Automatic Identification of Temporal Sequences in Chewing Sounds

Oliver Amft, Martin Kusserow and Gerhard Tröster

Wearable Computing Lab, ETH Zurich, Switzerland

Email: amft@ife.ee.ethz.ch

Abstract— Chewing is an essential part of food intake. The analysis and detection of food patterns is an important component of an automatic dietary monitoring system. However chewing is a time-variable process depending on food properties. We present an automated methodology to extract sub-sequences of similar chews from chewing sound recordings. The approach is based on a chew-accurate segmentation of the sound signal, a multi-objective evolutionary search for temporal partitions in the sequence using NSGA-II and a validation of the best solution by classification.

We evaluate the method on chewing sound recordings from a four participant study, eating foods with different rheological properties. The proposed methodology allows to determine the most appropriate partitioning of the sequences and extract relevant sound features at the same time. Potato chips and chocolate showed a two-phase structure, for lasagne and apples a single-phase structure was derived. The results led to the hypothesis that a sequential structure can be found in chewing sounds from brittle or rigid foods.

I. INTRODUCTION

Food intake is a vital aspect of human health. The prevalence of over- and under-consumption as well as unbalanced meal composition surges the risk of chronic diseases such as obesity. Consequently, assistive systems that monitor food intake from non-invasive sensors could provide a valuable tool for dietary monitoring in risk groups. Manual dietary monitoring systems that are currently used require frequent user interaction, e.g. to log food type and time in consumption questionnaires, scan or take pictures of the foods consumed. Besides the high post-processing efforts of professional services to analyze and verify the data, all of these methods require intensive collaboration by the user.

We believe that the automation of dietary monitoring could alleviate this problem by reporting daily schedule and tentative food consumption using a wearable system. Towards automatic dietary monitoring, a system was proposed [1] that consists of three sensing domains: 1) identification of arm and trunk movements that characterize the food intake from inertial sensors [2], e.g. using fork and knife, a spoon or hand-only movements, 2) detection of swallowing from a sensing collar [3] and 3) characterization of foods from chewing sounds. The latter sensing domain is the focus of this work.

The breakdown of foods during chewing generates sound emissions that are conducted by bones, skull and body tissue. For a wearable dietary monitor, sound from chewing can be recorded by a microphone or similar acoustic transducer in the ear canal or close to it. In this way, different foods have been classified from their acoustic profile during chewing [1], [4]. However chewing is a time-dependent process, modified by the changing properties of the food material in the mouth when fluids are pressed out, the food is mixed with saliva and a bolus is formed [5]. The intrinsic chewing movement pattern is generated by a brain stem pattern generator. The movement pattern is continuously modified by oral sensory feedback [6].

Chewing sequences start from an *initial bite* (shearing of a sample from a food piece with the incisors), followed by a variable number of rhythmic *chewing cycles* (compressing the sample using molars) until swallowing occurs. Recent studies confirmed that changes occur in movement as well as muscle activity of the masticatory apparatus within these sequences (see Section I-A for a discussion of related work). Changes in movement of the mandible were attributed to the modification of rheology parameters (hardness, fracturability, adhesiveness) of food [7]. Since both, movement and rheology changes during a chewing sequence, it can be hypothesized that different acoustic stages exist in the sequence as well. Such stages could relate to sensory changes occurring during the sequence. Moreover, if the existence of acoustic stages could be confirmed, sequential food classification models become feasible. Such models can achieve a better fit to the chewing sound pattern and potentially improve the scalability of food sound models.

In the present work we apply an automated methodology to extract sub-sequences of chewing sequences from ear-canal chewing sound recordings. Sub-sequences are defined as a series of cycles in the chewing sequence with similar acoustic properties. Since the temporal evolution of the acoustic pattern and hence chewing sound stages are largely unknown, an unsupervised search and optimization strategy was deployed to analyze the sub-sequences in a chewing sequence and select appropriate features for discriminating these structures at the same time. The solutions are qualified for compliance to a chewing sequence model and validated by classification on test data.

A. Related work in chewing monitoring

First insight into the sequence of chewing was obtained by kinematic and electromyographic (EMG) studies. At the beginning of the sequence hardness was found to control the frequency of chewing and activity of M. masseter and M. temporalis while in middle of the sequence product rheology described mandibular movement such as vertical ampli-

tude [7]. Similar results were found in earlier studies, where burst duration and mean voltage of the EMG as well as vertical mandibular movement decreased during the sequence [8]. Chewing of materials with different rheology indicated, that changes in the activity pattern of M. masseter occur on the chewing cycle level during a sequence, however sequence parameters did not differ for materials of varying rheology [9].

Based on the initial investigation by Drake [10] chewing sound has been assessed predominately to study auditory and sensory perception of material texture in food science [11], [12]. From observing audio waveforms it was assumed that a normal chewing sequence could be partitioned into two phases: initial gross cutting of the ingested material and subsequent conversion in fine grained particles [10]. This process is understood as a gradual decomposition of the material structure during chewing and is audible as a decline of the sound amplitude in brittle foods [1].

The loudness of a foodstuff during chewing depends mainly on the cell arrangement, impurities and existing cracks [12]. Naturally grown foods, e.g. apples or lettuce, contain more liquids compared to dry products, e.g. potato chips, that have air inclusions [13].

Most previous works that targeted classification of chewing sounds used a small share of chewing cycles from each sequence only or analyzed a sequence as one entity. None of these works used the complete chewing sequence, consequently phasing of the sequences was not investigated. De Belie et al. [4] used the sound spectrum of the initial bite to compute principal components and discriminate two classes of crispness in apples. More recently a classification of five snack food products was presented, based on the auditory signal of the initial bite and the first chew [14].

Analyzing the initial bite or the chewing cycle refers to two separate investigation techniques. It was found that bite and chewing cycles differ regarding movement and emitted sound [15], [16]. For our purpose of finding stages in the chewing sequence, all chewing cycles are included in this analysis, while the initial bite is manually excluded.

II. METHODS

This section summarizes the study procedure to acquire chewing data and presents the data analysis steps for extracting a partitioning from the chewing sequences. After summarizing the feature processing from chewing cycles, the multiobjective search strategy is presented. Finally the procedure to identify and test the best partitioning solution is detailed.

A. Chewing study

Four participants (male students, natural dentition, aged 20 to 30 years) without known chewing or swallowing abnormality were recruited for the study. During a pre-recording interview the lab and measurement environment was shown to each participant for familiarization. Two measurement sessions were carried out on separate days, with at least one day break in between. Each session was recorded around mid-day. Participants were asked to eat the following foods:

Potato chips ("Chio chips Ready salted", ∼25 pieces), meat lasagne (∼250 g), one apple ("Jangold", ∼100 g), 12 pieces of chocolate ("Coop lait", total: 40 g). The meat lasagne was a commercial deep-frozen version, heated in an oven for 40 minutes.

All participants were familiar with the food types based on cultural origin. None of the participants expressed a dislike for any of the foods nor problems to chew or swallow the selected foods. Participants were sitting comfortably on a chair close to a table carrying the food items and a glass of water. They were asked to chew and swallow normally and allowed to move, drink and speak during the recording sessions between the chewing sequences. The recording duration was not constraint since the participants were eating/drinking at their individual pace.

Chewing sound was recorded using a miniature microphone (Knowles, TM-24546) embedded into an ear-pad. The sound signal was amplified and sampled at 44 kHz, 16 bit. Surface EMG from left and right M. masseter was recorded at 2 kHz, 24 bit.

An observer controlled the recording procedure during each session and annotated the chewing sequences. In a postprocessing step all annotated sequences were reviewed, the start/end times adapted and swallowing events marked for exclusion by inspecting the signals. Synchronization marks in the data (set during the recordings) were used to align audio and EMG data streams.

B. Chewing segmentation and feature processing

From the continuous chewing sequence recordings individual chewing cycles were extracted using the muscle activity derived from EMG. Signals from left and right M. masseter were bandpass filtered and combined for independence from the chewing side. Peak muscle activity was used to estimate teeth clench (complete occlusion, teeth in full contact) and hence the previous onset of the combined EMG signal determined the beginning of the current and end of the previous chewing cycle. The first and last cycles were determined using the sequence annotation. Only chews segmented in the bounds of the annotation were used for further analysis.

Three feature search spaces were computed from the segmented chewing cycle. For search space one, the complete segmented chewing cycle was used, for search space two, only an initial part of the chewing cycle was used (EMG onset to peak) and for search space three both previous spaces were combined. Spaces one and two contained 65 audio features (130 for feature space three), computed from time and frequency domain. The time domain features included: length of the segment, extrema, fluctuation, zero crossings and the integrated signal. The spectral features included: total and band energy, fluctuation, centroid, bandwidth, rolloff as well as auto-correlation and cepstral coefficients.

For the subsequent investigation observations were derived by computing the features from every chewing cycle. The total set of observations was split into a search and testing set.
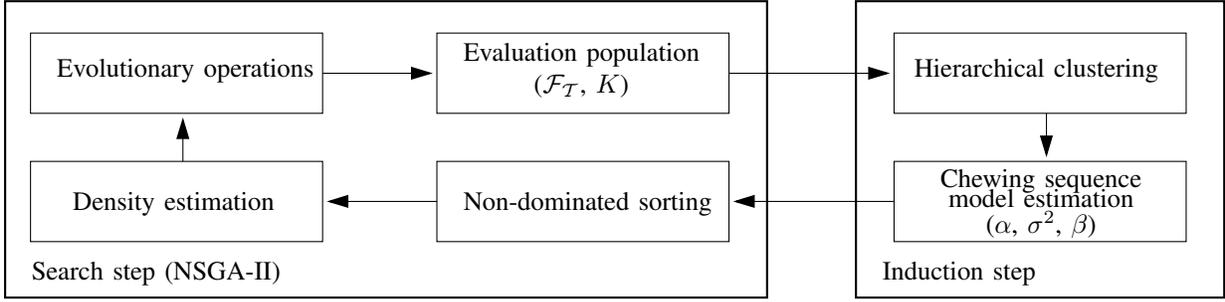
Fig. 1. Outline of the search procedure for feature selection and chewing sequence partitioning.

The testing set (10% of all observations) was used for result validation in the final classification.

C. Sequence partitioning using multi-objective search

Finding partitions in the temporal evolution of chewing sequences was regarded as a search problem that required the identification of related observation sub-sequences and the selection of discriminative features to model the partitioning. Fig. 1 provides an overview on the search framework, composed of a search and an induction step. In the search step potential features were selected. In the induction step, these features were used to determine clusters in the observation data. The clustering result was validated against a chewing sequence model.

1) Search step: We selected the Non-dominated Sorting Genetic Algorithm, version 2 (NSGA-II) [17] as search algorithm and feature selection wrapper. NSGA-II belongs to the family of search heuristics based on evolutionary algorithms that can accommodate multiple search goals. A detailed introduction to evolutionary algorithms can be found in [18].

The algorithm keeps a diverse population of individual solutions and aims at finding non-dominated (Pareto-optimal) solutions by applying evolutionary operations (selection, mutation and crossover). With the diversity of individuals in the population a high robustness is achieved: locked oscillations and single solutions at local optima are avoided. Moreover, the algorithm promotes elitism by maintaining Pareto-optimal solutions, once found, in the following generations. The algorithm achieved a good performance when compared to similar methods [19].

For each individual solution, a binary bit vector encodes the feature set and expected sub-sequences (clusters) in the chewing sequences. At initialization a uniform distribution of the bits was used to set the vector.

The genetic search was performed with a population size of 100 individuals, a mutation rate of 0.05 (uniform) and a crossover rate of 0.8 (single point). An independent search was performed for each of the three feature spaces. After 250 generations the feature selection was stopped and the Pareto-optimal solutions were evaluated and tested as described in II-D.

2) Induction step: As induction step for the feature selection we used a hierarchical clustering of the chewing observations in each sequence and assessed the partitioning result using validity parameters of a chewing sequence model. The validity parameters were used in the search step (as search goals) and during the result selection (as quality measure).

The expected number of sub-sequences (specified in the search step) was used as clustering target. We analyzed the range of 2 to 5 clusters.

The following chewing sequence model was defined to describe the search problem and to develop the model validity parameters. The model was applied for each food type individually.

A complete sequence of chewing cycles $\mathcal{S}_i$, $(\mathcal{S}_i \in \mathcal{S})$ of food type $\mathcal{T}$ consists of K unique sub-sequences, called phases $\mathcal{P}$. $\mathcal{S}$ is the set of all N sequences from $\mathcal{T}$. Each phase $\mathcal{P}_{i,j}$ consists of $M_{i,j}$ chewing observations described by feature vector $f_{i,j,n}$ from feature space $\mathcal{F}_\mathcal{T}$. The relation of the sets is described by Eq. 1.

$$
\begin{aligned}
\mathcal{S} &= \{\mathcal{S}_1 \ldots \mathcal{S}_N\} \\
\mathcal{S}_i &= \{\mathcal{P}_{i,1} \ldots \mathcal{P}_{i,K}\}, & i = 1 \ldots N \\
\mathcal{P}_{i,j} &= \{f_{i,j,1} \ldots f_{i,j,M_{i,j}}\}, & \forall i : j = 1 \ldots K
\end{aligned} \quad (1)
$$

Phases can neither overlap nor repeat in a single chewing sequence $\mathcal{S}_i$. An observation $f_{i,j,n}$ belongs to exactly one phase $\mathcal{P}_{i,j}$.

Based on this model formulation, the following model validity parameters were defined to describe a chewing sequence: *phase count* α, *phase size variance* σ^2 and *phase transitions* β. The parameters were defined in order to obtain a minimization problem.

1) The parameter *phase count* α relates the number of retrieved phases $\hat{K}$ and the number of expected phases K, normalized over all sequences of $\mathcal{S}$ (Eq. 2).

$$
\alpha = 1 - \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{K}}{K} \quad (2)
$$

The value of α is minimal (zero) if $\hat{K} = K$, i.e. if the number of phases found is equal to the number of expected phases in all sequences of $\mathcal{S}$.

2) The parameter *phase size variance* σ^2 depicts the variation in the number of chewing observations within each phase, normalized over all expected phases K and

chewing sequences N (Eq. 3). M_j is the mean share of observations in all sequences assigned to phase P_j.

$$\sigma^2 = \frac{1}{K} \sum_{j=1}^{K} \sum_{i=1}^{N} \frac{(\frac{M_{i,j}}{M_i} - M_j)^2}{N-1} \qquad (3)$$

$$M_i = \sum_{j=1}^{K} M_{i,j} \; ; \qquad M_j = \frac{1}{N} \sum_{i=1}^{N} \frac{M_{i,j}}{M_i} \qquad (4)$$

The parameter σ^2 reflects the stability of the obtained partitioning: if $\sigma^2 = 0$, the relative number of chewing observations in each phase is constant for all phases of all chewing sequences.

3) The parameter *phase transitions* β is defined as ratio of existing phase changes in each sequence and the expected number of phase changes $(K - 1)$, normalized over all sequences in $\mathcal{S}$ (Eq. 5 and 6).

$$\beta = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\sum_{n=1}^{M_i - 1} d_{i,n}}{K - 1} - 1 \right| \qquad (5)$$

$$d_{i,n} = \begin{cases} 1 : cluster(f_{i,n}) \neq cluster(f_{i,n+1}) \\ 0 : otherwise \end{cases} \qquad (6)$$

The parameter β measures the consistency of the phases in the sequences: the higher β, the less correspondence exist between the obtained grouping and the expected sequential partitioning (the more alternations exist). If $\beta = 0$, the number of transitions matches the expectation. If $\beta > 0$ the number of transitions does not match, because more or less than the expected transitions were obtained.

D. Result selection and testing

A selection strategy was developed to extract the best result from the solution space. Fig. 2 provides an overview on the applied selection steps.

All results from the independent search runs for each feature space were merged into a single solution space and their Pareto-optimal individuals were selected. Fig. 3 exemplarily visualizes the obtained solution space for the three model validity parameters.

In the following step all solutions that weakly matched the chewing sequence model were removed. Eq. 7 shows the constraints applied for α and σ^2. Individual solutions that did not comply with these bounds were removed from the solution space. Parameter p_{err} limits the share of sequences that do not conform to the chewing sequence model and was set to $p_{err} = 0.2$.

$$\alpha \leq \frac{p_{err}}{K} \; ; \qquad \sigma^2 \leq \frac{p_{err}}{2} \qquad (7)$$

The best candidate was chosen from the remaining solutions in a final selection step. The space was reduced by sequentially selecting the minimal value of each parameter α, σ^2, β and $|\mathcal{F}_\mathcal{T}|$. Individuals with a minimal α were chosen first, since the
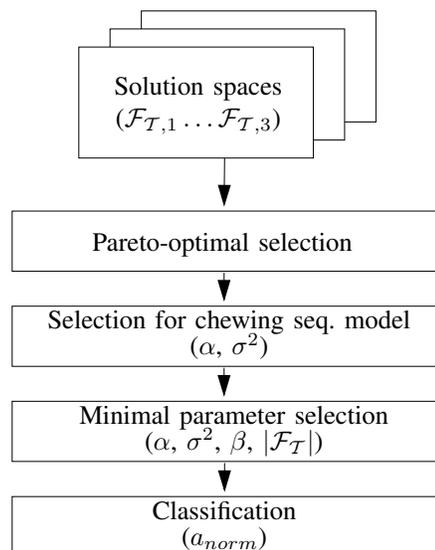

Fig. 2. Outline of the selection and testing procedure.

correct phase count is a prerequisite to achieve the partitioning target.

In order to verify the result, the identified solution was applied in a classification of phases in the testing observation set. Using the selected features for the best candidate, a nearest centroid classifier was trained on the training set using the candidate's clustering result as class association rule. For evaluating the testing result, the class associations were obtained by assuming the phase distribution of the best candidate. In this way, the solution was tested independently of the search procedure.

The partitioning into phases resulted in class skews (one phase contained more observations than another phase). Training a skewed classifier was avoided by selecting an equal number of training observations from all phases.

To compare classification results with an unequal number of test observations in each class, the normalized accuracy measure as used. For a multi-class classification the normalized accuracy was derived as mean of the class-relative accuracies:

$$a_{norm} = \frac{1}{K} \sum_{i=1}^{K} \frac{Recognized_i}{Relevant_i} \qquad (8)$$

where K is number of expected phases, $Recognized_i$ and $Relevant_i$ are the number of correctly identified observations and the total number of observations for each phase category.

III. RESULTS AND DISCUSSION

A. Chewing segmentation

In total 11480 chewing cycles were extracted for the analysis from 602 annotated chewing sequences. Tab. I summarizes the number of chewing sequences and cycles for each food product.
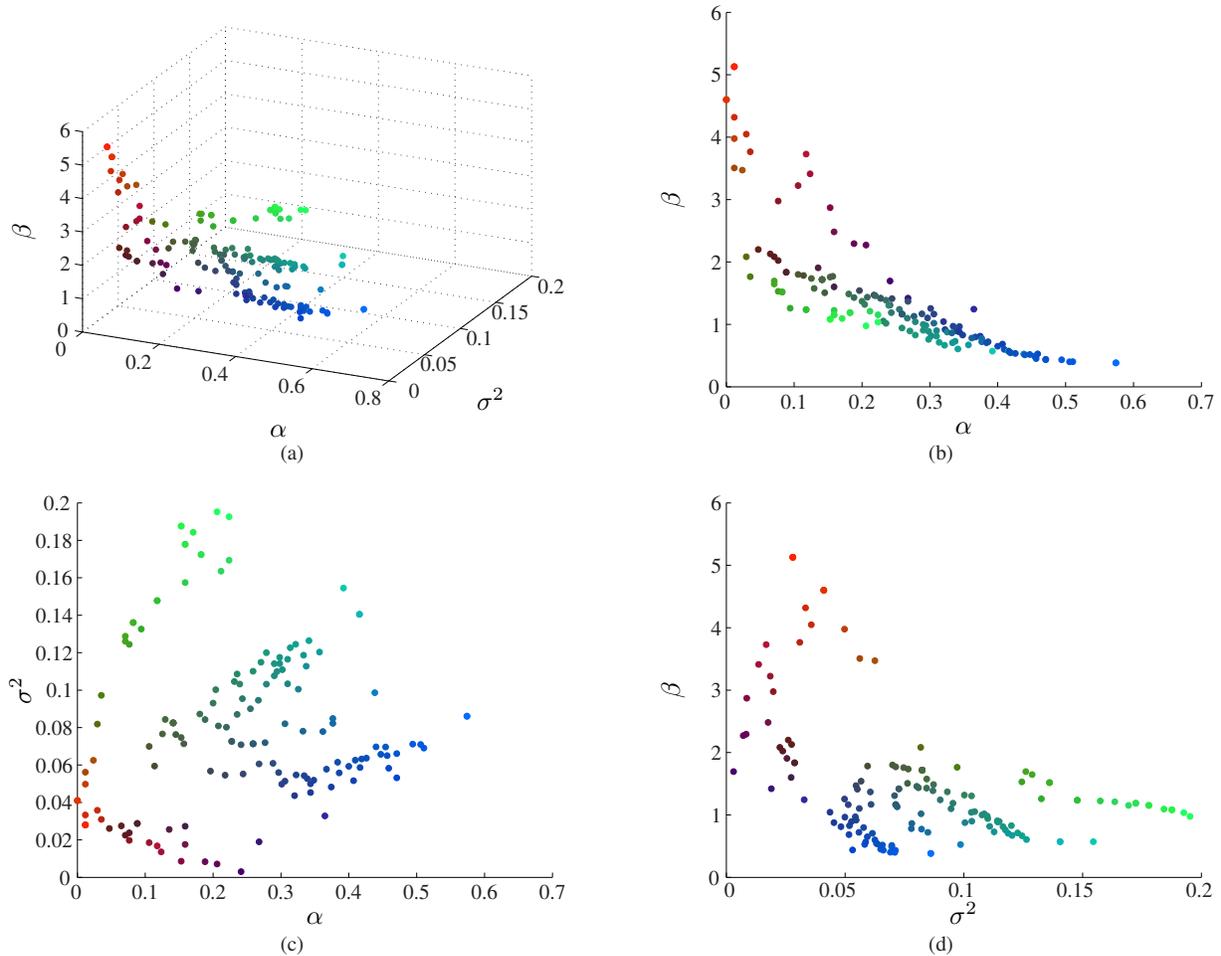
Fig. 3. Visualization of the solution space derived by the search procedure for food type *chocolate* after merging of the feature spaces. Each point corresponds to an individual solution. The optimal compliance with the chewing sequence model is at $\alpha = 0$, $\sigma^2 = 0$ and $\beta = 0$. The plots show: (a) All axes α, σ^2 and β; (b) Projection on α-β; (c) Projection on α-σ^2; (d) Projection on σ^2-β.

TABLE I

SUMMARY OF THE SEGMENTATION RESULTS.

Food type	Chewing sequences	Chewing cycles
Apple	102	2447
Potato chips	227	3015
Chocolate	90	1913
Lasagne	183	4105
Total	**602**	**11480**

B. Food-specific analysis

The partitioning was analyzed for every food type, including the observations from all four study participants. The solution space spanned by the model validity parameters is visualized in Fig. 3 for the food type *chocolate*.

The solution space diagram shows the distribution of the individuals around the optimal solution of the minimization problem ($\alpha = 0$, $\sigma^2 = 0$ and $\beta = 0$). While the optimal solution was not achieved in the search, good solutions were found, using the result selection procedure.

The result selection from a solution space is considered to be a critical step for multi-objective approaches [20]. The procedure applied in this work aimed at evaluating the search goals individually, instead of combining the goals in a single weighting function. This strategy was reasonable due to the exploratory analysis approach.

In the food-specific evaluation a two-phase structure was obtained for all foods. Fig. 4 shows the average distribution of the phases (occurrence ratios) in the chewing sequences for the individual foods. The occurrence ratios indicate the relative share and position of each phase in the chewing sequences. For all food types a short first phase (30-40% of the sequence length) was found, followed by a second longer phase. For *potato chips* the longest initial phase was obtained, when compared to the other foods.

This finding confirms the assumption of an initial chewing phase that differs from the remaining sequence made by observing sound amplitude vs. time plots [10]. Tab. II summarizes the phase ratios and quality of the solutions obtained. Both, search and testing performances of the selected
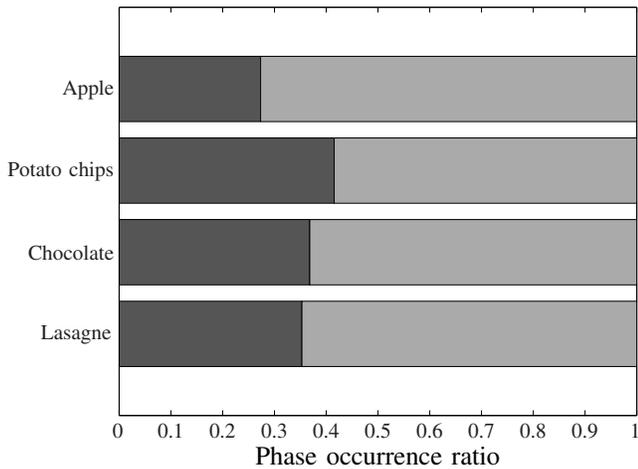
Fig. 4. Average distribution of phases in the chewing sequences for all four foods obtained in the food-specific analysis. The plots indicate the location of the phases, but do not show insertions inside individual phases. For food types *potato chips* and *chocolate* the partitioning was confirmed by the testing procedure.

TABLE II

RESULTS OF THE FOOD-SPECIFIC SEQUENCE PARTITIONING ANALYSIS.

Search parameter	Apple	Potato chips	Choco-late	Lasa-gne
Expected phases (K)	2	2	2	2
Nr of features ($|\mathcal{F}_\mathcal{T}|$)	16	24	22	20
Feature space	2	2	2	2
Sequence model parameter	Candidates after search			
Phase count (α)	0.04	0.00	0.00	0.07
P. size variance (σ^2)	0.07	0.02	0.04	0.09
Phase transitions (β)	3.45	1.35	4.60	5.55
	Testing data			
Phase count (α)	0.07	0.02	0.00	0.08
P. size variance (σ^2)	0.09	0.05	0.03	0.09
Phase transitions (β)	2.86	1.07	4.80	5.63

candidates are indicated by the model validity parameters of the chewing sequence model.

A good quality, according to the parameters α and σ^2 was found for all four food types. However, the high number of phase transitions (indicated by β) suggest that the phases had several insertions. This effect was very strong for the food type *lasagne*. The overall best quality for all parameters was achieved for *potato chips*.

The effect of applying the candidate solutions on testing data was very low on the model parameters: the values for all three model parameters remained in the same range. For all foods 16 to 24 features were selected from feature space two.

Whether the sequence structure of the candidates can be verified, was assessed by the classification performance on test data. Fig. 5 shows the classification result for all four foods, when applying the properties of the selected solution on the testing set. A normalized accuracy of $a_{norm} = 0.5$ would indicate a random classification. Therefore only the

range $a_{norm} = 0.5 \ldots 1.0$ is shown.

The classification confirms the partitioning for *potato chips* and *chocolate*: a normalized accuracy of approx. 80% was achieved, indicating that the two-phase assumption holds, even on the testing data.

For *apple* and *lasagne* a low accuracy was obtained, indicating that the foods cannot be classified with the two-phase search solution. Since phase counts larger than two were rejected during the search stage already, we concluded, that the foods do not exhibit a sub-sequence structure at all.
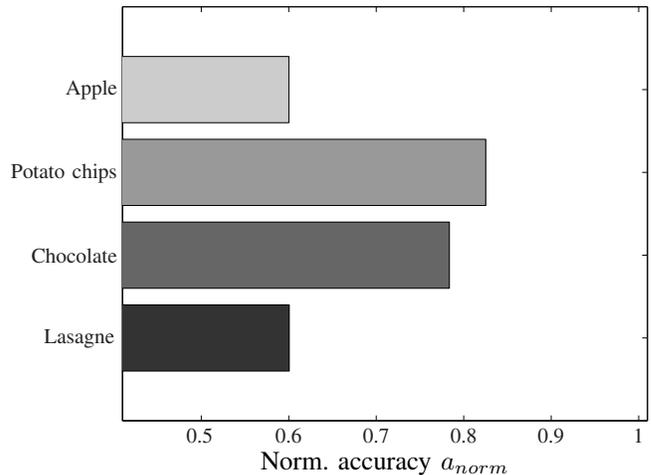

Fig. 5. Classification result for all four foods when applying the search solution on the testing set. (A normalized accuracy of $a_{norm} = 0.5$ would indicate a random classification. Therefore the plot shows the range $a_{norm} = 0.5 \ldots 1.0$ only.)

C. Participant-specific analysis

The search and solution selection was performed for each participant individually to analyze the person-related structure of the chewing sequences. Tab. III summarizes the results for all four foods and participants A–D. Except for food type *apple* from participant A, two-phase solutions were found for all combinations of food and participant using the selection procedure. For *apple* from participant A none of the solutions matched the required criteria.

The classification accuracy a_{norm} achieved during testing gives an indication whether the two-phase result can be generalized on the testing data. Except for *chocolate*, the result of the food-specific analysis were confirmed: For *potato chips* a two-phased partitioning was obtained with good classification rates ($0.79 \ldots 0.92$) for all participants. For *apple* and *lasagne* the rates are generally lower. While the performance for *apple* was above random (~ 0.6), the two-phase partitioning for *lasagne* often performed less well.

For *chocolate* participants A and C achieved a classification performance of 0.81 and 0.75 respectively. However, no phase structure was found for the remaining two participants and *chocolate*. In order to analyze this effect in more detail, the study should be extended by additional participants.

TABLE III

RESULTS OF THE PARTICIPANT-SPECIFIC SEQUENCE PARTITIONING ANALYSIS.

Food type	Participant	Expected phases K	Nr of features $\|\mathcal{F}_\mathcal{T}\|$	Feature space	Phase ratio 1	Phase ratio 2	Candidates α	Candidates σ^2	Candidates β	Testing α	Testing σ^2	Testing β	Norm. accuracy a_{norm}
Apple	A	—	—	—	—	—	—	—	—	—	—	—	—
	B	2	34	1	0.79	0.21	0.00	0.01	2.96	0.00	0.00	0.00	0.67
	C	2	29	2	0.44	0.56	0.00	0.04	10.80	0.00	0.16	9.00	0.61
	D	2	20	1	0.88	0.12	0.00	0.02	3.93	0.50	0.00	1.00	0.50
Potato chips	A	2	26	1	0.52	0.48	0.00	0.02	1.27	0.04	0.05	1.23	0.79
	B	2	61	3	0.72	0.28	0.00	0.01	2.09	0.00	0.00	3.20	0.79
	C	2	28	1	0.34	0.66	0.00	0.00	0.42	0.00	0.00	0.40	0.92
	D	2	65	3	0.84	0.16	0.00	0.01	1.72	0.00	0.01	3.60	0.82
Chocolate	A	2	17	2	0.46	0.54	0.00	0.01	2.09	0.00	0.06	1.50	0.81
	B	2	25	1	0.88	0.12	0.00	0.00	2.95	0.00	0.02	6.00	0.46
	C	2	13	1	0.16	0.84	0.00	0.00	1.95	0.00	0.00	2.33	0.75
	D	2	14	1	0.84	0.16	0.00	0.01	3.71	0.00	0.01	5.00	0.53
Lasagne	A	2	28	2	0.67	0.33	0.10	0.07	1.84	0.33	0.09	1.50	0.51
	B	2	26	2	0.55	0.45	0.05	0.08	3.70	0.11	0.14	2.56	0.56
	C	2	21	2	0.64	0.36	0.01	0.08	9.96	0.00	0.11	6.50	0.56
	D	2	27	1	0.91	0.09	0.00	0.00	2.60	0.25	0.03	4.50	0.57

D. Result discussion

The phase distribution and selected features of the participant-specific analysis showed a higher variability when compared to the food-specific analysis. It was assumed that this is a result of the lower number of test observations and hence less averaging effect in the participant-specific analysis.

In contrast to the food-specific analysis, applying the search solution on the testing data in the participant-specific analysis had a strong effect on the model validity parameters. For *apple* and *lasagne* this indicated the absence of sub-sequences, for the other food types this result was not expected. Although an overall large set of observations was available, the splitting into search and test set (test set size was 10% of all observations) may have led to a reduction of the result stability. Even to confirm the results of the food-specific analysis, a test with different observations should be performed. However, these result confirms the benefit of the applied testing procedure.

The modeling of phases in chewing sequences for an exploratory search is a challenging task. While the developed optimization goals (phase count, phase size variance and phase transitions) proved to be vital parameters for a valid partition, some limitations of the automatic phase extraction approach remain to be solved.

Regarding the search procedure, the indirect control by rating the solutions after the clustering step, produced invalid solutions, that could have been avoided. To this end clustering should consider the sequence of the observations. However, to achieve this, a highly domain specific grouping algorithm would be needed.

From the phase transitions result (parameter β), large values were found for both food- and participant-specific investigations. While the parameter correctly indicates a partitioning that does not correspond to the analyzed number on phases, it is sensitive to insertions. Single assignments of observations to a different phase, which is non-critical for the overall integrity of the partitioning and could be ignored, was not distinguishable from an completely invalid partitioning. This

aspect influenced the selection of optimal solutions and could mislead conclusions. Again with the help of our selection and testing procedure this issue was minimized, since β was used as the last of all three model parameters and all results were verified by the classification test.

IV. CONCLUSION

We presented an automatic method to extract partitions from chewing sequences, that follow a sequential order and can be identified using sound features. Our approach relied on a search and selection procedure, followed by a verification on test data. The search was performed using a NSGA-II wrapper for selecting appropriate features and expected sub-sequences in combination with an induction step. The induction was composed of hierarchical clustering of the chewing observations and analyzing the search quality with respect to a chewing sequence model.

Three parameters were derived that describe a chewing sequence model. The parameters were used for the qualification of the search results and the final selection of the valid solutions. In the present work the model parameters *phase count*, *phase size variance* and *phase transitions* were used.

The sequence structure of four food types was analyzed in recordings from four participants, regarding food-specific and participant-specific behavior. A two-phase structure was found from the food- and participant-specific analysis of the food types *potato chips* and *chocolate*. The search results were verified in a classification test, by assuming the retrieved features and the partition structure in a food sequence model. A classification accuracy of approx. 80% was achieved for both foods. A person-dependency was found for *chocolate*, where no valid phasing was obtained for two of the participants, while for the remaining tow a good classification was possible. We assumed that this variability was caused by the small and fixed test observation set, since the food-specific analysis of all participants returned a classification performance of 78%.

Overall, the food- and participant-specific evaluations returned at maximum two phases for all foods. A common distribution of the phases was found in the food-specific analysis. For all food types, a short first phase (30-40% of the sequence length) was derived, followed by a second longer phase. For the participant-specific analysis different distributions were obtained depending on both, food and person.

Out of the three model validity parameters described above, the most important goal was *phase count*, since the number of actually retrieved phases was vital for the application of the sequence model. Solutions that had large values for this goal ($\alpha \geq 0.2$), typically did not perform well. A selection procedure was designed that reflected this observation, by sequentially limiting the share of sequences that did not conform to the chewing sequence model.

The parameter *phase transitions* was found to be most difficult to minimize, since the phases contained insertions in many sequences of the foods. The insertions were attributed to the natural variability in the data that could not be captured by the clustering algorithm. In the current work, the impact of this issue was minimized by the selection and testing procedure. For further investigations alternate solutions to obtain a requested number of transitions should be evaluated.

The frequent selection of feature space two in the food-specific analysis indicated that the chewing cycles were not stationary and hence supportive information was extracted from the intra-chew signal variation (as captured by the feature space) when compared to the entire chew (feature space one).

For the food types *apple* and *lasagne* no stable partitioning was found, indicating that both foods have a non-sequential sound pattern that alternatively could be described by a single phase for each sequence. Following these findings, it can be hypothesized that out of the different foods analyzed, only dry and rigid foods have a clear sequential structure. Consequently, foods that are soft, e.g. *lasagne* or based on a fiber-structure as *apple* do not change their sound pattern in an ordered sequence.

We consider the classification rates for *potato chips* and *chocolate* as a comparably good result in relation previous investigations of food sound classification. For bite and first chew classifications from five snack foods up to 18% classification error were reported [14]. In our previous work on classifying foods from chewing recordings, rates between 66% and 86% were achieved, depending on the food type [1]. For an application in automatic dietary monitoring the result of the current work will support the development of food-adapted classifiers. For the above food types an independent model of the two phases could be helpful to boost the system performance, while for other (single-phase) foods one model is sufficient.

ACKNOWLEDGMENTS

REFERENCES

[1] O. Amft, M. Stäger, P. Lukowicz, and G. Tröster, "Analysis of chewing sounds for dietary monitoring," in *UbiComp 2005: Proceedings of the 7th International Conference on Ubiquitous Computing.*, M. Beigl, S. Intille, and J. Rekimoto, Eds. Springer: Lecture Notes in Computer Science, September 2005, pp. 56–72.

[2] O. Amft, H. Junker, and G. Tröster, "Detection of eating and drinking arm gestures using inertial body-worn sensors," in *ISWC 2005: IEEE Proceedings of the Ninth International Symposium on Wearable Computers.*, October 2005, pp. 160–163.

[3] O. Amft and G. Tröster, "Methods for detection and classification of normal swallowing from muscle activation and sound," in *Proceedings of the First International Conference on Pervasive Computing Technologies for Healthcare.* ICST, November 2006.

[4] N. DeBelie, V. De Smedt, and D. B. J., "Principal component analysis of chewing sounds to detect differences in apple crispness," *Postharvest Biol Technol*, vol. 18, pp. 109–119, 2000.

[5] P. J. Lillford, "The materials science of eating and food breakdown," *MRS Bulletin*, vol. 25, no. 12, pp. 38–43, December 2000.

[6] J. P. Lund and A. Kolta, "Generation of the central masticatory pattern and its modification by sensory feedback." *Dysphagia*, vol. 21, no. 3, pp. 167–174, Jul 2006.

[7] K. D. Foster, A. Woda, and M. A. Peyron, "Effect of texture of plastic and elastic model foods on the parameters of mastication." *J Neurophysiol*, vol. 95, no. 6, pp. 3469–3479, Jun 2006.

[8] K. Kohyama and L. Mioche, "Chewing behavior observed at different stages of mastication for six foods studied by electromyography and jaw kinematics in young and elderly subjects," *J Tex Stud*, vol. 35, no. 4, pp. 395–416, October 2004.

[9] I. Ashida, H. Iwamori, S.-Y. Kawakami, Y. Miyaoka, and A. Murayama, "Analysis of physiological parameters of masseter muscle activity during chewing of agars in healthy young males," *J Tex Stud*, vol. 38, no. 1, pp. 87–99, 2007.

[10] B. Drake, "Food crushing sounds. an introductory study," *J Food Sci*, vol. 28, no. 2, pp. 233–241, March 1963.

[11] Z. Vickers, "Relationships of chewing sounds to judgements of crispness crunchiness and hardness," *J Food Sci*, vol. 47, pp. 121–124, 1981.

[12] W. AlChakra, K. Allaf, and A. Jemai, "Characterization of brittle food products: Application of the acoustical emission method," *J Tex Stud*, vol. 27, no. 3, pp. 327–348, July 1996.

[13] J. Edmister and Z. Vickers, "Instrumental acoustical measures of crispness in foods," *J Tex Stud*, vol. 16, no. 2, pp. 153–167, 1985.

[14] N. DeBelie, M. Sivertsvik, and J. DeBaerdemaeker, "Differences in chewing sounds of dry-crisp snacks by multivariate data analysis," *J Sound Vib*, vol. 266, no. 3, pp. 625–643, September 2003.

[15] C. Lassauzay, M. A. Peyron, E. Albuisson, E. Dransfield, and A. Woda, "Variability of the masticatory process during chewing of elastic model foods." *Eur J Oral Sci*, vol. 108, no. 6, pp. 484–492, Dec 2000.

[16] C. Dacremont, B. Colas, and F. Sauvageot, "Contribution of air- and bone-conduction to the creation of sounds perceived during sensory evaluation of foods," *J Tex Stud*, vol. 22, no. 4, pp. 443–456, January 1991.

[17] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Proceedings of the Conference on Parallel Problem Solving from Nature VI*, ser. Lecture Notes in Computer Science, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, Eds., no. 1917, Paris, France, 2000, pp. 849–858.

[18] S. De, A. Ghosh, and S. K. Pal, *Genetic Algorithms for Pattern Recognition*. Boca Raton, FL, USA: CRC Press, 1996, ch. 1, pp. 1–23.

[19] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm," ETH Zürich, Gloriastrasse 35, CH-8092 Zurich, Switzerland, Tech. Rep. 103, 2001.

[20] A. A. Freitas, "A critical review of multi-objective optimization in data mining: a position paper," *SIGKDD Explor Newsl*, vol. 6, no. 2, pp. 77–86, 2004.