The SPLENDID chewing detection challenge

Vasileios Papapanagiotou[1], Christos Diou[1], Lingchuan Zhou[2]
Janet van den Boer[3], Monica Mars[3], Anastasios Delopoulos[1]

***Abstract*— Monitoring of eating behavior using wearable technology is receiving increased attention, driven by the recent advances in wearable devices and mobile phones. One particularly interesting aspect of eating behavior is the monitoring of chewing activity and eating occurrences. There are several chewing sensor types and chewing detection algorithms proposed in the bibliography, however no datasets are publicly available to facilitate evaluation and further research. In this paper, we present a multi-modal dataset of over 60 hours of recordings from 14 participants in semi-free living conditions, collected in the context of the SPLENDID project. The dataset includes raw signals from a photoplethysmography (PPG) sensor and a 3D accelerometer, and a set of extracted features from audio recordings; detailed annotations and ground truth are also provided both at eating event level and at individual chew level. We also provide a baseline evaluation method, and introduce the "challenge" of improving the baseline chewing detection algorithms. The dataset can be downloaded from `http://dx.doi.org/10.17026/dans-zxw-v8gy`, and supplementary code can be downloaded from `https://github.com/mug-auth/chewing-detection-challenge.git`.**

I. Introduction

Recently, automated monitoring of eating habits has received increased interest both by the research community (to study eating behavior) and for commercial applications (to support objective dietary monitoring). For example, in the case of the SPLENDID project [1], a prototype chewing sensor [2] is used to objectively measure eating occurrences and help users restrict their snacking [3].

Most approaches targeting the automatic detection of eating activity usually rely on detecting chews and/or swallows. Based on these, chewing bouts, and subsequently eating events (such as lunch, dinner, short snacks, etc) can be identified and reported [2]. One of the first and most common modalities that are deployed for detecting chewing and swallowing is audio; it is usually captured by microphones mounted either inside the ear [4]–[7], to capture chewing sounds, or around the throat [8], to capture swallowing sounds.

Additional sensors have also been employed to the task; in [9], [10] authors use strain sensors to detect chewing activity, while in [11] a photoplethysmography (PPG) sensor housed inside an ear hook is used to detect chewing activity through light fluctuations. In [2], this PPG sensor is combined with an in-ear microphone and a belt-mounted accelerometer to further improve chewing detection. In [12] audio and strain sensors are combined to detect both chews and swallows.

The effectiveness of such monitoring systems has been increasing, as new sensors and detection algorithms are being explored. However, development and evaluation of new algorithms and methods requires development data to be captured, annotated, and analyzed. Conducting data collection trials is a challenging task; it requires careful planning, financial resources, time, ethical permissions, trained supervising staff during the trial, data curation and laborious annotation work. Currently, to the best of our knowledge, no publicly available datasets from such chewing sensors exist.

In this paper, we introduce a publicly available dataset for chewing detection. The dataset has been recorded and processed in the context of EU-funded project SPLENDID in 2015 at Wageningen University. The recording equipment includes an in-ear housed microphone and a PPG sensor, and a belt-mounted 3D accelerometer. The raw signals from the PPG sensor and the accelerometer are publicly available. The original audio recordings, however are not, due to privacy restrictions. Instead, we provide a set of extracted features (both from audio and PPG signals) that were used in [2]; in addition, we include raw audio recordings from two members of the supervising staff (which include chewing sounds). In addition, detailed diaries are provided regarding the physical and eating activity of the participants, as well as timestamps for eating events. For each eating event, timestamps are also provided at individual chew level.

The rest of the paper is organized as follows: Section II provides a detailed description of the recording equipment. Section III provides information about the recording trials, e.g. information about the participants, the methodology, etc. Section IV describes the recorded signals, features, and statistics of the dataset, and Section V establishes the chewing detection challenge. Finally, Section VI concludes the paper.

II. Recording equipment

The prototype chewing sensor combines an audio microphone and a PPG sensor, housed together in a common ear hook (see Figure 1). The microphone is the FG-23329-D65 model from Knowles installed in a commercial ear-bud, so that it is placed inside the outer ear canal where body-generated sounds, including chewing sounds, are naturally amplified due to ear physiology, while external environmental sounds are somewhat dampened. The PPG sensor includes

*The work leading to these results has received funding from the European Community's Health, demographic change and well-being Programme under Grant Agreement No. 727688, 01/12/2016 - 30/11/2020 (http://bigoprogram.eu/).

[1]Multimedia Understanding Group, Information Processing Laboratory, Aristotle University of Thessaloniki, Greece

[2]CSEM SA, Switzerland

[3]Division of Human Nutrition, Wageningen University and Research, The Netherlands

(a) The data-logger with the 3D accelerometer connected to the audio and PPG chewing sensor

(b) Positioning of the chewing sensor

Fig. 1: The recording equipment

the BPW34FS photo-diode and the SFH4247 light-emitting-diode (LED), both from Osram. The photo-diode is placed inside the ear concha facing down and slightly backward, while the LED is placed behind the ear, facing towards the photo-diode. The PPG sensor adaptively amplifies the measured light intensity by taking into account ambient light levels. The variations in the light, and thus blood flow, contain information regarding chewing activity [2], [11].

Audio and PPG signals (including control signals for PPG) are recorded in the prototype data-logger (Figure 1a) along with the signals from the integrated 3D accelerometer (model LIS3DH by STMicroelectronics). These are the three modalities that were captured during the recording trials. More details about the hardware can be found in [2].

III. Data collection

The recording trials were conducted during June of 2015 at Wageningen University. In total, 22 individuals (19 female and 3 male) with mean age of 22.9 ± 1.9 years and mean body-mass-index (BMI) of $28 \pm 2.3\,kg/m^2$ participated; 19 of them participated during two different days, two weeks apart, and the remaining 3 participated during a single day. Each recording day lasted approximately 5 hours, split into 2 or 3 sessions. Due to hardware failures, a total of 26 such sessions from 14 participants have been collected lasting approximately 60 hours.

An overview of a recording day for a single participant is shown in Figure 2. Each day, three participants were using the equipment simultaneously. Upon arrival, the participants were introduced to the system and recording equipment, and were assisted in wearing the sensors, while the supervising staff ensured that the sensors were operating and recording properly. Then, the main recording started and soon the participants were seated for the first main meal, lunch. During the main meals, a variety of servings were available, and participants were free to select any combination and quantity multiple times. Once the meal was over, participants were free and could leave the university premises; no specific script was given to follow. They were instructed however to include at least three distinct eating activities/events (snacks), and at least four physical activities (including walking, running, playing outside, performing typical household tasks, etc) in their routine.

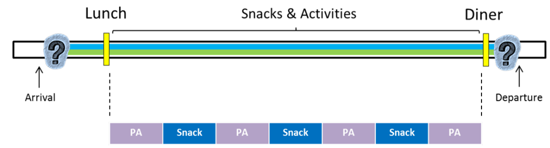


Fig. 2: Overview of a recording day; each subject was asked to perform at least four physical activity (PA) tasks and three snacking events.

TABLE I: Food types consumed during lunches

Type	Day 1	Day 2
Bread	Sliced bread, crackers	Soft buns, Baguette, rusk
Topping	Butter, jam, chocolate sprinkles, chocolate spread, peanut butter, cheese, sliced meat	Butter, jam, chocolate sprinkles, chocolate spread, peanut butter, cheese, sliced meat
Fruit	Grapes, banana, apple	Grapes, banana, apple
Drinks	Water, milk, orange juice	Water, milk, orange juice

The recording day concluded with the participants returning to the university and having the second main meal of the day, dinner. A different set of servings were available during dinners. More detailed information regarding the meals, snacks, and physical activity can be found in the dataset files and in [2].

IV. Collected signals & features

The dataset includes both recorded data as well as detailed ground truth annotations. The data is organized according to the recording procedure. For each participant, one or more sessions are available, as shown in Table IV. The table also lists the number of eating events and chews and total duration of the recordings.

TABLE II: Food types consumed during snacks

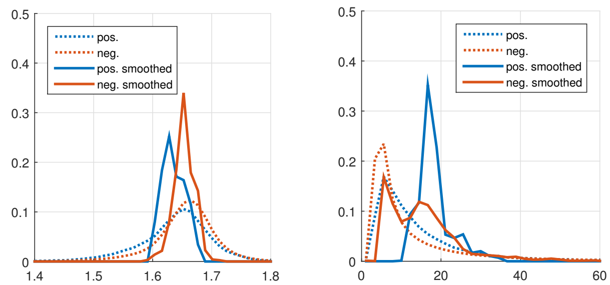
Type	Day 1	Day 2
Fruit	Grapes, banana, apple	Orange, strawberry, kiwi,
Cookie	Bastogne cookie, gingerbread, fruit biscuit	Hazelnut waffle, spongecake caramel waffle
Chips	-	Potato chips
Candy	Hard boiled candy, liquorice, twix bar, chewing gum	Lollipop, wine, gums mars bar
Drinks	Coffee, tea, hot chocolate, water, lemonade, orange juice, coke	Coffee, tea, hot chocolate, water, lemonade, orange juice, coke, milk

TABLE III: Food types consumed during dinners

Type	Day 1	Day 2
Potatoes	Boiled	Puree
Vegetables	French beans	Salad (lettuce, tomato, cucumber, boiled egg)
Meat	Meatball, wrapped in a slice of meat	Chicken schnitzel
Condiment	Gravy	Salad dressing
Dessert	Custard, vanilla & chocolate	Vanilla ice cream

TABLE IV: Dataset statistics

Participant	Sessions	Events	Chews	Duration (min)
11	1	4	797	154
31	2	8	2243	320
41	2	6	1171	287
42	1	6	904	164
43	1	3	427	68
51	2	7	1192	296
52	4	10	2269	398
53	2	7	1512	297
61	2	5	1202	316
62	1	2	807	104
63	1	4	670	162
65	2	5	522	210
71	4	16	2274	637
72	1	3	822	162



(a) Fractal Dimension (b) Skewness

Fig. 3: Normalised (per-class) histograms for chewing (pos.) and non-chewing (neg.) windows of example audio features (original and 1-min smoothed).

Regarding the recorded data, we provide the raw PPG signal, sampled at $\frac{64}{3}\,Hz$, along with two control signals sampled at the same frequency. These control signals consist of an index corresponding to the level of the current going through the LED, and the amplification gain of the A/D converter. The PPG signal values are stored as 15-bit unsigned integers. A set of ten features are also available corresponding to the non-normalized and normalized (histogram) time varying spectrum (TVS) in five log-bands ($0.0-1.0$, $1.0-1.8$, $1.8-3.3$, $3.3-5.9$ and $5.9-10.7$ Hz).

Audio was originally recorded at $48\,kHz$; it was subsequently downsampled at $2\,kHz$ and 15 features were extracted [2], including the fractal dimension (FD) [7], condition number (CN) of the 6×6 auto-correlation matrix, four 3$^{\text{rd}}$ and 4$^{\text{th}}$ order statistics and TVS in nine log bands ($0.0-4.0$, $4.0-7.4$, $7.4-15.8$, $15.8-31.6$, $31.6-63.0$, $63.0-125.9$, $125.9-251.2$, $251.2-501.2$, and $501.2-1000$ Hz). The recorded raw audio is not publicly available due to privacy restrictions. As a reference point regarding recording quality and conditions, two staff members tested the equipment a day before the first recording trial; their recordings are also available (both original recordings at $48\,kHz$ and downsampled at $2\,kHz$ in FLAC format, as well as the features). Finally, 3D accelerometer signals are also sampled at $\frac{64}{3}\,Hz$ and are stored as floating point numbers, measured in gs.

Figure 3 presents histograms of the FD and skewness for the positive (chewing) and negative (non-chewing) classes. Histograms are shown for both the original features as well as 1-minute smoothed features by a normalised hamming window. Each feature seems to be differently distributed for each class, however a classification model such as an SVM was shown to achieve high effectiveness. The histograms for the conditional number might indicate that this feature does not contain any information that can help classification, however we have observed that it in fact helps significantly to distinguish voice segments when combined with other features.

In Figure 4, a histogram of the mid-point of each eating event (meal/snack) is shown, in quarter-hour resolution. There are 24 meals during lunch time (from 12:00 to 13:00) and 20 meals during dinner (from 17:00 to 18:00). Snacks occur mostly around 13:45, 15:15, and 16:15.

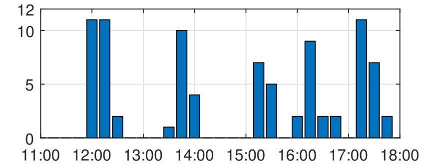


Fig. 4: Histogram of eating event times (quarter-hour resolution)

V. The SPLENDID chewing detection challenge

We present the following four challenges for the chewing detection task:

- PPG features: the first challenge is to identify and extract features from the PPG signal that can be used to effectively discriminate chewing from non-chewing. Currently, only spectral features are used [2], [11] that identify the "rhythmic" pattern of chews within a chewing bout.
- PPG-based detection: this challenge involves designing effective chewing detectors based on the features of the first challenge, or the ones provided with the dataset.
- Audio-based detection: further increasing the effectiveness of audio-based classifiers.
- Fusion-based detection: combining all available signals and features to improve detection.

We propose the three evaluation approaches presented in [2]; evaluation can also be performed at chew-level, based on the provided ground truth chew annotations. All ground truth is in the form of start and stop timestamps, for both events and chews. For the eating event level evaluation, a set of MATLAB scripts are also available through GitHub, for post-processing the detection classifiers in order to derive eating events, and for evaluating the results against the ground truth. The post-processing scripts aggregate the binary decisions of the classifiers from individual chews to chewing bouts by merging chews closer than $2\,sec$ apart and requiring a minimum chewing bout duration of $5\,sec$. Chewing bouts are then aggregated to eating events by merging bouts closer than $60\,sec$ and requiring that the resulting eating event is covered at least with 25% by chewing bouts.

TABLE V: Benchmark evaluation results for PPG-based and Fusion algorithms of [2]. For duration-based evaluation we prsent precision, recall, accuracy, weighted accuracy, and F1 Score, and for event-based the number of correct detections, missed detections, and false detections.

(a) Duration-based evaluation

		prec.	rec.	acc.	w. acc.	F1
LOSO	PPG (aver.)	0.341	0.814	0.753	0.767	0.448
	Fusion (aver.)	0.760	0.802	0.928	0.886	0.729
	PPG (cumul.)	0.278	0.801	0.710	0.749	0.413
	Fusion (cumul.)	0.641	0.805	0.918	0.870	0.714
Split	PPG (aver.)	0.148	0.802	0.511	0.620	0.250
	Fusion (aver.)	0.242	0.548	0.780	0.692	0.335
	PPG (cumul.)	0.200	0.785	0.190	0.540	0.319
	Fusion (cumul.)	0.227	0.714	0.208	0.528	0.433

(b) Event-based evaluation

		No. of CDs	No. of MDs	No. of FDs
LOSO	PPG	70	16	202
	Fusion	69	17	51
Split	PPG	9	7	11
	Fusion	9	7	41

Three methods for evaluating the detection algorithms are proposed. The first two, "average duration-based" and "cumulative duration-based" measure the duration during which the detector agrees or disagrees with the ground truth, thus partitioning each session into true-positive (TP) time, false-positive (FP) time, true-negative (TN) time, and false-negative (FN) time. The first method, average duration-based, computes five metrics, specifically precision, recall, accuracy and weighted accuracy, and F1 score, based on TP, FP, TN and FN, *for each participant*. It then averages each metric across participants. The second method, cumulative duration-based, does not average across participants; it instead computes the five metrics on the entire dataset duration, thus taking into account the variability of total recording duration across participants.

Finally, the third method, "event-based", performs a one-to-one matching among detected and ground truth events, taking into account their overlap duration which is required to be at least 75% of their combined duration for the detected event to be considered a correct detection (CD). Non-matched ground truth events count as missed detections (MDs) and non-matched detected events count as false detections (FDs).

For evaluating alrogithms that require a training step and a test step, one can follow the "Leave-One-Subject-Out" (LOSO) procedure [2], or use the train/test split proposed with the dataset. Table V presents the benchmark evaluation results of [2] for all three evaluation methods.

VI. Conclusions & future work

In this paper we have presented a dataset and evaluation framework for eating event and chewing detection, consisting of PPG, audio, and acceleration sensor data. The dataset has been recorded in semi-free living conditions, and extensive ground truth annotations of eating events and individual chews are provided.

Results of a baseline evaluation are also included for comparison purposes. These show that the combination of all three sensors can yield very high effectiveness for the eating event detection task. However, there is still room for improvement for the individual chew detection task and for PPG-only detection (which is important, given the low sampling rate and smaller form-factor of the PPG).

The dataset is available at `http://dx.doi.org/10.17026/dans-zxw-v8gy`, and supplementary code can be downloaded from `https://github.com/mug-auth/chewing-detection-challenge.git`.

References

[1] C. Maramis, C. Diou, I. Ioakeimidis, I. Lekka, G. Dudnik, M. Mars, N. Maglaveras, C. Bergh, and A. Delopoulos, "Preventing obesity and eating disorders through behavioural modifications: the splendid vision," in *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*. IEEE, 2014, pp. 7–10.

[2] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos, "A novel chewing detection system based on ppg, audio and accelerometry," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2016.

[3] C. Diou, I. Sarafis, I. Ioakimidis, and A. Delopoulos, "Data-driven assessments for sensor measurements of eating behavior," in *Proceedings of the International Conference on Biomedical and Health Informatics (BHI)*, 2017.

[4] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of chewing sounds for dietary monitoring," in *UbiComp 2005: Ubiquitous Computing*. Springer, 2005, pp. 56–72.

[5] O. Amft*, M. Kusserow, and G. Trster, "Bite weight prediction from acoustic recognition of chewing," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 6, pp. 1663–1672, June 2009.

[6] S. Päßler and W.-J. Fischer, "Evaluation of algorithms for chew event detection," in *Proceedings of the 7th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2012, pp. 20–26.

[7] V. Papapanagiotou, C. Diou, Z. Lingchuan, J. van den Boer, M. Mars, and A. Delopoulos, "Fractal nature of chewing sounds," in *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2015, vol. 9281, pp. 401–408. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-23222-5_49

[8] E. S. Sazonov*, O. Makeyev, S. Schuckers, P. Lopez-Meyer, E. L. Melanson, and M. R. Neuman, "Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 626–633, March 2010.

[9] E. S. Sazonov and J. M. Fontana, "A sensor system for automatic detection of food intake through non-invasive monitoring of chewing," *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1340–1348, May 2012.

[10] M. Farooq and E. Sazonov, "Comparative testing of piezoelectric and printed strain sensors in characterization of chewing," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 7538–7541.

[11] V. Papapanagiotou, C. Diou, L. Zhou, J. van den Boer, M. Mars, and A. Delopoulos, "A novel approach for chewing detection based on a wearable ppg sensor," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2016, pp. 6485–6488.

[12] E. Sazonov, S. Schuckers, P. Lopez-Meyer, O. Makeyev, N. Sazonova, E. L. Melanson, and M. Neuman, "Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior," *Physiological measurement*, vol. 29, no. 5, p. 525, 2008.