Proximity-based Eating Event Detection in Smart Eyeglasses with Expert and Data Models

Addythia Saphala
Chair of Digital Health
FAU Erlangen-Nürnberg, Germany
addythia.saphala@fau.de

Rui Zhang
Chair of Digital Health
FAU Erlangen-Nürnberg, Germany
rui.rui.zhang@fau.de

Oliver Amft
University of Freiburg, Germany
Hahn-Schickard, Germany
FAU Erlangen-Nürnberg
amft@computer.org

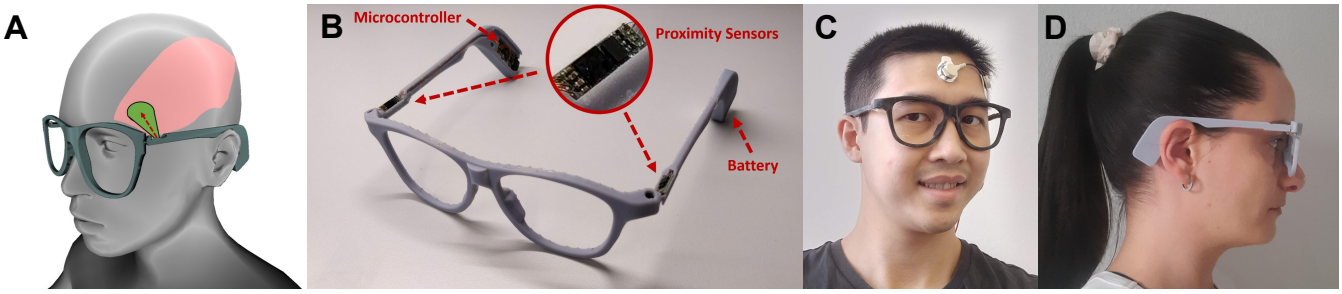

Figure 1: Overview on our smart eyeglasses with temple-embedded proximity sensors. A: Schematic of proximity signal capture. B: Smart eyeglasses integration. C: Lab-based EMG reference. D: Lateral view of worn smart eyeglasses.

ABSTRACT

We compare performances of an expert model-based approach and a data-based baseline for eating event detection using proximity sensor data of smart eyeglasses. Proximity sensors in smart eyeglasses can provide dynamic distance estimates of cyclic temporalis muscle contraction during chewing without skin contact. Our expert model is based on proximity signal preprocessing and two-threshold grid search. In contrast, baseline data models were based on One-class Support-Vector-Machines. We evaluate both models with in-lab and free-living data from 15 participants. Free-living data were obtained across one day of wearing smart eyeglasses with temple-integrated proximity sensors in unconstrained settings. Overall, the retrieval performance F1 score of the two-threshold-based algorithm for free-living data ranged between 0.6 and 0.7, and outperformed all tested SVM model configurations. While SVM models achieved maximum recall, precision was often below 0.5. We report head-side specific performances for a bilateral arrangement of the proximity sensors and detail performance characteristics in model parameter sweeps. We conclude that eating detection using proximity sensors in smart eyeglasses is a promising approach for unconstrained automated dietary monitoring. Nevertheless, further investigations are needed to deal with the proximity signal characteristics in everyday life monitoring.

ISWC '22, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9424-6/22/09…$15.00
https://doi.org/10.1145/3544794.3558476

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**; **Empirical studies in ubiquitous and mobile computing**; • **Computing methodologies** → *Machine learning*.

KEYWORDS

automatic dietary monitoring, eating detection, chewing detection, smart eyeglasses, wearable accessory

ACM Reference Format:
Addythia Saphala, Rui Zhang, and Oliver Amft. 2022. Proximity-based Eating Event Detection in Smart Eyeglasses with Expert and Data Models. In *The 2022 International Symposium on Wearable Computers (ISWC '22), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544794.3558476

1 INTRODUCTION

Automated dietary monitoring (ADM) systems intend to detect intake-related behaviour, thus could serve as a basis for continuous dietary assistance, e.g. in weight management programmes, or guide clinical decision-making for diagnosis and treatment of dietary-related behaviour disorders. Recently, eyeglasses-attached proximity sensors were proposed as a non-contact measurement principle to detect temporalis muscle activation during chewing [4]. Food chewing involves a cyclic contraction and release of temporalis muscles to elevate and relax the mandible bone. Mechanical excursions due to temporalis muscles contractions are visually observable at temporal head regions and can be captured by consumer-grade proximity sensors based on the principle of emitted and reflected infra-red (IR) light. IR-based proximity sensors can be conveniently embedded into eyeglasses temples that provide a frame to interpret distance changes between sensor and skin.

While wearable system integration yields an inconspicuous eyeglasses design to measure chewing activity, detection of eating events is an open challenge due to several issues. First, skin excursions are within 1-2 mm and eyeglasses temples have a typical distance of 5-20 mm to the skin, which puts IR-based proximity sensors at the limits of their operational ranges. Second, eyeglasses are not constantly in the same position, resulting in distance estimation artefacts. Third, IR-based distance estimation can be perturbed by variations in environmental lighting and sunlight effects.

In this work, we investigate proximity-based eating event detection in free-living and compare two detection algorithms. In particular, the paper provides the following contributions:

(1) We propose a regular-look smart eyeglasses design with bilaterally integrated IR-based proximity sensors in temples to capture skin excursion during chewing activity. We implemented in-lab and free-living studies with 15 participants wearing the smart eyeglasses for one day each.
(2) We analyse eating detection performance by comparing a two-threshold-based algorithm against an already established approach based on Support-Vector-Machines (SVM). Our expert-based algorithm was inspired by observing the low-amplitude, artefact-rich proximity signal data.
(3) We present detailed retrieval performance analyses of both eating detection methods.

2 RELATED WORK

Here, we focus on approaches that exploited muscle activation to detect chewing or eating events, especially those with a threshold-based algorithm and SVM classifiers.

Wang et al. [6] used a motion sensor to capture masticatory muscle contraction, detect eating, and count chewing cycles. They used a triaxial accelerometer attached to the temporalis area and recorded Electromyography (EMG) measurements as a reference. The authors reported that temporalis muscle contraction frequency corresponds to chewing frequency and thus could be used to count chewing cycles. The eating activity was detected using SVM, among other algorithms. Zhang and Amft [7] utilised smart eyeglasses with integrated EMG electrodes to detect chewing, eating events, and differentiated food types. They placed electrodes at temple ear bends of personalised eyeglasses to detect contractions of temporalis muscles. In-lab and free-living studies with 10 participants were performed. Chewing and eating events were detected with a threshold-based algorithm. Shin et al. [5] mounted an attachable *MyDJ* at an eyeglasses temple to detect food intake. They used a piezoelectric sensor to capture temporalis muscle contraction and an accelerometer to record the mechanical vibrations due to chewing activity. They performed free-living studies with 30 participants for a week, including a one-day data collection with a reference camera. Eating episodes were detected using DNN with an F1 score of 0.919. Bedri et al. [1] utilised inertial and optical sensors embedded into a pair of eyeglasses. A camera was used to capture food visuals. The authors conducted two studies to evaluate the system with 23 participants, resulting in an average F1 score of 0.89 for detecting eating and drinking episodes.

Selamat and Ali [3, 4] mounted an IR-based proximity sensor within a 3D-printed housing at an eyeglasses frame temple to detect chewing. In their experiment, proximity signals from the right temple region of one participant were recorded. The participant performed ten different activities, including eating, for a total dataset duration of 40 minutes. Push buttons were used to obtain labels for eating and chewing activity. The authors tested several SVM kernels on 40 time- and frequency-domain features derived from proximity sensor data. Most developments so far have either used threshold-based algorithms with expert-designed signal processing pipelines, or machine learning classifiers based on general feature sets, to detect intake activity. For proximity sensor data, only Selamat and Ali have described an algorithm to detect chewing based on SVMs [4]. In our present work, we compare expert-based (i.e. two-threshold) and data-based (i.e. SVM) algorithms for eating detection. In contrast to previous work, we investigate eating detection performance in free-living recordings from 15 participants. We build unobtrusive 3D-printed eyeglasses with frame-integrated sensors, microcontroller, and battery (see Fig. 1A).

3 METHODOLOGY

3.1 Smart Eyeglasses Platform

We embedded IR-based proximity sensors (Vishay VCNL4010) in regular-look 3D-printed eyeglasses. Figure 1B shows the bilateral configuration with proximity sensors oriented to the anterior temporalis muscle as illustrated in Fig. 1A. The proximity sensors detect nearby surfaces by measuring reflected infra-red light, where a higher intensity of reflected light correlates with a shorter distance. Measurement units were basic, non-calibrated counts. The proximity data were recorded at 50 Hz by a microcontroller (ST STM32F4) and stored in 512 MB flash memory integrated into the eyeglasses' temple end. Our smart eyeglasses did not require external components, which is essential for unconstrained free-living recordings. Figure 2 shows a timeseries example of low-pass filtered proximity sensor readings with the smart eyeglasses. The waveform shows typical signal drift and artefacts, besides cyclic chewing patterns.

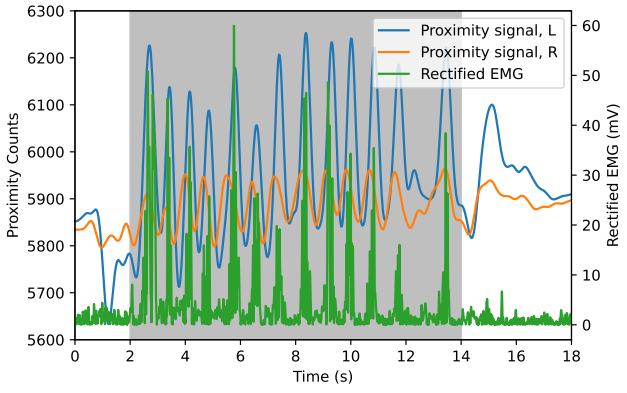

Figure 2: Low-pass filtered proximity sensor data of left (L) and right (R) head sides and rectified EMG. The highlighted region indicates one chewing sequence, i.e. one bite.

Table 1: List of feature used for the SVM-based algorithm.

Frame	Features
TD	Max, min, max-min, rms, median, variance, std, skew, kurtosis, interquartal range.
FD	Mean, bandpower, median, kurtosis, skew.
PSD	Max, min, mean, std, spectral entropy, spectral kurtosis, kurtosis, skew, median.
ESD	Sum, min, max, mean, four energy band within (0.94-2.17, 1.25-2.5, 0.5-2.5, 1-3) Hz, and 7 spectral energy at (1.3, 1.6, 1.9, 2.1, 2.4, 2.7, 3) Hz.

3.2 Two-Threshold-based Algorithm

The two-threshold-based algorithm was specifically designed to deal with the proximity signal properties, including low amplitude chewing patterns and motion-related artefacts.

Signal Processing. Proximity sensor readings of left and right head sides were filtered by a 2 Hz low-pass (see Fig. 2). Filtered signals within the highlighted area show cyclic distance variation that corresponds to chewing cycles, as confirmed by EMG data. We increase the signal contrast by deriving the signal gradient. To convert the pattern into an amplitude, we apply the Teager–Kaiser Energy Operator (TKEO) [2], followed by signal rectification. Finally, a moving average (window size: 5 sa) was applied. On the resulting envelope curve, we applied a two-threshold detector, where the upper threshold helped to reject large signal artefacts and the lower threshold was used to reject lower-amplitude signal noise.

Threshold Grid Search. Based on training data, a two-dimensional grid of lower and upper threshold values was constructed. To obtain optimal training performance, we searched the grid for lower ($0-1$, step size 0.01) and upper ($0-100$, step size 1) thresholds.

3.3 SVM-based Algorithm

Frame Segmentation & Feature Extraction. Hanning windows with 50% overlap were used to extract features from bandpass-filtered signals ($0.5-5\,Hz$). In total, 40 features were extracted, inspired by the work of Selamat et al. [4]. Table 1 lists all features derived from time domain (TD), frequency domain (FD), power spectral density (PSD) and energy spectral density (ESD). TD features were derived from sliding window signals, FD features after discrete-time Fourier transform (DTFT) of the windowed signal, PSD was derived as the magnitude of STFT in logarithmic (decibel) scale, and ESD from the absolute square of PSD.

One-Class SVMs. In this work, we utilised One-Class SVMs to account for imbalanced data (eating vs. no eating) and to reduce assumptions about non-eating data. We trained SVM models with linear, quadratic, and Radial Basis Function (RBF) kernels.

3.4 Eating Detection

To compensate for artefacts and gaps between chewing sequences, we employ a five seconds eating section removal on outputs of two-threshold and SVM algorithms. Gaps between chewing sequences are naturally reported as non-eating, especially for sliding window sizes below 30 s. Therefore, we utilise gap filling between chewing sequences, for gaps below 5 min., as described by Zhang et al. [8].

3.5 Study Methods and Statistics

Our study involved 15 participants (four females), 23 to 33 years old. Body Mass Indices (BMI) ranged between 18.9 kg/m^2 and 29.3 kg/m^2, avg. 23.93 kg/m^2. Each participant took part in both, in-lab and free-living study parts. The study was approved by an institutional ethical committee and participants provided written consent before the study began. Participants had no known dental or dietary-related disorder, nor vision impairment beyond refractive error.

In-Lab Study. Participants were invited to perform a set of scripted activities of daily living (ADL) in a lab setting. They were equipped with smart eyeglasses and an additional EMG recorder mounted at one temporal region (CamNtech Actiwave) to obtain a reference of temporalis muscle contraction (see Fig. 1C). A free choice of food items was provided, including baguette, carrot, apple, cheddar, chips, and beef jerky. Foods were prepared in pieces of 2 g, 4 g, and 6 g. An observer annotated activity and timing on a paper journal. Each study session was recorded on video. Subsequently, we used the manual observer annotation of chewing sequences, i.e. each bite, to create the ground truth, complemented by reviewing videos when there were doubts.

Free-Living Study. We instructed participants to wear smart eyeglasses for one day, from wake-up to bedtime. Two 90 min. charging breaks were unavoidable due to battery limitations. Therefore, we instructed participants to schedule charging breaks so that they would not conflict with their meal times or other logistic requirements. Participants were asked to wear the smart eyeglasses as much as possible during the day, with exceptions for potentially safety-critical or device-damaging activities, including showering or intensive sports. Use exceptions were also recommended if the smart eyeglasses could prevent wearing required items in specific tasks, e.g. a helmet, safety goggles, or prescription eyeglasses. Participants were asked to note down the start and end of eating events, and when they were not wearing the smart eyeglasses, all rounded to the minute. We review the participant annotations against signal recordings to generate the ground truth. If there were conflicts or doubts, we interviewed participants about that particular period and could resolve any mismatches.

Recording Statistics. In-lab recording included ~120 min of data, with 8 ± 1.5 min per participant, and a cumulative eating time of 22 min, 1.5 ± 1 min per participant. Free-living data was ~6912 min, 461 ± 86 min per participant, with a cumulative eating time of ~544 min, 36 ± 15 min per participant.

3.6 Evaluation Procedure

We used leave-one-participant-out (LOPO) cross-validation (CV), whereby all in-lab data except data of one participant was used for parameter fitting and model training, and free-living data of the left-out participant was used for testing both algorithms. Performance results across all LOPO-CV folds were kept for analysis. Free-living data was excluded from training since participant annotations provided eating events only, thus including non-chewing gaps. In-lab annotation specifically focused on chewing sequences.

We retrieved insertions, deletions, and true positives, which were used to calculate F1 scores, precision, and recall. For SVM model evaluations, we varied the sliding window size between 3 s and 60 s. Ground truth was adapted to the frame-based structure of the sliding windows. For the two-threshold algorithm, performance was further detailed by sweeping the lower threshold at a fixed, training-derived upper threshold to construct Precision-recall (PR)-curves. Ground truth was adapted to sample rate resolution.

4 RESULTS AND DISCUSSION

Figure 3 shows a comparison between the two-threshold algorithm and SVM models. Box plots illustrate medians, inter-quartile ranges (IQR), and outliers (> 1.5×IQR). The two-threshold-based algorithm showed the best median F1 score, between 0.6 and 0.72. Among SVM models, the quadratic kernel had a larger variance and underperformed both, RBF and linear kernels. Performance varied across individual participants for all models, while at substantially lower F1 scores for SVM models. Reasons include data artefacts of proximity sensors, e.g. eyeglasses movement at the head. Instead of training SVM models specific to a head side (L, R), SVM* models were trained from data of both head sides to investigate, if additional observations may improve performance. Among SVM* models, linear kernels showed the best performance, however, could not compete with the two-threshold algorithm results. We consider the head side separately here, to compare with published work. Data fusion between left and right sides, or with additional sensors may improve performance.

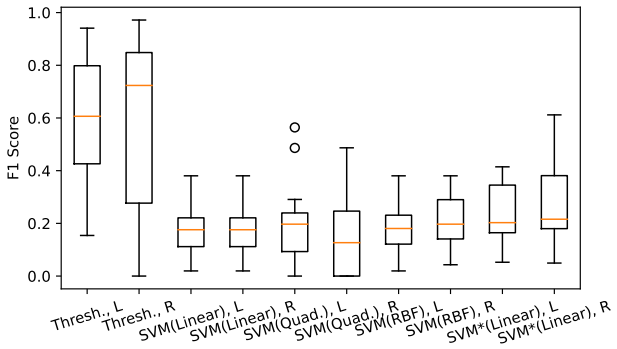

Figure 3: Performance comparison between two-threshold algorithm (Thresh.) and SVM models at window size of 3 s, for left (L) and right (R) head sides. SVM* models were trained from combined training data of both head sides. Box plots illustrate medians, IQR, and outliers.

To further analyse the two-threshold algorithm performance, we derived PR-curves per participant and average by sweeping the lower threshold on free-living data. Figure 4 illustrate that a top recall can be reached, and maximum precision reaches ~0.7.

To assess the One-Class-SVM performance, we analysed retrieval performance over sliding window size for RBF kernels. Figure 5 shows precision and recall across sliding windows. Larger window size helped to increase model precision from ~0.2 to ~0.3, however at expense of reduced recall, from ~1.0 to ~0.9. Larger windows may have eliminated detection fragmentation, but still could not substantially increase feature discrimination. Our SVM parameterisation closely resembles that of Selamat et al. [4]. In particular, the authors proposed a 3 s window to extract features. Our analysis confirms that SVM models yield top recall, even in free-living data. However, we could not compare retrieval precision as Selamat et al. reported accuracy only.

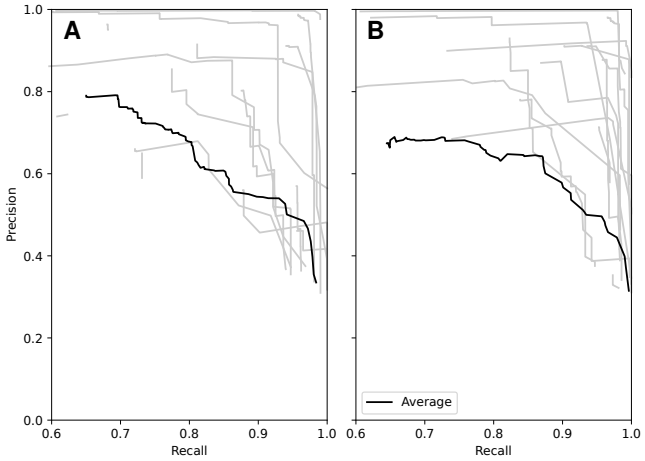

Figure 4: PR-curves of the two-threshold algorithm for sweeping lower thresholds per participant, and averaged curve. A: Left head side. B: Right head side.

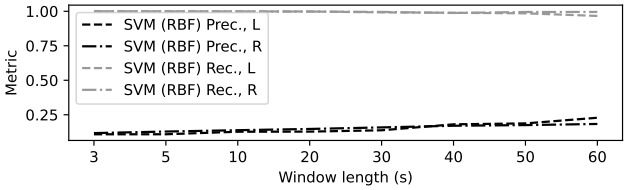

Figure 5: Precision and recall of SVMs with RBF kernel on free-living data for left (L) and right (R) head sides, trained for varying sliding window sizes.

5 CONCLUSION

We investigated temporalis muscle contraction to detect eating events in free-living using proximity sensors embedded in regular-look smart eyeglasses. The two-threshold algorithm, designed based on detailed signal analysis, outperformed in this study the SVM-based data models, with a f1-score ranging between 0.6 and 0.7. While the SVM models evaluated in this study could reach maximum recall, their precision was often below 0.5. Detecting eating events using proximity sensors embedded in smart eyeglasses is a promising approach for unconstrained ADM. Further research is needed to find algorithms that could increase retrieval performance for eating event detection, considering sensor data acquired under everyday life conditions.

REFERENCES

[1] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376869

[2] J.F. Kaiser. 1990. On a simple algorithm to calculate the 'energy' of a signal. In *International Conference on Acoustics, Speech, and Signal Processing*. 381–384 vol.1. https://doi.org/10.1109/ICASSP.1990.115702

[3] Nur Asmiza Selamat and Sawal Hamid Md. Ali. 2021. A Novel Approach of Chewing Detection based on Temporalis Muscle Movement using Proximity Sensor for Diet Monitoring. In *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. 12–17. https://doi.org/10.1109/IECBES48179.2021.9398736

[4] Nur Asmiza Selamat, Sawal Hamid Md Ali, Khairun Nisa' Minhad, and Jahariah Sampe. 2021. Feature Selection Analysis of Chewing Activity Based on Contactless Food Intake Detection. *International Journal of Integrated Engineering* 13, 5 (May 2021), 38–48. https://penerbit.uthm.edu.my/ojs/index.php/ijie/article/view/8620

[5] Jaemin Shin, Seungjoo Lee, Taesik Gong, Hyungjun Yoon, Hyunchul Roh, Andrea Bianchi, and Sung-Ju Lee. 2022. MyDJ: Sensing Food Intakes withnbsp;annbsp;Attachablenbsp;onnbsp;Yournbsp;Eyeglassnbsp;Frame. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 341, 17 pages. https://doi.org/10.1145/3491102.3502041

[6] Shuangquan Wang, Gang Zhou, Yongsen Ma, Lisha Hu, Zhenyu Chen, Yiqiang Chen, Hongyang Zhao, and Woosub Jung. 2018. Eating detection and chews counting through sensing mastication muscle contraction. *Smart Health* (2018).

[7] Rui Zhang and Oliver Amft. 2018. Monitoring Chewing and Eating in Free-Living Using Smart Eyeglasses. *IEEE Journal of Biomedical and Health Informatics* 22 (2018), 23–32. https://doi.org/10.1109/JBHI.2017.2698523

[8] Rui Zhang and Oliver Amft. 2020. Retrieval and Timing Performance of Chewing-Based Eating Event Detection in Wearable Sensors. *Sensors* 20, 2 (2020). https://doi.org/10.3390/s20020557