Food Quality and Preference 12 (2001) 19–28

Food Quality and Preference

www.elsevier.com/locate/foodqual

Training a sensory panel for TI: a case study

Christine Peyvieux [a], Garmt Dijksterhuis [b,*]

[a] *ENSBANA, Dijon, France*

[b] *KVL, Sensory Science Group, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK 1958 Frederiksberg C, Denmark*

Received in revised form 20 June 2000; accepted 7 July 2000

Abstract

Time–intensity (TI) is an increasingly used sensory method, however, no proper guidelines for training panellists with TI seem to have been set up, hence comparison of results from different TI studies is difficult. In this paper a three steps approach is proposed: (1) introduction to the method; (2) training for the TI task through a simple product assessment: basic taste beverages; (3) running a TI pilot experiment. The loadings from a principal components analysis (PCA) on TI data have proved to be useful in assessing panel agreement and providing information about individual differences. Pilot profiling proved useful for choosing the attributes to be used in the TI study. This minimises the chance of performing a TI study with irrelevant attributes. It is concluded that, following the steps in this study, most panellists are able to learn to perform a TI task reliably. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Time–intensity; Panel training; Meat; Procrustes

1. Introduction

The objective of this study was to suggest a way to train panellists for the time intensity (TI) method. Training panellists for sensory descriptive tests is important to get a consistent panel, that is panellists with a good discriminatory ability and sensitivity, who are reproducible in their assessments of identical samples and agree as a group with respect to their perception. For the traditional sensory methodologies training guidelines are available (e.g. Civille & Czczesniak, 1973; ISO, 1992, 1994; Mioche & Touraille, 1990) and in the literature there is reference to these, but for (TI) methodology no proper guidelines seem to have been set up so far. TI methodology is a method that permits the recording of the perception of an attribute and its change over time while a foodstuff is processed in the mouth. These measurements can last from less than one minute to several minutes, depending on the type of food or drink under study. There seems to be increasing understanding of the value of TI results. They provide information not available from the traditional sensory methods (see e.g. Duizer, Gullet & Findlay, 1993). However, lack of consistency in the way to train panellists can lead to problems when comparing the results of different TI studies. Some training methods reported are very long and fastidious, with repetition of the task for each product and each attribute until panellists show a 'good' consistency in their replications (see e.g. Duizer et al.). Other studies have mainly been focusing on the means to check the consistency of the panel in conventional profiling (Bloom, Duizer & Findlay, 1995; Dijksterhuis, 1995). Some of these issues concerning TI studies are reviewed in Dijksterhuis and Piggott (2000).

2. Materials and methods

The study was part of a larger experiment in which a TI method to assess the change in perceived texture attributes was tested. The product used in that experiment was sliced, salted and smoked pig meat, called casseler rib. In this paper only the part of the study focussing on the initial training-stage of the experiment is presented.

2.1. Samples

The samples were furnished by Meester B.V. and consisted of two kinds of casseler rib.

* Corresponding author.
E-mail address: gbd@kvl.dk (G. Dijksterhuis).

PII: S0950-3293(00)00024-0

1. Product A: casseler rib made with fresh meat that has a low ultimate pH (< 5.5).
2. Product C: casseler rib made with fresh meat that has a high ultimate pH (> 5.9).

We had to take into account that there is a difference in slice composition due to natural variation in the muscle, depending on whether the slices are cut in the neck or in the leg side of the pig. The so-called 'apple' of the muscle is bigger at the leg side than at the neck side.

The samples were kept in a chilled room at 4°C.

2.2. Training of the panellists

Ten panellists, seven men and three women selected from the ID-DLO internal institute sensory panel, were trained with respect to the TI methodology. The training was run in three steps.

1. Introducing the method to the panellists.
2. Familiarisation of the panellists with the computer system (Compusense Inc., 1996) and the TI task using basic taste solutions.
3. Training the panellists using the real product:
   A. sensory profiling,
   B. pilot TI study.

2.2.1. Introducing the method

The first step consisted of a short talk presenting the method to the panellists. No information was given about curve shape, thinking that this may influence the panellists when recording their perception, although no study was found that could support this hypothesis (cf. Pälsgård & Dijksterhuis, in press). The panellists were not shown their TI curves during and after the experiment. Another reason not to show curves during training is that papers about texture assessments using TI methodology showed that of texture and flavour TI curve shapes were not the same (see e.g. Butler, Poste, Mackie & Jones, 1996).

In this first step the panellists were also introduced to the computer system. General questions about the experiment and the procedure were answered, not revealing information about the product and TI curves.

2.2.2. Training with basic tastes

The second step was to familiarise the assessors with the actual use of the computer system and to train them to track a single sensory attribute. They were presented with ten millilitres of four basic taste beverages (sweet, salt, bitter and acid) of which the concentrations were chosen higher than every panellist's threshold (see Table 1). Because subjects reported to have difficulty tasting, a higher concentration was given subsequently. Over several sessions, each beverage was presented four times to the panellists.

Table 1
Highest threshold in the entire panel and concentrations (g/l) of the beverages given to the panellists in the basic taste TI training step

	Acid	Salt	Sweet	Bitter
Highest threshold	0.48	1.4	7.2	0.27
First concentration given	0.6	2.0	12	0.27
Second concentration given	0.93	2.87	33.22	0.5

2.2.3. Training with the real product

The third step was to train the panellists with the real product, as it is obviously more complex both in flavour and in texture than basic-taste solutions. This part took place after discussion by the panel to set up a list of attributes for this product (Civille & Szczesniak, 1973; Mioche & Touraille, 1990) and subsequent analysis of the line scale data from a small sensory profiling study (step 3A) of both products using the attributes from the list. This permitted us to choose attributes that were consistently used by the panel. Thus the TI training (step 3B) was carried out with two attributes only, one for flavour, and one for texture, on one kind of casseler rib. Panellists were trained until they showed at least two reasonably overlapping replicate TI curves out of three. We considered as good replicates, curves of which the intensity was not different for more than approximately 40% of the time.

2.3. Sample preparation and lay out of the experiment

Samples were put at room temperature at least half an hour before being evaluated and consisted of a thin slice of meat (approximately circular, 8–10 cm diameter) folded four times on a covered plate. Each assessor received slices that were taken from the same box and the same part of the *longissimus dorsi* muscle. Five panellists received their slices from the neck part of the muscle (panellists 2, 3, 5, 7, 9), and the other five from the leg part (1, 4, 6, 8, 10). The experiment was carried out with the attributes smoky, salty and juicy, on the two different casseler rib types. For each type×attribute combination three replicates were presented. Samples were evaluated under red light. The Compusense Five (version 2.2) computer program was used (Compusense Inc., 1996) for collection of the TI data. The panellists were shown an unstructured vertical line scale anchored with 'a little' at the bottom and 'very' at the top (Fig. 1).

No reference samples were given to the panellists. Assessors were asked to take the whole sample in the mouth. They were instructed to give a mouse-click on the *start* button when putting the sample in their mouth, and to move the cursor along the line according to the intensity of their perception. The software collected the position of the pointer along the scale every half a second during 2 min, thus collecting 240 data points for

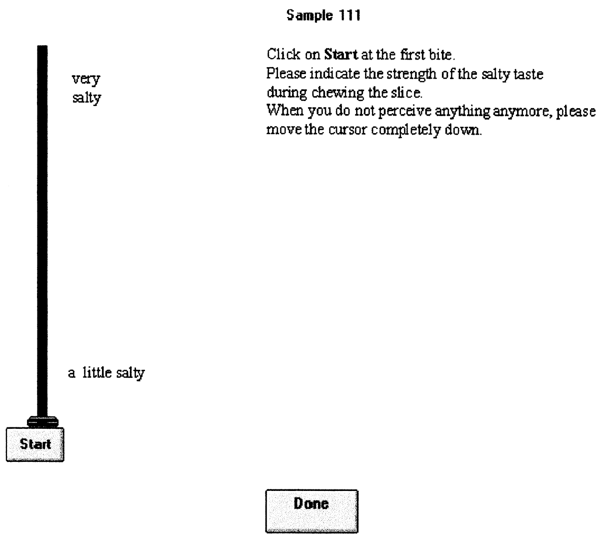

Fig. 1. Computer screen for TI evaluation of saltiness.

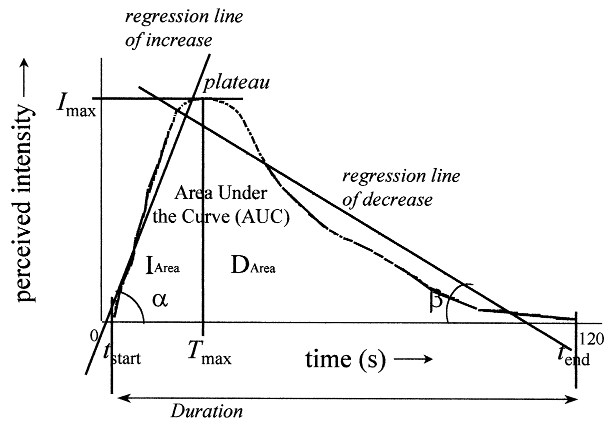

Fig. 2. Typical TI curve and a set of parameters commonly derived from it.

each TI judgement. Panellists were instructed to move the cursor back to zero, i.e. completely down, when they did not perceive anything of a flavour attribute. With texture attributes they were expected to move the cursor back to zero after they swallowed the sample.

For the ordinary sensory profiling line scale measurements, panellists were provided with unstructured horizontal line scales. Two labels, 'a little' and 'very' indicated the anchors of the scale. The samples were evaluated using the texture attributes fibrous, dry, watery and juicy. For flavour two attributes were used: smoky and salty.

A period of 1 min between two samples was allowed for the panellists to cleanse their palate with water and non-salted crackers.

2.4. Statistical analysis of the results

Line scale data were analysed using generalised Procrustes analysis (GPA) (Gower, 1975; see also Dijksterhuis & Gower, 1992). The TI assessments result in data which have the following characteristics (see e.g. Dijksterhuis, 1997):

- a common shape (see Fig. 2),
- a large amount of data points
- large inter-individual differences but intra-individual consistencies.

From the curves several parameters can be extracted (see Fig. 2). To check the consistency of the panel and hence the efficiency of the training, individual curves were drawn and analysed using non-centred principal component analyses (Dijksterhuis, 1993). The loading plots from these analyses can be used to study the consistency of the panel (see Dijksterhuis, Flipsen & Punter, 1994; Van Buuren, 1992).

3. Results and discussion

The main objective of this study was methodological, with respect to the training of a sensory panel for a TI task. The differences between the two types of products will not be highlighted in the remainder of the paper since they are secondary to the above-mentioned objective of the study.

3.1. Results of the training stage with basic taste stimuli

The raw curves from this step are presented in Figs. 3 and 4. Through these first curves two main problems are illustrated:

1. Initially not all assessments are made according to the theoretical and generally expected shape of a TI curve (see Fig. 3).
2. Panellists seem to have problems of replicating the intensity, i.e. the height of the curve (see Fig. 4)

Although the results from the first session did not all look like typical TI curves, replicates were quite close in general and we concluded that the panellists in principle have understood the task. No panellists were eliminated at this stage.

Tracking basic tastes is an easy way for panellists to learn the expected tracking task, easier than tracking one attribute from a composite of more attributes, which is the case with the rather complex product in this study.

3.2. Training with the real product

3.2.1. Results from the line scale profiling study

GPA was used to analyse the line scale profiling data in order to check the consistency of use of the attributes by the panel. This was needed in order to be able to select the most relevant attributes for the TI study. The GPA group average (often called ‘consensus’) explained 59.8% of the total variance, the first two dimensions of the GPA group average explained 36.2% variance (25.8 and 10.4% for the 1st and 2nd dimension, respectively). The group average is not shown here because the interest was mainly in the consistency of use of the attributes. Plots of the correlations of the attributes with the dimensions of the two-dimensional GPA group average are shown in Fig. 5. There is a plot for each attribute, in these plots the numbers refer to the panellists. Obviously, panellist 2 has a completely different interpretation of *fibrous* than most other panellists, further the group of panellists is not homogeneous in scoring *fibrous*. The plot for *smoky* shows a relatively homogeneous group of panellists, with the clear exception of panellists 3. The result for *dry* shows two panellists (1 and 8) opposing a group containing the other panellists. For *watery* there are panellist 4 and 9 who oppose a somewhat more heterogeneous group of panellists. Clearly the attribute *juicy* resulted in homogeneous scores, all panellist points are at the left of the centre of the plot. The same is true for *salty*. There appeared no clear single outlier for all the attributes.

It can be seen that the attributes *watery* and *juicy* point in a direction opposite to dryness, along the first dimension, which is an expected result, they correlate negatively.

For the TI study the most consistently used attributes *watery*, *juicy*, *salty* and *smoky* were retained. Dryness was eliminated because we thought that this attribute might be difficult to assess in thin slices.

3.2.2. *Training for TI with the real product*

During the training with the real product, most panellists showed that the quality of their curves increased, in that they were now more consistent over replications.

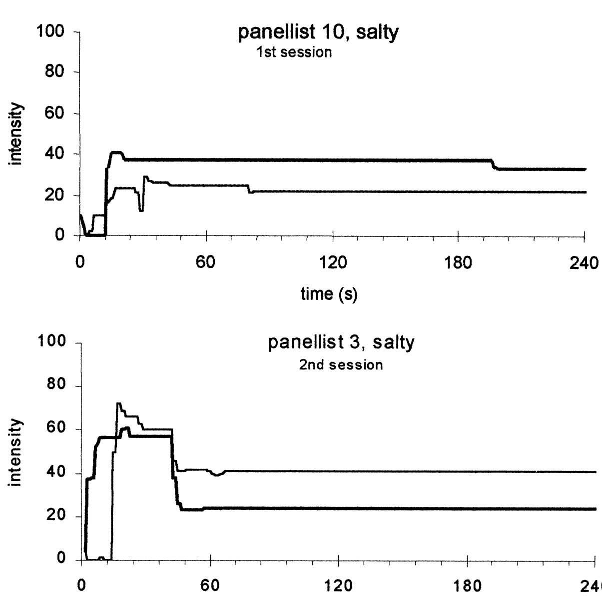

Fig. 3. Two replicates of two panellists’ TI curves from the 1st and 2nd basic taste sessions for salty, illustrating atypical TI curve shapes.

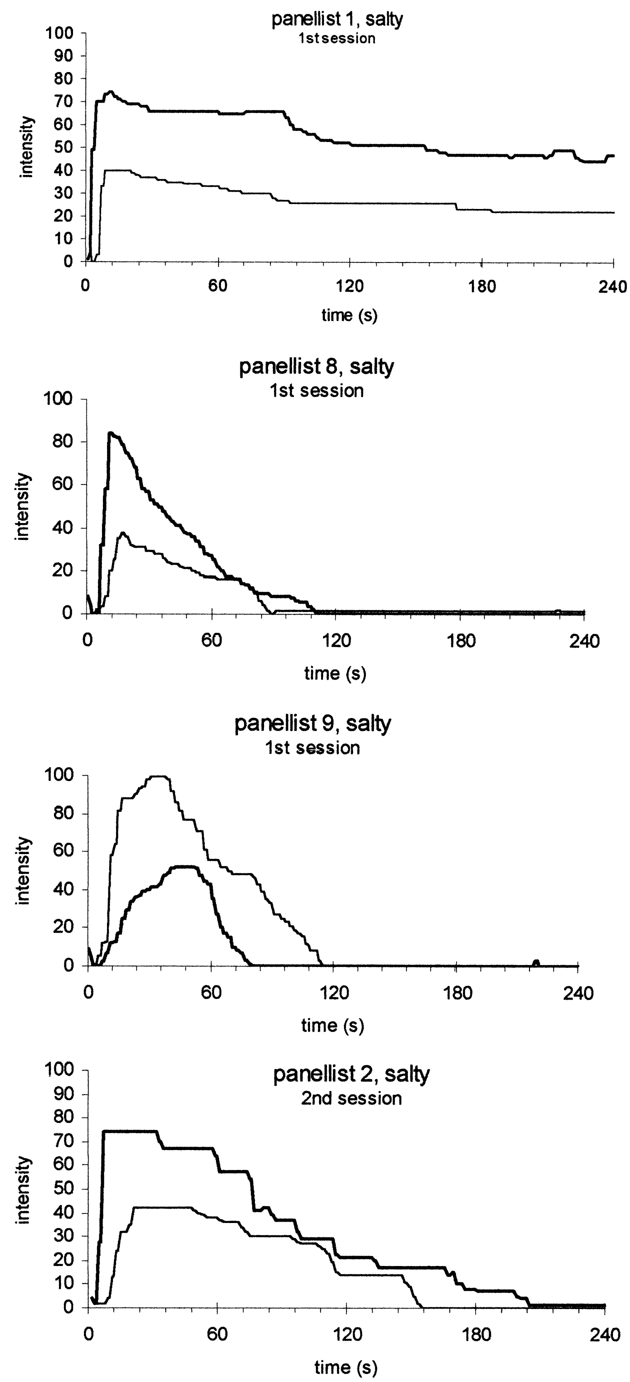

Fig. 4. Two replicates of some panellists’ TI curves from the 1st and 2nd basic taste sessions for salty, illustrating a replication problem.

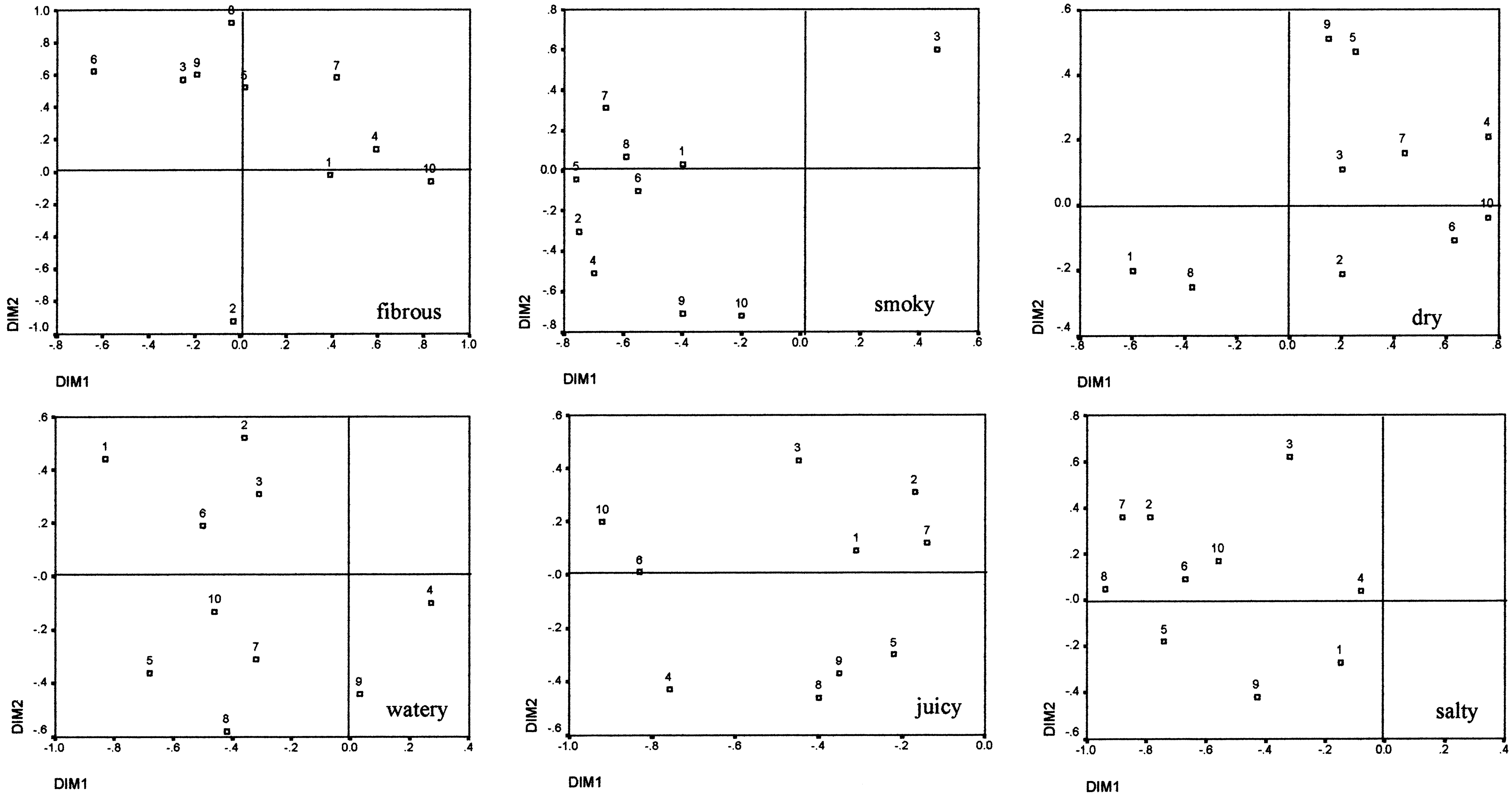

Fig. 5. Plot of correlations of attributes with the two first dimensions of the GPA solution. One plot per attribute, numbers 1,. . .,10 represent panellists.

Fig. 6 presents the curves of panellists 3 and 10, which were presented earlier producing an atypical curve (in Fig. 3), i.e. with a tail not returning to zero. These two panellists illustrate an improvement by producing TI curves that do return to zero. In Fig. 7 the curves from the panellists 1, 2, 8 and 9, which were shown in Fig. 4 to have trouble with consistently scoring intensity, are shown to have improved as well. These figures illustrate that panellists are able to learn the task of tracking a single attribute from a complex stimulus after some training.

To ratify that the panel was scoring consistently, a non-centred PCA on the curve data was carried out and the loading plots were studied (see Dijksterhuis, 1993; Van Buuren, 1992). Loadings represent the weight each panellist's first and second principal curves received (only the first and second dimensions are considered). A consistent panel is expected to have high loadings on the first principal first axis (see Fig. 8). For saltiness the loadings are all in the negative part of the first component which suggests a good homogeneity of the panel. The second component seems to divide the panel into two groups:

1. Panellists 1, 6, 10 at the negative side of the second component.
2. Panellists 2, 3, 4, 5, 7, 8, 9 form the opposite group.

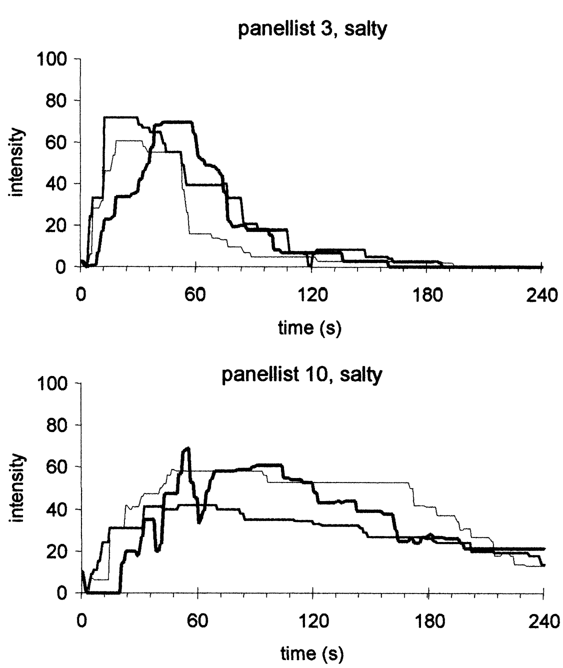

Fig. 6. Three replicate TI curves, from training with the real product, showing the improvement of the expected curve shape compared with the curves of the subjects 3 and 10 in Fig. 3.

In the first group panellists 6 and 10 show their replicates being quite close. Studying the individual curves it can be seen that these panellists produced curves that never return to zero, and present a slow decrease to the end of the curve (see Fig. 9). Thus, the second component seems to represent the way of decrease of the individual curves. This result is in accordance with the study of Dijksterhuis et al. (1994) who found that the second principal curve models the rate of rising and declining of the individual TI curves.

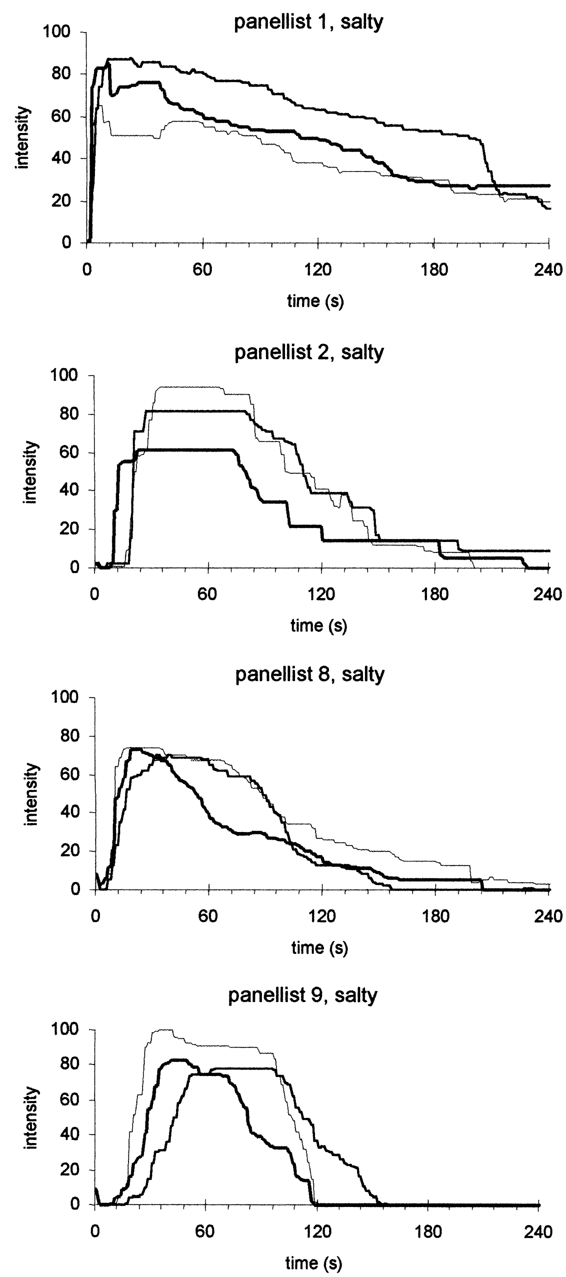

Fig. 7. Three replicate TI curves, from training with the real product, showing the improvement of the expected curve shape compared with the curves of the panellists 1, 2, 8 and 9 in Fig. 4.

Two outliers exist in this panel. Obviously from Fig. 8 one of the replicates from panellist 4 is astray (see Fig. 10); something seems to have gone wrong with the second replicate. Panellist 5 shows one odd replicate according to the first component, which can be seen in the individual curves of this panellist 5 (Fig. 10). The 2nd replicate shows a null intensity curve.

The loading plots provide a way to perform a check on the curves of a TI panel. With a high number of curves this is easier than printing and visually inspecting all the individual TI curves.

For the watery attribute, the loadings are in a small range, note they are all negative along the first component (Fig. 11). This indicates good homogeneity of the panel. The second component divides the panel into two main clusters.

1. Panellists 4, 5 and 7 at the negative side of the second component, present curves that rise slower than the curves from the other panellists.
2. Panellists 1, 2, 3, 6, 8, 9, 10 at the positive side of the second component.

The second component again seems to represent the rising of the curves.

Panellist 5 appears to be an outlier again from the rest of the panel for both components. Even if the range of loadings is small, the replicates for panellists 1, 2, 9 and 10 are not so close together compared with the results for salty. No effect of the different experimental sessions was found.

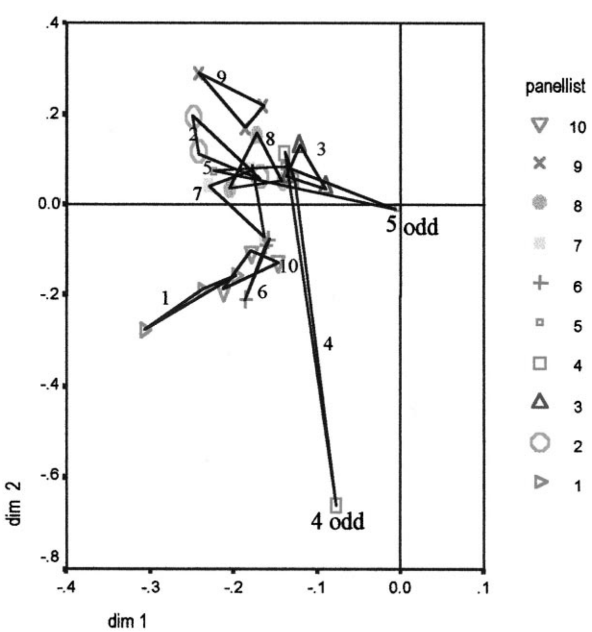

Fig. 8. Loading plots of the panellists and their replicates (connected by lines) in the training stage with the real product, for salty. Two 'odd' replicates (for panellists 5 and 4) indicated in the figure.

3.3. Results of the real TI experiment

The loadings from a PCA of the TI curves proved to be a convenient tool to summarise this information and to enable a check on the consistency of the panel and the identification of deviant curves or panellists. Some curves that were clearly aberrant, like having zero I_{max}, were removed a priori, this is mentioned in the corresponding figure captions.

An uncentred PCA was performed for each of the attributes separately, like was done on the TI curves from the training. The loading plots were inspected for consistency of the panel. For *salty* the resulting loading plot is presented in Fig. 12, for *smoky* in Fig. 13 and for *juicy* in Fig. 14. All the PCAs performed resulted in a solution which captured most variance in its first two dimensions.

For saltiness (see Fig. 12), the loadings on component 1 present a small range, are all negative and quite close

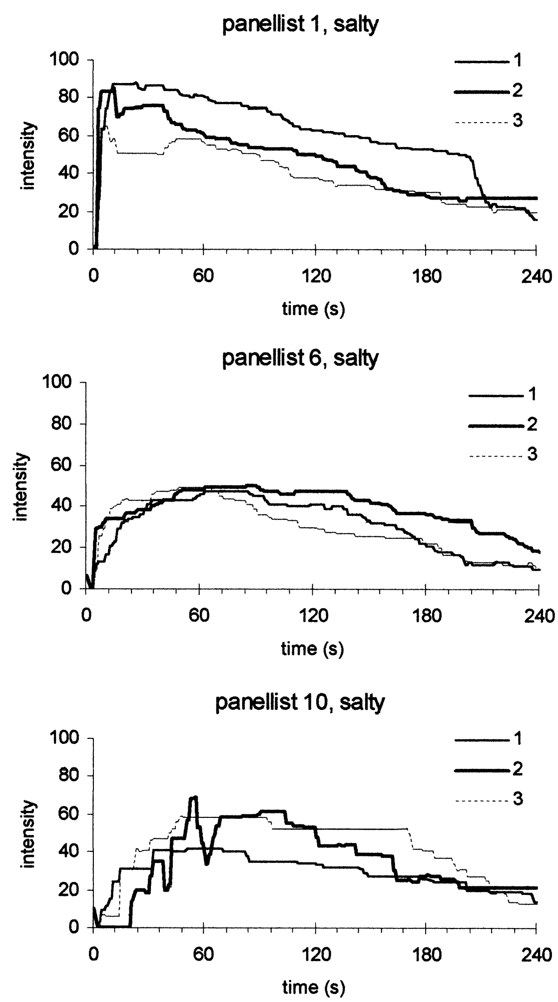

Fig. 9. TI curves from the training with the real product, illustrating the group of curves, for salty that do not return to zero.

to the axis. This indicates a good homogeneity of the panel. The second component separates the panel into two main groups: 2, 5, 6, 7, 8 and 1, 3, 4, 9. This would represent assessors who show an abrupt rising and decreasing of the two flanks of the curve and those whose curves have longer lingering down-flanks also present a lower I_{max} in general. It can be noticed that panellist 9 is an outlier compared to the rest of the panel.

For smoky (Fig. 13), the loadings are negative along the first component and in rather a small range. For product A assessors 3, 4 and 9 have lower scores on the second component, indicating differently decreasing curves. One of the replicates of assessors 1 and 6, for product C, is a bit out of range. No obvious systematic outliers appear from these plots. Two main groups are separated by the second component for *smoky* (Fig. 13): panellists 1, 2, 6, 7 present high loadings for the second principal curves and panellists 3, 4, 8, 9 low loadings.

For juiciness (Fig. 14) two clusters can be defined along the second component: panellists 1, 4, 7, 8, 9 form the first group and panellists 2, 3, 5, 6 form the second one. No real cluster can be defined along the first component

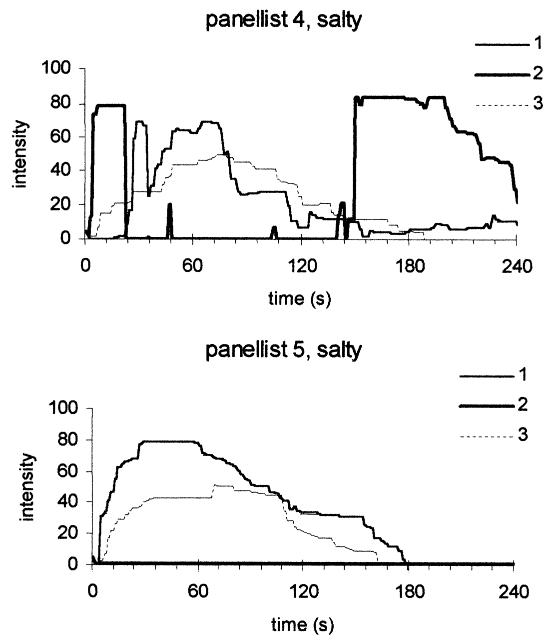

Fig. 10. Three replicate curves for panellists 4 and 5, salty, in the training with the real product.

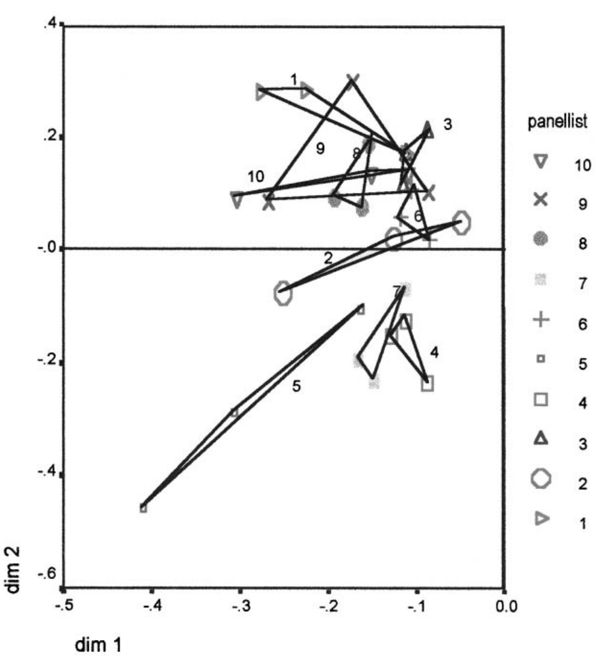

Fig. 11. Loading plots of the panellists and their replicates in the training stage with the real product for watery.

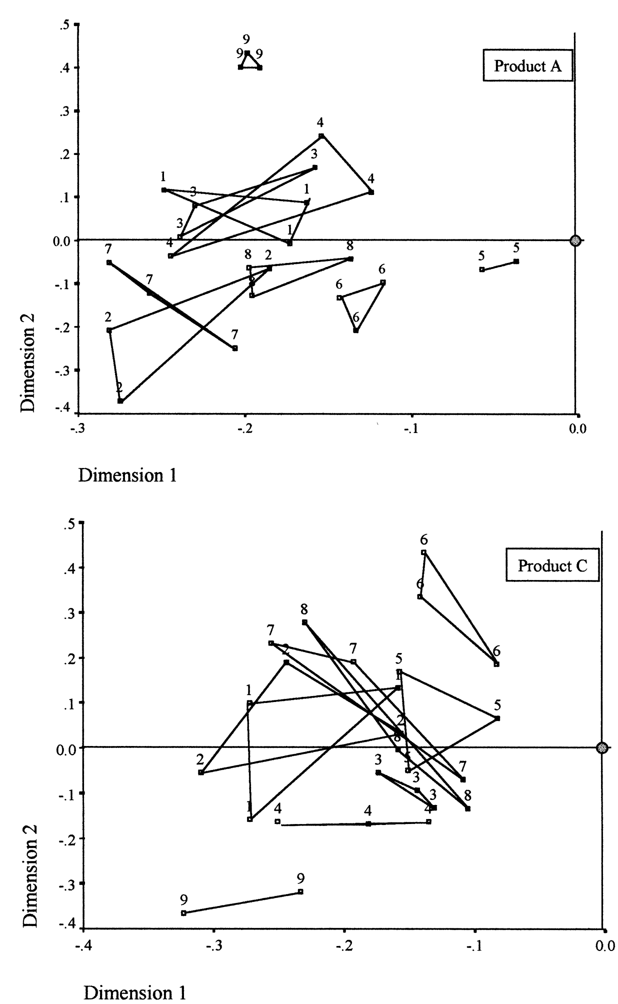

Fig. 12. Loading plots for the panellists for salty, for product A (one replicate for panellist 5 removed) and product C (one replicate for panellist 9 removed).

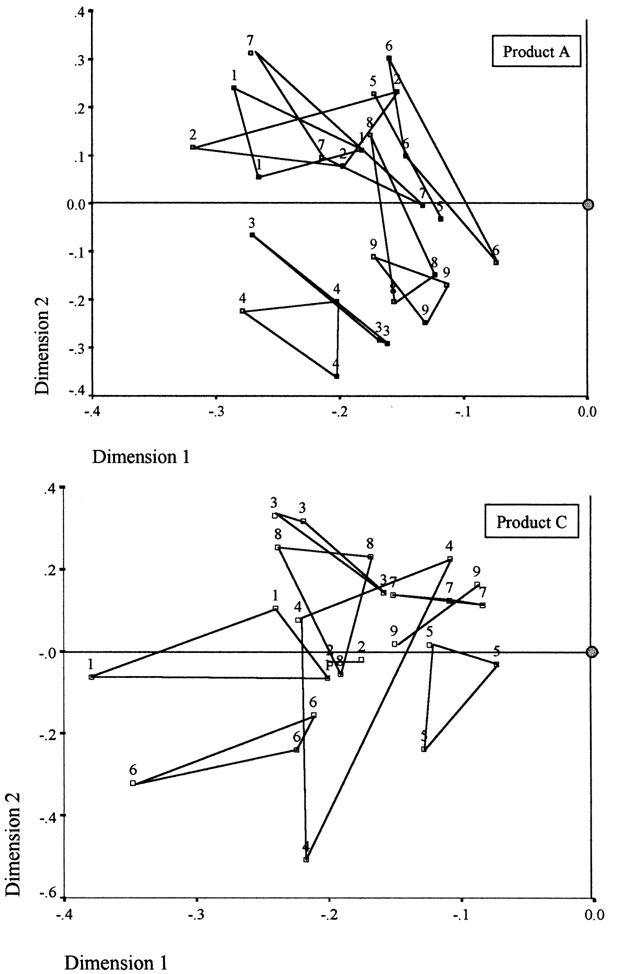

Fig. 13. Loading plots for the panellists for smoky, for product A (one replicate for panellist 5 removed) and product C (one replicate for panellist 9 and one for panellist 2 removed).

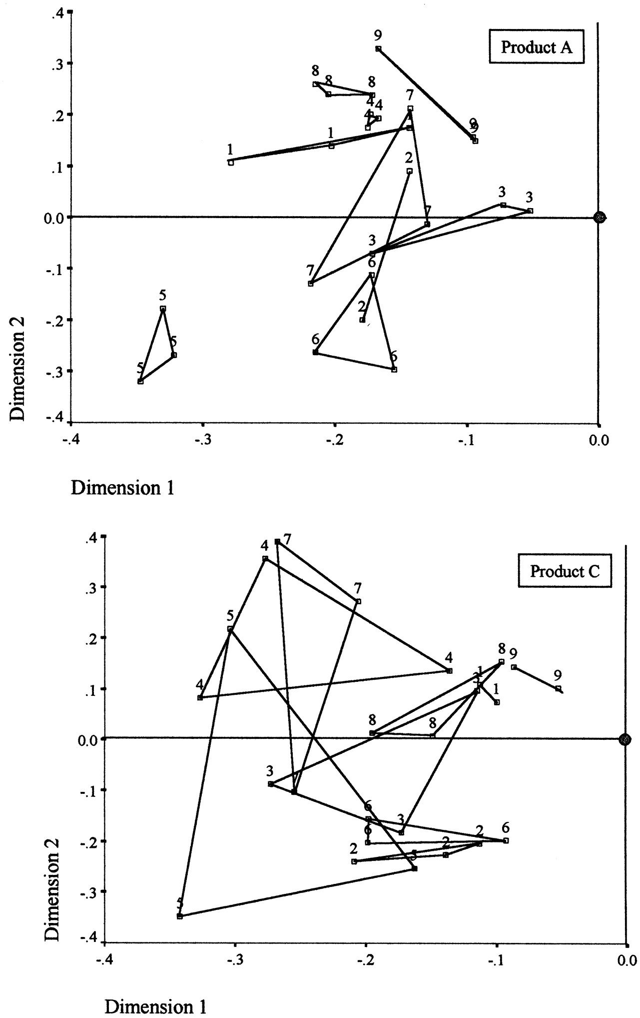

Fig. 14. Loading plots for the panellists for juicy, for product A (one replicate for panellist 2 removed) and product C (one replicate for panellist 1 and one for panellist 9 removed).

but it can be noticed that the panellist 5 is an outlier. The within consistency between the panellists is less than with the two other attributes. This was also seen with the raw curves and can be linked maybe to a problem for the assessors to distinguish between the amount of saliva flow and the release of juice from the product. The second group of panellists seem to correspond to the ones who record curves with a slower increase thus having longer curves than the others.

3.4. Results with respect to the two products

The loadings in Figs. 12–14 show a larger spread between the replicates for product C than for product A. The TI curves for the replicates of product A have been more alike than for product C. This could have been an effect of differences in the type of product, but also of inconsistency of the product C. However, this is secondary to the main objective of the study: training procedures for a sensory TI panel.

4. Conclusion

Performing a TI task is not a straightforward matter for sensory panellists. It is well known that large individual differences in TI curves exist between the panellists, each panellist has his-her own ‘signature’ in producing TI curves (van Buuren, 1992). Minimising individual differences would increase the gain in important information and aid interpreting the results.

It is recommended to perform a conventional profiling study prior to a TI experiment in order to check the feasibility of the attributes. One wants to be sure that the attributes used in the TI task are meaningful, i.e. discriminate between the products in the study, in addition the assessors must be able to use the attributes in a consistent way, i.e. they must not use the attributes to indicate different perceptions.

Running a GPA on profiling data is recommended for a first checking of possible outliers in the panel before going further on with the TI training, and for the choice of proper attributes to use in the TI assessments. It is advisable to add a pilot session with e.g. basic tastes to enable the panellists to become accustomed to the TI-task and thus reduce the training difficulty in learning the task with a more complex product. This could be especially useful when using panels on an ad hoc basis, i.e. a group of panellists used for a particular experiment, rather than having a specialised (TI) panel stand-by all the time. TI assessments of texture (juicy, watery) seem to be more difficult than TI assessments of flavour, which may impose a need for extra training. It is not easy for the panellists to distinguish between their own saliva flow and the moist from the product, especially after some chewing has taken place. This could be a topic for another study.

The use of PCA, and inspection of the loading plots in particular, is supported here as a means to check the panel consistency, as well as a source of information about the individual differences in the panel.

Acknowledgements

The research for this paper was carried out at ID-DLO, the Institute for Animal Science and Health, Lelystad, The Netherlands, while the first author visited the second. The work was part of the first author's master's thesis at ENSBANA. Roelof Rump and Gerdien Vonder of ID-DLO are acknowledged for their cooperation during the research reported here. Meester BV is thanked for providing the casseler rib used in the experiments. The final stage of completing this article was part of the FØTEK programme supported by the Danish Dairy Research Foundation (Danish Dairy Board) and the Danish Government.

References

Bloom, K., Duizer, L. M., & Findlay, C. J. (1995). An objective numerical method of assessing the reliability of time–intensity panellists. *Journal of Sensory Studies, 10*, 285–294.

Butler, G., Poste, L. M., Mackie, D. A., & Jones, A. (1996). Time-Intensity as a tool for the measurement of meat tenderness. *Food Quality and Preference, 7*, 193–204.

Civille, G. V., & Szczesniak, A. S. (1973). Guidelines to training a texture profile panel. *Journal of Texture Studies, 4*, 204–223.

Compusense Inc. (1996). Compusense Five version 2.2 user manual.

Dijksterhuis, G. B. (1993). Principal Component Analysis of Time-Intensity Bitterness Curves. *Journal of Sensory Studies, 8*, 317–328.

Dijksterhuis, G. B. (1997). *Multivariate data analysis in sensory and consumer science*. Trumbull, USA: Food & Nutrition Press.

Dijksterhuis, G. B. (1995). Assessing panel consonance. *Food Quality and Preference, 6*, 7–14.

Dijksterhuis, G. B., Flipsen, M., & Punter, P. H. (1994). Principal Component Analysis of TI curves: three methods compared. *Food Quality and Preference, 5*, 121–127.

Dijksterhuis, G. B., & Gower, J. C. (1992). The interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference, 3*, 67–87.

Dijksterhuis, G. B., Piggott, J. R. (2000). Dynamic sensory measurements. *Trends in Food Science and Technology*, submitted for publication.

Duizer, L. M., Gullet, E. A., & Findlay, C. J. (1993). Time–intensity methodology for beef tenderness perception. *Journal of Food Science, 58*(5), 943–947.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40*, 33–51.

ISO (1992). *International standard ISO 5492 — sensory analysis, vocabulary* (1st ed.) (pp. 6–9).

ISO (1994). *International standard ISO 11036 — sensory analysis, methodology, texture profile* (1st ed.).

Mioche, L., & Touraille, C. (1990). Texture profile set up for oral analysis of food. *Science Des Aliments, 10*, 697–711.

Pålsgård, E. & Dijksterhuis, G. B. (in press). The sensory perception of flavour release as a function of texture and time. *J. of Sensory Studies.*

Van Buuren, S. (1992). Analysing TI responses in sensory evaluation. *Food Technology* February, 101–104.