Training of a sensory panel and profiling of winter hardy and coloured carrot genotypes

Stine Kreutzmann a,*, Anette K. Thybo a, Wender L.P. Bredie b

a Department of Food Science, Danish Institute of Agricultural Sciences, Research Centre Aarslev, DK-5792 Aarslev, Denmark
b Department of Dairy and Food Science, The Royal Veterinary and Agricultural University, DK-1958 Frederiksberg C, Denmark

Received 31 August 2005; received in revised form 10 May 2006; accepted 26 May 2006
Available online 20 October 2006

Abstract

The aim of the study was to determine the performance of a sensory panel in evaluation of carrots by investigating the effect of training sessions on the reliability of the sensory attributes. A sensory panel (n = 10) was trained in profiling five carrot genotypes during a training session with seven sensory replications and 13 attributes. A significant effect of training on the sensory profile was determined for bitterness, green flavour and burning aftertaste, indicating that a learning process in judging these attributes was taking place. Terpene flavour, terpene aroma and burning aftertaste were the most reliable determined attributes and carrot flavour, soapiness flavour and nutty flavour were the least reliable attributes. The developed sensory profile was applied on 16 carrot types with a large variation in quality. Winter hardiness, carrot genotype and carrot colour formed distinct groups within the sensory profile, which indicates that the present sensory profile is relevant for assessing sensory quality of carrots. The winter hardy genotypes were characterised by having relatively high intensities in green aroma, carrot aroma, nutty flavour, carrot flavour and sweetness, while the reverse was true for the Nantes types and the coloured carrots with the red and white genotype as being the most extreme.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Daucus carota; Sensory profiling; Training; Reliability; Multivariate data analysis

1. Introduction

Quality is important in marketing of carrots, and flavour is one of the important factors in assessing carrot quality. Genetic variation and environmental conditions largely influence the sensory quality of carrots, as they vary in quality even when grown under controlled conditions and sorted with respect to size and shape (Baardseth et al., 1996; Hogstad, Risvik, & Steinholt, 1997; Rosenfeld, Aaby, & Lea, 2002; Seljåsen, Bengtsson, Hofstun, & Vogt, 2001; Simon, Peterson, & Lindsay, 1980). Carrots can vary in quality dependent on the amount of volatile flavour compounds and non-volatile bitter tastants and sugars, as these compounds influence the sensory perception of the volatile compounds, and thus the total impression of the sensory quality. Carrot genotypes with variation in colour (orange, red, yellow and white), winter hardiness and carrot genotype show a large diversity in quality (Alasalvar, Grigor, Zhang, Quantick, & Shahidi, 2001). Due to an increased consumer demand for diversity in vegetables (Jongen, 2000), carrot genotypes with different colours are now available on the market. However, there is a limited knowledge on sensory quality variation between these genotypes. To meet the needs of the consumers today (Jongen, 2000) development of a sensory map of carrots will provide information about sensory dissimilarities and/or similarities between genotypes.

When assessing quality of biological raw materials as carrots it is important to be aware of the large heterogeneity within genotypes. The different assessors participating in a sensory analysis may get “different” products, hence a training of the panel is very important and the development of a comprehensive set of attributes is essential to achieve a reliable sensory profile of carrots (Byrne, Bak,
Tens, 2000; N and assessor * product effects (Martens, Bredie, & Martin, 1999). Therefore many statistical tools have been used to investigate the effect of product, replicate, assessor and to study the operation of a panel and individual assessors. In order to evaluate discriminant ability and reproducibility, ANOVA is generally applied to the results to study the effect of judge, replicate and products for each sensory attribute. An alternative method is multivariate data analysis in which all attributes are analysed simultaneously providing a holistic interpretation of correlation and redundancy between the attributes. MANOVA, Generalized Procrustes Analysis (GPA), discriminant partial least squares regression (DPLSR), ANOVA partial least squares regression (APLSR) and principal component analysis (PCA) have been applied to study the relevance of each attribute in describing a relevant interpretable sensory variation and the effect of training (Byrne, Bak, et al., 1999; Byrne, Bredie & Martens, 1999; Dijksterhuis, 1995; Hunter & Muir, 1995; Noronha et al., 1995; Sinesio, Risvik, & Rød, 1990; Thybo & Martens, 1998, 2000; Wolters & Allchurch, 1994).

APLSR on a data set using sensory variables as Y-matrix and design variables as X-matrix can be used to relate the sensory structure to product design structure and vice versa. A simplified graphical method to study the reliability of assessors and attributes has been introduced by Martens et al. (2000) where the trivial level effect between the assessors is first removed. Then the remaining variation in the level-corrected sensory data (the total ‘signal’: S) is modelled in terms of a systematic part and a non-systematic residual part (the residual ‘noise’: N). To determine the reliability of assessors, products and attributes S/N plots can be interpreted.

The aim of the present work was to develop a reliable and relevant aroma, flavour and taste profile of carrots. In the study the effect of training sessions on the sensory attributes, and the reliability of each attribute was determined by univariate (ANOVA), and by principal component analysis (PCA) after multivariate modelling of the unstructured noise as assessor level differences by APLSR. The effect of training on the sensory profile determined if a learning process in judging was taking place. A sensory profile for 16 different carrot genotypes was established.

### 2. Materials and methods

#### 2.1. Experimental design

Sixteen different genotypes of carrots with various colours were chosen to represent nearly all-occurring variation in aroma, flavour, and taste of carrots produced in Denmark. The carrots were cultivated in different locations in Denmark on sandy loam soil. In order to include as many genotypes as possible, carrot genotypes from two years were selected. The carrots were harvested in the end of October 2002 and 2003 and stored at 1 °C until preparation. The different genotypes were: (1) 2003; the winter hardy genotypes ‘Navarre’, ‘Sirkana’, ‘Dordogne’, ‘Nepal’, ‘Eskimo’, ‘Artico’, ‘Noveno’ and ‘Nipomo’ (2) 2002; the Nantes types ‘Guerande’, ‘Adelaide’, ‘Bolero’, and the coloured carrots ‘NutriRed’ (red), ‘Yellowstone’ (yellow), and three varieties from screening trials with the codes Line 1 (white), Line 2 (yellow), Line 3 (yellow).

#### 2.2. Preparation of the samples

The carrots were washed, peeled and approximately 2 cm pieces of the root tip and from below the green zone on the leaf-end were discarded. The rest were shredded in julienne pieces of 4.5 mm and the samples were packed and stored at −24 °C for 1 month for the winter hardy genotypes and for 1 year for the coloured carrots and the Nantes types before analysis. All samples were analysed simultaneously. The stability of the samples frozen for 1 year was currently controlled by a minor descriptive analysis including basic tastes and terpene flavour.

#### 2.3. Sensory analysis

Sensory profiling analyses were performed, according to the international standards (ASTM STP 913, 1986) in a sensory evaluation laboratory. The panel consisting of 10 assessors (6 females/4 males, aged from 26–56 years) was screened for sensory ability (basic taste, odour detection, colour vision), as well as ability to communicate sensory descriptions of products as recommended in ISO 8586-1:1993. In addition, the panel was trained in terpene flavour by testing mixtures of three terpenes (α-pinene, caryophyllene, terpinolene) known to have a turpentine-like flavour (Burdock, 2002). They were trained in soapiness by introduction of a reference consisting of soft soap solution. In an introductory discussion with the panel the assessors agreed on a consensus list of attributes for sensory profiling and on the definition of each attribute (Table 1). The 13 attributes were: terpene aroma, carrot aroma, silage aroma, green aroma, hay aroma, terpene flavour, carrot flavour, sweetness, bitterness, green flavour, soapiness, nutty flavour, burning aftertaste. The sensory study was divided into two parts: (1) a training study with sensory profiling of five genotypes and seven sensory replicates for all assessors; and (2) a sensory profiling of the 16 different genotypes determined in three sensory replications. In the training trial, the five genotypes (Nipomo, Navarre, NutriRed, Line 2, and Adelaide) were selected to represent the largest variation within carrots. The five genotypes were served within one session. This was repeated seven times within 4 days, and the replicate sessions represented seven
training sessions. In training 1–4 the assessors got feedback after each session to improve their discriminability. The feedback was focusing on consensus in definition of the sensory attributes between assessors and on assessors’ way of using the scale. The overall aim was to provide an agreement between the assessors and their way of ranging the samples. Training 5–7 took place without feedback.

For each sensory trial the samples and replicates were served in coded containers in a totally random order. Each panellist was served 25 g grated raw carrot room temperature ± 2 °C in a covered transparent cup. To make sure that the panellists could not differentiate visually between the samples the sensory laboratory was illuminated by red light. All samples were supplied with three random digit numbers. The panellists evaluated the samples at individual speed by descriptive analysis on an unstructured 15-point line scale. Intensity ratings ranging from low (value 0) to high intensity (value 15) were registered on a direct computerised registration system (FIZZ, ver. 2.00M, Coutenon, F).

2.4. Data treatment

Statistical analyses were performed using SAS software (Release 8e, SAS Institute, Inc., Cary, NC, USA). A mixed model of analysis of variance (ANOVA) was applied to the data to study the signal to noise ratio ($F$-value) of the assessors and samples in a univariate way considering assessor * sample as random effects. At this time, level and range differences between individual assessor and sample are modelled (Brockhoff & Skovgaard, 1994). In the training study principal component analysis (PCA) and ANOVA partial least squares regression (APLSR) were used to study the discriminability and reliability of the panel. APLSR was performed on the raw data of 13 attributes from each of the seven replicates, with the $X$-matrix as 0/1 design variables (assessors, replicates) and the $Y$-matrix as sensory variables. The assessor level and replicate level differences were determined and removed. A second APLSR was performed on the corrected data to investigate the remaining systematic structure in the data. Signal (total variance) to noise (residual variance) values ($S/N$ values) were calculated as the total variances after 0 principal components (PC) divided by the residual variances after modelling the optimal number of PCs (Martens et al., 1998, 2000). $S/N$ values of assessors, samples, attributes and replicates determine the reliability and performance of each. For the sensory mapping of the 16 carrot varieties a PCA was applied on the sensory data corrected for assessor and replicate level differences as described above. Analyses were performed using Unscrambler (Windows Version 8.0 software package, CAMO A/S, Trondheim, N). Full cross validation was used as validation criterion.

3. Results and discussion

3.1. Assessors performances

In order to study the reliability and performance of the panel $S/N$ values for the attributes, assessors and samples are determined on data, where level effects between assessors and replicates are removed. The elements that are best modelled have high total signal (systematic variation) and low estimated residual variance (unstructured variation). Or in other words elements with high $S/N$ ratio have good discriminating properties and describe a sensory attribute, a sample or an assessor to give consistent results denominated as a reliable result (Martens et al., 1998). The $S/N$ plot of all sensory attributes (Fig. 1) illustrates the highest $S/N$ values for terpene aroma, terpene flavour, green aroma, green flavour and burning aftertaste and proved these attributes to be the most reliable attributes and carrot flavour, soapiness flavour, and nutty flavour to be the least reliable attributes. It is noticeable that carrot flavour seems to be not as easy to determine in contrast to terpene flavour. This is surprising since carrot flavour is likely to be determined by terpene content.

The $S/N$ plot of assessors and samples were studied for all sensory attributes. The $S/N$ values for assessors were...
revealed 83% of the total variation in the training data. The plots (Fig. 3). Three significant principal components (PC) ing study with seven replicates was visualised in a PCA bi-

PCA plot gives an interpretable discrimination of sensory quality of carrots and the progression in using the vocabulary. The first PC mainly described differences between samples in terpene flavour, terpene aroma, green flavour, burning aftertaste, hay aroma, and silage aroma versus sweetness, carrot aroma, green aroma, and nutty flavour. The second PC described differences in carrot flavour versus soapiness flavour between samples and finally the third PC indicated that a learning process in judging bitterness, green aroma and flavour and burning aftertaste had taken place (Fig. 4). During the training session the intensity range of these attributes in especially NutriRed and Line2 increases as a variation is observed for the 7 training sessions in PC3. To investigate the discriminative development of individual attributes over the replicative sessions a

Fig. 1. Signal/noise plot of 13 sensory attributes after APLSR modelling (a = aroma; f = flavour).

very similar (Fig. 2), which means that they nearly have got identical reliability. Assessor 10 revealed the most noise (residual variances) but still had S/N-values similar to the other assessors, as this assessor performed sensory data with high signal. Adelaide seemed to be the least reliable sample and Navarre to be the most reliable sample as Navarre had higher S/N-value than Adelaide. Even that the S/N values for samples and assessors are quite similar; the values are relatively low (S/N = 1.2–2.3), which may be reflected by the fact that the sensory investigation is a train-

Fig. 2. Signal/noise plot of ten assessors and five carrot samples after APLSR modelling. ■ = Assessors (ass1–ass10); ▼ = samples (Adelaide, NutriRed, Nipomo, Navarre, Line2).

Fig. 3. A training study: principal component analysis (bi-plot of PC1 and PC2) of 13 attributes, five carrot genotypes and seven replicates after modelling assessor level differences by APLSR. ■ = samples (Ni = Nipomo; Na = Navarre; Nu = NutriRed; Ad = Adelaide; LD = Line2; 1–7 = replicates); ▲ = sensory attributes (a = aroma; f = flavour).

The systematic structure in the sensory data from the training study with seven replicates was visualised in a PCA bi-

Fig. 4. A training study: principal component analysis (bi-plot of PC1 and PC3) of 13 attributes, five carrot genotypes and seven replicates after modelling assessor level differences by APLSR. ■ = samples (Ni = Nipomo; Na = Navarre; Nu = NutriRed; Ad = Adelaide; LD = Line2; 1–7 = replicates); ▲ = sensory attributes (a = aroma; f = flavour).
S/N analysis was performed on the 13 attributes over each of session 1–7. For the majority of the attributes an increase in the S/N ratio were apparent between the sessions. This was most obvious for attributes like green flavour, bitterness and burning aftertaste (Fig. 5) and suggests that training had a significant effect on the reliability of these attributes. An inverse effect of training was observed for soapiness, sweetness and carrot flavour (Fig. 5). According to soapiness this indicates that soapiness is a difficult attribute for the assessors to judge, strengthened of the relatively low intensity. According to sweetness and carrot flavour the missing effect of training might be due to the fact that focus was on the other attributes, which assumed more difficult attributes.

In summary, the S/N plots for each attribute illustrated some assessors to be good at evaluating some attributes and less good at others, and that the training had an effect on the majority of the attributes. S/N plots and PCA are valuable tools for evaluation of sensory panel performance, and have been applied for developing sensory profiles of potatoes, chicken and porcine meat (Byrne, Bak, et al., 1999; Byrne, Bredie & Martens, 1999; Thybo & Martens, 1998, 2000).

3.2. Sensory profiling

The developed sensory profile was used for sensory profiling of 16 carrot genotypes. A large variation in intensity of the sensory attributes, except for terpene, silage and hay aroma, and nutty flavour was determined (Tables 2 and 3). Results from univariate ANOVA (Table 2) showed the effects of samples and assessors to be significant for all sensory attributes. This means that genotype largely influence the variation in sensory attributes. The large differences are not caused by one extreme genotype but by the fact that the genotypes used spun as much variation as possible. Simon, Peterson, and Lindsay (1982) and Seljåsen et al. (2001) also found that genotype largely influenced the variation in sensory attributes of carrots. On the other hand, Gills, Resurreccion, Hurst, Reynolds, and Pathak (1999) reported limited differences between flavour attributes for carrot genotypes, namely in the perception of sweet taste only, and Martens, Rosenfeld, and Russwurm (1985) found differences between carrot genotypes only for fruit flavour. Rosenfeld, Risvik, Samuelsen, and Rødboth (1997) found the genetic variation, mainly related to the variation in sugar and texture related attributes. When comparing sensory studies on genotypes it is expected to find some differences since the genotypes determines the magnitude of the effects that can be observed. The result from ANOVA

![Fig. 5. Development in signal to noise ratio (S/N) for 6 selected attributes over replicates 1–7.](image-url)

<table>
<thead>
<tr>
<th>Sensory attribute</th>
<th>Sensory scores</th>
<th>Assessor</th>
<th>Sample</th>
<th>Assessor * sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terpene aroma</td>
<td>2.1–11.2</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Carrot aroma</td>
<td>2.9–7.7</td>
<td>***</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Silage aroma</td>
<td>0.8–5.8</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Green aroma</td>
<td>0.9–6.3</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Hay aroma</td>
<td>1.2–5.6</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Sweetness</td>
<td>3.9–11.4</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Bitterness</td>
<td>1.4–11.8</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terpene flavour</td>
<td>1.6–11.9</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Carrot flavour</td>
<td>3.2–9.7</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Green flavour</td>
<td>0.6–8.8</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Soapiness</td>
<td>1.5–7.2</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Nutty flavour</td>
<td>1.8–4.8</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Burning aftertaste</td>
<td>1.7–10.6</td>
<td>***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* *** ** * Significantly different using F test at P > 0.001, P > 0.01, P > 0.05, respectively using a mixed ANOVA model with assessor * sample as random effect.
Table 2 showed significant effects of assessors * samples, which illustrated disagreement among the assessors in scoring the samples for some of the attributes. This means that even after correction for level differences between assessors level and range some assessor differences exist, which is common in sensory studies (Brockhoff & Skovgaard, 1994; Næs & Langsrud, 1998). However, the panel was still detecting variety differences for all attributes after modeling the assessor differences.

A PCA was performed on mean values over assessors and sensory replicates for the 16 genotypes and the 13 attributes. The PCA bi-plot are visualized in Fig. 6, showing the first two PCs to explain 86% and 7% of the variation. A high correlation was observed between all attributes on the first PC. The attributes carrot aroma, nutty flavour, carrot flavour were positively correlated and inversely correlated to green flavour, burning aftertaste, bitterness, terpene aroma and flavour, soapiness, and silage aroma. The second PC determined information about odour attributes and sweetness, where green aroma was inversely correlated to hay aroma and sweetness. It is obvious that some of the attributes explaining the major proportion of the variation in sensory quality of carrots (Fig. 6, PC1) are highly correlated and expresses some redundancy. Most of the corresponding aroma and flavour attributes are highly correlated as e.g. for the attributes carrot, green and terpene aroma and flavour. Therefore the sensory profile could have been reduced to include only aroma or flavour attributes. In conclusion, the perception of aroma attributes seems to be good indicators of flavours for carrots. A high correlation is found for the "negatively" perceived attributes burning aftertaste, bitterness, terpene aroma and flavour, soapiness, and silage aroma. These correlations are also found by Varming et al.

![Fig. 6. Sensory mapping of 16 carrot genotypes and 13 attributes by principal component analysis (PCA bi-plot) after modelling assessor level differences by APLSR. ■ = samples; ▲ = sensory attributes (a = aroma; f = flavour).](image-url)
(2004). Rosenfeld et al. (2002) and Seljäsen, Hoflin, and Bengtsson (2000). Rosenfeld et al. (2002) examined sensory quality of carrot roots and found the attributes sweet taste, crispness, juiciness and acidic taste to correlate inversely to terpene odour, green odour, terpene flavour, bitter taste, bitter aftertaste and others. Varming et al. (2004) examined eating quality of raw carrots and found the attributes fruity taste, carrots aftertaste, nutty taste and sweet taste to correlate inversely to bitter taste, bitter aftertaste, earthy taste and earthy odour when carrying out a APLSR on the results.

Winter hardiness, carrot genotypes and carrot colour formed distinct groups (Fig. 6). The winter hardy genotypes ‘Navarre’, ‘Sirkana’, ‘Dordogne’, ‘Nepal’, ‘Eskimo’, ‘Artico’, ‘Noveno’ and ‘Nipomo’ were characterised by having significantly higher intensities in green aroma, carrot aroma, nutty flavour, carrot flavour and sweetness, while significantly lower intensity was found in the Nantes types (‘Guerande’, ‘Adelaide’, ‘Bolero’) and in the coloured carrots (‘NutriRed, Yellowstone’, ‘Line1’, ‘Line2’ and ‘Line3’) with the red (‘NutriRed’) and the white (‘Line1’) genotypes as being the most extreme (Fig. 6, Table 2). The most bitter genotypes were NutriRed (average score 11.8), Line3 (11.2), Line1 (10.0) and Line2 (10.0), and the sweetest genotypes were Navarre (11.4), Sirkana (11.3) and Novenco (11.1).

Significantly higher intensity in green flavour, burning aftertaste, terpene aroma, terpene flavour, silage aroma and low intensity in green aroma, carrot aroma, nutty flavour, carrot flavour and sweetness was found in the yellow, red and white genotypes. Alasalvar et al. (2001) examined four different coloured carrots for sensory quality. They found significant difference among the groups for some of the flavour attributes, including sweetness with the purple carrot to have the highest score, and that orange and white genotypes had higher levels of terpenes than the other genotypes. Furthermore they found no significant difference among the groups for bitterness and aftertaste. An important observation is that the genotype ‘Bolero’ being the most commercial genotype on the Danish market is actually described as having relatively high sweetness, hay aroma and relatively low carrot flavour. It is therefore obvious that many other carrot types match the quality of for instance ‘Bolero’. A sensory profile of carrots is therefore providing information of high relevance for marketing of other genotypes of carrots to fulfill demands of consumers today on larger diversity in quality in foods (Jongen, 2000).

4. Conclusion

A sensory vocabulary was developed, which described the sensory variation among a large variation in genotypes of carrots in terms of 13 aroma, flavour and taste attributes. The results showed that during training of a sensory panel a learning process is taking place. This learning process is more obvious for some attributes and some samples than others. In a sensory training it is therefore important to have tools for investigating such differences as by e.g. S/N plots. The developed profile was considered relevant for evaluation of carrots, which was demonstrated by a sensory mapping of 16 highly different carrot genotypes.

The overall sensory profile showed that taste and colour are highly related indicating that genes may code for taste and colour. As consumers’ today demands for large variation in quality of raw material to be used for various purposes in the meal, a sensory mapping of carrot genotypes will provide consumers such information.

Acknowledgements

The authors wish to thank Merete Edelenbos and Kai Grevesen for providing raw material for the present investigation and for skilful discussion. The investigation is a part of a Ph.D. study supported financially by The Danish Institute of Agricultural Sciences.

References


