Electronic Preprint for Journal of Information Processing Vol.24 No.2

Regular Paper

Toothbrushing Performance Evaluation Using Smartphone Audio Based on Hybrid HMM-recognition/SVM-regression Model

Joseph Korpela[1,a)] Ryosuke Miyaji[1,b)] Takuya Maekawa[1,c)]
Kazunori Nozaki[1,d)] Hiroo Tamagawa[1,e)]

Received: June 28, 2015, Accepted: December 7, 2015

Abstract: This paper presents a method for evaluating toothbrushing performance using audio data collected by a smartphone. This method first conducts activity recognition on the audio data to classify segments of the data into several classes based on the brushing location and type of brush stroke. These recognition results are then used to compute several independent variables which are used as input to an SVM regression model, with the dependent variables for the SVM model derived from evaluation scores assigned to each session of toothbrushing by a dentist who specializes in dental care instruction. Using this combination of audio-based activity recognition and SVM regression, our method is able to take smartphone audio data as input and output evaluation score estimates that closely correspond to the evaluation scores assigned by the dentist participating in our research.

Keywords: toothbrushing, healthcare, smartphone, audio

1. Introduction

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 [5], [10]. Despite oral health's significant impact on our overall well-being, there is evidence that a significant portion of the population brushes incorrectly [7]. Moreover, while proper toothbrushing can have a positive impact on oral health, improper toothbrushing 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 toothbrushing behavior using optical motion capture systems [2], [13] and embedded accelerometer sensors [11], [15], [16], [17], [25]. We introduce each of these in detail in the related work section.

Each of the systems described above relied on complex video equipment or custom-made sensing devices, requiring most users to purchase new equipment to use them. Our research proposes a method for evaluating toothbrushing 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 their front teeth. Our system can return scores for each area of the mouth and can also output a total evaluation score for the toothbrushing. In this research, we used a supervised machine 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.

We estimate scores using regression models built from the audio recognition results for toothbrushing actions. First, we label the audio time-series data with the toothbrushing actions that were being conducted during different periods using recognizers based on hidden Markov models (HMMs) [21]. For example, from 89 seconds to 110 seconds after the start of the audio could

[1] Osaka University, Suita, Osaka 565–0871, Japan
[a)] joseph.korpela@ist.osaka-u.ac.jp
[b)] ryosuke.miyaji@ist.osaka-u.ac.jp
[c)] maekawa@ist.osaka-u.ac.jp
[d)] knozaki@dent.osaka-u.ac.jp
[e)] tamagawa@dent.osaka-u.ac.jp

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

be labeled "brushing the outer surface of front teeth." Second, we use these labeled segments to calculate independent variables for the regression models used for estimating scores. For example, 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' toothbrushing.

The proposed method has the following features:

(1): In order to improve toothbrushing proficiency, it is necessary to single out deficiencies in the user's brushing habits. The proposed method has the ability to detect such deficiencies (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 HMMs for the recognition of toothbrushing actions, but the final goal of the research is to use the output from these models to estimate scores for toothbrushing for different areas of the mouth and/or evaluation criterion. The importance of the toothbrushing actions will vary depending on the score being estimated, e.g., toothbrushing 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 toothbrushing 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 studies that relate to environmental sound recognition and sensing toothbrushing. Then we propose a method for evaluating toothbrush performance using audio recorded by a smartphone. Finally, we evaluate our method with 94 sessions of toothbrushing data. To the best of our knowledge, this is the first study that attempts to evaluate toothbrushing performance based solely on audio data. The research contributions of this paper are: (1) We propose a method for evaluating toothbrushing performance using a machine learning approach. First, a dentist who specializes in toothbrushing instruction assigns scores to training data based on his evaluation of toothbrushing performance. Then, we use this training data to construct a regression model with score estimates close 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 toothbrushing performance. In this method, we automatically generate separate HMM sets for each score, with each set tailored to improve the accuracy of its corresponding score estimates. (3) We evaluate the proposed method using 94 sessions of toothbrushing audio data taken from 14 research participants.

2. Related Work

2.1 Environmental Sound Recognition

There are many ubicomp studies on environmental sound recognition. For example, in Ref. [3], bathroom activities such as showering, flushing, and urination were recognized using microphone data. Also, several studies recognize daily activities with microphones in smartphones by recognizing environmental sounds such as the sound of vacuuming and the sound of running water [19], [23].

2.2 Sensing Toothbrushing

In Ref. [14], Braun's SmartGuide was used to study the effects of real-time feedback on the quality of toothbrushing, in which they found a significant improvement in brushing habits when using this system. Other research has been conducted on the analysis of toothbrushing behavior using optical motion capture systems [2], [13] and embedded accelerometer sensors [11], [15], [16], [17], [25]. In particular, a system developed in Ref. [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, Ref. [11] used an embedded accelerometer to evaluate toothbrushing 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, which eliminates the need for most users to purchase any new equipment.

3. Toothbrushing Sensor Data

3.1 Assumed Environment

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

We extracted features from the raw audio data as vectors of mel-frequency cepstral coefficients (MFCCs). Although MFCCs were originally designed for use in speech recognition, they have also been successfully applied to environmental sound recognition [3]. **Figure 2** shows graphical representations of MFCC data derived from toothbrushing audio. As shown in the figure, the audio characteristics differ when brushing the back teeth from when

Fig. 1 Assumed setup for using a smartphone to record audio from toothbrushing.

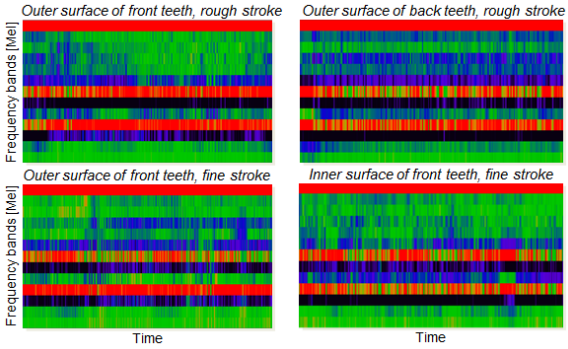

Fig. 2 MFCC representation of audio data from four toothbrushing activity classes.

Table 1 HMM classes used for audio recognition.

Outer front teeth, fine (FO-Fine)	Outer front teeth, rough (FO-Rough)
Outer back teeth, fine (BO-Fine)	Outer back teeth, rough (BO-Rough)
Inner front teeth (FI-Fine)	Inner back teeth (BI-Fine)
No toothbrushing activity (None)	

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 toothbrushing 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 toothbrushing.

3.2 Toothbrushing Activity

We use HMMs based on audio characteristics to recognize the seven toothbrushing activities listed in **Table 1**. Note that the classes *Inner front teeth, rough* and *Inner back teeth, rough* were not included in this study, as an insufficient amount of data was collected for these activities.

In this study, the term *inner* refers to the inner (i.e., lingual) surface, the term *outer* refers to the outer (i.e., facial) surface, the term *front teeth* refers to the incisors and canine teeth, and the term *back teeth* refers to the molars. The term *rough* indicates that the stroke used when brushing was too forceful, while the term *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 toothbrushing activities in Table 1 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 toothbrushing.

During our investigation, a limitation was found in using audio data to classify toothbrushing activities. While audio data can be used to differentiate between brushing the front vs. the back of the mouth and between brushing the inner surface vs. the outer surface of the teeth, it cannot be used for more symmetric differentiations such as the left vs. right side or upper vs. lower teeth. Because of this limitation, some issues can arise when scoring

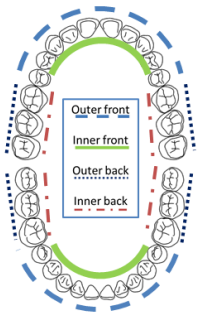

Fig. 3 Four regions of the mouth used during evaluation of toothbrushing performance.

a user's toothbrushing. For example, in the case where a user brushes their upper front teeth for a long duration, but not their lower front teeth, the evaluation score for his/her toothbrushing 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.

3.3 Toothbrushing Evaluation by a Dentist

Using the audio data collected as described above, we applied a machine learning approach to evaluate and estimate a score for the user's toothbrushing performance. To do this, we needed training data that could be used to generate score estimates. In this research, a dentist prepared such training data, allowing for an evaluation of toothbrushing performance that is based on an actual dentist's evaluation. One typical method used by dentists for evaluating toothbrushing 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 toothbrushing used in the days preceding the test, plaque tests may not be an ideal test for evaluating isolated sessions of toothbrushing.

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 Fig. 1, we recorded video data for each session of toothbrushing using a smartphone. A dentist then evaluated the toothbrushing performance using the video data, and assigned evaluation scores for each session of toothbrushing, assigning scores for each of the four regions of the mouth depicted in **Fig. 3**. These scores were then combined with audio data extracted from the videos to build the score estimation models. Because the dentist evaluated the toothbrushing 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.

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, we award up to 2 points 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.

3.4 Relationship between Video-based Evaluation Scores and Plaque Test Scores

During this study, an experiment was conducted to examine whether our video-based evaluation can adequately evaluate how well a user is removing plaque. Since plaque tests are a standard test used in evaluating plaque removal, we conducted an experiment in which we evaluated the toothbrushing of several users, with each session of toothbrushing evaluated using both a plaque test and a video-based evaluation. In this experiment, 14 subjects were videoed while brushing their teeth using the setup depicted in Fig. 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 conduct a video-based evaluation. In this way, we were then able to measure the correlation between our video-based evaluation scores and plaque test scores, giving us an indication of how well our method is evaluating plaque removal.

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 Fig. 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 brushed their teeth and a plaque score was calculated. Finally, all videos were evaluated using this study's criteria to assign scores, and the video-based score for each session was compared to the corresponding plaque score for each session.

Figure 4 shows the relationship between plaque scores and the video-based evaluation scores. 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 use a 24-point scale with higher values indicating better toothbrushing 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 toothbrushing, there was a shift between the plaque scores and the video-based scores.

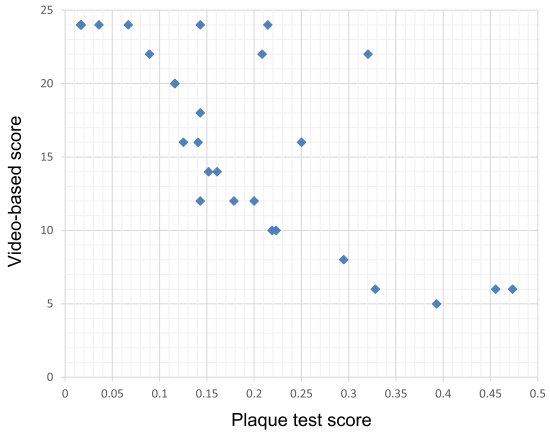

Fig. 4 Comparison of plaque test scores and video-based evaluation scores from 28 sessions of toothbrushing.

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 toothbrushing, the plaque score and the scores used by this study were strongly correlated. Based on this, we believe that our study's scoring method is able to assign scores that closely correspond to the de facto standard plaque score without applying plaque indicator liquid. Using this method, we are able to easily prepare the large amount of toothbrushing scores needed for a machine learning approach through the use of video data.

4. Proposed Method

4.1 Naïve Architecture

The basic procedure used in this study starts with using HMMs to recognize toothbrushing activities in audio data. We then generate independent (explanatory) variables from those recognition results, using these independent variables to build regression models for estimating scores for sessions of toothbrushing activity. Both the audio data used to train the HMMs and the evaluation scores used to train the regression models are collected in advance, without the need for any labeled training data from the target users. In our method, we adapt the HMMs to each target user using raw unlabeled audio collected from the end user during normal use.

In our simplest implementation, we use an HMM set that has seven HMMs (corresponding to all seven classes from Table 1) to recognize toothbrushing activities in the audio data, which corresponds to *HMM-7* in **Fig. 5**. The architecture for this implementation is depicted in **Fig. 6**. It starts by using *HMM-7* to perform audio recognition, and then uses the output from *HMM-7* to generate the independent variables for its regression model. Finally, the regression model outputs a score from 0 to 24, representing an overall evaluation of the user's brushing across all four areas of the mouth.

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 a more complex version of the naïve 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:

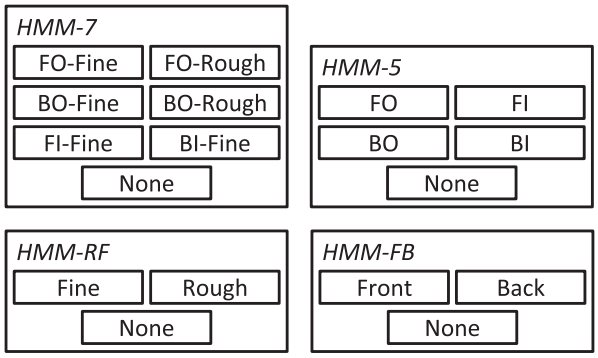

Fig. 5 The four basic HMM sets with varying granularity. HMM-7 has all seven toothbrushing activity classes. HMM-5 removes the distinction between fine stroke and rough stroke, resulting in five classes. HMM-RF and HMM-FB both simplify the classes down to three each.

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

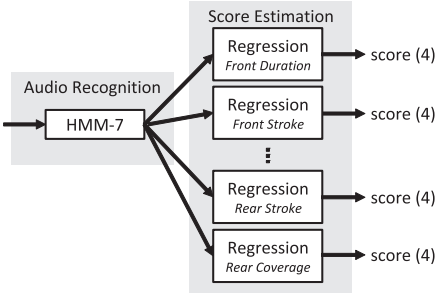

Fig. 7 Naïve architecture: Uses the same seven-class HMM set (HMM-7) to generate independent variables for all regression models.

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 naïve architectures described above have the following issues:

- Accurate classification of toothbrushing activities into seven classes is difficult, and in the case of some architectures it is unnecessary. For example, the architecture in Fig. 7 estimates scores for only two regions, *front teeth* and *back teeth*. In this case, distinguishing between all seven toothbrushing activities may be unnecessary for score estimation, and more accurate estimates are possible by using a coarser set of classes without the *inner surface* and *outer surface* distinction.
- Each of the regression models estimates scores using the classification results from the same HMM set, but the usefulness of the toothbrushing 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 a coarser set of classes depending on the needs of the regression model, more accurate results can be achieved.

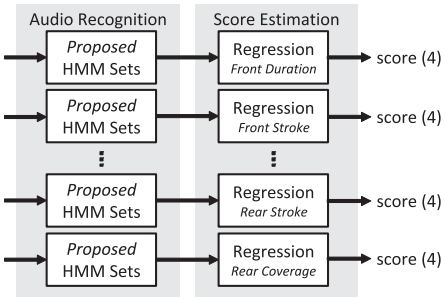

Fig. 8 Proposed architecture: Uses a separate group of eight HMM sets for each regression model, with each group made up of the four basic HMM sets (HMM-7, HMM-5, HMM-RF, and HMM-FB) and four HMM sets that have been tailored to the regression model.

4.2 Overview of Proposed Approach

In the proposed method, we solve the problems with the naïve architectures described above by preparing separate HMM sets for each of the regression models used for score estimation, as is shown in **Fig. 8**. For example, in Fig. 8, we prepare a specialized HMM set for the regression model that estimates a score for *Front Duration*. Each of the HMM sets generated is specialized to its regression model, in order to increase the estimation accuracy of the regression model. Specifically, we automatically discover which toothbrushing activities are useful for estimating the score in question and then generate an HMM set that focus on only those classes. When doing so, we ignore activity classes that are not considered useful for estimating the score. The recognition results from this reduced model set are then 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 toothbrushing activity classes are important when estimating each score.
(2) Generate HMM sets for accurately recognizing those important classes.
(3) Build a regression model for estimating each score using the recognition results of those HMM sets.

Each of these steps is explained in detail below.

4.3 Discovering Useful Toothbrushing Activity Classes

As discussed above, the usefulness of toothbrushing activity classes varies according to different evaluation criteria. In this study, we use regression models to estimate the evaluation criteria, extracting 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 [22], 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 then determining the k nearest neighbors for each of those m instances. 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 [22]:

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

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 toothbrushing activity classes C for a given evaluation criterion. The usefulness U_c of a toothbrushing 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 toothbrushing activity class c:

$$U_c = \sum_{f \in \boldsymbol{F}_c} W(f).$$

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.

4.4 Tailoring HMM Sets to Improve Score Estimates

Using the method described in the previous subsection, we can determine which toothbrushing activity classes are useful for estimating scores for a given evaluation criterion. Using this information, we can determine which classes are most useful and make an HMM set using only those useful classes. As mentioned previously, in the naïve approach there are two issues that arise from using the same HMM 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. We address the first of these issues by generating four basic HMM sets that have varying granularity (see Fig. 5 for a graphical depiction of these sets):

- ***HMM-7***: A seven-class HMM set generated using all seven toothbrushing activity classes.
- ***HMM-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).
- ***HMM-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*.
- ***HMM-RF***: A three-class HMM set for distinguishing between stroke types, generated using the classes *rough stroke*, *fine stroke*, and *no activity*.

We address the second of the issues with the naïve architectures by generating a tailored 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 toothbrushing activity classes included in a basic HMM set. We then only include the class c in the new model set if $U_c \geq T$. 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-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 classes: 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 HMM sets, the four basic HMM sets from Fig. 5 and four tailored HMM sets that are generated from each of the four basic HMM sets (HMM-7, HMM-5, HMM-RF, and HMM-FB) and are tailored to the score being estimated.

4.5 Toothbrushing Activity Recognition

Using the method described in the previous subsection to select the classes used in each of our HMM sets, we then generate the HMMs used for toothbrushing activity recognition.

4.5.1 Feature Extraction

In this study, we use MFCCs to recognize toothbrushing activities, as MFCCs have been reported to be one of the better transformation schemes for environmental sound recognition [3], [4]. We extract MFCCs using the hidden Markov model toolkit (HTK) [26]. We compute a 12-order MFCC, along with the log energy for the window and the corresponding 13-order delta and 13-order acceleration coefficients, giving a vector of 39 values in total.

4.5.2 Recognition with HMMs

Our method uses HMMs to recognize toothbrushing activity classes in audio data, with our HMMs implemented using HTK [26]. **Figure 9** shows an example of the 10-state left-to-right

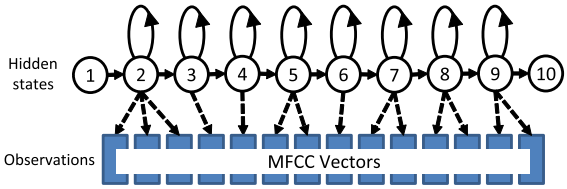

Fig. 9 10-state left-to-right HMM used to represent toothbrushing activities. The first and last states are non-emitting. The dashed lines show an example of how hidden states could be matched with individual MFCC vectors using such an HMM.

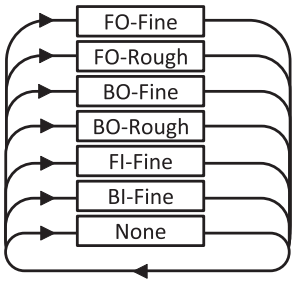

Fig. 10 Isolated word grammar used for toothbrushing activities in our HMMs.

HMM used for each class, e.g., FO-Fine. The observed variables for the models are the vectors of 39 MFCC-based coefficients. These models are tied together across sessions of toothbrushing audio using the isolated word grammar depicted in **Fig. 10**. Using this grammar, our system allows for any sequence of toothbrushing activities to occur in a session of toothbrushing, so users do not need to brush their teeth in a predetermined order.

As was mentioned in the introduction, the model of toothbrush and the shape of the user's mouth can affect the sound made when brushing his/her teeth, so the audio obtained for the toothbrushing activities will differ per user. In order to cope with this issue, we also employ an unsupervised version of the maximum likelihood linear regression (MLLR) adaptation method [9], [18] to shift the output distributions of the initial toothbrushing 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. MLLR adaptation works by creating a transformation matrix which can be used to transform a user-independent HMM 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 toothbrushing 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.

In order to perform MLLR on the unlabeled data collected from the end-users, we first perform recognition of the target user's audio using unadapted HMMs. We then use the labels estimated by the unadapted HMMs to label the target user's audio, and run MLLR using the now labeled target-user audio, giving us our adapted HMMs. We then use these HMMs to recognize toothbrushing activities over full sessions of audio data using the Viterbi algorithm [21], finding the most probable sequence of toothbrushing activity classes across the session. Using these recognition results, we can then compute the independent variables used in the regression models.

4.6 Estimating Scores

4.6.1 Computing Independent Variables

Using the adapted HMMs, it is possible to recognize which toothbrushing activities were conducted in a session of audio data. For example, by using *HMM-7*, it is possible to detect that the activity *FO-Rough* 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 sets, excluding the None and Others classes. Each of these variables is computed as the total duration of its corresponding toothbrushing 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. 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).

4.6.2 Estimating a Score for Each Criterion

Finally, using these independent variables, we estimate the evaluation scores using regression analysis. We first perform dimensionality reduction using the WEKA implementation of the Random Projection algorithm [6], [12] to reduce the number of variables down to 10. Using these 10 independent variables, we then estimate scores using WEKA's support vector machine (SVM) algorithm for regression, using the sequential minimal optimization (SMO) algorithm to analytically solve the dual optimization problem ($W(\alpha)$) that is used to train the SVM [12], [24]:

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} y_i y_j k(\vec{x}_i, \vec{x}_j) \alpha_i \alpha_j,$$
$$0 \le \alpha_i \le C, \qquad \forall i,$$

$$\sum_{i=1}^{\ell} y_i \alpha_i = 0,$$

where each α is a Lagrange multiplier, $\vec{x}_i$ is the ith of ℓ input vectors, y_i is the target value for the ith input vector, $k(\vec{x}_i, \vec{x}_j)$ is a kernel function, and C is a complexity constant (a tunable parameter) [20].

5. Evaluation

5.1 Data Set

In this study, we gathered a total of 94 sessions of toothbrushing audio from 14 participants. All audio data was collected with a smartphone microphone using the setup depicted in Fig. 1. The audio was collected as WAV files with a sampling rate of 44.1 kHz. The average time for each session of toothbrushing was approximately 94 seconds.

The study was conducted over the course of three months, with the data collected in a quiet environment, either in the bathroom of our graduate school building or in the bathroom in a participant's own home. Both environments included normal background noises, such as the sound of air conditioning units and fans. All participants used manual toothbrushes, either their own toothbrush or a toothbrush provided by our lab. During the course of the experiment, each participant received instruction from a dentist on proper toothbrushing technique. All sessions were evaluated using video data as was described in Section 3.3. **Figure 11** shows the distribution of scores for the sessions. The audio data was labeled using the corresponding video data for each session, with each label corresponding to one of the classes of toothbrushing activity described in Section 3.2.

Additionally, this study included an investigation on the effects of background noise on toothbrushing activity recognition. In this study, we collected five sessions of audio while running a hair dryer in the background near the smartphone used to collect the audio. We tried to cope with the background noise by employing Cepstral Mean Normalization (CMN), which is an additive noise cancellation technique that is used widely in speech recognition studies [8], [26].

5.2 Experiment Parameters

The 39 MFCC-based coefficients were computed based on a 26-channel filterbank over a window of 50 ms with 50% overlap, windowed using a Hamming window. We applied energy normalization, cepstral liftering with a coefficient of 22, and a first order preemphasis with a coefficient of 0.97.

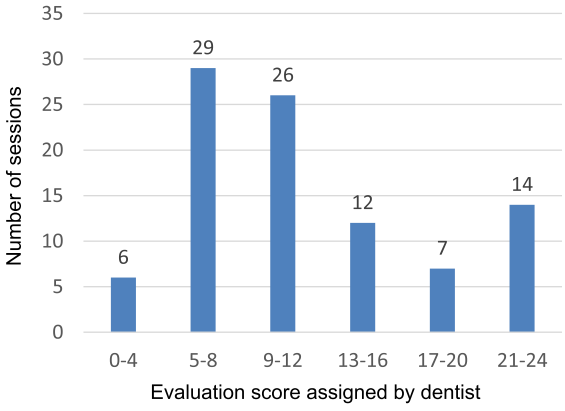

Fig. 11 Distribution of scores for the dataset.

The HMMs used in this study were 10-state left-to-right models with output distributions represented by 32 Gaussian mixture densities.

The target number of dimensions used for the random projections algorithm was 10. Additionally, we used a random seed of 42 and the *Sparse1* distribution to calculate the projection matrix, with *Sparse1* computed as $r_{ij} = \sqrt{3} \times (\pm 1$ with probability $\frac{1}{6}$ each, 0 with probability $\frac{2}{3})$.

For the SMO regression algorithm, we normalized all attributes, set the complexity constant to 1.0, set the tolerance parameter to 0.001, set the epsilon parameter in the epsilon-insensitive loss function to 0.001, set the epsilon round off error to 1.0×10^{-12}, set the random number seed to 1, and used the linear kernel:

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^{\mathsf{T}} \mathbf{y},$$

with a cache size of 250,007.

5.3 Evaluation Methodology

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

- ***Avg***: A baseline method in which we estimated a user's scores using the average scores for all other users. This method was chosen to represent a low baseline for comparison with the other methods.
- ***SHMM***: A naïve approach in which we prepared only a single HMM set (*HMM-7*). Otherwise this method was the same as the proposed method. This method represents a straightforward implementation of our proposed system for toothbrushing evaluation.
- ***SHMM100***: A modified version of the *SHMM* method in which we built the regression models using corrected labels instead of actual audio recognition results, i.e., this method simulates 100% recognition accuracy for *HMM-7*. This method represents an upper bound on performance for a straightforward implementation of our proposed system.
- ***MHMM***: A baseline method in which we prepared the four basic HMM sets from Fig. 5, but did not prepare any tailored HMM sets. Otherwise this method was the same as the proposed method. This method was chosen as a means to compare the effectiveness of tailoring HMM sets to specific scores versus only providing multiple HMM sets with varying granularity.
- ***Proposed***: The proposed method, in which we prepared groups of eight HMM sets for each of the scores, made up of the four basic HMM sets from Figure 5 and the four tailored HMM sets generated from those basic HMM sets.

Additionally, we prepared six evaluation architectures to use when testing these methods. Showing the results across several architectures allows us to better evaluate the overall performance of the methods, since each architecture represents a trade-off between estimation accuracy and usefulness of the scores provided.

- ***Total* (24)**: Estimated a single score (24-point scale) that represents the total score for all toothbrushing activity in the session.

Table 2 Recognition results for basic HMM sets used in this study.

	precision	recall	F-measure
HMM-7	0.457	0.455	0.451
HMM-5	0.485	0.506	0.491
HMM-FB	0.658	0.654	0.652
HMM-RF	0.677	0.692	0.684

- *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 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). Accuracy was measured using the mean absolute error (MAE) and error ratio. The MAE was calculated using:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i|,$$

where e_i is the error for the ith estimate.

The error ratio was calculated using $MAE/MaxScore$, where $MaxScore$ is the maximum score possible for a given architecture. For example, $MaxScore = 24$ for the *Total* architecture and $MaxScore = 12$ for the *FB* architecture.

5.4 Audio Recognition Results

5.4.1 Audio Recognition by Basic HMMs

Table 2 shows the recognition results for each of the basic HMM sets used in this study (see Fig. 5), using the macro-averaged F-measure as the performance metric. Both *HMM-7* and *HMM-5* had similar results, with F-measures of 0.451 and 0.491 respectively. Both *HMM-FB* and *HMM-RF* had comparable results, achieving average F-measures of 0.652 and 0.684, respectively. In all cases, recognition accuracy is well below 100%, but still high enough to gain a significant amount of information about the location and brush stroke that corresponds to the audio data.

5.4.2 Audio Recognition by Tailored HMMs

In order to confirm the effectiveness of our proposed method when creating tailored HMM sets for audio recognition, we first compare the recognition results for the tailored HMM sets with those for the basic HMM sets. **Table 3** shows the percent change in average precision, recall, and F-measure when switching from the four basic HMM sets to the tailored HMM sets generated using our proposed method. As is seen in these results, in *HMM-7* and *HMM-5* there was a large improvement in recall, but a slight deterioration in precision, resulting in an increase in F-measure for both, with *HMM-7* having the highest increase in F-measure. On the other hand, little change in performance was seen for *HMM-FB* and *HMM-RF*. For both *HMM-FB* and *HMM-RF*, there were only three classes initially (including the None class). Therefore, there was little difference between the basic three-class HMM sets and the HMM sets generated by the proposed method, as it was rare that one of the two classes other than *None* was judged useless.

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

	Δ precision	Δ recall	Δ F-measure
HMM-7	−1.9	10.8	4.1
HMM-5	−5.9	9.2	0.3
HMM-FB	0.0	0.0	0.0
HMM-RF	0.0	0.3	0.2

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

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

5.4.3 Effects of Noise

Table 4 shows an overall degradation of performance due to noise, with the F-measures for all sets reduced by at least 29%. Furthermore, the F-measure for *HMM-RF* dropped below 0.33, with the set apparently no longer able to distinguish between the classes. The reduction in F-measure for *HMM-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.

5.5 Score Estimation Results

5.5.1 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, *SHMM100* shows the results when the toothbrushing 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 *SHMM100*. Additionally, when comparing the *SHMM100* results to *SHMM*, *SHMM100* 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 aver-

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
SHMM100	3.10	1.58	2.61	1.04	1.41	0.58	**1.72**
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**
Proposed w/o var	4.25	1.53	2.74	0.95	1.38	0.56	1.90

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
SHMM100	12.9	19.8	21.7	26.1	23.6	29.2	22.2
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**
Proposed w/o var	17.7	19.2	22.8	23.6	23.0	27.8	22.4

age MAE across the architectures as *SHMM100*. In comparison to *SHMM100*, 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 HMMs 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 *SHMM100*.

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*.

5.5.2 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.

5.5.3 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.

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.

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. When analyzing the results of audio recognition, we found that the recognition accuracy for BI-Fine was low, which most likely had a large influence on the regression results.

5.5.4 Effectiveness of Independent Variables

This section discusses the independent variables that were useful for estimating various scores. We determined the usefulness for these variables using the RReliefF algorithm described earlier. **Table 7** shows that the variable for the total length of time spent brushing the teeth with a fine stroke was found to be useful for the *Total* architecture. Its usefulness was likely because it pro-

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

vides 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 *Coverage* scores. When estimating *Stroke* scores, 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 for *CSD*. However, in the case of the *FB* architecture, 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, computed from *HMM-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.

6. Conclusion

This paper presented a new method for evaluating toothbrushing performance using audio collected from a smartphone. By requiring only audio data from end users, this method enables users to evaluate their toothbrushing with little effort. Nevertheless, the use of audio data for evaluation does have its challenges. Even after attempts to increase performance through adapting models to users and tailoring classes to specific evaluation scores, audio recognition accuracy still ranged from 49% to 68%. However, despite the difficulties in recognizing toothbrushing activities in audio data, our experiments indicate that our proposed method is still able to compensate for the low recognition accuracy, with our method closely matching the performance of a baseline method that simulated 100% accuracy in audio recognition results.

The dentists participating in this study consider a system that does not require specialized equipment and provides feedback on toothbrushing performance with average error rates as low as 22% to be significant, as no equivalent system is currently available. They believe that error rates in the range of 20% to 30% are acceptable, and that the proposed method can be applied to real-life applications. Furthermore, as we are able to collect more data to use for training such a system, we believe that we will be able to provide even more effective toothbrushing guidance with this method.

Overall, our experiments indicate that our method can achieve acceptable performance when used to evaluate toothbrushing. While our current 20% to 30% error rates are believed to be acceptable for real-life applications, we plan to work to further improve on our method to reduce these error rates. As a part of our future work, we plan to employ deep learning techniques to discover useful features tailored for recognizing toothbrushing audio.

References

[1] Addy, M. and Hunter, M.: Can tooth brushing damage your health? Effects on oral and dental tissues, *International Dental Journal*, Vol.53, No.S3, pp.177–186 (2003).

[2] 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, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.363–372, ACM (2008).

[3] Chen, J., Kam, A., Zhang, J., Liu, N. and Shue, L.: Bathroom activity monitoring based on sound, *Pervasive 2005*, pp.47–61 (2005).

[4] Cowling, M.: Non-speech environmental sound recognition system for autonomous surveillance, PhD Thesis, Griffith University (2004).

[5] Fiske, J., Davis, D., Frances, C. and Gelbier, S.: The emotional effects of tooth loss in edentulous people, *British Dental Journal*, Vol.184, No.2, pp.90–93 (1998).

[6] Fradkin, D. and Madigan, D.: Experiments with random projections for machine learning, *KDD 2003*, pp.517–522 (2003).

[7] Ganss, C., Schlueter, N., Preiss, S. and Klimek, J.: Tooth brushing habits in uninstructed adults-frequency, technique, duration and force, *Clinical Oral Investigations*, Vol.13, No.2, pp.203–208 (2009).

[8] Garner, P.N.: Cepstral normalisation and the signal to noise ratio spectrum in automatic speech recognition, *Speech Communication*, Vol.53, No.8, pp.991–1001 (2011).

[9] Gauvain, J. and Lee, C.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains, *IEEE Trans. Speech and Audio Processing*, Vol.2, No.2, pp.291–298 (2002).

[10] 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*, Vol.8, No.126, p.552 (2010).

[11] 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*, Vol.84, No.2, pp.186–195 (2013).

[12] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H.: The WEKA data mining software: An update, *ACM SIGKDD Explorations Newsletter*, Vol.11, No.1, pp.10–18 (2009).

[13] 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*, pp.1–12 (2014).

[14] 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, *J. Contemp. Dent. Pract.*, Vol.9, No.7, pp.1–8 (2008).

[15] 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*, pp.1–6 (2014).

[16] 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*, Vol.96, No.2, pp.125–132 (2009).

[17] 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 Trans. Biomedical Engineering*, Vol.59, No.3, pp.872–881 (2012).

[18] Leggetter, C. and Woodland, P.: Maximum likelihood linear regres-

sion for speaker adaptation of continuous density hidden Markov models, *Computer Speech & Language*, Vol.9, No.2, pp.171–185 (1995).
[19] Lu, H., Pan, W., Lane, N., Choudhury, T. and Campbell, A.: SoundSense: Scalable sound sensing for people-centric applications on mobile phones, *MobiSys 2009*, pp.165–178 (2009).
[20] Platt, J. et al.: Fast training of support vector machines using sequential minimal optimization, *Advances in Kernel Methods — Support Vector Learning*, Vol.3 (1999).
[21] Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE*, Vol.77, No.2, pp.257–286 (1989).
[22] Robnik-Šikonja, M. and Kononenko, I.: An adaptation of Relief for attribute estimation in regression, *ICML 1997*, pp.296–304 (1997).
[23] Rossi, M., Feese, S., Amft, O., Braune, N., Martis, S. and Troster, G.: AmbientSense: A real-time ambient sound recognition system for smartphones, *IEEE International Conference on Pervasive Computing and Communications Workshops* (*PERCOM 2013 Workshops*), pp.230–235 (2013).
[24] Shevade, S., Keerthi, S., Bhattacharyya, C. and Murthy, K.: Improvements to the SMO algorithm for SVM regression, *IEEE Trans. Neural Networks*, Vol.11, No.5, pp.1188–1193 (2002).
[25] 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*, pp.1–9 (2014).
[26] 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).

Joseph Korpela received his M.S. in Computer Science from UCLA in 2014. He is currently a Ph.D. student at Osaka University, Japan. His research interests include ubiquitous computing and human activity recognition. He is a member of IPSJ, IEEJ, and ACM.

Ryosuke Miyaji was a graduate student at Osaka University, Japan. His research interests include activity recognition. He graduated in 2015 and is now working for Morikita Publishing Co., Ltd.

Takuya Maekawa was born in 1980. He is an associate professor at Osaka University, Japan. His research interests include ubiquitous computing and mobile sensing. He has a Ph.D. in Information Science and Technology from Osaka University. He is a member of IPSJ, IEEJ, DBSJ, ACM, and IEEE.

Kazunori Nozaki received his D.D.S. from Hokkaido University in 2000, and his Ph.D. from Osaka University in 2004. He was a post doctoral fellow at Osaka University's Cybermedia Center from 2004 to 2009. He received his Ph.D. in Informatics from Osaka University in 2010. He was a specially appointed lecturer for the Center for Advanced Medical Engineering and Informatics from 2009 to 2012, and a specially appointed lecturer for the Graduate School of Engineering Science of Osaka University. He is currently an assistant professor at the Osaka University Dental Hospital. His current research work is focused on using Artificial Intelligence to utilize physical knowledge for oral surgery. He was awarded the HPC Analytics Challenge Finalist in 2006 and Best Research Award of JBMES in 2010. He is a member of JAMI and JSAI.

Hiroo Tamagawa is an associate professor at the Osaka University Dental Hospital. He received his D.D.S. and Ph.D. from Osaka University in 1979 and 1986, respectively. His current research interests are in hospital information systems and standardization in dentistry.