Evaluating Tooth Brushing Performance With Smartphone Sound Data

Joseph Korpela, Ryosuke Miyaji, Takuya Maekawa
{joseph.korpela,maekawa}@ist.osaka-u.ac.jp
Graduate School of Information Science and Technology, Osaka University

Kazunori Nozaki, Hiroo Tamagawa
{knozaki,tamagawa}@dent.osaka-u.ac.jp
Osaka University Dental Hospital

ABSTRACT

This paper presents a new method for evaluating tooth brushing performance using audio collected from a smartphone. To do this, we use hidden Markov models (HMMs) to recognize audio data that include various types of tooth brushing actions, such as *brushing the outer surface of the front teeth* and *brushing the inner surface of the back teeth*. We then use the output of the HMMs to build regression models to estimate tooth brushing performance scores, such as *stroke quality of brushing for the back inner teeth* and *duration of brushing for the front teeth*. The scores used to train these regression models are obtained from a dentist who specializes in dental care instruction, with the resulting regression models estimating performance scores that closely correspond to the scores assigned by the dentist.

Categories and Subject Descriptors: J.3 Life and Medical Sciences: Medical information systems; H.3.4 Information storage and retrieval: Systems and software.

Keywords: Tooth brushing; healthcare; smartphone; sound.

INTRODUCTION

A large variety of health care research is being conducted through the use of sensor technologies. For example, many researchers have tried to record lifelog data such as amount of exercise, amount of sleep, and amount of food intake, and used the data for health control and dietary control [31, 7, 26]. Of the many fields of health care that can utilize sensor technology, this research focuses on oral health care. Oral health care is an important topic, as teeth must last a lifetime and cannot be replaced. While prosthetics such as dentures do exist, research indicates that tooth loss still carries a significant impact on one's quality of life, both physically and emotionally [15, 10]. Despite oral health's significant impact on our overall well-being, there is evidence that a significant portion of the population brushes incorrectly [12, 13, 34]. Moreover, while proper tooth brushing can have a positive impact on oral health [12], improper tooth brushing can not only fall short in maintaining oral health, it can have a damaging effect [1].

In recent years, several health care applications have been developed that focus on oral health. For example, Braun[1] has released a commercial product called SmartGuide that uses an embedded sensor to detect the force exerted on the teeth during brushing and uses a timing display on a smartphone screen to both prompt users to cycle through different regions of the mouth and provide immediate feedback when the user applies too much pressure. Other research has been conducted on the analysis of tooth brushing behavior using optical motion capture systems [18, 5] and embedded accelerometer sensors [21, 16, 22, 20, 32]. We introduce them in the related work section in detail.

Since the systems described above relied either on embedding a sensor into the toothbrush or the use of video equipment, their costs were high. Our research proposes a method for evaluating tooth brushing performance built around an off-the-shelf smartphone, which is readily available to the average person. In our proposed system, the user only needs to brush their teeth in the vicinity of their smartphone, e.g., by placing the smartphone on the sink next to them when brushing. The smartphone captures the audio data from their brushing, and then evaluates the performance of the brushing through analysis of that data. For example, it can return a score representing whether the user properly brushed his front teeth. Our system can return scores for each area of the mouth and can also output a total evaluation score for the tooth brushing. In this research, we used a supervised learning technique to conduct the brushing evaluation. Specifically, a dentist provided evaluation scores for the training data, and those scores along with the corresponding audio features were used to construct a recognition model for use in scoring test data. By using training data that has been prepared by a dentist with the necessary specialized knowledge, we were able to build a recognition model that is based on that dentist's knowledge.

Such attempts to quantify daily activities are extremely important in promoting a healthy lifestyle. Take for example applications such as Nike+[2] that records a daily journal of a person's running distance and routes, and allows them to track progress toward their goals and to compete with their friends and share accomplishments via social networking sites. By using the method proposed in this research, a user could record the scores for their tooth brushing on a regular basis,

[1] Braun Oral-B: http://www.oralb.com/products/electric-toothbrush/bluetooth-toothbrush.aspx

[2] Nike+: http://nikeplus.nike.com

UbiComp '15, September 7-11, 2015, Osaka, Japan.
Copyright 2015 ©ACM 978-1-4503-3574-4/15/09...$15.00.
http://dx.doi.org/10.1145/2750858.2804259

allowing the user to track information such as changes in the scores, in order to help motivate them to better brush their teeth. Also, by introducing an element of competition with the user's friends, we believe it is possible to further increase that motivation.

In our proposed method, we estimate scores using regression models built from the results of recognizing tooth brushing actions. First, we label the audio time series data with the tooth brushing actions that were being conducted during different periods. For example, from 89 seconds to 110 seconds after the start of the audio could be labeled "brushing the outer surface of front teeth." Second, we use labeled segments to calculate the independent variables for the regression models used for estimating scores. These independent variables can be values such as the total time for segments labeled as "brushing the outer surface of front teeth." Lastly, we use the regression models to estimate scores for the users' tooth brushing.

The proposed method has the following features.
(1): In order to improve tooth brushing proficiency, it is necessary to point out the weak points in the user's tooth brushing. The proposed method has the ability to detect such weak points (such as "front teeth were not thoroughly brushed"). Specifically, our method outputs a score for each region, e.g., front teeth and back teeth, and also for each evaluation criterion, e.g., stroke and coverage of brushing.
(2): The proposed method includes recognizers based on hidden Markov models (HMMs) [27] for the recognition of tooth brushing actions, but the final goal of the research is to use the output from these models to estimate scores for tooth brushing for different areas of the mouth and/or evaluation criterion. The importance of the tooth brushing actions will vary depending on the score being estimated, e.g., tooth brushing actions corresponding to the front teeth will be more important when estimating scores for the front teeth. In this study, we generate HMM sets that maximize the recognition of the important classes for each score type, using the output of these targeted HMM sets to estimate the scores.
(3): Because the characteristics of the audio obtained from tooth brushing differs between different users and different toothbrush models, the proposed method includes the capability to cope with these differences using model adaptation.

In the rest of this paper, we first introduce related work. Then we propose a method for evaluating toothbrush performance using audio recorded by a smartphone. After that we evaluate our method with 94 sessions of tooth brushing data. To the best of our knowledge, this is the first study that attempts to evaluate tooth brushing performance by solely using sound data. The research contributions of this paper are that: (1) We propose a method for evaluating tooth brushing performance using a machine learning approach. First, a dentist who specializes in tooth brushing instruction assigns scores to training data based on his evaluation of tooth brushing performance. Then, we use this training data to construct a regression model that is capable of estimating scores that match closely to those assigned by the dentist. (2) We propose a method for generating the HMM sets used as the basis for estimating scores for the various criterion related to tooth brushing performance. In this method, we automatically generate separate HMM sets for each score, with each set tailored to increase the performance of its corresponding score. (3) We evaluate the proposed method using 94 sessions of tooth brushing audio data taken from 14 research participants.

RELATED WORK

Environmental sound recognition

There are many ubicomp studies on environmental sound recognition. For example, in [6], bathroom activities such as showering, flushing, and urination are recognized using microphone data. Also, several studies recognize daily activities with microphones in smartphones [25, 29] by recognizing environmental sounds such as vacuuming sounds and the sound of running water.

Gamification in ubicomp

Gamification can be described as the use of game design elements in non-game contexts [9, 33]. For example, Foursquare has a gamified check-in system where users compete for *ownership* of spaces by the frequency that they visit them [24]. Game players are reported to exhibit several emotions [3, 2], such as the desire to attain first place, and gamification permits us to motivate players to accomplish a certain task by stimulating these emotions. The mobile game BudBurst uses a gamified approach to encourage users to collect plant life-stage data [17]. Players gain points and levels within the game by finding and making qualitative observations on plants. After collecting sufficient points, users obtain ranks ranging from "Sprout," with the fewest points, through "Seedling" and "Thriving" to "Deep-Rooted" as the highest rank. In EyeSpy [4], players tag geographic locations with photos or text, gaining in-game points for each location tagged. As listed above, many studies and systems quantify real-world activities to motivate users. In this study, we attempt to quantify toothbrushing activity.

Sensing toothbrushing

In [19], 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 [18, 5] and embedded accelerometer sensors [21, 16, 22, 20, 32]. In particular, a system developed in [5] 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 this research indicated a significant improvement in brushing performance as a result of the feedback. Similarly, [16] 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, in this paper we propose a low-cost system built around an off-the-shelf smartphone.

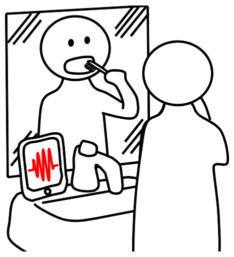
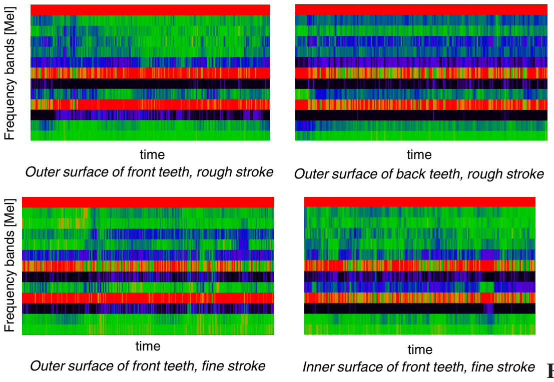


Figure 1. Assumed setup for using a smartphone to record audio from tooth brushing.



Outer surface of front teeth, rough stroke
Outer surface of back teeth, rough stroke
Outer surface of front teeth, fine stroke
Inner surface of front teeth, fine stroke

Figure 2. MFCC representation of audio data from four tooth brushing activity classes.

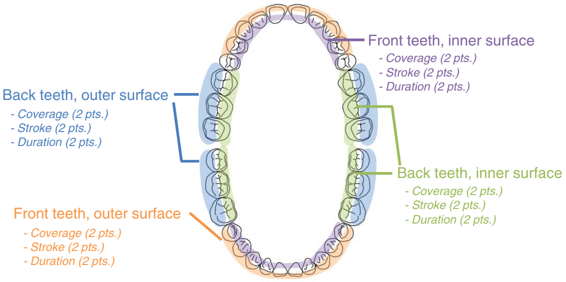


Figure 3. Four regions of the mouth used during evaluation of tooth brushing performance.

TOOTH BRUSHING SENSOR DATA

Assumed environment

In our method, users record the sound of their tooth brushing using their smartphone. Figure 1 shows the assumed setup, where the user places his/her smartphone next to the sink when recording the sound of his/her tooth brushing.

In this research, we extracted features from the raw audio data as vectors of mel-frequency cepstral coefficients (MFCCs). MFCCs were originally designed for use in speech recognition, dividing the frequency spectrum using logarithmically spaced bands, modeling the way in which humans perceive differences in frequencies. Although designed for speech recognition, they have also been successfully applied to environmental sound recognition [6]. Figure 2 shows graphical representations of MFCC data derived from tooth brushing audio. As shown in the figure, the audio characteristics differ when brushing the back teeth from when brushing the front teeth. Similarly, the characteristics also differ depending on the technique (or strength) of the brushing stroke. The quality of a participant's tooth brushing is dependent on their stroke technique and on how evenly they brush all areas of the mouth, e.g., a participant who uses too forceful of a stroke will be at higher risk of damaging their gums and teeth. By using these characteristics of the audio data to recognize which regions of the mouth were brushed along with the brushing technique used, we can facilitate the evaluation of the user's tooth brushing.

Tooth brushing activity

In this study, we used HMMs based on audio characteristics to recognize the following seven activities (referred to as "tooth brushing activities"). The performance of the participant's tooth brushing was then evaluated based on the output from those HMMs.

- Outer surface of front teeth, rough stroke (FO-Rough)
- Outer surface of front teeth, fine stroke (FO-Fine)
- Outer surface of back teeth, rough stroke (BO-Rough)
- Outer surface of back teeth, fine stroke (BO-Fine)
- Inner surface of front teeth, fine stroke (FI-Fine)
- Inner surface of back teeth, fine stroke (BI-Fine)
- No tooth brushing activity (None)

The following two activities were not included in this study, since an insufficient amount of data for these activities were present in the data collected.

- Inner surface of front teeth, rough stroke (FI-Rough)
- Inner surface of back teeth, rough stroke (BI-Rough)

Here, "inner surface" refers to the lingual surface, "outer surface" refers to the facial surface, "front teeth" refers to the incisors and canine teeth, and "back teeth" refers to the molars. The term "rough" indicates that the stroke used when brushing was too forceful, while "fine" indicates that a smaller, lighter stroke was used. (Dentists recommend that a fine stroke, used in brushing methods such as the horizontal scrub and Fones methods, be used when brushing one's teeth, as such a stroke is effective in removing plaque, while a rougher stroke increases the risk of damaging the teeth and gums.) The seven tooth brushing activities listed above were chosen after discussion between the computer science researchers and the dentists participating in this study. They were chosen because they can be differentiated when performing recognition by means of audio data and are important when evaluating the effectiveness of a person's tooth brushing.

During our investigation, a limitation was found in using audio to classify tooth brushing activities. It is difficult for our audio-based approach to distinguish between audio from the left side and the right side of the mouth, e.g., left back teeth vs. right back teeth. It is also difficult to distinguish between audio corresponding to upper teeth and lower teeth, e.g., upper front teeth vs. lower front teeth. Because of this limitation, some issues can arise when scoring a user's tooth brushing. For example, in the case where a user brushes their upper front teeth for a long time, but not their lower front teeth, the scoring of that tooth brushing should be reduced. However, if no distinction can be made between upper front teeth and lower front teeth, then the resulting score can be incorrect. The section entitled *Computing independent variables* contains a detailed discussion on ways to address this issue.

Tooth brushing evaluation by a dentist

Using the audio data collected as described above, we then applied a machine learning approach to evaluate and estimate a score for the user's tooth brushing performance. To do this, we needed training data that could be used to generate score estimates. In this research, a dentist (researching tooth brushing instruction) prepared such training data, allowing for an evaluation of tooth brushing performance that is based on an actual dentist's evaluation. One typical method used by dentists for evaluating tooth brushing is a plaque test. In a plaque test, a dentist applies a plaque indicator liquid to the patient's teeth. This liquid reacts to the patient's plaque, staining it so that the plaque is easily visible. This highlights the plaque left remaining after brushing, which the dentist then uses as the basis for scoring how well the patient brushed. While plaque tests are a typical method of evaluation, preparing a large amount of training data for machine learning using plaque tests would be costly. Additionally, because the scores derived from plaque tests are influenced by the foods eaten prior to testing, the condition of the patient's saliva, and the methods of tooth brushing used in the days preceding the test, plaque tests may not be an ideal test for evaluating isolated sessions of tooth brushing.

Because plaque tests are unsuitable for a machine learning approach to evaluation, we instead evaluated the brushing based on video data. Using the setup illustrated in Figure 1, we recorded video data for each session of tooth brushing using a smartphone. A dentist then evaluated the tooth brushing performance using the video data, and assigned evaluation scores for each session of tooth brushing. These scores were then combined with audio data extracted from the videos to build the score estimation models. Because the dentist evaluated the tooth brushing performance based only on video data, the resulting score was independent of other factors such as what was eaten prior to the test or the condition of the subject's saliva. During evaluation, the dentist assigned scores for each of the four regions of the mouth depicted in Figure 3. The evaluation of each of these four regions was conducted based on the following three criteria:

- **Coverage**: Did the brushing evenly cover the entire region?
- **Stroke**: Was the motion of the brush a fine stroke (good) or a rough stroke (poor)?
- **Duration**: Was the region brushed for a sufficient amount of time?

Researchers in the field of dental care instruction consider each of these criteria to be important for plaque removal. For a given region, up to 2 points is awarded for each of these criteria, with 2 points awarded if a criterion is fully satisfied, giving a maximum score of 6 points per region. Combining the scores for all four regions gives a maximum score of 24 points per session.

Relationship between evaluation criteria and plaque scores

During this study, an experiment was carried out to verify the connection between our evaluation criteria and plaque scores. In this experiment, 14 subjects were videoed while brushing their teeth using the setup depicted in Figure 1. After brushing their teeth, a dentist then performed a plaque test on each subject, applying a plaque indicator liquid to each subject's teeth and calculating a score based on the test results. After this, the videos were then used to determine the scores for the criteria used by this study. The experiment was conducted over two days, using the following procedure. On the first day, the subjects brushed their teeth using the setup depicted in Figure 1. Then, a dentist performed a plaque test on each subject, calculating a plaque score based on the results. On the second day, a dentist instructed the subjects on how to properly brush their teeth. This instruction was deemed necessary to facilitate the collection of data with high performance scores, after observing that many of the participants achieved poor performance scores on the first day. After the instruction, the subjects again brushed their teeth and another plaque score was calculated. Finally, all videos were evaluated using this study's criteria to assign scores, and these scores were compared to the plaque scores.

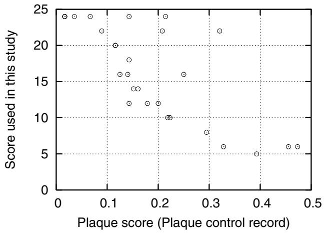


Figure 4. Correlation between plaque scores and the scores used in this study.

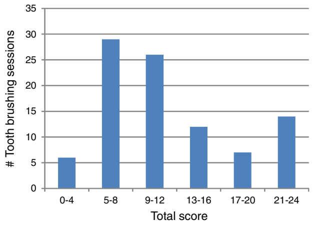


Figure 5. Distribution of scores for the dataset.

Figure 4 shows the relationship between plaque scores and the scores for this study's criteria. For plaque scores, the score decreases as the amount of plaque left after brushing decreases, with low scores indicating good brushing behavior. On the other hand, the scores used in this study are on a 24-point scale with higher values indicating better tooth brushing behavior. Figure 4 shows that the plaque scores and the scores used in this study have a strong negative correlation, with a correlation coefficient of -0.76. However, in several instances of tooth brushing, there was a shift between the plaque scores and the scores assigned by this study's criteria. One possible explanation for this shift is the additional outside influences that affect only the plaque score, such as the effects of foods eaten prior to the test. Nevertheless, for most sessions of tooth brushing, the plaque score and the scores used by this study were strongly associated. Therefore, by using this study's scoring method, it is possible to assign scores that closely correspond to the de facto standard plaque score without applying plaque indicator liquid. Furthermore, based on these results, it is possible to easily prepare large amounts of tooth brushing scores using video data.

Data set

In this study, we gathered a total of 94 sessions of tooth brushing from 14 participants. The average time for each session of tooth brushing was approximately 94 seconds. The participants used either their own toothbrush, or a toothbrush which we provided. The study was conducted over the course of three months, with the data collected in a quiet environment, either in our graduate school building or in the participant's own home. In addition, during the course of the experiment, each participant received instruction from a dentist on the

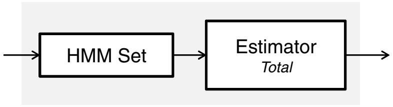


Figure 6. Simple architecture for estimating a total score per session.

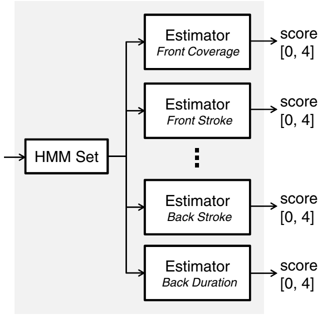


Figure 7. Architecture for providing detailed assessments by estimating six scores, one for each evaluation criterion for both the front and back of the mouth.

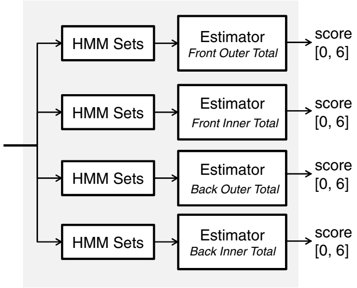


Figure 8. Proposed architecture in which specialized HMM model sets are generated for each score estimator.

proper tooth brushing technique. All sessions were evaluated using video data as was described in the section entitled *Tooth brushing evaluation by a dentist*. Figure 5 shows the distribution of scores for the sessions.

We labeled the audio data using the corresponding video for each session. Each label corresponds to one of the classes of tooth brushing activity, along with the start and end time for that instance of activity. As mentioned in the section entitled *Tooth brushing activity*, the tooth brushing activities “inner surface of front teeth, rough stroke” and “inner surface of back teeth, rough stroke” were not included as these classes were too rare in the data collected.

PROPOSED METHOD

Naive architecture

The basic procedure used in this study starts with using audio data to recognize the seven tooth brushing activities listed in the *Tooth brushing activity* section. We then extract independent (explanatory) variables from those recognition results, using these independent variables to build regression models for estimating scores for sessions of tooth brushing activity. Figure 6 shows our simplest implementation of this architecture. In this architecture, we use an HMM set consisting of seven HMM models, one for each of the seven tooth brushing activities, to recognize tooth brushing activities in the audio data. We then use the results from these seven models to create the independent variables for the regression model. Finally, the regression model outputs a score from 0 to 24, representing the total score for each session.

However, in order to provide a user with an assessment of various aspects of their brushing, it is necessary to estimate the scores in more detail. Figure 7 shows such an architecture that estimates six separate scores (each ranging from 0 to 4), three scores each for the front teeth and back teeth, with those three scores corresponding to the three criteria: coverage, stroke, and duration. For example, the *Front Coverage* score represents the total coverage score for both the upper front teeth and the lower front teeth, including both the inner and outer surfaces. For even more detailed scores, an architecture could further differentiate between the inner and outer surfaces to give 12 separate scores (each ranging from 0 to 2), corresponding to the three criteria for each of the four regions of the mouth, e.g., *Front Inner Duration* and *Front Outer Duration*. (Note that, in general, an architecture that provides scores in finer granularity has higher estimation errors. Therefore, the decision of which architecture to apply to an application should consider both the granularity of scores required and the estimation accuracy required.) However, the simple architectures described above have the following problems:

- Accurate classification of tooth brushing activities into seven classes is difficult, and in the case of some architectures it is unnecessary. Take for example the architecture shown in Figure 7 that estimates scores for only two regions, *front teeth* and *back teeth*. In this case, distinguishing between all seven tooth brushing activities may be unnecessary for score estimation, and more accurate estimates are possible by using a coarser set of models without the *inner surface* and *outer surface* distinction.
- Each of the regression models estimates scores using the classification results from an HMM set, but the usefulness of the tooth brushing activity classes varies for the different regression models. For example, when estimating the *coverage* score for the back teeth, activities related to the front teeth have less importance while activities such as BI-Fine and BO-Fine should be recognized as accurately as possible. By using coarser models depending on the needs of the regression model, more accurate results can be achieved.

Overview of proposed approach

In the proposed method, we solve the problems with the naive architectures described above by preparing separate HMM model sets for each of the regression models used for score estimation, as is shown in Figure 8. For example, in Figure 8, we prepare a specialized HMM set for the regression model that estimates a score for *Front Inner Total*. (Note that *Total* means the sum of the three scores for the evaluation criteria.) Each of the HMM model sets generated is specialized to its regression model, in order to increase the estimation accuracy of the regression model. For example, when preparing the HMM model set used as the basis for estimating the *coverage* score for the back teeth, we generate an HMM model set that performs well when recognizing tooth brushing activities related to the back teeth. Specifically, we automatically discover which tooth brushing activities are useful for estimating the score in question and then generate an HMM model set that focuses on only those classes. When doing so, we ignore any activity classes that are not considered useful for estimating the score. For example, when estimating the total score for the front teeth, we may not need to consider the recognition results related to the back teeth, and so could omit the models related to the back teeth from the HMM set. The recognition results from this reduced model set would then be used to build the regression model for that score.

We can divide the procedure for constructing the architecture for score estimation into three steps: (1) Identify which tooth

brushing activity classes are important when estimating each score. (2) Generate HMM model sets for recognizing those important classes. (3) Build a regression model for estimating each score using the recognition results of those HMM model sets.
Each of these steps is explained in detail below.

Discovering useful tooth brushing activity classes

As discussed above, the usefulness of tooth brushing activity classes vary for the different evaluation criteria. In this study, we use regression models to estimate the evaluation criteria, extracting the independent variables from the audio recognition results, e.g., an independent variable for the total duration of segments recognized as belonging to the FI-Fine class. In order to determine the usefulness of the activity classes, we first use the training data to evaluate the usefulness of each of the independent variables in estimating each of the evaluation criteria. Using the results of this evaluation, we can then determine which tooth brushing activity classes are useful for each of the evaluation criteria. For example, if we determine that many of the independent variables calculated using the results from the FO-Fine class are useful for estimating a given score, then we consider the FO-Fine class to be useful for estimating that score.

We start by evaluating the independent variables using the RReliefF algorithm [28], a feature selection algorithm which is used to determine the relevance of features to a given regression task. Given n instances of data, each with a set of feature values $\boldsymbol{F}$ (independent variables) along with a predicted value (dependent variable), RReliefF works by randomly selecting m of the n instances and for each of those m instances determining the k nearest neighbors. The ith feature f_i is assigned a weight based on the degree to which the value for f_i for each random instance differs from the values for f_i for the random instance's k nearest neighbors, relative to how much the predicted value for the random instance differs from those of its k nearest neighbors. In simpler terms, a feature's weight is increased if it discriminates between neighboring instances with differing predicted values and is decreased if it separates neighboring instances with similar predicted values. These weights indicate the importance of the feature f_i to the regression task and approximate the difference of probabilities [28]:

$$W(f_i) = \Pr(FD_i \mid PD) - \Pr(FD_i \mid PS),$$

where FD_i means that the values for f_i for neighboring instances differ, PD means that the predicted values for neighboring instances differ, and PS means that the predicted values for neighboring instances are similar. Using the weights calculated by RReliefF for each of the features, we can then determine the usefulness of the set of tooth brushing activity classes C for a given evaluation criterion. The usefulness U_c of a tooth brushing activity class $c \in C$ is calculated by summing the weights for $\boldsymbol{F_c}$, where $\boldsymbol{F_c}$ is the subset of $\boldsymbol{F}$ consisting of the features that are computed using recognition results from the tooth brushing activity class c: $U_c = \sum\limits_{f \in \boldsymbol{F_c}} W(f)$.
Since the weights output by RReliefF can be either positive or negative, we first perform feature scaling on all weights $W(f_i)$ so that they fall in the range $[0, 1]$ prior to computing U_c. After computing U_c, we then normalize the values in U_c to sum to 1.

Tailoring HMM sets to improve score estimates

Using the method described in the previous subsection, we can determine which tooth brushing activity classes are useful for estimating scores for a given evaluation criterion. Using this information, we determine the HMM set used to estimate that criterion using only the identified useful classes. As mentioned previously, in the naive approach there are two issues that arise from using the same HMM model set when estimating all the evaluation criteria: (1) Depending on the architecture being used, it may not be necessary to recognize the activities on as fine a scale as with all seven activity classes. (2) Depending on the score being estimated, the ideal set of activity classes to use in the HMM set may not include all seven classes. In our proposed method, we will address the first of these issues by generating the following four basic HMM sets which have varying granularity:

- *HMM set 7*: A seven-class HMM set generated using all seven tooth brushing activity classes.
- *HMM set 5*: A five-class HMM set generated using the classes *outer surface of front teeth*, *outer surface of back teeth*, *inner surface of front teeth*, *inner surface of back teeth*, and *no activity* (None), with no distinction made between *fine stroke* and *rough stroke*.
- *HMM set FB*: A three-class HMM set for distinguishing between the front and back teeth, generated using the classes *front teeth*, *back teeth*, and *no activity*, with no distinction made between *fine stroke* and *rough stroke* nor between *outer surface* and *inner surface*.
- *HMM set RF*: A three-class HMM set for distinguishing between stroke types, generated using the classes *rough stroke*, *fine stroke*, and *no activity*, with no distinction made between *front teeth* and *back teeth* nor between *outer surface* and *inner surface*.

Furthermore, in this study we address the second of the issues with the naive architectures by generating a new HMM set from each basic HMM set, using the method described in the previous subsection to compute the usefulness U_c of each class $c \in C$ as the basis for generating HMM sets tailored for estimating each score. We determine which classes to include by setting a threshold $T =^1 /_{|C|}$, where $|C|$ is the total number of tooth brushing activity classes included in a basic HMM set. We then only include the class c in the new model set if $U_c \geq T$. Any class for which $U_c < T$ is then ignored. Thus, in our proposed method, we attempt to improve the recognition performance for the useful activity classes by ignoring unnecessary activity classes. For example, starting with the *HMM set 7* above, in the case where the classes FO-Fine, FO-Rough, BO-Fine, BO-Rough, and None are determined to be unnecessary, we would combine those classes into a single Others class and create a three-class HMM set consisting of the models: FI-Fine, BI-Fine, and Others. By doing so, we can then increase the recognition performance of the more useful classes FI-Fine and BI-Fine.

In our proposed method, we then estimate scores using a combination of eight total HMM sets, the four basic HMM sets *HMM set 7*, *HMM set 5*, *HMM set FB*, and *HMM set RF*, along with four sets generated automatically from those four basic sets.

Toothbrushing activity recognition

Using the method described in the previous subsection to select the classes used in each of our HMM model sets, we then generate HMM models [35] used for tooth brushing activity recognition.

Feature extraction

In this study, we use MFCCs to recognize tooth brushing activities, as MFCCs have been reported to be one of the better transformation schemes for environmental sound recognition [6, 8]. We compute a 12-order MFCC over a window of 50 ms with 50% overlap, windowed using a Hamming window. Along with this 12-order MFCC, we compute the log energy for the window, along with the corresponding 13-order delta and 13-order acceleration coefficients for the MFCC and log energy coefficients, giving a vector of 39 values in total.

Recognition with HMMs

Our method recognizes tooth brushing activity classes in audio data using HMMs. The model for each class is a 10-state left-to-right HMM with output distributions represented by 32 Gaussian mixture densities. The observed variables for the models are the vectors of 39 MFCC-based coefficients. We use these HMM model sets to recognize tooth brushing activities over full sessions of audio data using the Viterbi algorithm [27], finding the most probable sequence of tooth brushing activity classes across the session. Using these recognition results, we can then compute the independent variables used in the regression models.

User adaptation

As was mentioned in the introduction, the model of tooth brush and the shape of the user's mouth can affect the sound made when brushing his/her teeth, so the audio obtained for the tooth brushing activities will differ per user. In order to cope with this issue, we employ the maximum likelihood linear regression (MLLR) adaptation method [23, 14]. MLLR adaptation works by creating a transformation matrix which can be used to transform a user-independent HMM model set, which is trained on other users' labeled data, to more closely match the target user's unlabeled data. That is, we shift the output distributions of the initial tooth brushing activity models (HMMs) using the target user's data, so that each state in the HMMs is more likely to generate the target user's data. A new estimation of the adapted mean vector $\hat{\boldsymbol{\mu}}$ is given by

$$\hat{\boldsymbol{\mu}} = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} = \boldsymbol{W}\boldsymbol{\xi},$$

where $\boldsymbol{\mu}$ is the initial mean vector for the output distributions, $\boldsymbol{A}$ is a $k \times k$ transformation matrix, where k is the number of dimensions of the feature vector ($k = 39$), $\boldsymbol{b}$ is a bias vector, $\boldsymbol{W}$ is a $k \times (k+1)$ transformation matrix that is decomposed into $\boldsymbol{W} = [\boldsymbol{b}\ \boldsymbol{A}]$, and $\boldsymbol{\xi}$ is the extended mean vector $\boldsymbol{\xi} = [1\ \mu_1\ \mu_2 \cdots \mu_k]^T$. Using this equation, we can estimate a $\boldsymbol{W}$ that reduces the mismatch between the initial models and the user's unlabeled data using the EM technique.

Estimating scores

Computing independent variables

Using the adapted HMM models, it is possible to recognize which tooth brushing activities were conducted in a session of audio data. For example, by using *HMM set 7*, it is possible to detect that the activity *Outer surface of front teeth, rough stroke* was conducted in the interval from 3.4 sec to 8.9 sec from the start of the audio. Using recognition results such as this, we can compute independent variables for use in the regression models for score estimation. For the first set of independent variables, we create a variable for each of the activity classes in our HMM model sets, excluding the None and Others classes. Each of these variables is computed as the total duration of its corresponding tooth brushing activity in the recognition results.

We then compute a second set of independent variables to help cope with a limitation we encounter when estimating scores using audio data. This limitation comes from the difficulty in distinguishing between the upper and lower teeth and between the right and left sides of the mouth. Because of this limitation, it is difficult to determine whether an activity was conducted evenly across both the upper and lower teeth or across both the back-left and back-right sides of the mouth. However, take for example the case where a user brushes only their upper teeth. In this case, we expect that features extracted from the audio data will not vary greatly over the course of the activity. On the other hand, if the user had brushed both the upper and lower teeth, then we would expect the features to vary more. Based on this idea, we generate additional independent variables corresponding to the variance of feature values across a given activity, generating one such independent variable for each of the features (MFCCs).

Estimating a score for each criterion

Finally, using the independent variables generated, we estimate the evaluation scores using regression analysis. We first perform dimensionality reduction using the Random Projection algorithm [11] to reduce the number of variables down to 10. Using these 10 independent variables, we then run regression analysis using the SMO algorithm [30] to estimate the scores.

EVALUATION

Evaluation methodology

In order to investigate the effectiveness of the proposed method, we prepared the following methods:

- ***Avg***: A naive approach in which we estimated a user's scores using the average scores for other users.
- ***SHMM***: As shown in Figure 6, we prepared only a single HMM set (*HMM set 7*). Otherwise this method was the same as the proposed method.
- ***MHMM***: We prepared four basic HMM sets: *HMM set 7*, *HMM set 5*, *HMM set FB*, and *HMM set RF*. Otherwise this method was the same as the proposed method.
- ***Proposed***: The proposed method, in which we prepared a separate group of eight HMM sets for each of the scores.
- ***Corrected***: A variation of the *SHMM* method in which we

Table 1. Recognition results for basic HMM sets when using MLLR adaptation.

	prec.	recall	F-meas.
HMM set 7	0.457	0.455	0.451
HMM set 5	0.485	0.506	0.491
HMM set FB	0.658	0.654	0.652
HMM set RF	0.677	0.692	0.684

Table 2. Recognition results for basic HMM sets without MLLR adaptation.

	prec.	recall	F-meas.
HMM set 7	0.310	0.335	0.313
HMM set 5	0.347	0.392	0.358
HMM set FB	0.525	0.591	0.545
HMM set RF	0.515	0.585	0.534

Table 3. Increase in recognition accuracy (%) for useful classes from the basic HMM sets to HMM sets generated by the proposed method.

	prec.	recall	F-meas.
HMM set 7	-1.9	10.8	4.1
HMM set 5	-5.9	9.2	0.3
HMM set FB	0.0	0.0	0.0
HMM set RF	0.0	0.3	0.2

built the regression models using corrected labels, i.e., this method assumed 100% recognition accuracy for *HMM set 7*.

Additionally, the following six evaluation architectures were prepared for use with those methods:

- *Total* (24): Estimated a single score (24-point scale) that represents the total score for all tooth brushing activity in the session.
- *CSD* (8): Estimated three scores (8-point scale), one for each of the evaluation criteria: coverage, stroke, and duration. For example, a single score was output for stroke, representing the stroke quality for the entire session.
- *FB* (12): Estimated two scores (12-point scale), one for the front teeth and one for the back teeth.
- *FB* x *CSD* (4): Estimated six scores (4-point scale), corresponding to each of the three evaluation criteria for both the front teeth and back teeth. For example, a score was output for the duration criterion for the front teeth.
- *IO* x *FB* (6): Estimated four scores (6-point scale), one for each region of the mouth: outer surface of front teeth, inner surface of front teeth, outer surface of back teeth, and inner surface of back teeth.
- *IO* x *FB* x *CSD* (2): Estimated 12 scores (2-point scale), corresponding to each of the three evaluation criteria for each region of the mouth. For example, a score was output for the duration criterion for the outer surface of back teeth.

All methods were evaluated using their mean absolute error (MAE) and error ratio ($MAE/Maximum\ Score$), with the evaluation done using leave-one-user-out cross validation. That is, when using a user's data as the test data, the training data consisted of the data collected from other users. However, when conducting MLLR adaptation, the adaptation data consisted of the current user's data (excluding the session being tested).

Tooth brushing activity recognition results

We first evaluated the tooth brushing activity recognition performance.

Activity recognition performance

Table 1 shows the recognition results for each of the basic HMM sets used in this study. Both *HMM set 7* and *HMM set 5* had similar results, with F-measures of 0.451 and 0.491 respectively. Overall, both sets had higher recognition accuracy for the back teeth than the front teeth. Also, the worst performing of the classes were the classes related to the inner surface of the teeth, with both sets performing poorly at distinguishing between the back and front teeth for the inner surface. Figure 9 shows an example of this in the confusion matrix for *HMM set 5*, with *inner surface of front teeth* showing low recall with many instances classified as *inner surface of back teeth* instead. As for *HMM set 7*, the F-measures for FI-Fine and BI-Fine were poor.

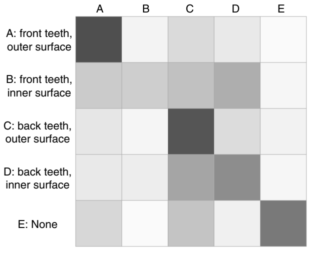


Figure 9. Visual confusion matrix for *HMM set 5*.

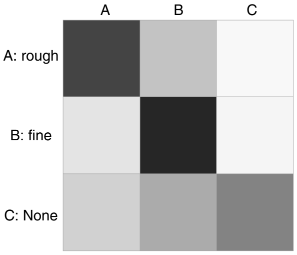


Figure 10. Visual confusion matrix for *HMM set RF*.

Table 4. Recognition results for basic HMM sets when background noise is present.

	precision	recall	F-measure
HMM set 7	0.171	0.219	0.187
HMM set 5	0.344	0.359	0.349
HMM set FB	0.516	0.576	0.524
HMM set RF	0.366	0.406	0.326

Both *HMM set FB* and *HMM set RF* had comparable results, achieving average F-measures of 0.652 and 0.684, respectively. For both sets, the main contributing factor preventing higher accuracy was low accuracy for the None class, with the None class showing less than 70% of the accuracy of the other classes in both sets. Figure 9 shows the confusion matrix for *HMM set RF*. As shown in these results, we could successfully classify *rough* and *fine* instances, achieving F-measures of 0.738 and 0.856, respectively.

Effectiveness of adaptation

Table 2 shows the results of an investigation into the results for each of the basic HMM sets used in this study when no adaptation was used. Comparing Table 2 with Table 1, the effectiveness of adaptation appears to increase with the increasing granularity of the models. For example, the F-measure for the seven-model set increased by about 31% with adaptation, while the F-measures for the 3-model sets only increased by about 19% on average.

Recognition performance of useful activity classes

In this section, we compare the recognition accuracies of useful classes from HMM sets generated by the proposed method with the recognition accuracies of the same classes from basic HMM sets. Table 3 shows the percent improvement for the average precision, recall, and F-measure. As is seen in these results, in *HMM set 7* and *HMM set 5* there was a large improvement in recall, but a slight deterioration in precision, resulting in an increase in F-measure for both. For *HMM set 7*, the increase in F-measure was larger. On the other hand,

Table 5. Mean absolute error (MAE) of score estimates for each architecture (columns) for each method (rows).

	Total	*CSD*	*FB*	*FB* x *CSD*	*IO* x *FB*	*IO* x *FB* x *CSD*	Average
Avg	5.48	2.03	3.16	1.16	1.98	0.79	2.43
SHMM	4.07	1.81	2.78	1.13	1.66	0.64	2.02
MHMM	3.99	1.53	2.56	0.95	1.43	0.55	1.84
Proposed	3.32	1.49	2.52	0.93	1.45	0.58	**1.72**
Corrected	3.10	1.58	2.61	1.04	1.41	0.58	**1.72**
Proposed w/o var	4.25	1.53	2.74	0.95	1.38	0.56	1.90
Proposed w/o adapt	3.71	1.67	2.49	1.06	1.51	0.61	1.84

Table 6. Error ratio (%) of score estimates for each architecture (columns) for each method (rows).

	Total	*CSD*	*FB*	*FB* x *CSD*	*IO* x *FB*	*IO* x *FB* x *CSD*	Average
Avg	22.9	25.4	26.3	29.0	33.1	39.3	29.3
SHMM	16.9	22.7	23.1	28.2	27.7	31.8	25.1
MHMM	16.6	19.1	21.3	23.8	23.8	27.5	22.0
Proposed	13.8	18.6	21.0	23.3	24.1	29.1	**21.7**
Corrected	12.9	19.8	21.7	26.1	23.6	29.2	22.2
Proposed w/o var	17.7	19.2	22.8	23.6	23.0	27.8	22.4
Proposed w/o adapt	15.5	20.8	20.8	26.6	25.1	30.4	23.2

Table 7. Useful independent variables (top-4) in *Total* and *CSD* architectures.

Total	*Total duration of fine stroke*
	Total duration of back teeth
	Variance of back inner teeth
	Variance of back inner teeth w/ fine
Coverage	*Variance of back inner teeth*
	Variance of back inner teeth w/ fine
	Total duration of fine stroke
	Total duration of front inner teeth w/ fine
Stroke	*Total duration of back teeth*
	Total duration of fine stroke
	Total duration of back outer teeth w/ fine
	Variance of back inner teeth w/ fine
Duration	*Total duration of fine stroke*
	Total duration of back teeth
	Total duration of back outer teeth w/ fine
	Variance of front inner teeth w/ fine

little change was seen in *HMM set FB* and *HMM set RF*. For both *HMM set FB* and *HMM set RF*, there were only three classes initially (including the None class). Therefore, there was little difference between the basic HMM set and the sets generated by the proposed method, as it was rare that one of the two classes other than None was judged useless.

Effects of noise

In this study, we also investigated the effects of background noise on tooth brushing activity recognition, collecting five sessions of audio while running a hair dryer in the background near the smartphone used to collect the audio. We tried to reduce the effects of the noise by employing Cepstral Mean Normalization (CMN), which is an additive noise cancellation technique widely used in speech recognition studies.

Table 4 shows an overall degradation of performance due to the noise, with the F-measures for all sets reduced by at least 29%. Furthermore, the F-measure for *HMM set RF* dropped below 0.33, with the set apparently no longer able to distinguish between the classes. The reduction in F-measure for *HMM set 5* appears to be from an inability to distinguish between the inside and outside surfaces when noise was present, while it still appeared well able to distinguish between the front and back teeth. If the proposed method is to be implemented in a real-life application, it will be important to give clear instructions on avoiding excessive background noise while running the application. (Note that because our data were collected in the participants' homes and our laboratory rooms, our data included small daily-life background noises such as the sounds of fans, air conditioners, and closing doors.)

Score estimation results

Score estimation error

Table 5 shows the mean absolute error (MAE) for each architecture using each of the prepared methods. When looking at these results, *Corrected* shows the results when the tooth brushing activity recognition was assumed to have 100% accuracy, and so this is assumed to be the lower bound on score estimation accuracy for a straightforward architecture. Here we observe that the error for the *Total* architecture for *Avg* was about 1.8 times as high as that of *Corrected*. Additionally, when comparing the *Corrected* results to *SHMM*, *Corrected* again showed lower error rates, with an MAE 0.97 points lower than that of *SHMM* for *Total*. Comparing the error for *Total* for *MHMM* to *SHMM*, *MHMM* had an MAE that was 0.18 points lower.

Using *Proposed*, the MAE for *Total* was reduced by 0.75 points from that of *SHMM*. In addition, using *Proposed*, we were able to reduce the MAE for *Total* by over 2 points in comparison to *Avg*. Moreover, *Proposed* was able to achieve the same average MAE across the architectures as *Corrected*. In comparison to *Corrected*, which had a recognition accuracy of 100%, the recognition accuracy for the HMM results in *Proposed* was much lower. However, by preparing HMM sets that were built using HMM models considered useful to each recognition task, *Proposed* was able to compensate for its lower recognition accuracy. Looking across all the architectures shown in Table 5, *Proposed* achieved a much lower MAE than *Avg* for all the architectures, achieving accuracies similar to those of *Corrected*.

Table 6 shows the error ratios for the estimates for each architecture using each of the prepared methods. Here, error ratios are computed as the MAE divided by the maximum score, e.g., an MAE of 2.4 for a 24-point scale would have an error ratio of 10%. It can be seen that overall the *Proposed* method reduced error ratios by about 7.6% on average from those of *Avg*. Additionally, *Proposed* reduced error rates by 3.4% on average compared to *SHMM* and by 0.3% on average compared to *MHMM*. Using a McNemar test on the results of *SHMM* and *Proposed* for their estimates across all architectures, the improvement was found to be statistically significant ($p < 0.001$).

Effectiveness of variance variables

In Tables 5 and 6, *Proposed w/o var* shows the accuracy of *Proposed* when we omitted the independent variables corresponding to the variances of feature values. Without the variance variables, the average MAE increased by about 0.18 points (0.7% in terms of error ratios). As was discussed above, by using the features' variance, we were able to capture the variation in the toothbrush's locations. We believe that including this variance improved the regression results beyond what is achieved through using the HMM results alone, because the audio-based HMM results could not distinguish certain location distinctions such as upper teeth vs lower teeth. In the case of the *CSD* architecture, incorporating the features' variance reduced the MAE for the *Coverage* score from 1.65 to 1.53 and reduced the MAE for the *Stroke* score from 1.63 to 1.55. On the other hand, the MAE for the *Duration* score did increase from 1.32 to 1.38. Despite that small increase, a large performance improvement was observed overall by use of variance in this architecture.

Differences in results between architectures

As can be seen in Table 6, the error ratio for the *Total* architecture was reduced down to 13.8% using the *Proposed* method, but as we look at architectures that estimated scores on a finer granularity, we see that the estimation accuracy degraded. For example, upon reaching the fine-scale *IO* x *FB* x *CSD* architecture, which estimates scores on a 2-point scale, the error ratio reached 29.1%. Such an architecture restricts the correct scores to the discrete values 0, 1, and 2, which increases the error ratio for estimates.

Here we introduce detailed results for the *Proposed* method in the various architectures. However, since the *Total* and *CSD* architectures were already discussed above, this will focus on the other architectures. In the *FB* architecture, the MAE for the front teeth score was 2.17 while the MAE for the back teeth score was 2.88. This is in contrast to the HMM recognition results, where accuracies for classes related to the back teeth were mostly higher than those for classes related to the front teeth. On the other hand, in the *FB* x *CSD* architecture, the average MAE for the three scores related to the front teeth was 0.95 while the average MAE for the three scores for the back teeth was 0.90, a reverse of the situation with *FB*. The results in Table 6 show that despite the fact that *FB* x *CSD* provided more detailed estimates than did *FB*, the error ratio does not change significantly. Based on these results, we believe that it probably was not possible to generate a good regression model in *FB* to estimate the score obtained by summing the scores for the three criteria. The performance of *FB* is discussed further in a following section.

In *FB* x *CSD*, the *Duration* score averaged across the back and front teeth had an MAE of 0.74. On the other hand, for *Stroke* the averaged score had an MAE of 1.08 and for *Coverage* it was 1.04. Just as with the *CSD* architecture, the *Duration* score's MAE is lower than those of the other criteria, since *Duration* can be computed directly from the lengths of each activity. As for the *IO* x *FB* architecture, the accuracies for scores related to the *inner surface of back teeth* were the worst. Among the results for the *IO* x *FB* x *CSD*, the MAE for the scores related to *Stroke* were as high as 0.95. On the other hand, the MAEs for *Duration* and *Coverage* were 0.51 and 0.73 respectively. As can be seen in the accuracies from the results of HMM recognition, the accuracy for recognition of BI-Fine was low, which most likely had a large influence on the aforementioned regression results.

Effectiveness of adaptation

In Tables 5 and 6, the *Proposed w/o adapt* method shows the accuracy of the *Proposed* method when MLLR adaptation was not performed. Comparing the accuracy of *Proposed w/o adapt* to *Proposed*, it can be seen that we could improve the estimation accuracies to some extent by performing user adaptation. Looking at Table 6, we were able to reduce the average error ratio by about 1.5%.

Effectiveness of independent variables

This section briefly discusses the independent variables that were useful for estimating various scores. We determined the usefulness for these variables using the RReliefF algorithm described earlier. As shown in Table 7, in the *Total* architecture, the variable for the total length of time brushing the teeth with a fine stroke was found to be useful. Its usefulness was likely because it provides essential information related to both *Stroke* and *Duration*. For the *CSD* architecture, the variances of MFCC features across various brushing locations were useful for estimating scores for *Coverage*. When estimating scores for *Stroke*, the useful variables were the total times for fine strokes for various brushing locations. For *Duration*, the useful variables corresponded to total times brushing at the various locations.

The results for the other architectures tended to be similar to those described for *CSD*. However, in the case of *FB*, there were a number of variables judged by RReliefF to be useful that were only indirectly related to the score being calculated. For example, when estimating scores for the front teeth, variables such as the total time spent brushing teeth with a fine stroke, which are computed from *HMM set RF* results, were found to be useful. It appears that in many cases, if the total time spent brushing with a fine stroke was long, then the total time spent brushing the front teeth with a fine stroke was also long. However, we believe that the inclusion of such indirectly related independent variables had a negative effect on the *FB* architecture, contributing to its poor performance.

Effect of distance from smartphone

In *Total*, for many of the subjects, we were able to reduce the MAE for estimation to below 4. However, there was a single subject that had an MAE of 10. Additionally, this subject's recognition accuracy for *HMM set 7* was only 32.8%. When reviewing the video taken for this subject, it was found that the subject was separated too far from the smartphone when recording the audio and that the volume of the recorded audio was low. If the proposed method is to be implemented in a real-life application, then we believe it will be important to give clear instructions to aspects such as how close the user should be to the smartphone when recording the data.

CONCLUSION

This paper presented a new method for evaluating tooth brushing performance using audio collected from a smartphone. As a part of our future work, we plan to employ deep learning techniques to discover useful features tailored for recognizing tooth brushing audio.

REFERENCES

1. Addy, M., and Hunter, M. Can tooth brushing damage your health? effects on oral and dental tissues. *International Dental Journal 53*, S3 (2003), 177–186.
2. Bartle, R. Bartle test of gamer psychology. gamerdna. Tech. rep., Retrieved 2010-12-08 from http://www. gamerdna. com/quizzes/bartle-test-of-gamerpsychology.
3. Bartle, R. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research 1*, 1 (1996), 19.
4. Bell, M., Reeves, S., Brown, B., Sherwood, S., MacMillan, D., Ferguson, J., and Chalmers, M. EyeSpy: supporting navigation through play. In *CHI 2009* (2009), 123–132.
5. Chang, Y.-C., Lo, J.-L., Huang, C.-J., Hsu, N.-Y., Chu, H.-H., Wang, H.-Y., Chi, P.-Y., and Hsieh, Y.-L. Playful toothbrush: ubicomp technology for teaching tooth brushing to kindergarten children. In *CHI 2008* (2008), 363–372.
6. Chen, J., Kam, A., Zhang, J., Liu, N., and Shue, L. Bathroom activity monitoring based on sound. In *Pervasive 2005* (2005), 47–61.
7. Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T., and Campbell, A. T. Unobtrusive sleep monitoring using smartphones. In *PervasiveHealth 2013* (2013), 145–152.
8. Cowling, M. *Non-speech environmental sound recognition system for autonomous surveillance*. PhD thesis, Griffith University, 2004.
9. Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From game design elements to gamefulness: defining gamification. In *15th International Academic MindTrek Conference: Envisioning Future Media Environments* (2011), 9–15.
10. Fiske, J., Davis, D., Frances, C., and Gelbier, S. The emotional effects of tooth loss in edentulous people. *British Dental Journal 184*, 2 (1998), 90–93.
11. Fradkin, D., and Madigan, D. Experiments with random projections for machine learning. In *KDD 2003* (2003), 517–522.
12. Gallagher, A., Sowinski, J., Bowman, J., Barrett, K., Lowe, S., Patel, K., Bosma, M. L., and Creeth, J. E. The effect of brushing time and dentifrice on dental plaque removal in vivo. *American Dental Hygienists Association 83*, 3 (2009), 111–116.
13. Ganss, C., Schlueter, N., Preiss, S., and Klimek, J. Tooth brushing habits in uninstructed adults? frequency, technique, duration and force. *Clinical Oral Investigations 13*, 2 (2009), 203–208.
14. Gauvain, J., and Lee, C. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Trans. on Speech and Audio Processing 2*, 2 (2002), 291–298.
15. Gerritsen, A. E., Allen, P. F., Witter, D. J., Bronkhorst, E. M., and Creugers, N. Tooth loss and oral health-related quality of life: a systematic review and meta-analysis. *Health Qual Life Outcomes 8*, 126 (2010), 552.
16. Graetz, C., Bielfeldt, J., Wolff, L., Springer, C., Fawzy El-Sayed, K. M., Sälzer, S., Badri-Höher, S., and Dörfer, C. E. Toothbrushing education via a smart software visualization system. *Journal of Periodontology 84*, 2 (2013), 186–195.
17. Han, K., Graham, E. A., Vassallo, D., and Estrin, D. Enhancing motivation in a mobile participatory sensing project through gaming. In *IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT 2011) and IEEE Third Inernational Conference on Social Computing (SocialCom 2011)* (2011), 1443–1448.
18. Inada, E., Saitoh, I., Yu, Y., Tomiyama, D., Murakami, D., Takemoto, Y., Morizono, K., Iwasaki, T., Iwase, Y., and Yamasaki, Y. Quantitative evaluation of toothbrush and arm-joint motion during tooth brushing. *Clinical Oral Investigations* (2014), 1–12.
19. Janusz, K., Nelson, B., Bartizek, R. D., Walters, P. A., and Biesbrock, A. Impact of a novel power toothbrush with smartguide technology on brushing pressure and thoroughness. *The Journal of Contemporary Dental Practice 9*, 7 (2008), 1–8.
20. Kim, K.-D., Jeong, J.-S., Lee, H. N., Gu, Y., Kim, K.-S., Lee, J.-W., and Park, W. Efficacy of computer-assisted, 3d motion-capture toothbrushing instruction. *Clinical Oral Investigations* (2014), 1–6.
21. Kim, K.-S., Yoon, T.-H., Lee, J.-W., and Kim, D.-J. Interactive toothbrushing education by a smart toothbrush system via 3d visualization. *Computer Methods and Programs in Biomedicine 96*, 2 (2009), 125–132.
22. Lee, Y.-J., Lee, P.-J., Kim, K.-S., Park, W., Kim, K.-D., Hwang, D., and Lee, J.-W. Toothbrushing region detection using three-axis accelerometer and magnetic sensor. *IEEE Transactions on Biomedical Engineering 59*, 3 (2012), 872–881.
23. Leggetter, C., and Woodland, P. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech & Language 9*, 2 (1995), 171–185.
24. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *CHI 2011* (2011), 2409–2418.
25. Lu, H., Pan, W., Lane, N., Choudhury, T., and Campbell, A. SoundSense: scalable sound sensing for people-centric applications on mobile phones. In *MobiSys 2009* (2009), 165–178.

26. Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., and Hong, J. I. Toss'n'turn: smartphone as sleep and sleep quality detector. In *CHI 2014* (2014), 477–486.

27. Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 2 (1989), 257–286.

28. Robnik-Šikonja, M., and Kononenko, I. An adaptation of relief for attribute estimation in regression. In *ICML 1997* (1997), 296–304.

29. Rossi, M., Feese, S., Amft, O., Braune, N., Martis, S., and Troster, G. AmbientSense: A real-time ambient sound recognition system for smartphones. In *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM 2013 Workshops)* (2013), 230–235.

30. Shevade, S., Keerthi, S., Bhattacharyya, C., and Murthy, K. Improvements to the SMO algorithm for SVM regression. *IEEE Trans. on Neural Networks 11*, 5 (2002), 1188–1193.

31. Thomaz, E., Parnami, A., Bidwell, J., Essa, I., and Abowd, G. D. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In *Ubicomp 2013* (2013), 739–748.

32. Tosaka, Y., Nakakura-Ohshima, K., Murakami, N., Ishii, R., Saitoh, I., Iwase, Y., Yoshihara, A., Ohuchi, A., and Hayasaki, H. Analysis of tooth brushing cycles. *Clinical Oral Investigations* (2014), 1–9.

33. Von Ahn, L., and Dabbish, L. Designing games with a purpose. *Communications of the ACM 51*, 8 (2008), 58–67.

34. Winterfeld, T., Schlueter, N., Harnacke, D., Illig, J., Margraf-Stiksrud, J., Deinzer, R., and Ganss, C. Toothbrushing and flossing behaviour in young adults? a video observation. *Clinical Oral Investigations* (2014), 1–8.

35. Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., et al. *The HTK book*, vol. 2. Entropic Cambridge Research Laboratory Cambridge, 1997.