INVESTIGATION OF TECHNIQUES FOR EVALUATING TOOTHBRUSHING PERFORMANCE USING AUDIO DATA

Joseph Korpela, ***Advisor:*** *Takuya Maekawa*

Osaka University

ABSTRACT

This work explores the use of audio recognition techniques to evaluate how well users brush their teeth using only audio data collected from smartphones. We investigate two main approaches: a two-tiered approach that first performs activity recognition on the audio data and then uses the output of activity recognition to build a regression model for skill assessment, and an approach that uses deep neural networks to build a regression model for skill assessment that takes raw audio data as input.

1. INTRODUCTION

One important application of ubiquitous computing to our daily lives is its use in healthcare. Using sensor data captured from smartphones and wearable devices, it is possible to evaluate how well users perform activities that relate to their health, including how well they brush their teeth. Evaluating toothbrushing is important, because not only does proper brushing improve a person's overall health, but also most people do not brush properly. Therefore, techniques are needed to provide feedback to users so they can improve their brushing.

The traditional method of evaluation is to be observed by a specialist while brushing your teeth. However, this is an extremely costly method of evaluation, and as such limits access to feedback severely. Recognizing this, prior work has been done on evaluating toothbrushing through the use of ubiquitous computing. However, while these studies have reduced the cost of feedback, they relied on the use of specialized equipment, limiting their accessibility.

In this work, we attempt to offer low-cost and accessible feedback to toothbrushing performance evaluation. We do this by only requiring that users capture audio data while brushing their teeth, with our audio captured using a regular smartphone. This audio data is then evaluated using approaches derived from automatic speech recognition and environmental sound recognition studies. We investigate two main approaches: a two-tiered approach that first performs activity recognition on the audio data and then uses the output of activity recognition to build a regression model for skill assessment, and an approach that uses deep neural networks (DNN) to build a regression model for skill assessment that takes raw audio data as input.

2. RELATED WORK

Several previous studies have been done on analyzing toothbrushing through the use of ubiquitous computing. In [1], Braun's SmartGuide was used to study the effects of real-time feedback on the quality of tooth brushing, in which they found a significant improvement in brushing habits when using this system. Other research has been conducted on the analysis of tooth brushing behavior using optical motion capture systems [2, 3] and embedded accelerometer sensors [4, 5, 6]. In particular, a system developed in [2] used an optical recognition system that encouraged children to brush their teeth by providing feedback on their performance by means of a cartoon display. Regions of the mouth that were adequately brushed were depicted as free of plaque in the cartoon, giving the children simple feedback on their performance. The results of their research indicated a significant improvement in brushing performance as a result of the feedback. Similarly, [4] used an embedded accelerometer to evaluate tooth brushing performance, using graphical feedback to motivate better performance. In each of these systems, specialized hardware was required, such as a specialized toothbrush or an accelerometer. In contrast, we propose a low-cost system built around an off-the-shelf smartphone, which eliminates the need for most users to purchase any new equipment.

3. PROPOSED APPROACHES

3.1. Intermediate activity recognition

Our first approach to evaluating toothbrushing using smartphone audio starts with an intermediate audio recognition process. In this process, we recognize several classes of toothbrushing activities in the audio using an environmental sound recognition technique based on hidden Markov models. These classes correspond to where in the mouth the user was brushing, i.e., front vs. back teeth and inner vs. outer surface, along with the type of brush stroke being used, i.e., rough vs. fine stroke. The output of that process is then used to generate independent variables for support vector machine based regression models. These regression models are trained using evaluation scores assigned by a dentist who participated in our work, giving us regression models that estimate how an expert would have evaluated the users.

Table 1. Error ratios (%) for our *Proposed* method and the baseline method *Avg*.

	Proposed	*Avg*
Total	13.8	22.9
CSD	18.6	25.4

3.2. Deep neural networks

In our first approach to evaluating toothbrushing performance using audio data, we needed to build and train an additional recognition model for our intermediate audio recognition process. This intermediate process added a significant burden when designing and training the skill assessment model, as we needed to design handcrafted features for audio recognition and needed to annotate the mouth locations and brush-stroke types throughout the audio. In our second approach, we are attempting to exploit the ability of DNNs to abstract complex features from raw data to perform skill assessment without an intermediate activity recognition process.

4. RESULTS

The results from our first approach to evaluating toothbrushing performance using audio data were presented in [7]. We evaluated our method using 94 sessions of toothbrushing audio taken from 14 participants. Evaluation was done using leave-one-user-out cross validation, with models adapted to users using unsupervised user adaptation.

Table 1 shows an excerpt from those results. This table includes results for the *Proposed* method along with results for a baseline method *Avg*. The *Avg* method estimates scores for a user as the average score for all other users. This table includes results from two scoring systems: *Total* and *CSD*. The *Total* score system estimates a single score on a 24-pt scale that gives an overall evaluation of one session of toothbrushing. The *CSD* score system estimates three scores per session of toothbrushing, each on a 8-pt scale. These scores evaluate the users toothbrushing for three different criteria: coverage, stroke, and duration. Coverage refers to how well the user covered their entire mouth, stroke refers to whether they used a proper brush stroke, and duration refers to whether they brushed for a sufficient amount of time. For both scoring systems, we were able to achieve an error ratio significantly lower than that of the baseline. Additionally, the dentists participating in the study felt that these error ratios are acceptable for many use cases of our method.

5. ONGOING WORK

Having completed our study on our first approach to evaluating toothbrushing using audio data, we are now focusing on the second approach. We are currently investigating the performance of several DNN architectures, including deep-feed-forward networks, convolutional neural networks, and long-short-term-memory recurrent neural networks. We are also investigating the use of autoencoders as a means of automated feature extraction from audio data. Our goal is improve on our results in the first approach and to make a larger scale study more cost efficient by removing much of the work needed to process the audio collected.

6. ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP16J01917.

7. REFERENCES

[1] Karen Janusz, Bruce Nelson, Robert D Bartizek, Patricia A Walters, and AR Biesbrock, "Impact of a novel power toothbrush with smartguide technology on brushing pressure and thoroughness," *Journal of Contemporary Dental Practice*, vol. 9, no. 7, pp. 1–8, 2008.

[2] Yu-Chen Chang, Jin-Ling Lo, Chao-Ju Huang, Nan-Yi Hsu, Hao-Hua Chu, Hsin-Yen Wang, Pei-Yu Chi, and Ya-Lin Hsieh, "Playful toothbrush: ubicomp technology for teaching tooth brushing to kindergarten children," in *SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2008, pp. 363–372.

[3] Emi Inada, Issei Saitoh, Yong Yu, Daisuke Tomiyama, Daisuke Murakami, Yoshihiko Takemoto, Ken Morizono, Tomonori Iwasaki, Yoko Iwase, and Youichi Yamasaki, "Quantitative evaluation of toothbrush and arm-joint motion during tooth brushing," *Clinical Oral Investigations*, pp. 1–12, 2014.

[4] Christian Graetz, Jule Bielfeldt, Lars Wolff, Claudia Springer, Karim M Fawzy El-Sayed, Sonja Sälzer, Sabah Badri-Höher, and Christof E Dörfer, "Toothbrushing education via a smart software visualization system," *Journal of Periodontology*, vol. 84, no. 2, pp. 186–195, 2013.

[5] Kyeong-Seop Kim, Tae-Ho Yoon, Jeong-Whan Lee, and Dong-Jun Kim, "Interactive toothbrushing education by a smart toothbrush system via 3d visualization," *Computer Methods and Programs in Biomedicine*, vol. 96, no. 2, pp. 125–132, 2009.

[6] Young-Jae Lee, Pil-Jae Lee, Kyeong-Seop Kim, Wonse Park, Kee-Deog Kim, Dosik Hwang, and Jeong-Whan Lee, "Toothbrushing region detection using three-axis accelerometer and magnetic sensor," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 872–881, 2012.

[7] Joseph Korpela, Ryosuke Miyaji, Takuya Maekawa, Kazunori Nozaki, and Hiroo Tamagawa, "Evaluating tooth brushing performance with smartphone sound data," in *2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 109–120.