12 units with 1 dimensional Root Mean Square (RMS), 13 dimensional Δ and 13 dimensional ΔΔ. Additionally, the cross-correlation was estimated at 3, 7, and 15 point shifts of the two-channel input.

The detection performance was evaluated through the performing of a ninefold cross validation with 16 subjects of learning data and 2 subjects of test data. The mean absolute percentage error (MAPE) of left and right chewing performed in the eating of one cracker was used as an evaluation metric. MAPE is calculated by the following equation, where A_k is the number of correct answers, F_k is the number of estimates, and N is the number of frames for evaluation.

$$\mathrm{MAPE} = \frac{100}{N}\sum_{k}^{N}\left|\frac{A_k - F_k}{A_k}\right|$$

The Recall, Precision, and F-measure of the event unit were also employed for evaluation.

B. Experiment Results and Discussion

Table I demonstrates the average error value per frame of the chewing side when one cracker was eaten. The MFCC of the left signal evinced a large error because it did not include the information of the chewing side. The proposed method (a combination of the MFCC of the sum of the signals and a cross-correlation at the 7-point shift) exhibited the highest performance.

TABLE I. Average error value per frame of chewing side: Three-class (left chewing/right chewing/others) detection

Feature Type	MAPE (%)	
	Before Smoothing	***After Smoothing***
MFCC of the Left Signal	97.51	81.26
Concatenation of 2 MFCCs of Two-ch Signals	70.07	59.99
MFCC of Sum of Two-ch Signals	65.57	55.43
Proposed* Method (15 Point)	48.13	27.81
Proposed* Method (7 Point)	**43.24**	**26.64**
Proposed* Method (3 Point)	55.02	48.53

* Proposed: MFCC of sum of two-ch + Cross-correlation

Table II displays the detection performance of left and right chewing after smoothing. The highest performance was recorded using the proposed method.

TABLE II. Detection performance of left and right chewing after smoothing: Three-class (left chewing/right chewing/others) detection

Feature Type	Recall	Precision	F-measure
MFCC of the Left Signal	0.47	0.62	0.49
MFCC of Sum of Two-ch Signals	0.63	0.80	0.72
Proposed* Method (7 Point)	**0.74**	**0.82**	**0.78**

* Proposed: MFCC of sum of two-ch + Cross-correlation

Table III indicates the detection performance of left and right chewing according to food items in terms of the proposed method. The difference of detection accuracy vis-à-vis the food item variation was small.

TABLE III. Detection performance of left and right chewing by food type using the proposed method

Food Type	Recall	Precision	F-measure
Crackers	0.72	0.79	0.75
Chewing Gum	0.78	0.80	0.79
Cabbage	0.73	0.88	0.80

Fig. 3 illustrates the changes recorded in the F-measure over time in left and right chewing of specific food items. The chewing time was normalized for each food by dividing the whole chewing time into five equal parts. Then, the average of the F-measure for left and right chewing was estimated. It can be said that the chewing of gum evidenced little change in the F-measure with time, and the chewing sound was stable. Conversely, the F-measure for crackers and cabbage tended to be higher in the first half of chewing and lower in the latter half. Although the left or right correct labels were accorded to each chewing side according to the subject's declaration, it was considered that the latter half of the chewing was affected by the fact that the chewing position became unclear due to the expansion of the food.

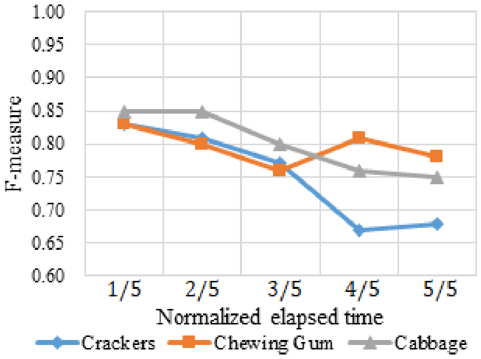

Fig. 3. Relationship between normalized elapsed time and F-measure using the proposed method

IV. Conclusion

The present study posits a mechanism for automatic detection of left and right chewing. The results of the experiment reveal that the method using the combination of the MFCC of the sum of signals and cross-correlation obtained high accuracy, and the objective of the automatic detection of left and right chewing was achieved to some extent.

Future investigations will be undertaken to verify whether the joint use of attention in CTC could enable detection reflecting a complicated action history that includes front chewing and swallowing. Furthermore, the practicability will be examined and the evaluation will be conducted using the actual eating sounds without specifying the food item.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Number 18H03260.

References

[1] Jumpei Ando, Takato Saito, Satoshi Kawasaki, Masaji Katagiri, Daizo Ikeda, Hiroshi Mineno, Takashi Tsunakawa, Masafumi Nishida, Masafumi Nishimura, "Dietary and Conversational Behavior Monitoring by Using Sound Information," NCSP 2018, pp.675-678, 2018.

[2] A. Graves, S. Fernandez, F. Gomez, J. Schmid huber, "Connectionist Temporal Classication:Labelling Unsegmented Sequence Data with Re current Neural Networks," Proc. Int. Conf. on Machine Learning, pp. 369-376, 2006.

[3] Muhammad Mehedi Billah,Taiju Abe, Akihiro Nakamura, Takato Saito, Daizo Ikeda, Hiroshi Mineno, Masafumi Nishimura, "Estimation of Number of Chewing Strokes and Swallowing Events by Using LSTM-CTC and Throat Microphone," Proc. of GCCE 2019, pp.944-945, 2019.

[4] Ryosuke Shogetsu, Tsutomu Terada, Masahiko Tsukamoto, "Consideration to Prevent Bias Chews and Increse Chewing Count with Sensor," IEICE technical report, pp.1-6, 2018. (in Japanese)