Food Quality and Preference 38 (2014) 24–29

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Panel performance for Temporal Dominance of Sensations

Melissa Lepage [a], Tara Neville [a], Andreas Rytz [a], Pascal Schlich [b], Nathalie Martin [a], Nicolas Pineau [a,*]

[a] Nestlé Research Centre, P.O. Box 44, Vers-chez-les-blanc 1000, Lausanne 26, Switzerland
[b] Centre des Sciences du Goût et de l'Alimentation, Dijon, France

ARTICLE INFO

Article history:
Received 23 December 2013
Received in revised form 14 April 2014
Accepted 3 May 2014
Available online 24 May 2014

Keywords:
Temporal Dominance of Sensations (TDS)
Panel performance
Panel discrimination
Panel agreement
Subject discrimination
Subject agreement

ABSTRACT

Given the particular nature of TDS data, specific analyses were developed and published over the past years to characterize the evolution of the dominant sensations in a product, to compare products to each other and to map the entire product space. Statistical approaches were also proposed to assess panel discrimination ability and agreement between subjects. To continue further on this direction, a general framework is proposed to evaluate panel and subject performance in TDS context. A protocol for testing products to evaluate the performance is proposed (before the measurement phase). Seven indicators are then described and summarized in a single table to facilitate the interpretation of the results by the panel leader. Three of the indicators describe panel and subject behavior in terms of attribute selection. The four remaining indicators describe the discrimination ability and the agreement, both at panel and subject levels. These indicators are inspired from usual ANOVA calculations used for descriptive profiling, but tested according to a permutation approach to overcome issues due to the nature of TDS data (non-independency between recorded values).

Introduction

With more than 20 papers published in scientific journals in the past 5 years, the Temporal Dominance of Sensation (TDS) methodology became an established method to evaluate the temporality of the sensory sensations. Focusing on the question of the statistical analysis of TDS data, several works proposed methodologies to represent TDS sequences at panel level (e.g. the TDS curves in Pineau et al., 2009) and map the products (e.g. PCA of parameters in Pineau et al., 2009, PCA trajectories in Lenfant, Loret, Pineau, Hartmann, & Martin, 2009, or canonical variate analysis in Meillon, Urbano, & Schlich, 2009 and Albert, Salvador, Schlich, & Fiszman, 2012) to identify product differences. Lenfant et al. (2009) also address the point of data standardisation to account for inter-individual evaluation durations due to variable mastication durations. This standardization ensures an equal weight for each evaluation and authors observed that it provides a better alignment between product sequences across subjects. This standardisation is therefore described as very useful when the tasting protocol does not fix the evaluation period, but it has limited interest when the protocol is time bounded.

Focusing on publications related to the performance of the panel, statistical tests were developed to test product differences (binomial approach in Pineau et al., 2009; permutation test in Meyners & Pineau, 2010) and to test panel and subject agreement (Meyners, 2011; permutation approach). As reported by the authors, the main advantage of the permutation approaches is their independency of assumptions, in particular regarding data distribution and auto-correlation. On the other hand, permutation tests can be time demanding and it is not always easy to run the appropriate permutations to fit initial design structure. A recent paper from Dinella, Masi, Naes, and Monteleone (2013) proposes a methodology to assess product differences in the TDS context according to a stepwise approach. The first step consists in the data visualization of the TDS curves to identify attributes, product pairs and/or subjects that are worth of investigation. Then, an ANOVA approach is proposed to assess product differences and panel performance. Compared to the permutation approach, the latter is based on more assumptions that might not fit perfectly with TDS data, but the authors confirm the validity of the approach since the distributions of the residuals fit with their assumption.

Considering the panel performance scope in sensory analysis, the ISO norm 11132:2012 proposes two alternative frameworks to check panel performance: the use of a dedicated session prior to product evaluation or directly the use of product evaluation sessions as such (as long as products were evaluated in replicates). In the context of highly trained panels dedicated to one specific product category and doing many evaluations every week, we believe having a dedicated session to check panel performance is more

* Corresponding author. Tel.: +41 21 785 9342.
E-mail address: nicolas.pineau@rdls.nestle.com (N. Pineau).

http://dx.doi.org/10.1016/j.foodqual.2014.05.002
0950-3293/

appropriate as already described in Labbe, Rytz, and Hugi (2004) and detailed in Lawless and Heymann (2010).

In this paper, we therefore propose a testing protocol for assessing panel performance in TDS experiments together with the corresponding data analyses and keys for interpretation. The analysis is a single step approach that does not require data previsualization and selection of attribute, product and/or subject of interest prior to the analysis. Results are summarized into a single table based on the calculation of seven performance indicators. Three of the indicators describe panel and subject behavior in terms of attribute selection. The four remaining indicators describe the discrimination ability and the agreement, both at panel and subject levels. These indicators are inspired from usual ANOVA calculations used for descriptive profiling, but tested according to a permutation approach to overcome issues due to the nature of TDS data (non-independency between recorded values) that might occur when all attributes, products and subjects have to be considered.

Materials & methods

TDS data

An experiment will be comprised of n_s subjects, n_r replications, n_p products, n_a attributes and n_t timepoints.

Most commonly, a TDS evaluation represents an assessment of a product over mastication time or aftertaste period. Subjects select which attribute they perceive as dominant using computerized technology. By convention, a subject can only select one dominant attribute per timepoint. The selected attribute is considered dominant until the subject changes his/her selection or terminates the evaluation. The sequence of selections are recorded as binary responses and unfolded into a data matrix of dimension n_a by n_t.

To illustrate the approach described below, the dataset with the following characteristics was used. A panel of 16 subjects trained to the TDS methodology evaluated four chocolate wafer bars in four replicates (4 sessions) according to a list of 9 texture attributes. The list of attributes was generated according to previous work on the same product space (descriptive profiling not published) and refined by the trained subjects through two training sessions.

TDS data pre-processing

All results presented in this article refer to standardized TDS data to ensure an equal weight to each evaluation. This also provides an easier framework to data computation since the length of each evaluation is the same, but the same approach could be applied to non-standardized data. The choice for data standardization or not depends on the nature of the products, the objective of the study and the tasting protocol (Lenfant et al., 2009) and this choice is not in the scope of the present paper.

In the context of panel performance for TDS, the main objective consists in assessing whether the panel is able to detect product differences in a consistent way across subjects, but it is not about getting a detailed understanding of the nature of these product differences. Consequently, there is no need for a high level of granularity in the temporal response (typically $n_t = 100$). As proposed by Pineau, Neville, and Lepage (2011) and further used by Dinella et al. (2013), time points can therefore be aggregated into time intervals to simplify results interpretation. In this article, we choose to aggregate TDS data into three time periods ($n_t = 3$) of equal length to respectively represent the beginning, the middle and the end of the evaluation period (Fig. 1). For each attribute and time period, the aggregated value represents the dominance duration of the attribute in the given time period divided by the duration of the time period. For each attribute, assuming data are coded as "0" if the attribute is not dominant for a given time point and "1" if it is dominant, this can be simply calculated as the average of the 0/1 values over the time points within the given time period. For each evaluation of one product, TDS data can therefore be displayed in an attribute*time period matrix, as displayed in Fig. 1. Each value represents the frequency of selection of an ttribute in the given period (value between 0 and 1), the sum of all values for a given period being always equal to 1. For instance, the data from the example of evaluation displayed in Fig. 1 for time period $T1$ is summarized as: attribute crispiness was dominant for 29% of the period $T1$ (crispiness dominance = $2/7 \approx 0.29$) and attribute hardness was dominant for 71% (hardness dominance = $5/7 \approx 0.71$). Other attributes were never selected as dominant for this period so the value is 0.

Panel performance protocol

The same product presentation order is used for each subject and each replicate in order to get unbiased estimate of subject effect, subject*product interaction and subject repeatability. Since the focus of a priori panel performance assessment is on the subject performance, the fact that product and order effects are confounded with this protocol is not an issue.

The minimum number of replicates needed to evaluate individual performance is two. However, in the context of TDS evaluation, data collected are selections of attributes, i.e. 0/1 data. This is not an issue at panel level given the total number of evaluations, but at subject level this would result in a low level of precision in the estimate of their performance. It is therefore recommended to use more replicates, at least during the performance evaluation, say three to four.

The attribute list used for the performance evaluation is the same as the one expected to be used for the product evaluation. As highlighted in Pineau et al. (2012), the selection of the attributes to be part of the list is of high importance since the TDS task consists in a choice among these terms. This article recommends the use of a relatively small number of attributes, say about 10.

As described in Lawless and Heymann (2010) and Labbe et al. (2004) about the good practices for panel performance in descriptive profiling, it is recommended to use a small number of products per session to minimize subject's boredom and time resources (say four to six, so that they can all be tested in one session, i.e. replicate = session). The selected products have to ensure a good covering of the product space that will be tested in the product evaluation phase.

Three panel behavior indicators

Three indicators have been defined to characterize the way the panel is using the attributes in the list:

- At panel level and for each attribute, the *maximum frequency of selection (% max freq)*, i.e. the maximum dominance rate observed on a TDS curve (Pineau et al., 2009) across all products. This indicates how often an attribute is selected at maximum for one of the products. Very low value for this indicator evidence attributes that are questionable in the list. Typically, attributes with maximum frequency of selection below the significance level (i.e. the significance level is never reached for any product) should be questioned and potentially removed from the list.
- For each subject, the *average number of attribute selection per evaluation (n Att Sel/eval)*. This indicator gives some clues about the differences in individual behaviors in front of the attribute