Recognition of Food-Texture Attributes Using an In-Ear Microphone

Vasileios Papapanagiotou[1(✉)], Christos Diou[1,2], Janet van den Boer[3],
Monica Mars[4], and Anastasios Delopoulos[1]

[1] Multimedia Understanding Group, Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
vassilis@mug.ee.auth.gr, adelo@eng.auth.gr
[2] Department of Informatics and Telematics, Harokopio University of Athens, Kallithea, Greece
cdiou@hua.gr
[3] Department of Biomedical Signals and Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands
j.h.w.vandenboer@utwente.nl
[4] Division of Human Nutrition and Health, Wageningen University, Wageningen, The Netherlands
monica.mars@wur.nl

Abstract. Food texture is a complex property; various sensory attributes such as perceived crispiness and wetness have been identified as ways to quantify it. Objective and automatic recognition of these attributes has applications in multiple fields, including health sciences and food engineering. In this work we use an in-ear microphone, commonly used for chewing detection, and propose algorithms for recognizing three food-texture attributes, specifically crispiness, wetness (moisture), and chewiness. We use binary SVMs, one for each attribute, and propose two algorithms: one that recognizes each texture attribute at the chew level and one at the chewing-bout level. We evaluate the proposed algorithms using leave-one-subject-out cross-validation on a dataset with 9 subjects. We also evaluate them using leave-one-food-type-out cross-validation, in order to examine the generalization of our approach to new, unknown food types. Our approach performs very well in recognizing crispiness (0.95 weighted accuracy on new subjects and 0.93 on new food types) and demonstrates promising results for objective and automatic recognition of wetness and chewiness.

Keywords: Food texture · Dietary monitoring · Wearables

1   Introduction

Chewing is one of the main ways of how we perceive food texture [25]. While there is a huge variety of products with different food textures (e.g. crispy products like potato chips, hard and moist products like apples and cucumbers), some textures are generally perceived as more pleasant and desirable than others [11]. It is clear from several studies that food texture and structure are becoming more important in understanding eating behavior, especially in food intake regulation and weight management [1,17,29,31]. Taking food structure into account in dietary advice can may support the prevention and dietary treatment of overweight and obesity [6,9,14]. However, the effects still need to be supported by longer-term studies outside the laboratory. The current, existing dietary assessment methods, such as diaries and recalls, rely on memory and do not provided detailed information about the texture of foods [7].

Currently, there is a strong effort in creating automated tools for dietary monitoring, in the context of understand and preventing obesity and eating disorders [22,27]. One of the first approaches has been to automatically detect chewing based on the audio captured by an in-ear microphone [4]. Such audio signals have also been used to extract information such as the food type [3,23]. There are also alternative types of approaches for identifying food types; for example, leveraging photos that people take with their mobile phones can be used to detect food-relevant photos and then subsequently to perform image segmentation and identify the food type [10,13,18]. Alternatively, user input can be requested when an automatic eating detection system detects eating activity [5].

Identifying food-content–relevant information such as food type from audio signals is commonly formulated as a multi-class problem where each class is a different food type from a list of pre-determined food types [2,4]. The selection of food types is usually aimed at creating a diverse set with different textures, such as crispy and non-crispy food types; crispiness, however, is only one of the attributes of texture. In the early work of [19], a set of texture reference scales were introduced that include multiple attributes. Table 1 presents three groups of attributes: attributes related to surface attributes and springiness, attributes assessed during mastication (chewing), and those assessed during manual manipulation. Out of the three groups, the attributes assessed during mastication are the ones of interest to this work. The more recent work of [30] presents a more extensive and complete review of the state of knowledge for food texture. According to that work, it is commonly accepted that "texture is the sensory and functional manifestation of the structural, mechanical and surface properties of foods detected through the senses of vision, hearing, touch and kinesthetics". As a result, no single modality sensor (such as a microphone) can completely identify texture.

It is worth mentioning that there are additional, non-medical fields where research in understanding human eating behavior is also useful. More specifically, in the field of food design and engineering, understanding how people perceive certain attributes of food (such as crispiness) has been found to correlate with freshness (in particular for the case of apples [12,24]) which in turn is the most

Table 1. Food-texture attributes as presented and organized in [19], and their correspondance with the food-attributes used in this work.

Attributes	Crisp	Wet	Chewy
Attributes related to surface attributes and springiness			
Wetness	✗	✓	✗
Adhesiveness to lips	✗	✗	✗
Roughness	✓	✗	✗
Self-adhesiveness	✗	✗	✓
Springiness	✓	✗	✗
Attributes assessed during mastication			
Cohesiveness of mass	✓	✗	✗
Moisture absorption	✗	✓	✗
Adhesiveness to teeth	✗	✗	✓
Attributes assessed during manual manipulation			
Manual adhesiveness	✗	✗	✗

important factor in consumer's choices [16,26]. Thus, providing tools that can objectively measure such food attributes cannot only help to assess eating behavior and food intake, but can also help food technologists to design food with more desirable and pleasant characteristics.

In this work we use an in-ear microphone that is part of the wearable, prototype sensor developed in the context of the SPLENDID project and focus on the audio captured during chewing in order to recognize three attributes of food texture from a variety of food types. The three attributes are crispiness, wetness, and chewiness. Each one corresponds to each of the attributes "assessed during mastication" according to [19] (Table 1). The attributes "related to surface attributes and springiness" can also be loosely mapped to these three attributes. The attributes that are "assessed during manual manipulation" are not of interest to this work (but are only included in the table for completeness).

We propose an algorithm for recognizing each of the three food-texture attributes of this work (i.e. crispiness, wetness, and chewiness) given a single chew based on three individual binary SVMs (one for each attribute). We also propose a modified version of the algorithm that operates on entire chewing bouts. We evaluate the generalization both in new subjects and in new food types.

2   Attribute Recognition Algorithms

The algorithms presented in this work require audio of at least 8 kHz sampling rate. The original audio recordings of the dataset used in this work (see Sect. 3) have been sampled at 48 kHz. We have experimented with down-sampled versions of the signal, in particular 2, 4, 8, 16, 32, as well as the original 48 kHz. We have observed that down-sampling as low as 8 kHz does not cause any noticeable drop

in recognition effectiveness (by repeating the experiments presented in this paper for all different sampling frequencies), however, down-sampling lower than 8 kHz does. In all the following, we use the 8 kHz down-sampled versions of the audio signals.

We also apply a high-pass Butterworth filter to the down-sampled signal to remove low frequency components. The filter is of 9-th order with a cut-off frequency at approximately 20 Hz. We propose two algorithms: the first recognizes attributes for each chew individually, while the second one for entire chewing bouts.

2.1   Chew-Level Algorithm

A feature vector is first extracted from each chew; note that start and stop time-moment annotations for chews have been made available by manual extraction based on acoustic and visual observation of the captured audio signals (in a fully automated application, a chew-detection algorithm such as [22] can be used to obtain them). The extracted features consist of signal energy in 11 log-scale frequency bands based on time-varying spectrum (TVS) estimation, fractal dimension (FD), condition number (CN) of the auto-correlation matrix, and higher order statistics (i.e. third and fourth-order moments). These features have been used in [22] for chewing detection and are thus a good starting point for food-texture–attribute recognition. Since the audio sampling frequency is only 2 kHz in [22] (compared to 8 kHz of this work) we have added two more bands in the TVS, in particular 2 to 4 and 4 to 8 kHz. It is also worth noting that each chew has different duration (average of 0.56 s and standard deviation of 0.15 s in the data-set used in this work) in contrast to windows of fixed length often used in signal processing. The features we have selected, however, are invariant to the length of the audio segment used to extract them.

Before the classification stage each feature is standardized by subtracting the mean and then dividing by the standard deviation (the mean and standard deviation of each feature are estimated over the available training set for each case). The multi-label classifier we use is an array of three binary SVMs; each SVM is related to one of the three food-texture attributes: crispiness, wetness, and chewiness. We use a radial-basis function (RBF) kernel. Parameters C and γ are selected automatically using 5-fold cross-validation on the training set. The optimal values are chosen using Bayesian optimization [15]; care to escape local minima of the objective function is also taken using a threshold for the standard deviation of the posterior objective function [8,28].

As a result, each chew can be classified individually for each of the three attributes. As the food type does not change within a bout, we can obtain one decision per bout based on the chews that belong to it. One way to do so is to use majority voting across the chews of a bout (for each attribute). Another way is to consider only the first n chews of a bout, since processing of the food in the mouth transforms it into a wet bolus, thus altering the attributes of the food in its unprocessed state. All these evaluation methods are presented in Sect. 4.

2.2   Chewing-Bout–Level Algorithm

Chewing-bout–level detection shares the same pre-processing steps with chew-level detection: the audio signal is down-sampled and the same high-pass is applied to remove low-frequency components.

Bout segments are then obtained based on the chews that belong to each bout. A bout audio segment starts at the start time of its first chew and stops at the stop time of its last chew. The average bout duration in the data-set used in this work is 15.22 s with a standard deviation of 10.7 s. Since bouts are significantly longer than chews and also contain non-chewing sounds between each chew (see Fig. 1) we do not directly extract the features over the entire bout duration. Instead, we obtain overlapping windows of 0.5 s length and 0.1 s step from each bout and extract the features from each window separately. Thus, we obtain one list of feature vectors for each bout; each list contains, in general, a different amount of vectors.
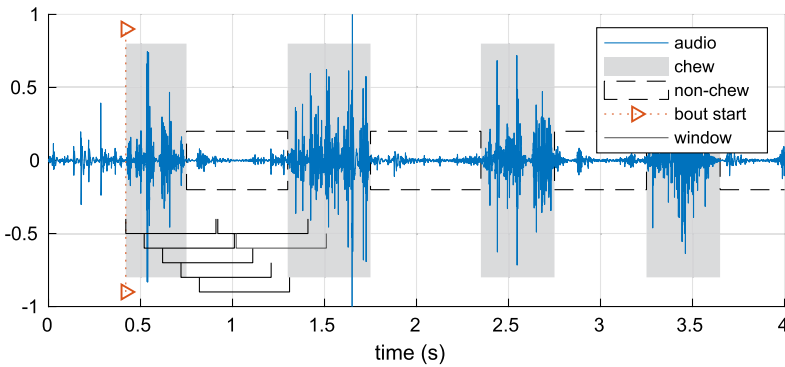
Fig. 1. An example of a chewing bout. The first four chews are marked by the gray rectangles. When entire chewing bouts are used, the audio between two successive chews is also used.

We then use a bag-of-words (BoW) approach to obtain a single feature vector of fixed length for each bout. In particular, given a set of bouts, we obtain all of the feature vectors from each bout and use k-means to select a set of k centroid vectors. Once we obtain the k centroid vectors we can transform any new bout into a feature vector of fixed length (equal to k) by computing the normalized histogram of the bout's feature vectors against the set of k centroid vectors. Each feature vector is assigned to one of the k centroid (i.e. hard assignment).

Using the BoW approach offers many advantages. It allows to use the same features of the previous, chew-level algorithm (Sect. 2.1) on short windows (with similar duration to that of chews). In addition it allows to extract feature vectors of fixed length from audio segments with (highly) varying length. Finally, it also allows to handle the non-chewing sounds that occur between successive chews within each bout: window-based feature vectors that correspond to such

in-between audio segments will likely be similar and will be clustered together; corresponding cluster centers are equally present in signals of different food types and the SVM models are expected to learn to ignore them.

The BoW features are then standardized similarly to Sect. 2.1. Classification is performed in exactly the same way as in chew-level recognition: we use an array of three binary SVM classifiers with RBF kernel and hyper-parameter selection using Bayesian optimization.

3   Data-Set and Evaluation Metrics

The data-set used to evaluate the proposed algorithms has been collected at Wageningen University, Netherlands, in the context of the EU-funded SPLEN-DID project[1] and is the same data-set as the one we use in [21]. The recording apparatus is an in-ear microphone (FG-23329-D65 model manufactured by Knowles) connected via wire to a computer audio interface. The sensor housing and recording has been done by CSEM S.A. [22]. In this work, the first version of the SPLENDID sensor is used; more details about this version and the future versions of the sensor can be found in [5]. In total, 21 subjects were enrolled for the data collection trials, however, signals from only 9 could be used in this work due to problems with data acquisition (such as incorrect sensor placement or corrupted audio due to hardware/software malfunction). Each subject consumed a variety of food types (complete list can be found in [5]).

We have selected 9 different food types that we can clearly annotate their attributes. Not all 9 subjects have consumed all 9 food types. Table 2 lists these food types along with their attribute values we have assigned them. This data-set of 9 participants and 9 food types is the one we use to evaluate this work.

Table 2. List of food types and their attributes for the evaluation data-set.

	Crispy	Wet	Chewy
Apple	✓	✓	✗
Banana	✗	✓	✗
Bread	✗	✗	✗
Candy bar	✗	✗	✓
Cookie	✓	✗	✗
Lettuce	✓	✓	✗
Potato chips	✓	✗	✗
Strawberry	✗	✓	✗
Toffee	✗	✗	✓

For this evaluation data-set, we have manually created ground truth on chew and chewing bout levels (with start and stop time-moments) based on the available experimental logs as well as audio and visual inspection of the captured

---

[1] https://splendid-program.eu/.

signals. It contains 4,989 chews with a total duration of 46.31 min which belong to 238 chewing bouts; total duration of bouts is almost 1 h and is greater than total duration of chews because bouts also contain the audio segments between each successive pair of chews. Tables 3 and 4 show duration statistics for the chews and chewing bouts.

Table 3. Statistics of chews duration across the food types for the evaluation data-set.

	Instances	Total (min)	Average (s)	Std (s)
Apple	1,091	9.74	0.54	0.13
Banana	294	2.67	0.54	0.16
Bread	1,260	11.72	0.56	0.13
Candy bar	415	4.26	0.62	0.17
Cookie	359	3.07	0.51	0.14
Lettuce	492	4.06	0.50	0.11
Potato chips	386	3.45	0.54	0.12
Strawberry	236	2.04	0.52	0.13
Toffee	456	5.30	0.70	0.17
Total	4,989	46.31	0.56	0.15

Table 4. Statistics of bouts duration across the food types for the evaluation data-set.

	Instances	Total (min)	Average (s)	Std (s)
Apple	48	12.12	15.15	5.23
Banana	25	3.57	8.58	2.91
Bread	37	15.03	24.37	15.69
Candy bar	22	5.51	15.02	7.97
Cookie	16	4.03	15.10	5.31
Lettuce	28	5.31	11.38	5.39
Potato chips	29	4.40	9.11	3.12
Strawberry	17	2.63	9.30	5.82
Toffee	16	6.23	23.35	11.37
Total	238	58.83	14.83	9.84

We evaluate each food attribute, namely crispiness, wetness, and chewiness, as binary classification problems. We regard crispy, wet, and chewy as the positive classes, and non-crispy, dry, and non-chewy as the negative ones respectively.

To account for class imbalance, which is particularly large for chewiness, we calculate weighted accuracy as

$$\frac{w \cdot TP + TN}{w \cdot (TP + FN) + TN + FP} \tag{1}$$

where $w = (TN + FP)/(TP + FN)$ is the ratio of priors.

4   Evaluation and Results

The chew-level algorithm can also be modified to operate on bout-level by taking into account each bout's chews. We explore how the bout-level decision is affected by the number of chews that are taken into account. We train models both in the typical leave-one-subject-out (LOSO) fashion as well as in leave-one-food-type-out (LOFTO) fashion. In both types of experiments we use the entire available training data (from the non–left-out subjects or food types) to both obtain the BoW centroid vectors and train the SVM classifiers.

Table 5 presents per-chew classification results for each of the three attributes; the BoW centroids, SVM models, and SVM hyper parameters have been trained in LOSO fashion. Crispiness is the attribute that the algorithm identifies more effectively. The majority voting approach consistently improves results by 2 to 5%.

While it makes sense to assume that food type as such does not change during a bout, the attributes of the food within the mouth do as food is grinded and lubricated during oral processing/chewing. Given that we are interested in identifying the attributes that the food has in its unprocessed form, we can consider only the first few chews of each bout in the majority voting stage, during which the food type's original attributes are still retained to some degree. Figure 2 shows the recognition effectiveness for each attribute when considering only the first n chews of each bout for $n = 1$ to 20. Recognition of crispiness exhibits high effectiveness, however, the highest (weighted) accuracy is obtained by considering the first 6 to 10 chews. Almost the same range of 5 to 10 chews seems to be the best choice for recognizing wetness as well. Chewiness results are somewhat different since considering only the first few chews seems to yield erratic effectiveness. The situation improves as more chews are taken into account.
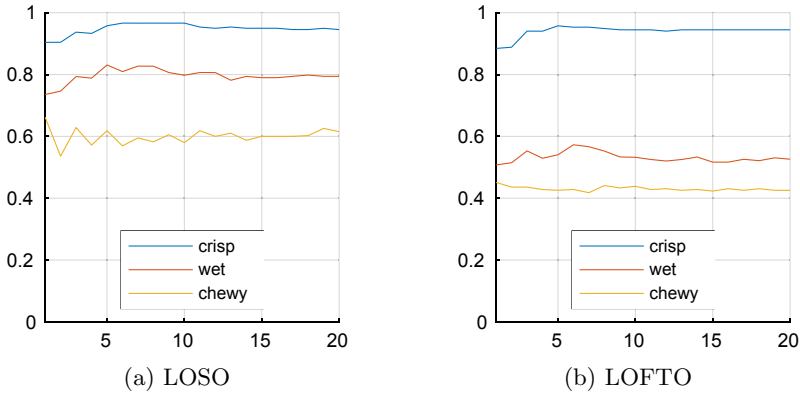
Table 6 presents results for the LOFTO experiments. Crispiness is the most easily recognizable attribute again, however, wetness does not generalize well across different food types. Majority voting improves results for crispiness and chewiness only. Looking at Fig. 2, the highest effectiveness for wetness is achieved by considering only a few chews (6 to 7 chews).

Table 7 presents LOSO and LOFTO results for the bout-level algorithm. Comparing these results with chew-level results we can see that the bout-level algorithm achieves almost similar results for LOSO: slightly lower weighted accuracy for crispiness (but still quite high), and almost the same for wetness. Recognition accuracy for chewiness is worse.

Table 5. LOSO results for chew-level recognition.

	Prior	Weighted accuracy	w
Chew level			
Crispy (avg)	0.4707	0.9068	1.1968
Crispy (sum)	0.4666	0.9017	1.1430
Wet (avg)	0.4280	0.7516	1.4588
Wet (sum)	0.4235	0.7503	1.3611
Chewy (avg)	0.1741	0.5994	4.9418
Chewy (sum)	0.1746	0.6212	4.7279
Majority voting per bout			
Crispy (avg)	0.4943	0.9519	1.1141
Crispy (sum)	0.5063	0.9496	0.9752
Wet (avg)	0.4850	0.7978	1.1018
Wet (sum)	0.4937	0.7900	1.0254
Chewy (avg)	0.1666	0.6296	5.9804
Chewy (sum)	0.1632	0.6154	5.1282

(a) LOSO (b) LOFTO

Fig. 2. Weighted accuracy for each attribute for the LOSO and LOFTO experiments.

Table 6. LOFTO results for chew-level recognition.

	Prior	Weighted accuracy	w
Chew-level			
Crispy (sum)	0.4666	0.8987	1.1430
Wet (sum)	0.4235	0.5481	1.3611
Chewy (sum)	0.1746	0.3957	4.7279
Majority voting per bout			
Crispy (sum)	0.4957	0.9446	1.0172
Wet (sum)	0.4829	0.5046	1.0708
Chewy (sum)	0.1667	0.4179	5.0000

On the other hand, the bout-level algorithm seems to be able to better generalize across different food types. Looking at the results of the LOFTO experiment, the algorithm achieves only slightly lower weighted accuracy (compared to the chew-level algorithm with majority voting) for crispiness and improves significantly for wetness and chewiness.

Table 7. Results for bout-level recognition.

	Prior	Weighted accuracy	w
LOSO			
Crispy (avg)	0.4967	0.9541	1.1030
Crispy (sum)	0.5084	0.9534	0.9669
Wet (avg)	0.4869	0.7865	1.0880
Wet (sum)	0.4958	0.7900	1.0169
Chewy (avg)	0.1625	0.5200	6.0749
Chewy (sum)	0.1597	0.5238	5.2632
LOFTO			
Crispy (sum)	0.5084	0.9288	0.9669
Wet (sum)	0.4958	0.6422	1.0169
Chewy (sum)	0.1597	0.4970	5.2632

5   Conclusions

In this work we have proposed two algorithms for automatically and objectively recognizing three attributes of food texture from audio signals captured by an in-ear microphone. The algorithms combine feature extraction and binary SVMs and operate both for single chews and entire chewing bouts. We have examined their ability to generalize not only to new subjects (LOSO) but also to new food types (LOFTO). With the use of these algorithms, the SPLENDID sensor was able to recognize 3 important food-texture attributes affecting eating behavior. In particular, crispiness was recognized with weighted accuracy of at least 0.9 across all experiments. Results for recognizing wetness and chewiness are promising but there is a large margin for improvement: introducing more suitable features in the algorithms could possibly help improve the overall effectiveness of the proposed algorithms. Including more food types in the training set could also potentially improve recognition, however, certain food types with attributes that are not easy to annotate (e.g. food types that are neither completely dry nor wet) might not be suitable for a crisp-label classification problem and thus require alternative methods. As a result, this type of digital devices will make it possible to further study the objective exposure to different food textures in relation to eating behavior and longer term outcomes, such as weight change, in a real life setting.

References

1. Aguayo-Mendoza, M.G., Ketel, E.C., van der Linden, E., Forde, C.G., Piqueras-Fiszman, B., Stieger, M.: Oral processing behavior of drinkable, spoonable and chewable foods is primarily determined by rheological and mechanical food properties. Food Qual. Pref. **71**, 87–95 (2019). https://doi.org/10.1016/j.foodqual.2018.06.006

2. Amft, O., Kusserow, M., Troster, G.: Bite weight prediction from acoustic recognition of chewing. IEEE Trans. Biomed. Eng. **56**(6), 1663–1672 (2009)

3. Amft, O.: A wearable earpad sensor for chewing monitoring. In: SENSORS, 2010 IEEE, pp. 222–227 (2010). https://doi.org/10.1109/ICSENS.2010.5690449

4. Amft, O., Stäger, M., Lukowicz, P., Tröster, G.: Analysis of chewing sounds for dietary monitoring. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 56–72. Springer, Heidelberg (2005). https://doi.org/10.1007/11551201_4

5. van den Boer, J., et al.: The splendid eating detection sensor: development and feasibility study. JMIR Mhealth Uhealth, p. e170. (2018). https://doi.org/10.2196/mhealth.9781

6. van den Boer, J., Werts, M., Siebelink, E., de Graaf, C., Mars, M.: The availability of slow and fast calories in the dutch diet: the current situation and opportunities for interventions. Foods **6**(10), 87 (2017). https://doi.org/10.3390/foods6100087

7. Brouwer-Brolsma, E.M., et al.: Dietary intake assessment: from traditional paper-pencil qestionnaires to technology-based tools. In: Athanasiadis, I.N., Frysinger, StP, Schimak, G., Knibbe, W.J. (eds.) ISESS 2020. IAICT, vol. 554, pp. 7–23. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39815-6_2

8. Bull, A.D.: Convergence rates of efficient global optimization algorithms. arXiv e-prints arXiv:1101.3501 (2011)

9. Campbell, C.L., Wagoner, T.B., Foegeding, E.A.: Designing foods for satiety: the roles of food structure and oral processing in satiation and satiety. Food Structure **13**, 1–12 (2017). https://doi.org/10.1016/j.foostr.2016.08.002

10. Christodoulidis, S., Anthimopoulos, M., Mougiakakou, S.: Food recognition for dietary assessment using deep convolutional neural networks. In: Murino, V., Puppo, E., Sona, D., Cristani, M., Sansone, C. (eds.) ICIAP 2015. LNCS, vol. 9281, pp. 458–465. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23222-5_56

11. Cox, D.N., et al.: Sensory and hedonic judgments of common foods by lean consumers and consumers with obesity. Obesity Res. **6**, 438–447 (1998). https://doi.org/10.1002/j.1550-8528.1998.tb00376.x

12. Daillant-Spinnler, B., MacFie, H.J.H., Beyts, P.K., Hedderley, D.: Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. Food Qual. Pref. **7**(2), 113–126 (1996). https://doi.org/10.1016/0950-3293(95)00043-7

13. Dehais, J., Anthimopoulos, M., Mougiakakou, S.: Food image segmentation for dietary assessment. In: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management. MADiMa '16, ACM, New York, NY, USA, pp. 23–28 (2016). /DOIurl10.1145/2986035.2986047

14. Forde, C.G.: From perception to ingestion; the role of sensory properties in energy selection, eating behaviour and food intake. Food Qual. Pref. **66**, 171–177 (2018). https://doi.org/10.1016/j.foodqual.2018.01.010

15. Gelbart, M.A., Snoek, J., Adams, R.P.: Bayesian optimization with unknown constraints. arXiv e-prints arXiv:1403.5607 (2014)

16. Harker, R.F., Gunson, A.F., Jaeger, S.R.: The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. Postharvest Biol. Technol. **28**(3), 333–347 (2003). https://doi.org/10.1016/S0925-5214(02)00215-6

17. Lasschuijt, M.P., Mars, M., Stieger, M., Miquel-Kergoat, S., De Graaf, C., Smeets, P.A.M.: Comparison of oro-sensory exposure duration and intensity manipulations on satiation. Physiol. Behav. **176**, 76–83 (2017). https://doi.org/10.1016/j.physbeh.2017.02.003

18. Lu, Y., Allegra, D., Anthimopoulos, M., Stanco, F., Farinella, G.M., Mougiakakou, S.: A multi-task learning approach for meal assessment. arXiv e-prints arXiv:1806.10343 (2018)

19. Muñoz, A.M.: Development and application of texture reference scales. J. Sens. Stud. **1**(1), 55–83 (1986). https://doi.org/10.1111/j.1745-459X.1986.tb00159.x

20. Papapanagiotou, V.: Modeling and automatically measuring human eating behavior. Ph.D. thesis, Department Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki (2019)

21. Papapanagiotou, V., Diou, C., Lingchuan, Z., van den Boer, J., Mars, M., Delopoulos, A.: Fractal nature of chewing sounds. In: Murino, V., Puppo, E., Sona, D., Cristani, M., Sansone, C. (eds.) ICIAP 2015. LNCS, vol. 9281, pp. 401–408. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23222-5_49

22. Papapanagiotou, V., Diou, C., Zhou, L., van den Boer, J., Mars, M., Delopoulos, A.: A novel chewing detection system based on PPG, audio and accelerometry. IEEE J. Biomed. Health Inform. **21**(3), 607–618 (2017). https://doi.org/10.1109/JBHI.2016.2625271

23. Päßler, S., Wolff, M., Fischer, W.: Food intake monitoring: an acoustical approach to automated food intake activity detection and classification of consumed food. Physiol. Meas. **33**(6), 1073–1093 (2012). https://doi.org/10.1088/0967-3334/33/6/1073

24. Péneau, S., et al.: Relating consumer evaluation of apple freshness to sensory and physico-chemical measurements. J. Sens. Stud. **22**(3), 313–335 (2007). https://doi.org/10.1111/j.1745-459X.2007.00112.x

25. Pereira, L.J., van der Bilt, A.: The influence of oral processing, food perception and social aspects on food consumption: a review. J. Oral Rehabil. **43**(8), 630–648 (2016). https://doi.org/10.1111/joor.12395

26. Péneau, S., Hoehn, E., Roth, H.R., Escher, F., Nuessli, J.: Importance and consumer perception of freshness of apples. Food Qual. Pref. **17**(1), 9–19 (2006). https://doi.org/10.1016/j.foodqual.2005.05.002

27. Sazonov, E.S., Makeyev, O., Schuckers, S., Lopez-Meyer, P., Melanson, E.L., Neuman, M.R.: Automatic detection of swallowing events by acoustical means for applications of monitoring of ingestive behavior. IEEE Trans. Biomed. Eng. **57**(3), 626–633 (2010). https://doi.org/10.1109/TBME.2009.2033037

28. Snoek, J., Larochelle, H., Adams, R.P.: Practical bayesian optimization of machine learning algorithms. arXiv e-prints arXiv:1206.2944 (2012)

29. StribiŢcaia, E., Evans, C.E.L., Gibbons, C., Blundell, J., Sarkar, A.: Food texture influences on satiety: systematic review and meta-analysis. Sci. Rep. **10**(1), 12929 (2020). https://doi.org/10.1038/s41598-020-69504-y
30. Szczesniak, A.S.: Texture is a sensory property. Food Qual. Pref. **13**(4), 215–225 (2002). https://doi.org/10.1016/S0950-3293(01)00039-8
31. Zijlstra, N., de Wijk, R., Mars, M., Stafleu, A., de Graaf, C.: Effect of bite size and oral processing time of a semisolid food on satiation. Am. J. Clin. Nutr. **90**(2), 269–275 (2009). https://doi.org/10.3945/ajcn.2009.27694