Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 266 (2003) 625–643
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Differences in chewing sounds of dry-crisp snacks by multivariate data analysis

N. De Belie[a,*], M. Sivertsvik[b], J. De Baerdemaeker[c]

[a] *Magnel Laboratory for Concrete Research, Department of Structural Engineering, Ghent University, Technologiepark Zwijnaarde 904, B-9052 Ghent, Belgium*
[b] *NORCONSERV, Institute of Fish Processing and Preservation Technology, P.O. Box 327, 4002 Stavanger, Norway*
[c] *Laboratory of Agricultural Machinery and Processing, Department of Agro-Engineering and Agro-Economics, Catholic University Leuven, Kasteel Arenberg 30, B-3001 Leuven, Belgium*

Received 13 January 2003

Abstract

Chewing sounds of different types of dry-crisp snacks (two types of potato chips, prawn crackers, cornflakes and low calorie snacks from extruded starch) were analysed to assess differences in sound emission patterns. The emitted sounds were recorded by a microphone placed over the ear canal. The first bite and the first subsequent chew were selected from the time signal and a fast Fourier transformation provided the power spectra. Different multivariate analysis techniques were used for classification of the snack groups. This included principal component analysis (PCA) and unfold partial least-squares (PLS) algorithms, as well as multi-way techniques such as three-way PLS, three-way PCA (Tucker3), and parallel factor analysis (PARAFAC) on the first bite and subsequent chew. The models were evaluated by calculating the classification errors and the root mean square error of prediction (RMSEP) for independent validation sets.

It appeared that the logarithm of the power spectra obtained from the chewing sounds could be used successfully to distinguish the different snack groups. When different chewers were used, recalibration of the models was necessary. Multi-way models distinguished better between chewing sounds of different snack groups than PCA on bite or chew separately and than unfold PLS. From all three-way models applied, N-PLS with three components showed the best classification capabilities, resulting in classification errors of 14–18%. The major amount of incorrect classifications was due to one type of potato chips that had a very irregular shape, resulting in a wide variation of the emitted sounds.

*Corresponding author. Tel.: +32-9-264-5522; fax: +32-9-264-5845.
E-mail address: nele.debelie@ugent.be (N. De Belie).

0022-460X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0022-460X(03)00589-3

1. Introduction

Drake was the first researcher to study food crushing sounds in order to obtain more information on food texture [1]. He found that the breakdown of the structure of foodstuff during normal chewing was paralleled by a decline in the average amplitude of successive bursts of mastication sounds. Crisper products would also produce a louder noise. The team of Malcolm Bourne and Zata Vickers started to use food crushing sounds as a measure of crispness. Other than possessing a broad frequency range and very irregular amplitude versus time picture, there appeared to be no frequency characteristic unique to crisp sound [2]. The crisper samples showed higher sound amplitude and/or a greater density of sound occurrence. Most of the sound energy fell in the range of 0–10 kHz [3]. Although they first hypothesized that crispness was an auditory sensation [2], they later on suggested that the stimulus for crispness might be vibratory [4].

A cellular model was introduced to typify crisp structures: each sound burst corresponds to the rupture of a cell or group of cells [5]. The mechanism of sound production differs between wet-crisp and dry-crisp food. Wet-crisp food is comprised of living plant cells, having a certain internal pressure, the turgor. When a turgid cell bursts, its contents expand rapidly, producing a sound. In a dry-crisp product, the cells are usually filled with air and the cell walls are brittle. The sound pressure wave is generated by the snap back of the remainder of the cell walls after they have been bent and broken. Stiffer or less dense food would have a higher vibration frequency and produce a typical crisp higher pitched sound.

Lee et al. found significant differences in the sounds produced when eating fresh or stale potato and tortilla chips [6]. Sounds from fresh samples were typically louder and displayed greater amounts of higher frequency components. Other authors also found that crisp food generated high-pitched sounds [5,7,8]. However, in a later study of Lee et al. [9] there were no statistical differences in the sound levels, although they observed statistical differences in the sensory evaluations, between fresh and stale samples of potato and tortilla chips. A 'double hump' intensity versus frequency spectrum was associated with the first chew of the fresh samples, with the first hump maximum at 3–4 kHz and the second at about 6 kHz. Although some authors only consider the frequency range up to 3500 Hz [10], Lee et al. think frequencies up to 8000 Hz may provide useful information. Al Chakra et al. [11] recorded the sound emission of cylindrical pasta samples compressed with a jack. He filtered the signal in high pass at 100 Hz in order to eliminate low-frequency noises from origins different from the actual rupture sounds and low pass at 20 kHz (audible domain). However, all the information characterizing the signals was located below 12 kHz.

When studying chewing sounds, it is important to know how characteristics of jaw, teeth and soft tissues in the mouth influence the perceived sounds. Kapur [12] found that the resonance frequency of the jaw was about 160 Hz. The bone conducted sound travelling through teeth and jaws to the ear is therefore amplified at this frequency. Drake had also situated the mouth opening sounds in a broad region around 160 Hz [1]. On the other hand the soft tissues in the mouth tend to absorb or damp especially the higher frequencies of the sound [13]. Dacremont et al. [14] studied the contribution of air and bone conduction to the transmission of chewing noises to the inner ear. Each panelist had to reconstitute sounds he truly heard during eating, by mixing the air and bone conduction records together. It appeared that for a bite, bone and air conduction had a similar importance, whereas for a chew bone conduction had to be more attenuated to match the

perceived noise. Both bone and air conduction records had to be attenuated around 160 Hz and air conduction records had to be amplified around 3500 Hz.

Different parameters have been used with varying success to judge a chewing sound. This includes the number of sound bursts n in a bite or chew, the mean height or amplitude of the bursts A, and the product $n.A$ and $n.A/duration$ [15]; the frequency spectrum by fast Fourier transformation (FFT) followed by data reduction and averaging of different spectra [6,11]; frequency contents in a combination of different frequency bands [16]; mean sound pressure, pressure level and intensity in different frequency ranges [10]; fractal analysis [16–19]. Vickers mentions that the number of sound bursts provided most of the information useful for indicating crackliness [15]. Edminster and Vickers found that $\log n.A$ showed the largest correlation with auditory crispness judgements [20]. Vickers [21] and Seymour and Hamann [10] stated that the combination of mechanical and acoustic parameters provided the best correlation with sensory crispness of dry-crisp products. Duizer et al. [19] collected acoustic recordings of the sound produced during biting an extruded snack food. According to this research, sensory characteristics of crispness, pitch and crumbliness were significantly correlated to the fractal dimensions of the amplitude–time data. However, in research on Royal Gala apples De Belie et al. [16] found that changes in fractal dimension could not be correlated with the sensory crispness.

An alternative way of analyzing the dataset is through exploration using advanced multivariate techniques (chemometrics). The general objectives of these techniques are data reduction and easier interpretation [22]. They are appropriate to extract information from extensive data sets (e.g. a 22050 point FFT on a sound of 1 s registered at a sampling rate of 22050 Hz results in a power spectrum with 11025 frequency components), showing a high degree of covariance. De Belie et al. [23] have shown that application of principal component analysis (PCA) on the logarithm of the power spectra allowed them to distinguish groups of Cox's Orange Pippin apples, which showed crispness differences due to storage under different conditions.

In the current research chewing sounds of different types of dry crisp snacks (potato chips, prawn crackers, cornflakes and low calorie snacks of extruded starch) were investigated with PCA and with more advanced multi-way techniques including N-way partial least squares (N-PLS), multi-way PCA (Tucker3), and PARAFAC (parallel factor analysis) on the first bite and subsequent chews.

2. Materials and methods

2.1. Dry-crisp products

Chewing sounds of different types of commercially obtained dry-crisp snacks (two types of potato chips, prawn crackers, cornflakes and low calorie snacks from extruded starch) were recorded to assess differences in sound emission patterns. The first type of potato chips (type A), the prawn crackers and the extruded snacks were regularly shaped and they could be placed between the incisors in a standard way for recording of the first bite. The second type of potato chips (type B) was irregularly shaped and large chips with an average mass of 1.30 g (range 1.06–1.48) were selected for the experiment. Potato chips, prawn crackers and extruded snacks were bitten in half with the incisors and than chewed with the molars. The bite and different chews

were recorded at a chewing rate of one chew per second. This chewing rate could be easily controlled with a stopwatch and was also used by Lee et al. [9]. For the cornflakes, 1.7–1.8 g of material was placed in the mouth and chewed immediately. Two researchers each tested 20 samples of each of the different snacks, except for the cornflakes, which were only chewed by the first researcher and the extruded snacks, which were only chewed by the second researcher.

2.2. Sound recording

The chewing sounds were recorded by placing a microphone (ECM-2005, Monacor) over the ear canal at the side where the snacks would be chewed. The sampling rate amounted to 22050 samples per second and sampling size was 16 bits. An example of a time spectrum for potato chips type A is shown in Fig. 1. Different sound bursts can be discerned in the bite, which is typical for a dry crisp product. The first bite and the first subsequent chew were selected from the signal in the Creative Wave Studio programme (Creative Technology Ltd.). The selected sound waves were imported in a program written using Matlab® and a 22050 ($= N$) point FFT was performed. This implies that for sound waves shorter than 22050 samples (1 s), the wave was padded with trailing zeros to length 22050. The frequency resolution was therefore 1 Hz for all sound waves. The power spectrum was obtained by squaring the magnitude of the first 11025 ($= N/2$) frequency components. As the Fourier transformation of a real signal is symmetric, the power at a positive frequency is the same as the power at the corresponding negative frequency (contained in the second half of the original Fourier spectrum). To eliminate the DC component and low-frequency noises from the signal, only the frequencies larger than 100 Hz were considered for further processing.

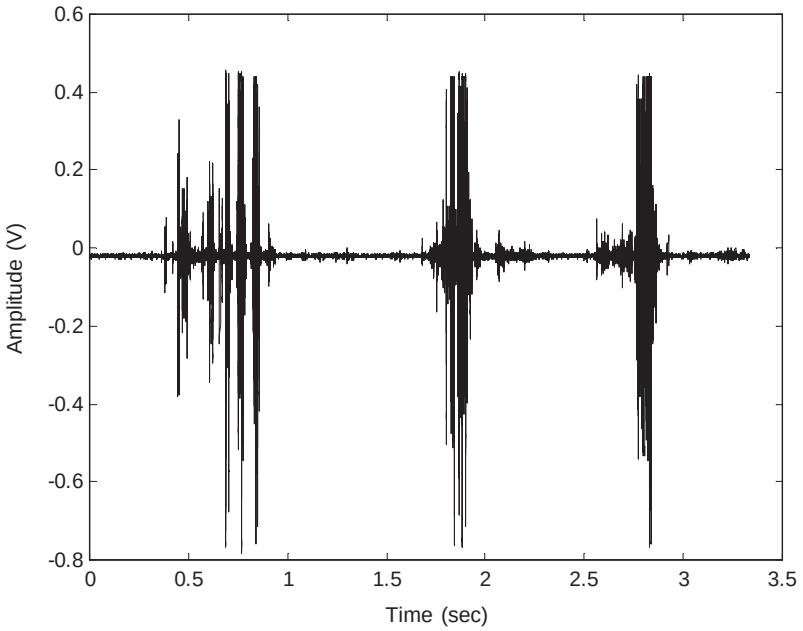

Fig. 1. Time signal for potato chips type A.

2.3. Principal component analysis

The power spectra of the different bites and chews were analyzed with different multivariate analysis techniques, in order to evaluate if clear differences in chewing sounds exist between different snack groups. First, PCA was carried out on the pure power spectra of bite and chew separately, as well as on their logarithm. Algebraically, principal components (PCs) are linear combinations of the original variables (e.g., the frequency values) so that the largest part of the variance is concentrated on the first PCs. These first few PCs can then be used to replace the original variables, leading to data reduction and easier interpretation without much loss of information. The spectra were split into a training set and a validation set, one-fourth of the spectra being located in the validation set. Both training and validation set contained equal numbers of spectra from the different snacks groups. All spectra were mean-centered by subtraction of the average spectrum of the training set and normalized to compensate for possible differences in recording volume, for instance due to slightly different positions of the microphone in the ear during different recordings. Therefore, differences in total sound level will not take part in the PCA, and only relative differences between sound levels at different frequencies will have an influence. PCA was carried out on the mean-centered and normalized training set data. Scree plots were made as a useful visual aid to determine the appropriate number of principal components. With the eigenvalues ordered from largest to smallest, a scree plot is a plot of the magnitude of an eigenvalue (= the variance of the PC) versus its number. The number of components can be taken to be the point at which a distinct bend in the scree plot occurs, such that the remaining eigenvalues are then relatively small and all about the same size. Both training and validation set spectra were then expressed in terms of those principal components. The nominal variable referring to the groups was for this analysis to be replaced by a binary coding system. If it is known for the snacks in the training set to which 'group' they belong, a calibration matrix can be generated, based on the PCs. The calibration matrix provides the best possible prediction of group belonging or crispness value for the data in the training set. The existence of an independent validation set then permits the performance of a calibration to be evaluated.

2.4. Multi-way data analysis

The data under consideration can be regarded as three-way data, with the sample number as the first mode, the frequency as second mode, and the number of bite or chew in one chewing procedure as third mode. Therefore, the chewing sounds of the second researcher (of whom bite and chew were available for four different snack groups) were also explored with the multi-way methods PARAFAC, Tucker3, and N-PLS. Kiers [24] shows that PARAFAC can be considered as a constrained version of Tucker3, and Tucker3 a constrained version of two-way PCA. Any data that can be modelled adequately by PARAFAC can thus also be modelled by Tucker3 or two-way PCA, but PARAFAC uses the fewest degrees of freedom [25]. In the sense that it uses most degrees of freedom the PCA model can be considered the most complex and flexible model, while PARAFAC is the most simple and restricted model. The reason for using multi-way models is not to obtain better fit, but rather more adequate, robust and interpretable models. A characteristic feature of PARAFAC is the uniqueness of the solution, i.e., there is no problem of rotational freedom. This means that, if the data is indeed tri-linear, pure spectra can be recovered.

The advantage of Tucker3 is that different numbers of components can be retained in each mode [26]. N-PLS in turn is an extension of PLS to multi-way data. The most important feature of PLS and N-PLS is that the decomposition is accomplished such that the successively computed score vectors have the property of maximum covariance with the unexplained part of the dependent variable [27], which is in this case the snack type. Multi-way data analysis was performed using the N-way toolbox for Matlab [28].

Because it resulted from the PCA, and also from former analysis on chewing sound spectra [23], that a better classification was obtained when working with the logarithm of the power spectra, the multi-way analysis was performed on the logarithmic data. It is also known that human hearing can be characterized by a logarithmic scale.

It appeared that the multi-way algorithms could not handle the original spectra with 10 926 frequency components for both bite and chew. Therefore, the spectra were smoothed with a Savitsky–Golay smoothing algorithm applying a second order polynomial and a filter width of 11 points. Then data reduction was carried out by selecting one out of every 10 values. The reduced data set had dimensions $80 \times 1093 \times 2$, the three modes representing the 80 samples (20 replicates $\times 4$ snack groups), 1093 frequencies (from 101 to 11025 Hz in steps of 10 Hz) and bite/chew. PARAFAC was carried out on the complete data set to track outliers, which were deleted. To allow evaluation of the classification capabilities of the different applied models, the data were split into a training set containing 12 samples from each snack group (48 samples in total), and a validation set containing seven samples from each group (28 in total). The calculations were made with two different selected training sets. The training set was used to construct the best possible regression between the response variable (binary coded snack groups) and the loadings of the first mode (samples). This regression was afterwards used to calculate the loadings for the samples (scores) in the validation set and predict their group belonging. The data were centered across the first mode. Centering was performed on the training set and the calculated mean values were used to preprocess the validation set.

PARAFAC models with one to eight components were constructed. In fact only two real trilinear components can be extracted from the data, since the third mode of the data array equals two. However, more (non-trilinear) components can be used to explain additional variation in the data. To evaluate the validity of the model as a function of the number of components, the core consistency, sum of squared residuals (SSR), explained variation and Tucker congruence coefficient were evaluated [28]. Three-way and unfold PLS with one to 10 components were calculated. For Tucker3 the explained variation as a function of the model dimensionality was investigated for all possible combinations of one to 10 components in mode 1 (samples) and mode 2 (frequency) and one to two components in mode 3 (chew and bite). All models were used to predict validation set classes. Also the root mean square error of prediction (RMSEP) was calculated as the sum of squares of deviations from 1 in the correct class and from 0 in the other classes. The multi-way models were compared with unfold PLS and principal component regression (PCR) on the same training and validation sets.

3. Results and discussion

On a linear scale, humps could be seen in the power spectra around 200 Hz and, especially for the bites, around 900 Hz. Only the frequencies below 2000 Hz were of any importance. However,

since the sensation of the human ear is proportional to the logarithm of the intensity, higher frequencies can also be important for crispness evaluation. On a logarithmic scale for the first chewer a minimum occurred between 5000 and 6000 Hz, followed by a maximum at about 7000 Hz. For the second chewer the minima and maxima were located between 6000 and 7000 Hz and around 8200 Hz, respectively. In previous experiments on apples in which researcher one acted as chewer [23], high amounts of frequency components between 700 and 900 Hz were partly attributed to general chewing movements and teeth contact. The presence of a minimum in the spectra around 3000 Hz, which was noticed in the previous experiments and attributed to the damping properties of the soft tissues in the mouth [13,14,16,23], was less obvious in the current tests.

On average for the bite, on a linear scale the power spectra of the chips type A showed, compared to the other snacks, a lower amount of frequency components below 500 Hz and a higher amount of components above this value, resulting in a higher pitched sound. On a logarithmic scale (for the higher frequencies), both types of potato chips showed the same typical spectrum, with a pronounced hump around 7000 or 8200 Hz depending on the chewer. This hump was on average less pronounced for prawn crackers and extruded snacks. For the chew, the amount of frequency components below 500 Hz decreased from cornflakes, extruded snacks and prawn crackers over chips type B to chips type A, while the amount of frequency components above 500 Hz decreased from cornflakes and chips type A over chips type B and extruded snacks to prawn crackers. The mean bite power spectra of researcher two are shown in Fig. 2.

3.1. Principal component analysis

The first 15 principal components calculated from the power spectra explained together about 65% of the total training set variance for chewer one. Using those 15 PCs implied a large data reduction from the original 11 025 variables. A distinct bend occurred in the scree plot at the second PC, indicating that PC1 has a major influence on the total population variance, explaining about 19%. For the chew, the first 15 PCs and PC1 alone explained 65% and 23% of the variance, respectively. Similar results were found for chewer two. In Fig. 3 the first two PC values are plotted for the different bite power spectra of chewer 2. It appeared that prawn crackers and extruded snacks could be relatively well separated from the chips, especially when the logarithm of the power spectra was used in the PCA. However, the two types of chips were difficult to distinguish from each other. Prediction of group belonging, using a calibration matrix based on the first 15 PCs, gave also much better results for the logarithmic data, than when the original signals were used (Table 1) and for both researchers more information was obtained from the bite than from the chew sound wave. Using the logarithm of researcher one's bite spectra, chips type A and prawn crackers were all correctly classified, but three out of five type B chips in the validation set were incorrectly attributed to the type A chips group. Application of the logarithm of researcher two's bite spectra resulted in correct classification of all type A chips and extruded snacks, while one prawn cracker and again three type B chips were incorrectly classified. The irregular shape of the type B chips apparently resulted in a wide variation of emitted sounds, which made it very difficult to separate them from acoustic emissions of other snacks. It was clear from comparison of the raw spectra that different people produced different chewing sounds, and

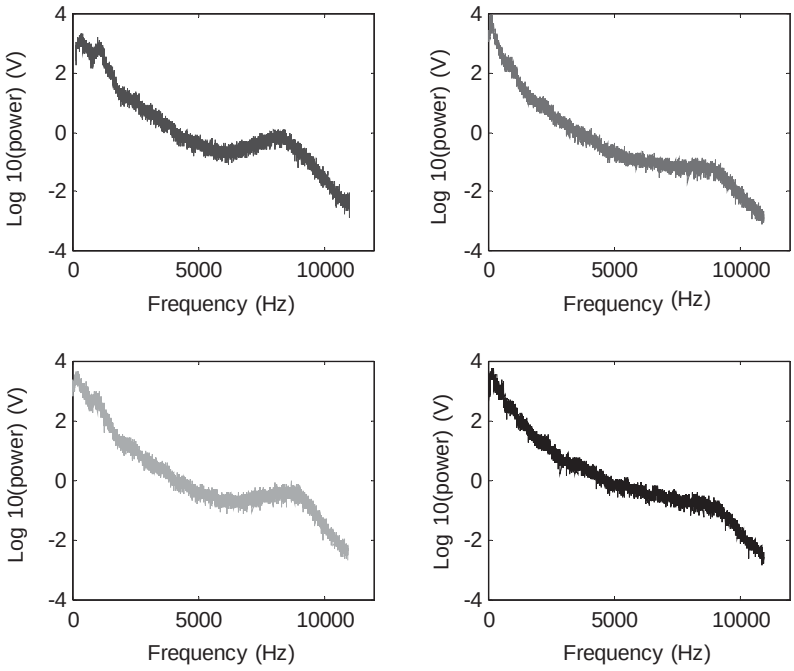

Fig. 2. Average logarithm of the bite power spectra of researcher two for chips type A (top left), prawn crackers (top right), chips type B (bottom left), extruded snacks (bottom right).

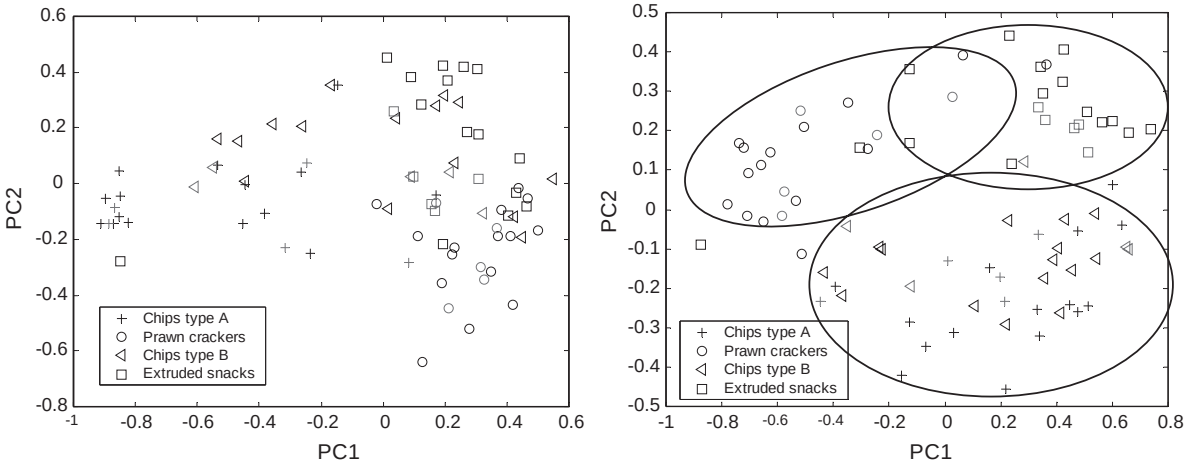

Fig. 3. Second versus first PC for the bite power spectra of researcher two on linear (left) and logarithmic (right) scale for samples from training and validation set; clustering of chips, prawn crackers and extruded snacks is indicated.

that the system will have to be trained separately for all possible 'chewers'. However, once the model is calibrated for another person, his chewing sounds can be used successfully for evaluation of dry-crisp snacks.

Table 1
Classification errors for dry-crisp snacks based on PCA of chewing sounds for researcher one: number of sound waves incorrectly attributed to another snack group on a total of 15 training (T) and five validation (V) samples per snack group

	Chips A	Chips B	Prawn crackers	Cornflakes
Bite spectra				
T	2	7	3	
V	0	3	2	
Chew spectra				
T	1	3	2	1
V	0	4	3	1
Log bite spectra				
T	0	1	0	
V	0	3	0	
Log chew spectra				
T	3	7	0	0
V	0	4	0	1

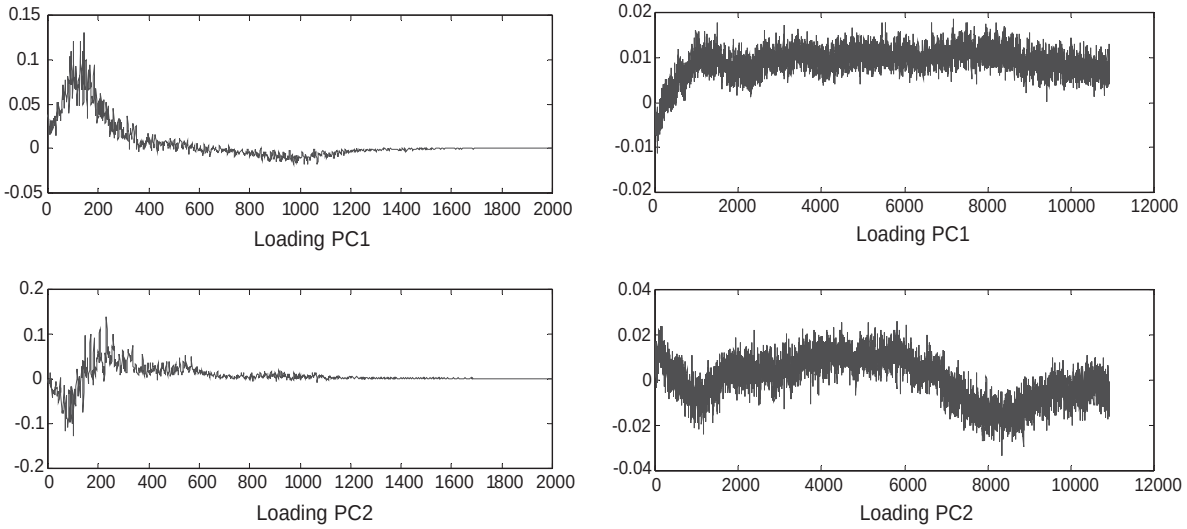

Fig. 4. Loadings of the first two principal components for researcher two, calculated on the pure power spectra (left) and on their logarithm (right).

From the PC-loadings, Fig. 4, it can be seen that especially the frequencies between 100 and 500 Hz (with a peak for PC1 around 200 Hz) contributed to PC1 and PC2 derived from the power spectra. This could be expected from the peaks in the power spectra, which are more or less located at those frequencies. The PC-loadings derived from the logarithm of the power spectra were less clear indicating the importance of certain frequencies and nearly the whole spectrum was taken into account.

3.2. PARAFAC

A PARAFAC model with three components on the complete non-centered data set showed that there were two outliers: samples 31 and 63, Fig. 5. Plotting of the spectra confirmed that sample 63 was an outlier for the bite and sample 31 for the chew. These outliers were deleted. Also one sample from each of the two other snack groups was deleted (to allow easier data processing). A new PARAFAC with three components was carried out on the remaining 76 samples (non-centered). The core consistency equalled 100% and the model was able to explain 91.6% of the overall variation with a sum of squared residuals (SSR) of 31 438. The loadings are shown in Fig. 6. The first two components seem to show a high correlation, especially in the first mode. However, calculation of Tuckers congruence coefficient did not point at any real problems, indicating that this is a valid model. In the frequency mode these components look like actual bite/chew spectra, while the third component may represent deviations from these spectra for the different snacks. A four-component PARAFAC model explained 92.04% of the total variance. It therefore gives only little improvement compared to a three-component model, taking also the increased complexity of the model into account. Furthermore, Tuckers congruence coefficient showed that some factors in this model were highly correlated.

To judge the classification capabilities of PARAFAC, models with one to eight components were calculated on the centered training set data. Different parameters giving an indication of the value of the model are presented in Table 2. It appeared that a model with three components was not to be preferred in this case because of the low core consistency and values in the Tucker congruence coefficient that were close to –1. However, when the number of components increased to four or five, these problems seemed to be eliminated and the explained variation of

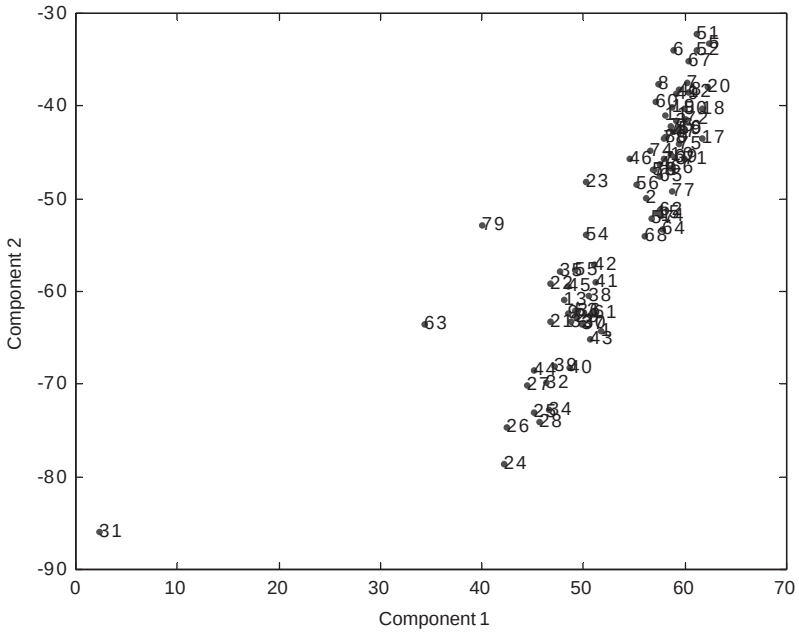

Fig. 5. A PARAFAC with three components on the complete data set revealed that samples 31 and 63 were outliers.

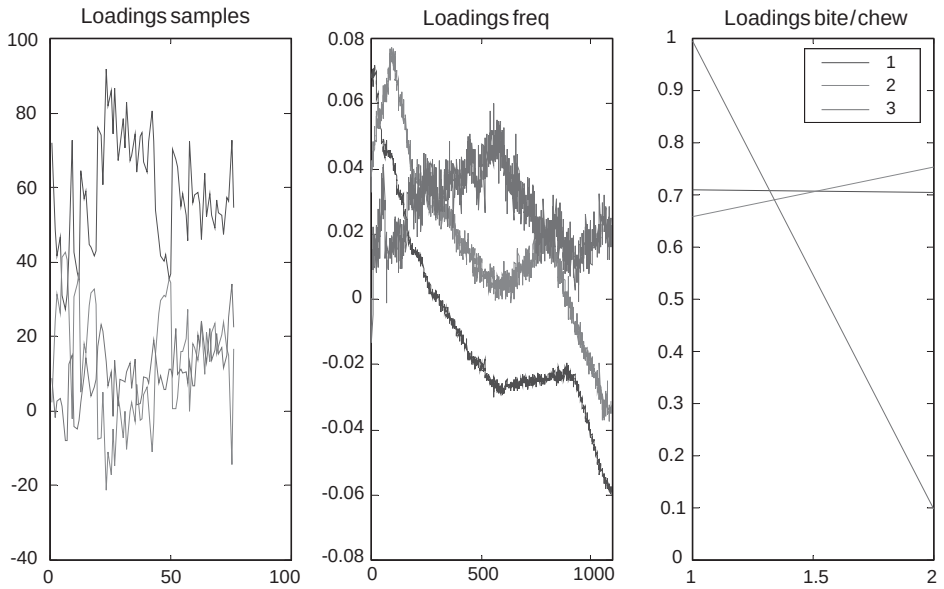

Fig. 6. Loadings of a PARAFAC model with three components (non-centered data).

Table 2
Core consistency (Corco), sum of squares of residuals (SSR), explained variation (Explv) and TCC, which is the value in the Tucker congruence coefficient that is closest to –1 (values close to –1 indicate degenerate solutions)

	Number of PARAFAC components							
	1	2	3	4	5	6	7	8
Corco (%)	100	100	31	97	98	25.21	86.76	38.32
SSR	21420	19504	18232	17636	17062	16615	16230	15898
Explv (%)	28.8	35.2	39.4	41.4	43.3	44.8	46.1	47.2
TCC	1.00	–0.30	–0.93	–0.81	–0.59	–1.00	–0.90	–0.93

four or five-component models were respectively, 6.2% and 8.1% higher than that of the two-component model.

It could also be seen that by centering the data the variance explained by a three-component PARAFAC model is reduced from 91.6% to 39.4%. This is due to the offset in the non-centered data, accounting for a major part of the variation. This offset can be easily explained by one of the PARAFAC components, thereby increasing the explained variation. It should be taken into account that the total variance of the raw non-centered data equaled 381840, while the total variance of the centered data just equaled 30098. This implied that the residual variance left in the three-component PARAFAC model of the non-centered data (SSR = 31438) was actually larger than the total variance in the centered data, which emphasizes the need for centering.

Table 3 shows the classification errors amongst others for the different PARAFAC models. The PARAFAC model with five components clearly provided the best prediction with only 14% of the validation set samples classified in the wrong snack group. For the training set (not shown) the

Table 3
Prediction of 28 validation samples by different multivariate models in function of the number of model components, showing the number of incorrect classifications per snack type, the total classification error (Err) and the mean RMSEP

No. of components	No. of incorrect classifications per snack type				Err (%)	Mean RMSEP
	Chips A	Prawn crackers	Chips B	Extruded snacks		
PCA bite						
1	7	0	7	0	50	0.4660
2	2	0	7	0	32	0.4172
3	0	0	4	0	**14**	0.3863
4	1	0	5	0	21	0.3939
5	1	0	4	0	18	0.3951
15	1	0	4	0	18	0.3926
PCA chew						
1	2	0	7	7	57	0.4819
2	2	1	7	0	36	0.4484
3	2	0	7	0	32	0.4210
4	2	0	7	0	32	0.4178
5	2	0	7	0	32	0.4141
15	2	0	7	0	32	0.4138
PARAFAC						
1	2	0	7	7	57	0.4752
2	3	0	7	0	36	0.4493
3	1	0	7	0	29	0.4027
4	2	0	6	0	29	0.4099
5	0	0	4	0	**14**	0.3839
6	0	0	4	0	14	0.3813
7	1	0	3	0	14	0.3834
8	1	0	4	0	18	0.3875
Tucker3						
2 2 1	2	1	7	0	36	0.4252
2 2 2	3	0	7	0	36	0.4493
3 3 1	1	0	4	0	**18**	0.3885
3 3 2	1	0	7	0	29	0.4029
4 4 1	2	0	4	0	21	0.3890
4 4 2	1	0	6	0	25	0.3998
5 5 1	2	0	4	0	21	0.3877
10 10 1	2	0	3	0	18	0.3897
10 10 2	1	0	4	0	18	0.3916
15 15 2	1	0	4	0	18	0.3881
N-way PLS						
1	2	0	7	7	57	0.4759
2	1	0	7	0	29	0.4221
3	1	0	3	0	**14**	0.3878
4	0	0	4	0	14	0.3820
5	2	0	4	0	21	0.3854
6	2	0	4	0	21	0.3878

Table 3 (*continued*)

No. of components	No. of incorrect classifications per snack type				Err (%)	Mean RMSEP
	Chips A	Prawn crackers	Chips B	Extruded snacks		
7	2	0	3	0	18	0.3883
8	1	0	3	0	14	0.3854
9	2	0	3	0	18	0.3878
10	2	0	3	0	18	0.3878
Unfold PLS						
1	2	0	7	7	57	0.4746
2	2	0	7	0	32	0.4199
3	2	0	4	0	21	0.3923
4	1	0	4	0	**18**	0.3848
5	1	0	5	0	21	0.3850
6	1	0	5	0	21	0.3862
7	1	0	5	0	21	0.3865
8	2	0	5	0	25	0.3858
9	2	0	5	0	25	0.3852
10	1	0	5	0	21	0.3856

For each type of model the preferred number of components is indicated in bold.

classification error amounted to 8% for the five-component model and was further reduced for models with six, seven or eight components to 4%, 4% and 2%, respectively. However, increasing the number of components did not improve the classification error or RMSEP for the validation set, showing that the extra components only describe training set noise.

A plot of the residuals of the five-component model showed that there was no obvious structure remaining. For the validation set the error was entirely due to misclassification of the potato chips type B: three type B chips were classified as type A chips and one as extruded snack. As stated higher the irregular shape of this snack, resulting in a high variation between spectra, was mainly responsible for the difficult classification. The RMSEP pointed in the same direction. Although the RMSEP of type B chips was the same as for the other snacks for the training set (0.37), it was significantly higher for the validation set (0.48 for chips B compared to 0.36, 0.37 and 0.33 for chips A, prawn crackers and extruded snacks, respectively). It can also be remarked that the classification of the validation set samples obtained with the 'degenerate' three-component model was as good as a for a four-component model.

In comparison to the PCA on the first bite only, no improvement of the prediction was obtained by using the PARAFAC model. This would indicate that little extra information is comprised within the first chew.

3.3. Tucker3

A steady increase in explained variation in function of the model dimensionality was observed, with no clear steps indicating large improvements in explained variation. To begin with, Tucker3

models with simple core dimensions such as (2 2 1), (3 3 1) and (2 2 2) were chosen for further investigation. The (2 2 1) Tucker3 model explained 32.3% of the variation, while the (3 3 1) and (2 2 2) Tucker3 models explained 34.3% and 35.2%. For the models with only one component in the third mode the core was diagonal and no rotation was necessary. The (2 2 2) model core was rotated to obtain maximum diagonality, but this did not result in major changes of the core, since one entry (1 1 1) in the core was much more important than all the other entries (Table 4). The performance of these Tucker3 models for snack classification can be judged from Table 3. The RMSE for prediction of the snack type for the training samples was higher for the (2 2 2) Tucker3 model (chips A: 0.41, prawn crackers: 0.40, chips B: 0.50, extruded snacks: 0.44) compared to the (3 3 1) Tucker3 model (chips A: 0.31, prawn crackers: 0.39, chips B: 0.38, extruded snacks: 0.39). A similar trend can be seen for prediction of the 28 validation samples with RMSEP = 0.45, 0.41, 0.51, 0.42 in the (2 2 2)-model and RMSEP = 0.37, 0.39, 0.46, 0.34 in the (3 3 1)-model for the four snack types, respectively. This indicated a better prediction for the (3 3 1)-model compared to the (2 2 2)-model, which was also confirmed by the classification errors of the different snack types (Table 3). Again the high RMSEP of chips B in the validation set draws the attention, as well as the high amount of misclassifications of this snack. From the score plot of the (3 3 1)-model (Fig. 7) the large overlap of sound spectra from the two chips groups can also be seen.

We also wanted to see if higher order Tucker3 models would give better classification of the validation set. Therefore Tucker3 models with (4 4 1), (5 5 1), (4 4 2), (10 10 1), (10 10 2) and (15 15 2) components were constructed. The explained variance increased with model dimensionality (Table 5). The significance of the (1 1 1) entry of the core decreased with increasing dimensionality, but even for the (15 15 2) model the (1 1 1) entry of the core was more than 6 times larger than the second most important entry. The core for these higher dimensionalities was not rotated. The RMSEP for the training set decreased with increasing dimensionality. The RMSEP for the validation set was more or less constant for all models (0.39–0.40) and comparable with the value obtained for the (3 3 1) model (Table 3). The utilization of the higher dimensionality models did not improve the classification of the validation samples compared to the (3 3 1) model. The (10 10 1), (10 10 2) and (15 15 2) models resulted in an equal amount of incorrect classifications as the (3 3 1) model (18%). Of all Tucker3 models a (3 3 1)-model should therefore be preferred as being the simplest model providing the best classification (Table 3). However, it should be noted that overall the best Tucker3 model gave results somewhat inferior to the five-component PARAFAC model.

Table 4
The four most important elements in the core for (2 2 1), (3 3 1) and (2 2 2) Tucker3 models and for a (2 2 2) model with the core rotated to obtain maximum diagonality, with their explained variation (Explv)

(2 2 1) core		(3 3 1) core		(2 2 2) core		(2 2 2) maxdia core	
Core entry	Explv (%)	Core entry	Explv (%)	Core entry	Explv (%)	Core entry	Explv (%)
(1 1 1)	89.2	(1 1 1)	84.0	(1 1 1)	81.0	(1 1 1)	80.9
(2 2 1)	10.8	(2 2 1)	10.2	(2 1 2)	12.8	(2 1 2)	10.6
(2 1 1)	0	(3 3 1)	5.7	(2 2 1)	2.9	(2 2 2)	5.2
(1 2 1)	0	(3 1 1)	0	(2 2 2)	2.4	(2 2 1)	2.4

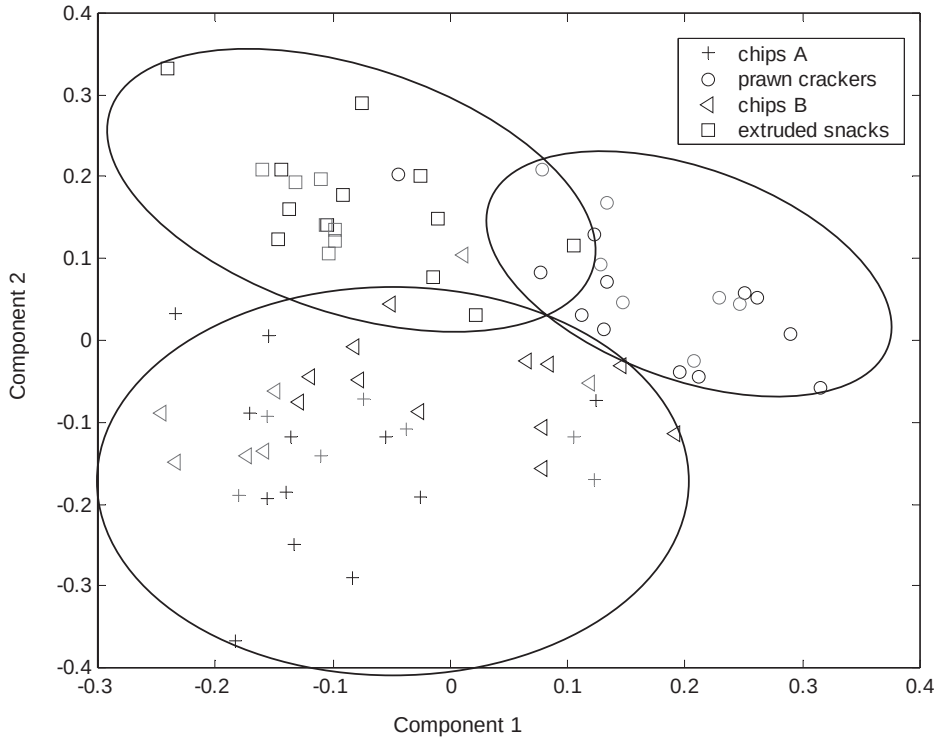

Fig. 7. Score plot of component one versus two for the training and validation samples, resulting from a (3 3 1) Tucker3 model; grouping of chips, prawn crackers and extruded snacks is indicated.

Table 5
The variation explained by Tucker3 models with higher dimensionality (explv) and the three most important elements in the core with their explained variation

Model dimensions	(4 4 1)		(5 5 1)		(4 4 2)		(10 10 1)		(10 10 2)		(15 15 2)	
Explv (%)	35.8		36.8		41.4		41.2		49.4		54.4	
Core entries with explv (%)	(1 1 1)	80.6	(1 1 1)	78.4	(1 1 1)	68.8	(1 1 1)	69.9	(1 1 1)	57.8	(1 1 1)	52.5
	(2 2 1)	9.8	(2 2 1)	9.5	(2 1 2)	9.4	(2 2 1)	8.4	(2 1 2)	8.3	(2 1 2)	7.6
	(3 3 1)	5.5	(3 3 1)	5.3	(3 2 1)	4.4	(3 3 1)	4.8	(3 2 1)	3.8	(3 2 1)	3.5

3.4. N-PLS and unfold PLS

The three-dimensional data array was also modelled with N-PLS (three-way PLS in this case). Models with one to 10 components were compared for modelling centered as well as non-centered data. The percentage variation of the measured and response variables (spectra and snack groups) captured by the models is shown in Table 6. As for PARAFAC, it was expected that the explained variation would decrease after centering. For the centered data, the explained variation of the response variable increased quite a lot up to five components, to approach asymptotically to a value of about 75% for more than five components. For the training set the amount of incorrectly

Table 6
Percentage variation captured by an N-PLS model on non-centered and centered data

	Non-centered		Centered	
Number of components	Measured variable	Response variable	Measured variable	Response variable
1	86.67	24.95	28.10	14.95
2	90.10	40.11	32.25	34.45
3	90.73	59.34	34.34	54.21
4	91.00	79.17	37.57	57.99
5	91.77	83.02	38.77	68.21
6	91.91	92.70	39.58	72.00
7	92.02	96.37	40.65	73.34
8	92.18	97.81	41.51	74.23
9	92.34	99.08	42.72	74.55
10	92.46	99.56	43.74	74.78

classified samples decreased steadily with an increasing number of components, to reach 0% for a model with five or more components. The RMSEP for the training set was similar for all snack groups. It evolved from 0.46 for a one-component model to 0.28 for a five-component model and then approached a value of 0.25 for a 10-component model. For the validation set the average RMSEP stabilized around 0.38–0.39 for models with three or more components (Table 3). Considering the amount of incorrect classifications a three-component model should be preferred as being the simplest model with a minimum number of classification errors. The errors were again mainly due to misclassification of the chips type B. The four-component N-PLS gave in fact exactly the same errors as the five-component PARAFAC model, the same chips B being classified as chips A and extruded snacks. On the other hand, these two models were able to completely distinguish all chips type A, prawn crackers and extruded snacks from each other. In Fig. 8 the components in the first mode of a three-component N-PLS model are plotted against each other. Especially in the plot of component three versus component two, clustering of the different snack groups can be seen.

Also unfold PLS models with one to 10 components were created and the classification results were compared with those from N-way PLS. It appeared that the training samples were classified in exactly the same way by the two models, with only very minor differences in the RMSEP. The number of classification errors for the validation set was always higher or the same for unfold PLS than for N-PLS (Table 3), the best unfold PLS model being the one with four components (18% classification error).

3.5. *Change of training and validation set*

The same models were constructed for a different choice of training and validation set. The results will not be discussed here in detail, but the classification errors and RMSEP of the models with optimum number of components are summarized in Table 7. Except for the PCA on the first chew only, the same number of components as previously was selected for all models, which points out the consistency of the optimum number. In comparison with the first choice of training/validation set higher classification errors were obtained with all models. However, in this

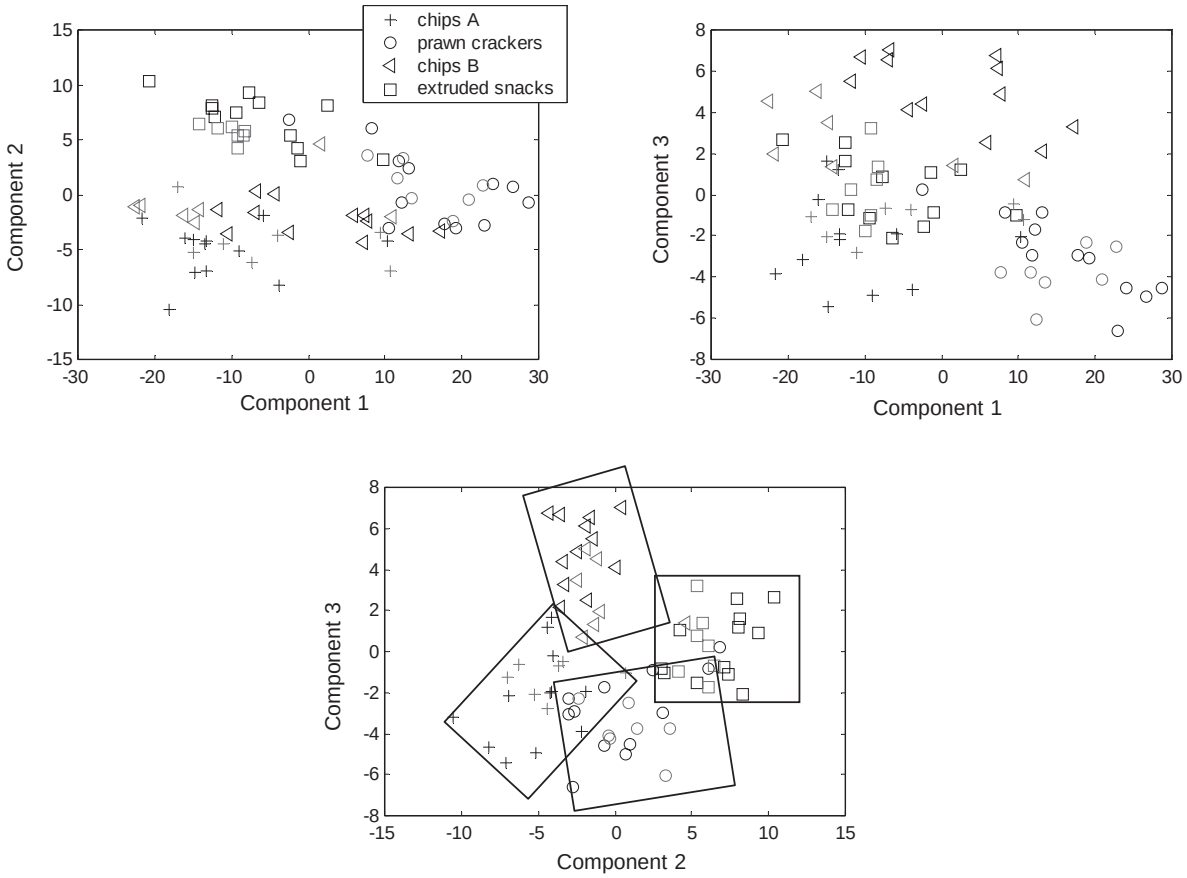

Fig. 8. Score plot of a three-component N-PLS model for the training and validation samples; clustering of the four snack groups is shown.

Table 7
Prediction of 28 validation samples by different multivariate models with optimum number of components for the second choice of training/validation set, showing the number of incorrect classifications of the different snack types, the total classification error and the mean RMSEP

Model	Number of components	Chips A	Prawn crackers	Chips B	Extruded snacks	Classification error (%)	Mean RMSEP
PCA bite	3	1	2	4	3	36	0.4152
PCA chew	4	2	0	6	2	36	0.4255
PARAFAC	5	0	2	3	2	25	0.3995
Tucker3	(3 3 1)	1	1	3	1	21	0.3972
N-way PLS	3	1	1	2	1	18	0.3898
Unfold PLS	3	1	1	3	2	25	0.3986

case it becomes clear that the multi-way models distinguish better between chewing sounds of different snack groups, than PCA on bite or chew separately and than unfold PLS, which combines information from bite and chew, but not in a three-way structure. From all models, N-PLS showed the best classification capabilities, with a classification error of 18%. This could be expected since the algorithm produces score vectors that in a tri-linear sense have maximum covariance with the unexplained part of the dependent variable [27].

4. Conclusions

The sounds emitted during chewing of dry-crisp snacks could be successfully applied to distinguish different snack groups, such as potato chips, prawn crackers, corn flakes, snacks from extruded starch, using FFT and multivariate data analysis techniques. The classification was improved by taking the logarithm of the power spectra for further analysis. Different people produced different sound spectra, which makes recalibration of the model necessary when a new chewer is used as 'measuring instrument'. A procedure for transfer of calibrations would therefore be helpful. Multi-way models distinguished better between chewing sounds of different snack groups than PCA on bite or chew separately and than unfold PLS. From all three-way models applied, N-PLS with three components showed the best classification capabilities, resulting in classification errors of 14–18%. The major amount of incorrect classifications was due to one type of potato chips that had a very irregular shape, resulting in a wide variation of the emitted sounds.

Acknowledgements

Nele De Belie carried out this research project as a Postdoctoral Fellow of the Fund for Scientific Research—Flanders (FWO). The multi-way data analysis was performed during a 6 month stay with the Department of Food Technology of the Royal Veterinary and Agricultural University (KVL) at Copenhagen, Denmark. She would like to thank Prof. Lars Munck for providing the opportunity to work within his group. The expert advice of Prof. Rasmus Bro and Claus Andersson in dealing with multi-way algorithms is gratefully acknowledged.

References

[1] B.K. Drake, Food crushing sounds, An introductory study, *Journal of Food Science* 28 (1963) 233–241.
[2] Z.M. Vickers, M.C. Bourne, A psychacoustical theory of crispness, *Journal of Food Science* 41 (1976) 1158–1164.
[3] Z. Vickers, Crispness of cereals, *Advances in Cereal Science and Technology* 9 (1986) 1–19.
[4] C.M. Christensen, Z.M. Vickers, Relationships of chewing sounds to judgments of food crispness, *Journal of Food Science* 46 (1981) 574.
[5] Z. Vickers, Crispness and crunchiness—a difference in pitch?, *Journal of Texture Studies* 15 (1984) 157–163.
[6] W.E. Lee III, A.E. Deibel, C.T. Glembin, E.G. Munday, Analysis of food crushing sounds during mastication: frequency–time studies, *Journal of Texture Studies* 19 (1988) 27–38.
[7] Z. Vickers, The relationships of pitch, loudness and eating techique to judgments of the crispness and crunchiness of food sounds, *Journal of Texture Studies* 16 (1985) 85–95.

[8] C. Dacremont, Spectral composition of eating sounds generated by crispy, crunchy and crackly foods, *Journal of Texture Studies* 26 (1995) 27–43.
[9] W.E. Lee III, M.A. Schweitzer, G.M. Morgan, D.C. Shepherd, Analysis of food crushing sound during mastication: total sound level studies, *Journal of Texture Studies* 21 (1990) 165–178.
[10] S.K. Seymour, D.D. Hamann, Crispness and crunchiness of selected low moisture foods, *Journal of Texture Studies* 19 (1988) 79–95.
[11] W. Al Chakra, K. Allaf, A.B. Jemai, Characterization of brittle food products: application of the acoustical emission method, *Journal of Texture Studies* 27 (1996) 327–348.
[12] K. Kapur, Frequency spectrographic analysis of bone conducted chewing sounds in persons with natural and artificial dentitions, *Journal of Texture Studies* 2 (1971) 50.
[13] Z. Vickers, Sound perception and food quality, *Journal of Food Quality* 14 (1991) 87–96.
[14] C. Dacremont, B. Colas, F. Sauvageot, Contribution of air- and bone-conduction to the creation of sounds perceived during sensory evaluation of foods, *Journal of Texture Studies* 22 (1991) 443–456.
[15] Z. Vickers, Crackliness: relationships of auditory judgments to tactile judgments and instrumental acoustical measurements, *Journal of Texture Studies* 15 (1983) 49–58.
[16] N. De Belie, F.R. Harker, J. De Baerdemaeker, Analysis of chewing sounds as a measure of apple crispness, in: *The XIV Memorial CIGR World Congress*, Tsukuba, Japan, 28 November–1 December 2000.
[17] A.H. Barrett, A.V. Cardello, L.L. Lesher, I.A. Taub, Cellularity mechanical failure and textural perception of corn meal extrudates, *Journal of Texture Studies* 25 (1994) 77.
[18] R. Tesch, M.D. Normand, M. Peleg, On the apparent fractal dimension of sound bursts in acoustic signatures of two crunchy foods, *Journal of Texture Studies* 26 (1995) 685.
[19] L.M. Duizer, O.H. Campanella, G.R.G. Barnes, Sensory, instrumental and acoustic characteristics of extruded snack food products, *Journal of Texture Studies* 29 (1998) 397–411.
[20] J.A. Edminster, Z.M. Vickers, Instrumental acoustical measures of crispness in foods, *Journal of Texture Studies* 16 (1985) 153–167.
[21] Z. Vickers, Sensory, acoustical and force-deformation measurements of potato chips crispness, *Journal of Food Science* 52 (1987) 138–140.
[22] R.A. Johnson, D.W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1998.
[23] N. De Belie, V. De Smedt, J. De Baerdemaeker, Principal component analysis of chewing sounds to detect differences in apple crispness, *Postharvest Biology and Technology* 18 (2) (2000) 109–119.
[24] H.A.L. Kiers, Hierarchical relations among three-way methods, *Psychometrika* 56 (1991) 449.
[25] R. Bro, PARAFAC, Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems* 38 (1997) 149–171.
[26] R. Henrion, N-way principal component analysis. Theory, algorithms and applications, *Chemometrics and Intelligent Laboratory Systems* 25 (1994) 1–23.
[27] R. Bro, Multiway calibration, Multilinear PLS, *Journal of Chemometrics* 10 (1996) 47–61.
[28] R. Bro, C.A. Andersson, The N-way toolbox for MATLAB Ver. 1.04, November 1999. Food Technology MLI-KVL Denmark. http://www.models.kvl.dk/source/nwaytoolbox.