YouBrush: Leveraging Edge-Based Machine Learning in Oral Care

Esteban Echeverri[1](✉) , Griffin Going[1] , Rahat Ibn Rafiq[1] ,
Jonathan Engelsma[1] , and Venu Vasudevan[2]

[1] Grand Valley State University, Allendale, MI, USA
{echevere.mail,goinggr,rafiqr,engeljo}@gvsu.edu
[2] P&G,Cincinnati, USA
vasudevan.v@pg.com

Abstract. A disconnect is frequent regarding the length of time a person claims to have brushed their teeth and the actual duration; the recommended brushing duration is 2 min. This paper seeks to bridge this particular disconnect. We introduce YouBrush,—a low-latency, low-friction, and responsive mobile application—to improve oral care regimens in users. YouBrush is an IOS mobile application that democratizes features previously available only to intelligent toothbrush users by incorporating a highly accurate deep learning brushing detection model—developed by Apple's createML—on the device. The machine learning model, running on the edge, allows for a low-latency, highly responsive scripted-coaching brushing experience for the user. Moreover, we craft in-app gamification techniques to further user interaction, stickiness, and oral care adherence.

Keywords: Mobile Development · Oral Care · Sound

1 Introduction

Mobile phones and embedded sensors are becoming increasingly influential in various health care use-cases. Several researchers have demonstrated the success of using sensor technologies to monitor the exercise, dietary, and sleep regimen of subjects [7,38]. This paper uses mobile phones' audio-sensory capability to improve oral care regimens. Oral health care is of paramount importance—researchers report it significantly impacts one's quality of life [15]. Notwithstanding the proven impact of oral care on physical and emotional well-being [15], studies demonstrate that a substantial portion of the population brushes with an incorrect technique, such as improper toothbrushing [13].

Several mobile applications and researches focusing on healthcare have come to the fore [12] in recent years. The majority of these researches leverage optical motion capture systems [18], accelerometer sensors installed in smart toothbrushes [20], and audio sensors [21]. Positive feedback and gamification techniques such as leaderboards have been influential in improving subjects' daily oral care

regimen of subjects [12, 22]. Other oral care use-cases such as disease detection [5, 36] have seen a significant application of deep learning approaches. Researches have shown that real-time positive feedback and a highly responsive coaching mechanism can significantly improve a user's experience with a mobile application that seeks to improve a particular behavior [10, 33]. To this aim, in this paper, we present YouBrush, which employs edge-based deep learning, mobile phones' audio sensory capability, and a gamification mechanism to facilitate the toothbrushing regimen of a subject. First, we collect and label brushing audio data, and bathroom ambient sounds to construct a robust and diverse training dataset. Then, we use Apple's off-the-shelf createML [1] to train and develop a deep learning model that accurately detects brushing events using the dataset. Next, we implement a gamification and dynamic feedback mechanism to augment user experience and engagement. Finally, we incorporate the deep learning and gamification modules in the YouBrush mobile application, which we develop in Swift programming language [4]. The edge-based brushing audio detection enables YouBrush to provide a seamless, highly responsive scripted-coaching brushing experience for the user—facilitating an engaging user experience.

Concerning other oral care aid applications with specific deference to non-smart toothbrushes, a weakness is present in the user trust requirement. Such applications often, for example, present themselves as a simple two-minute timer when in the context of an actual brushing/oral care session. Thus, the application has little power to ensure that the user spends the entire two-minute period brushing their teeth. Instead, it must simply assume that the user does what they say they are doing or will do. The aforementioned disconnect between user perception and reality makes this a less-than-ideal implementation. Of the primary goals of YouBrush, one of the most important is to replace this requirement with an accountability mechanic. This is done by using the edge-based brushing audio detection to drive a two-minute timer, forcing the user to actively participate in the brushing activity to push the timer forward instead of simply observing the timer. YouBrush relies heavily on this inversion, creating an environment in which the user's actions and habits drive application progress and logic where viable instead of the application solely driving the user.

Attempts such as YouBrush that log and evaluate daily activity are critical in promoting a healthy lifestyle [21]. Fitbit[1], an activity-tracker company in the wearable technology space, journals a person's running, walking, sleeping habit, and heart rate, for example. This information-logging is a tremendous help for the user to self-evaluate their daily progress. We argue that YouBrush has the potential to serve a similar requirement to evaluate toothbrushing performance. Moreover, the presence of a competitive leaderboard can motivate the users further to be stricter followers of daily oral-care regimens.

We organize the rest of the paper as follows. *First*, we introduce the related work in four subareas related to our research: oral care and mobile applications, oral care and machine learning, oral care and gamification, and brushing sensing. *Second*, we present the YouBrush application design and implementation details.

---

[1] https://www.fitbit.com/global/us/home.

Third, we describe the methodologies used to collect, augment and label the audio datasets we use to train our deep learning model to detect brushing events. *Finally*, we detail the tools and techniques used to train and implement the deep learning model and present the performance results.

2   Related Work

2.1   Oral Health Care and Mobile Apps

Mobile-based healthcare applications have spiked in recent years, and oral healthcare is no exception. Dentists, orthodontists, and oral health care evangelists have sought to improve their patients' oral care regimens by leveraging this ubiquitous nature of the mobile paradigm. Such efforts include health-risk information provision, self-monitoring of behavior and behavioral outcomes, prompting barrier identification, setting action and coping plans, and reviewing behavioral goals [30]. Researchers have also proposed mobile-based solutions to focus on diseases and treatments, such as oral mucositis [23], facilitation of the recovery process of patients who just went through an orthodontic treatment [37], promotion of oral hygiene among adolescents going through such treatments [34]. Recent works have also put under the microscope ways to ensure regular engagement of oral-care-related mobile apps—for example, motivating users to brush their teeth for 2 min using music [40].

2.2   Oral Care and Gamification

Gamification is defined as applying game design elements in non-game contexts [11]. Several Oral Care mobile applications employ gamification techniques such as badges, leaderboards, and levels to increase user engagement and activity [12]. In addition, researchers have recently investigated the effect of positive feedback through mobile applications and gamification techniques to gauge the improvement of oral care regimens among subjects [22].

2.3   Brushing Sensing

A recent study has found that real-time feedback can significantly improve the quality of brushing [18]. In [18], the authors evaluated the ability of a power brush with a wireless remote display to improve brushing force and thoroughness. Optical motion capture systems [18], accelerometer sensors embedded in brushing devices [20,39], and brushing audio sounds [21] have also been leveraged to analyze brushing behaviors. Research has shown that proper graphical user feedback during brushing—a cartoon display for children to show regions of the mouth that were adequately brushed to be displayed as plaque-free [6] or smart digital visual toothbrush monitoring and training system (DTS) in terms of correct brushing motion and grip axis orientation [16]—can motivate better oral care regimen. In [21], the authors proposed a low-cost system built around

an off-the-shelf smartphone's microphone to evaluate toothbrushing performance using hidden Markov models. However, in this paper, we leverage gamification and deep learning in addition to audio sensors to evaluate brushing performance. Audio and deep learning, in recent years, have seen several applications in areas such as speech [8], music [25], and environmental sound detection [27].

3   Application Design and Implementation

YouBrush revolves around three primary aims: Ensure regular brushing, quality of brushing technique, and every brushing event lasts at least 2 min [31]. For this purpose, we proceed to the application design and implementation phase of YouBrush. We use Swift to develop the frontend and use Firebase[2] as the backend of our application. Next, we focus on four design choices imperative for the implementation phase of YouBrush: handling live audio, in-app data presentation, privacy and latency concerns, and constructing the in-app game mechanics. The following subsections describe the implementation of each of the above-mentioned design choices (Table 1).

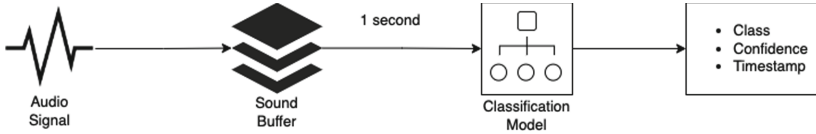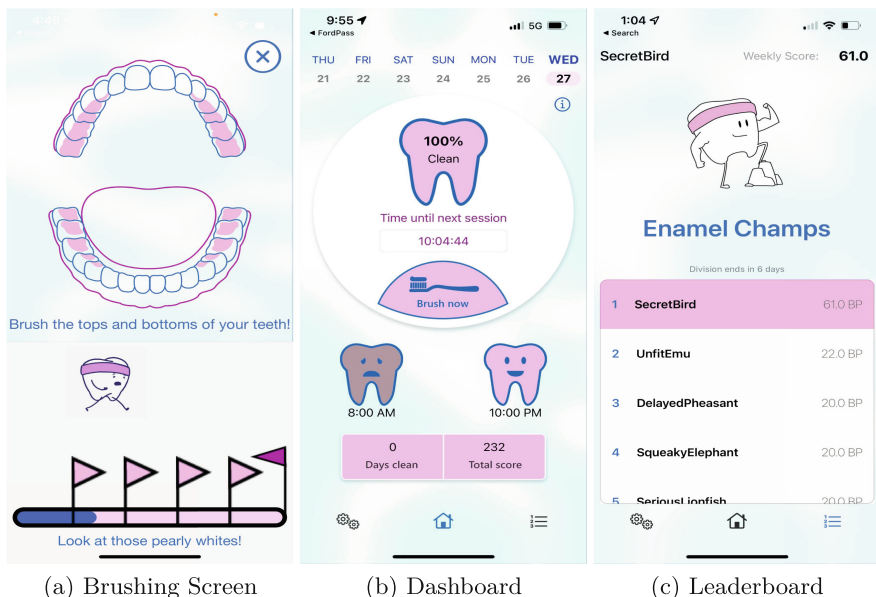### 3.1   Audio Processing

Fig. 1. Audio pipeline

Figure 1 depicts the pipeline we develop to process the audio stream whenever the microphone detects an audio signal. First, a sound buffer stores the audio stream and chunks them into 1-s audio—this chunking helps us later in the audio classification phase. Second, the pipeline sends the 1-s audio samples to the classification module for brushing event detection, enabling the app to accurately timestamp a brushing event's start and end point.

3.2   GUI Design

Brushing Screen: Figure 1a depicts the brushing screen. This screen serves the following. First, it provides the user with immediate goals after a brushing event is initiated. The immediate goals represent different stages of a brushing session—suggesting to the user which areas of their teeth (top and bottom of their teeth, for example, in Fig. 1a) they should be brushing. These immediate

---
[2] https://firebase.google.com/.

Table 1. YouBrush Application

<table>
<tr><td>(a) Brushing Screen</td><td>(b) Dashboard</td><td>(c) Leaderboard</td></tr>
</table>

goals represent real-time scripted coaching to improve brushing quality. Second, when the user finally completes the final goal, the app marks the brushing session as complete, ensuring a two-minute brushing session.

Dashboard Screen: Figure 1b shows the dashboard—the screen the users see after they log in. We highlight four points on the screen. The brown and pink-colored teeth at the bottom denote an incomplete (or did not use the app when brushing) and complete brushing session at the specified time, respectively. On the top, a timer lets the user know when is their next brushing session. The pink-colored tooth represents how clean their teeth are at the moment—an estimation based on their last completed brushing event. Finally, at the very bottom, the app provides a total score that we calculate based on user consistency. We use this score to facilitate the in-app game mechanics—to be described in Sect. 3.4.

3.3   Privacy and Latency Concerns

We implement the audio processing pipeline on edge to ensure minimal latency and privacy concerns. This design choice allows us an efficient minimal-latency real-time scripted coaching mechanism when the user initiates a brushing event,—an attribute imperative for an engaging user experience—and also enables us to bypass the users' privacy concerns. We do not keep audio from the brushing sessions on the server; all the audio analysis is performed on the device. Additionally, no audio is saved or stored on the user's device beyond the circular eight-kilobyte buffer from which the model draws audio to classify, which

itself is not preserved beyond the scope of the view in which it resides. Table 2 lists all the information we collect for a particular user. The set of data that we collect can be divided into two types: information related to a brushing event, such as its timestamp, and information related to the in-app game mechanics.

Table 2. Backend Values

Value	Usage
firstBrushingTime	time of the first (often morning) brushing session. Default: 8AM
secondBrushingTime	time of the second (often midday) brushing session. Default :2PM
thirdBrushingTime	time of the third (often night) brushing session. Default :10PM
brushingStreak	consecutive number of completed brushing sessions
dailyStreak	consecutive number of days for which all brushing sessions were completed.
division	Leaderboard division the user is currently in
totalScore	total score accumulated by the user (lifetime)
weeklyScore	weekly score accumulated by the user

3.4   Gamification

Gamification techniques such as leaderboards and levels have been shown to increase user engagement in oral care apps in several recent studies [12]. For this purpose, we implement a leaderboard functionality in YouBrush, as depicted in Fig. 1c. We give the user a score based on the duration and regularity of their brushing – both daily and across days – and rank them in divisions against simulated users [19]. The leaderboard view in Fig. 1c shows a user in their current position within a leaderboard full of simulated users; these simulated users scale upwards in competitive difficulty as they ascend to higher divisions. Note that the leaderboard view is personalized; every user's view of the leaderboard and the set of simulated users differs from everyone else.

4   Data Collection and Labeling

The organization of this section is as follows. First, we delineate the methodology applied to create the brushing audio dataset used to train and evaluate the eventual machine learning algorithm for brushing detection. Second, we discuss the ethical considerations of collecting audio data for our research. Finally, we introduce a labeling tool we develop to facilitate audio labeling using video recordings.

4.1   Methodology

In addition to recording the team members' audio while brushing—a dataset that we will refer to as self-owned data from now on—we also take advantage of two public sources of audio data: Audioset [14] and Freesound.org [3]. Google's audioset is a large-scale collection of human-labeled 10-s sound clips drawn from YouTube

videos. To collect all the audio data, the authors worked with human annotators who verified the presence of sounds they heard within YouTube segments. To nominate segments for annotation, they relied on YouTube metadata and content-based search [14]. Freesound is a collaborative repository of CC-licensed audio samples and a non-profit organization with more than 500,000 sounds and 8 million registered users. Sounds are uploaded to the website by its users and cover a wide range of subjects, from field recordings to synthesized sounds [3]. We ensure to collect audio samples labeled as different classes of bathroom-ambient sounds such as sink, tub, toilet, shower, negative, and silence. We define the tag *silence* as little or no sound and *negative* as any sound samples that do not belong to the labels mentioned above. We provide a detailed description of the audio classes in Sect. 5. Table 3 presents detailed statistics of our dataset, specifying individual audio classes and the length of audio samples for them.

Table 3. Dataset statistics

Class	Number of files	Total Audio Samples	Audio length	Source
Brushing	205	3945	65.75 min	Self-owned and freesound.org
Silence	49	488	8.133 min	Audioset
Sink	68	1478	24.633 min	Self-owned
Toilet	69	387	6.45 min	Self-owned
Tub	20	1522	25.367 min	Self-owned
Negative	25	2046	34.1 min	Self-owned
Shower	14	1811	30.183 min	Self-owned

4.2 Ethical Consideration

Crawling and analyzing data from a private environment such as the bathroom entails serious ethical implications. However, in this research, we only concern ourselves with publicly available audio data sets and refrain from interacting with human subjects besides the ones on our team. Therefore, this research conforms to the standard ethics guidelines to protect users [32].

4.3 Labeling

We implement an audio-video labeling tool to construct the self-owned dataset, as shown in Fig. 2. A typical usage flow of the labeling tool is as follows. First, the subject uses their mobile phone to videotape their brushing event. The subject then uploads the video recording to the labeling tool. Upon uploading, the user can move the scrubbing slider to select a portion of the video corresponding to a class of audio specified in the drop-down list located just below the slider. Next, the user clicks "export segments", which prompts the tool to extract just

the audio segment from the video file along with the associated label, duration of the audio segment, and start and end timestamp in the parent video file. The purpose of this labeling tool is threefold. First, The labeling tool allows us to bypass any need to store videos of users brushing on the server; we did not store any videos of the users participating in the data collection process. Instead, we only collect the audio extracted from the brushing videos, thereby protecting the users' privacy accordingly [42]. Second, The tool allows us to label the brushing audios with added granularity for future brushing behavior analysis. For example, one particular brushing sub-event audio can be tagged by the user as brushing the front teeth or the back teeth, using the "data label" drop-down list shown in Fig. 2. Third, the tool allows the audio labels to be as accurate as possible. Brushing is a continuous event that ranges from 40 s to two minutes or more, containing not just brushing sounds but sometimes silences, bathroom ambient sounds, or even human speech. The tool's scrubbing slider allows us to extract the exact audio segment containing a brushing event, devoid of any other foreign audios, as much as possible.
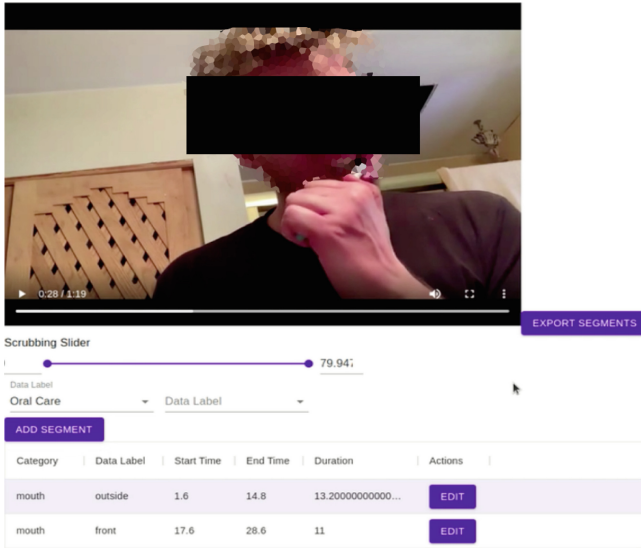

Fig. 2. Labeling Tool used to construct self-owned brushing dataset

5   Machine Learning Implementation

As mentioned in Sect. 3, YouBrush listens to the live audio when the user initiates a brushing event. We present a two-phase classification engine to ensure efficient processing and effective classification of the live audio stream. In the first phase, we constantly take a fixed size of 1 s from the live audio chunk-stream and use our Machine Learning Classifier (MLC) to classify them independently.
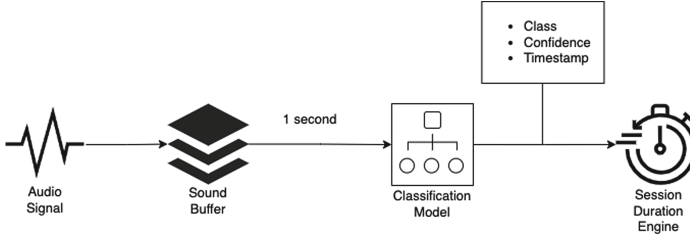
Fig. 3. Second Phase

The second phase uses the classification labels—instead of the audio—to infer the total duration of the user's brushing event. Figure 3 shows our two-phase classification engine. Domestic environments produce a wide variety of audio sounds [35] and a brushing event is no exception. This challenge necessitates us to consider—to train and construct our audio MLC—other surrounding sounds such as running-shower, running-sink, etc. In addition to brushing audio. Table 4 lists the classes of audio we consider to develop our brushing MLC.

Table 4. Classes of Audio Considered

Label	Description
Brushing	Traditional Toothbrush Brushing
Shower	Shower Running
Sink	Sink Running
Toilet	Toilet Flushing
Tub	Tub Filling
Silence	Plain Silence
Negative	Others (Fan Running etc.)

We organize the rest of the section into two subsections. The first sub-section, *Preprocessing*, delineates the approaches we follow to pre-process the audio samples to construct the training data. The second sub-section, *Model Creation*, describes and evaluates our brushing MLC.

5.1 Preprocessing

We convert all sound files to single channel wav files with a sampling rate of 16 kHz. We provide two reasons for this conversion choice. First, researchers have found such conversions effective in environmental sound classification tasks using Deep Learning Models [9]. Second, to develop our brushing MLC, we will use Apple's off-the-shelf CreateML platform (described in more detail in Sect. 5.2); and createML recommends audio file's sample rate to be 16 kHz [2]. Besides the

rate conversion, we also consider the following aspects of processing our data before starting model training.

Noise Reduction: Upon collecting the audio samples (Table 3), we first use the log-MMSE algorithm to reduce our dataset's noise. We argue that cleaning the noise from the dataset is an essential preprocessing step—most brushing events tend to occur when noises such as speech, running exhaust fan, and running sink are present. We choose log-MMSE because recent studies have found log-MMSE to be significantly effective in reducing noise when the actual noise is unknown [17,41].

Data Augmentation: One of the main drawbacks of deep learning approaches is the need for a significant amount of training data [24]. Therefore, researchers have turned to data augmentation as the go-to approach to tackle the training data to implement deep learning models [28,29]. We leverage Python's librosa library [26] to implement the data augmentation step when creating the training data. Table 5 lists the augmentation techniques we leverage to construct our training dataset.

Table 5. Audio Augmentation Techniques Performed

Method	Description
Pitch Modification	Using a factor to modify pitch
Stretching	Time-stretch an audio series by a fixed rate.
Changing Decibels	Modifying the decibels from an audio file, changing how loud it sounds
Noise Injection	Combining two different audios from the original dataset to get a new sample, for example combining sink running audio files with brushing files

MFCC Features: To visually analyze audio samples and engineer features for model training, we choose Mel-Frequency Cepstral Coefficient (MFCC) [43]. Mel-Frequency Cepstrum (MFC) is a representation of the short-term power spectrum of a sound based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. MFCCs are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum") [43]. MFCC as audio features have outperformed Fourier Transform (FT), Homomorphic Cepstral Coefficients (HCC), and others for non-speech audio use-cases [43].

5.2   Model Creation

Upon completing the audio preprocessing and preliminary analysis, we partition our dataset into training and test data, using an 80–20 split. Additionally, we

Table 6. Model results

Class	Precision %	Recall %
Brushing	97	92
Shower	100	96
Sink	95	95
Toilet	83	71
Tub	90	100
Silence	77	76
Negative	75	97

inject a large amount of publicly-sourced audio data into the test set to ensure testing for generalization at a high level. We use Apple's off-the-shelf tool CreateML to create and train custom machine learning models [1]. CreateML can be used for various use cases, such as recognizing images, classifying audios, extracting meaning from text, or finding relationships between numerical values. It allows for modifying a few aspects of the metadata during model creation (Maximum number of iterations and the overlap factor) and creates a coreml file to be used on any swift application. Because we use Swift to develop YouBrush (Sect. 3), Table 6 shows the precision and recall results on our test data. We use the CreateML-setup to take at most 25 iterations to finish the training and use an overlap factor of 50%. One characteristic of CreateML is that it configures all the internal setup of the model, for example, how many layers it will have and what type of layer it will use; everything is based on the use-case selected. We select the use-case of Sound Classification. After the training phase, the deep learning model has the following configuration: 6 convolution layers, 6 ReLU activation layers, 4 Pooling layers, and one flattened layer, followed by the softmax layer. Table 6 shows the precision and recall of the model's performance for every audio class in the test dataset. As it can be seen, for the brushing class, the model achieves precision and recall scores of 97% and 97%, respectively, using MFCC features.

5.3 Model Testing

Recognizing that the physical domain in which the model operates is home to many potential confounding sounds, we give special consideration to injecting such sounds into the training data as background noise to ensure we account for these. Additionally, the "Negative" class is composed of such sounds (e.g., bathroom exhaust fans, speech) to better define the boundary betwixt the brushing class and such confounding sounds concerning their tendency to occur in tandem. Finally, when injecting publicly-sourced data into our test sets, we give special deference to those where common confounding sounds are dominant.

Of particular concern is the evaluation scenarios by which we must test the model. Table 6 displays testing results concerning individual samples, wherein a sample is one single second of audio (achieved by decomposing full audio samples of greater length). These results offer a valuable evaluation of the model at a glance when quickly iterating, but of greater importance is the distribution of these results as they relate to samples of greater length. Consider the following: we sample ten seconds of a user engaged in their brushing session. Let's assume we classify seven seconds as brushing and wrongly classify the rest as some other arbitrary class. Based on the YouBrush implementation, the way these three misclassified seconds are distributed throughout the sample is significant. Namely, suppose these misclassified second-long samples occur disjointly from one another (e.g., are separated by correctly classified periods). In that case, the interruption to the user experience is lessened and can even be smoothed over by assuming certain levels of uncertainty (which we will discuss shortly). If the inverse occurs and these misclassified second-long samples occur consecutively with respect to time (e.g., are not separated by any period of properly-classified brushing samples), the user experience is degraded, and both trust and interest in the experience provided by the application quickly break down. For this reason, we separate model testing into two focuses: traditional sample-based testing, in which we focus on achieving accuracy for singular samples, and longer-form implementation-based testing, in which we focus on accuracy over time for sets of ordered samples. Within implementation-based tests, more extended periods of inaccuracy are more heavily penalized. Implementation-based tests are built to model real-life brushing sessions and therefore focus on improving the user experience.

5.4   Model Implementation

Finally, we address the subject of model uncertainty and how YouBrush addresses such shortcomings at the implementation level. Recognizing that failing closed (e.g. failing to properly classify moments in which the user is brushing) creates a negative experience for the user, the YouBrush implementation has a set of built-in fail-safes to make failing opening (e.g. considering moments of uncertainty to be aurally contiguous with a set of previously classified samples) the default in uncertain periods of sound. The first of these is simple: the brushing class does not require a high level of confidence. This allows samples nearer the class boundary to nonetheless be considered as truth.

   The second of these is a code-level concession offered to the model based on the assumption that each short-term sample taken (ranging from periods of a half-second to a full second) is aurally similar to those before and thus does not represent a significant departure from the pattern established in a small set of previous samples on their own (with respect to the order in which samples are taken). Based on this assumption, we define a variable amount of time that a departure from a previously established pattern must meet to break the said pattern. Within the YouBrush-specific implementation, only the brushing sound class is of interest, and thus the only pattern eligible is that of

consecutive brushing sounds. Therefore, once a short pattern of brushing sound is established, model uncertainty (by way of misclassification of brushing audio) must occur consecutively several times before the user experience is adversely affected. Returning to the previous example wherein seven of ten seconds of audio are properly classified as brushing, with the other three incorrectly classified as a non-brushing class, we address both hypothetical distributions of these incorrectly classified samples wherein they occur either disjointedly or as a set; in both circumstances, the threshold for pattern departure is not met, and model uncertainty is abstracted away from the user experience.

6 Conclusion

In this paper, we denote our experience on how to develop a mobile application, YouBrush. We delineate the design choices we make to ensure a real-time scripted coaching brushing experience for the users with traditional toothbrushes, using an on-device highly accurate, low-latency deep learning model for brushing audio detection. To ensure the machine learning model's robustness, we utilize an in-house labeling tool that facilitates precise audio labeling through video-ques. Next, to encourage consistency, regularity, and user engagement, we incorporate in-app game mechanics and real-time positive feedback mechanisms during brushing. In the future, we plan to release YouBrush in the AppStore and investigate its efficacy among the users, performing elaborate user behavior analysis.

References

1. createml. https://developer.apple.com/documentation/createml. Accessed Apr 2022
2. createml sound classifier. https://developer.apple.com/documentation/createml/mlsoundclassifier. Accessed Apr 2022
3. freesound.org. https://en.wikipedia.org/wiki/Freesound. Accessed 25 July 2022
4. Swift programminmg language. https://developer.apple.com/swift/. Accessed 28 July 2022
5. Camalan, S., et al.: Convolutional neural network-based clinical predictors of oral dysplasia: class activation map analysis of deep learning results. Cancers **13**(6), 1291 (2021)
6. Chang, Y.C., et al.: Playful toothbrush: ubicomp technology for teaching tooth brushing to kindergarten children. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 363–372 (2008)
7. Chen, Z., et al.: Unobtrusive sleep monitoring using smartphones. In: 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, pp. 145–152. IEEE (2013)

8. Chiu, C.C., et al.: State-of-the-art speech recognition with sequence-to-sequence models. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4774–4778. IEEE (2018)

9. Chong, D., Zou, Y., Wang, W.: Multi-channel convolutional neural networks with multi-level feature fusion for environmental sound classification. In: Kompatsiaris, I., Huet, B., Mezaris, V., Gurrin, C., Cheng, W.-H., Vrochidis, S. (eds.) MMM 2019. LNCS, vol. 11296, pp. 157–168. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05716-9_13

10. Croyère, N., Belloir, M.N., Chantler, L., McEwan, L.: Oral care in nursing practice: a pragmatic representation. Int. J. Palliat. Nurs. **18**(9), 435–440 (2012)

11. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15 (2011)

12. Fijačko, N., et al.: The effects of gamification and oral self-care on oral hygiene in children: systematic search in app stores and evaluation of apps. JMIR Mhealth Uhealth **8**(7), e16365 (2020)

13. Ganss, C., Schlueter, N., Preiss, S., Klimek, J.: Tooth brushing habits in uninstructed adults-frequency, technique, duration and force. Clin. Oral Invest. **13**(2), 203–208 (2009)

14. Gemmeke, J.F., et al.: Audio set: an ontology and human-labeled dataset for audio events. In: Proceedings of IEEE ICASSP 2017. New Orleans, LA (2017)

15. Gerritsen, A.E., Allen, P.F., Witter, D.J., Bronkhorst, E.M., Creugers, N.H.: Tooth loss and oral health-related quality of life: a systematic review and meta-analysis. Health Qual. Life Outcomes **8**(1), 1–11 (2010)

16. Graetz, C., et al.: Toothbrushing education via a smart software visualization system. J. Periodontol. **84**(2), 186–195 (2013)

17. Hu, Y., Loizou, P.C.: Subjective comparison and evaluation of speech enhancement algorithms. Speech Commun. **49**(7), 588–601 (2007). https://doi.org/10.1016/j.specom.2006.12.006, https://www.sciencedirect.com/science/article/pii/S0167639306001920. speech Enhancement

18. Janusz, K., Nelson, B., Bartizek, R.D., Walters, P.A., Biesbrock, A.R.: Impact of a novel power toothbrush with smartguide technology on brushing pressure and thoroughness. J. Contemp. Dent. Pract. **9**(7), 1–8 (2008)

19. Karatassis, I., Fuhr, N.: Gamification for websail. In: GamifIR@ SIGIR (2016)

20. Kim, K.S., Yoon, T.H., Lee, J.W., Kim, D.J.: Interactive toothbrushing education by a smart toothbrush system via 3D visualization. Comput. Methods Programs Biomed. **96**(2), 125–132 (2009)

21. Korpela, J., Miyaji, R., Maekawa, T., Nozaki, K., Tamagawa, H.: Evaluating tooth brushing performance with smartphone sound data. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 109–120 (2015)

22. Kumarajeewa, R., Jayarathne, P.: Improving children's oral hygiene habits in Sri Lanka via gamification. Asian J. Manag. Stud. **2**(1), 98–113 (2022)

23. Lin, T.H., Wang, Y.M., Huang, C.Y.: Effects of a mobile oral care app on oral mucositis, pain, nutritional status, and quality of life in patients with head and neck cancer: a quasi-experimental study. Int. J. Nurs. Pract. **28**(4), e13042 (2022). https://doi.org/10.1111/ijn.13042

24. Marcus, G.: Deep learning: a critical appraisal. arXiv preprint arXiv:1801.00631 (2018)

25. McFee, B., Ellis, D.P.: Better beat tracking through robust onset aggregation. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2154–2158. IEEE (2014)
26. McFee, B., et al.: librosa: audio and music signal analysis in python. In: Proceedings of the 14th python in science conference, vol. 8, pp. 18–25. Citeseer (2015)
27. Mesaros, A.: Detection and classification of acoustic scenes and events: outcome of the DCASE 2016 challenge. IEEE/ACM Trans. Audio Speech Lang. Process. **26**(2), 379–393 (2017)
28. Mushtaq, Z., Su, S.F.: Environmental sound classification using a regularized deep convolutional neural network with data augmentation. Appl. Acoust. **167**, 107389 (2020)
29. Nanni, L., Maguolo, G., Paci, M.: Data augmentation approaches for improving animal audio classification. Eco. Inform. **57**, 101084 (2020)
30. Patil, S., et al.: Effectiveness of mobile phone applications in improving oral hygiene care and outcomes in orthodontic patients. J. Oral Biol. Craniofac. Res. **11**(1), 26–32 (2021). https://doi.org/10.1016/j.jobcr.2020.11.004, https://www.sciencedirect.com/science/article/pii/S2212426820301652
31. Raypole, C.: 5 toothbrushing FAQs. https://www.healthline.com/health/how-long-should-you-brush-your-teeth Accessed 13 July 2022
32. Rivers, C.M., Lewis, B.L.: Ethical research standards in a world of big data. F1000Research, **3**(38), 38 (2014)
33. Schäfer, F., Nicholson, J., Gerritsen, N., Wright, R., Gillam, D., Hall, C.: The effect of oral care feed-back devices on plaque removal and attitudes towards oral care. Int. Dent. J. **53**(S6P1), 404–408 (2003)
34. Scheerman, J.F.M., et al..: The effect of using a mobile application ("WhiteTeeth") on improving oral hygiene: a randomized controlled trial. Int. J. Dent. Hyg. **18**(1), 73–83 (2020). https://doi.org/10.1111/idh.12415. https://onlinelibrary.wiley.com/doi/abs/10.1111/idh.12415
35. Serizel, R., Turpault, N., Eghbal-Zadeh, H., Shah, A.P.: Large-scale weakly labeled semi-supervised sound event detection in domestic environments. arXiv preprint arXiv:1807.10501 (2018)
36. Sultan, A.S., Elgharib, M.A., Tavares, T., Jessri, M., Basile, J.R.: The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. J. Oral Pathol. Med. **49**(9), 849–856 (2020)
37. Tayebi, A., et al.: Mobile app for comprehensive management of orthodontic patients with fixed appliances. J. Orofacial Orthop./Fortschritte der Kieferorthopädie (2022). https://doi.org/10.1007/s00056-021-00370-7
38. Thomaz, E., Parnami, A., Bidwell, J., Essa, I., Abowd, G.D.: Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 739–748 (2013)
39. Tosaka, Y.: Analysis of tooth brushing cycles. Clin. Oral Invest. **18**(8), 2045–2053 (2014). https://doi.org/10.1007/s00784-013-1172-3
40. Underwood, B., Birdsall, J., Kay, E.: The use of a mobile app to motivate evidence-based oral hygiene behaviour. Br. Dent. J. **219**(4), E2–E2 (2015). https://doi.org/10.1038/sj.bdj.2015.660
41. Wang, A., Yu, L., Lan, Y., Zhou, W., et al.: Analysis and low-power hardware implementation of a noise reduction algorithm. In: 2021 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), pp. 22–26. IEEE (2021)

42. Wickramasuriya, J., Datt, M., Mehrotra, S., Venkatasubramanian, N.: Privacy protecting data collection in media spaces. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 48–55 (2004)
43. Wikipedia contributors: Mel-frequency cepstrum. https://en.wikipedia.org/wiki/Mel-frequency_cepstrum. Accessed 13 July 2022