Food Quality and Preference 47 (2016) 79–90

Contents lists available at ScienceDirect

Food Quality and Preference

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

Temporal Check-All-That-Apply (TCATA): A novel dynamic method for characterizing products

John C. Castura [a,*], Lucía Antúnez [b], Ana Giménez [b], Gastón Ares [b]

[a] Compusense Inc., 255 Speedvale Ave. W., Guelph, Ontario N1H 1C5, Canada
[b] Facultad de Química, Universidad de la República, Gral. Flores 2124, C.P. 11800 Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form 14 May 2015
Accepted 24 June 2015
Available online 25 June 2015

Keywords:
Temporal Check-All-That-Apply
TCATA
Temporal sensory method
CATA questions
Yogurt

ABSTRACT

Temporal Check-All-That-Apply (TCATA) is introduced as a new dynamic method for describing multidimensional sensory properties of products as they evolve over time. TCATA extends the Check-All-That-Apply (CATA) method. Selection and deselection of attributes are tracked continuously over time, permitting assessors to characterize the evolution of sensory changes in products. TCATA is presented using results from trained panel evaluations of yogurt products. Data are also used to illustrate approaches for exploratory data analysis. Raw data from each assessor are represented using indicator charts. Panel data are aggregated into TCATA product plots. Reference lines are added to provide additional guidance. Product pairwise comparisons are made in TCATA difference plots, emphasizing differences that are less likely to have arisen from chance. Correspondence analysis (CA) is used to visualize product trajectories over time in a sensory space, providing a summary multivariate understanding of the dynamic sensory properties. CA conducted on the TCATA yogurt data highlight the importance of the dynamic profile, and suggest that understanding the complexity of products requires investigation of temporal changes. Results indicate that the TCATA method has potential for evaluating temporal aspects of sensory perception but further research is required to identify methodological issues and to refine the methodology.

1. Introduction

One of the challenges in sensory evaluation is characterizing the rapidly evolving sensations that comprise the dynamic properties of products. Temporal sensory evaluation of foods and beverages is a complex endeavour, but also one that is relevant for understanding how products are perceived in the mouth during consumption. It comprises the evaluation of many in-mouth sensations (olfactory, gustatory, and trigeminal, as well as sound, texture, and temperature) that are not static but rather evolve over time (see Lawless & Heymann, 2010, Ch. 8). Various temporal sensory methods are available for characterization of dynamic product properties. In this manuscript, the Temporal Check-All-That-Apply (TCATA) method is introduced as a novel temporal sensory method, which could be used by assessors to characterize products.

Historically, temporal sensory evaluations have focused on attribute intensities, and have been conducted within the framework of analytical sensory evaluation. The best example is continuous time intensity (TI), which enables intensity measurement of one relevant attribute at a time. TI data are usually represented using curves, which show changes in attribute intensity between onset and extinction times (ASTM, 2013). Although Dual-Attribute Time Intensity (Duizer, Bloom, & Findlay, 1996) has been proposed for evaluation of two attributes simultaneously, continuous measurement is not possible for three or more attributes.

The traditional approach for measurement of three or more attributes involves descriptive analysis at specific, discrete time points during consumption (Lawless & Heymann, 2010, Ch. 8). Various methods have been proposed which pre-establish time windows for responses, e.g., Time-Scanning Descriptive Analysis (Seo, Lee, Jung, & Hwang, 2009), Multi-Attribute Time Intensity (MATI; Kuesten, Bi, & Feng, 2013), Time Related Profiling (Kostyra, Baryłko-Pikielna, Dąbrowska, & Wasiak-Zys, 2008), and Sequential profiling (Methven et al., 2010). Other methods involve determination of sensory properties at points of consumption, such as the intensity variation descriptive method (Gordin, 1987), or in sync with chew strokes, such as Progressive Profiling (Jack, Piggott, & Paterson, 1994).

* Corresponding author.
E-mail address: jcastura@compusense.com (J.C. Castura).

http://dx.doi.org/10.1016/j.foodqual.2015.06.017
0950-3293/© 2015 The Authors. Published by Elsevier Ltd.

The Flavour Profile method (Caul, 1957) captures several product concepts, one being the order of elicitation of flavour characteristics. In a food product, the early development of appropriate sensations is closely tied to the emergence of a few attributes at the beginning of the evaluation. The Temporal Order of Sensations (TOS; Pecore, Rathjen-Nowak, & Tamminen, 2011) method captures the order in which the first few (e.g. 3) attributes emerge with each bite (or sip) of a multi-bite evaluation. TOS data are strictly ordinal, allowing the possibility of paper-based data collection.

Temporal Dominance of Sensations (TDS; Pineau, Cordelle, & Schlich, 2003; Pineau et al., 2009) was originally proposed as a multi-attribute temporal sensory method that scaled the intensities of a sequence of *dominant* attributes. Pineau et al. (2012, p. 164) propose a variant of TDS in which the dominant attribute is selected, without scaling its intensity, an approach often followed by other researchers (e.g., Thomas, Visalli, Cordelle, & Schlich, 2015; Zorn, Alcaire, Vidal, Giménez, & Ares, 2014). Time-Quality Tracking (Halpern, 1991; Zwillinger & Halpern, 1991) is an earlier method that also captures a sequence of attributes without intensities.

Adams, Williams, Lancaster, and Foley (2007) proposed using Check-All-That-Apply (CATA) questions to allow consumers to indicate their sensory perception of the samples that were also being evaluated hedonically. CATA questions provide multivariate binary data which properly indicate the applicability of provided descriptors to the samples. CATA questions do not directly capture intensities. If an attribute is present in one product, and also occurs in another product at a different intensity level, it is possible that an assessor will check the attribute for both samples. There is no scale to permit an assessor to characterize the level difference. Nonetheless, evidence presented so far in the literature has shown good correlation between CATA frequencies and attribute intensities (Bruzzone, Ares, & Giménez, 2012; Reinbach, Giacalone, Ribeiro, Bredie, & Frøst, 2014). CATA questions are being used increasingly in consumer questionnaires, and have been the subject of numerous methodological investigations published in this journal (Ares & Jaeger, 2013; Ares et al., 2013, 2014; Jaeger et al., 2014). Methods for evaluating reproducibility and repeatability of CATA assessors have been proposed (Jaeger et al., 2013; Worch & Piqueras-Fiszman, 2015). Meyners and Castura (2014) provide a review of methodological considerations and Meyners, Castura, and Carr (2013) give statistical approaches for CATA questions.

This manuscript presents TCATA as a temporal extension of CATA. TCATA extends the use of CATA questions by allowing continuous selection and deselection of attributes based on applicability of the attribute to describe the sample. The layout of the TCATA question is much like a CATA question. The assessors' task is to indicate and continually update the attributes that apply to the sample moment to moment, that is, to track the sensations in the sample as it changes over time. Assessors are permitted to check attributes at times whenever applicable, and to uncheck attributes (or leave attributes not checked) whenever not applicable. Multiple attributes can be selected simultaneously, which may permit description of sensations that arise either sequentially or concurrently.

In the present work the TCATA method is presented, along with results from a TCATA study involving a trained panel.

2. Materials and methods

2.1. Temporal Check-All-That-Apply (TCATA)

In a TCATA evaluation of a single sample by an assessor in a particular session, the computerized data collection system displays the entire list of attributes on the screen. The layout of a TCATA question is much like a CATA question (cf. Meyners & Castura, 2014). Fixed order of attributes is avoided to reduce confounding effects related to attribute order and position; in the study presented herein, attributes were presented according to an experimental design to balance biases associated with attribute position. The number of attributes was kept to a maximum of 10, consistent with TDS methodological recommendations (Pineau et al., 2012), which seems tentatively appropriate for TCATA.

Assessors were instructed to review the attributes to facilitate the task of locating attributes during the TCATA evaluation. They were instructed to click a Start button concurrently with putting the sample into the mouth, and to immediately commence tracking changes in the sample by checking and unchecking words, such that the words that were selected described the sample in that moment. At any time between clicking Start and the end of the evaluation time, assessors were free to check any unselected attribute, or to uncheck any selected attribute. In a TCATA evaluation, it is possible that some attributes are never checked, that other attributes are checked but never unchecked, and that other attributes are checked and unchecked one or more times, ending in either the checked or the not checked state. Note that checking, unchecking, or not checking one attribute does not affect the possibility of checking, unchecking, or not checking any other attribute. Multiple attributes can be selected simultaneously according to when the attribute is considered applicable to describe the sample. It is important to stress that TCATA does not rely on the concept of dominance: assessors select attributes that are deemed applicable for describing each sample at each time slice.

Computer records are maintained and indicate each change in the checked status of each attribute, e.g., when an assessor checks a not-selected attribute or unchecks a selected attribute. The attribute and all time(s) that the attribute was checked and/or unchecked are recorded. Compusense at-hand 5.6 (Compusense Inc., Guelph, Ontario, Canada) used the JavaScript `Date` object method `getTime()` to obtain timestamps for the evaluation start and for each event, which are subsequently reported in the software at 10 ms (0.01 s) intervals. Data form a multivariate binary time series.

Data in this TCATA study are considered to be a time series of applicability data because assessors are instructed to describe samples in a manner similar to CATA. For this type of data, no explicit assumptions are made regarding the type attentional process that might result in selection (or deselection) of attributes. For example, it is possible that a sensation (or the lack of a sensation) will capture the assessor's attention, leading the assessor to update or maintain the current checked state of the corresponding attribute, as applicable. It is also possible that seeing an attribute on the screen, thinking of an attribute, or other process might direct the attention of the assessor to consider the applicability of that particular attribute for describing the sample being evaluated, and that the assessor subsequently updates or maintains the checked state of that attribute in response.

2.2. Stirred yogurt study

2.2.1. Yogurt products

Six commercial strawberry yogurt products, available in supermarkets in Montevideo, Uruguay were selected for inclusion in the test. All products were stirred yogurts with blended fruit (Greek yogurts were not included). Products differed in fat content (0% and 3.2%) and sugar content (9–17%). Some of the products were formulated with sweeteners. Samples were purchased and maintained in storage under refrigeration temperatures (4 °C ± 1 °C), and removed from the refrigerator as needed immediately prior to sensory evaluation, and dispensed into plastic serving cups. Samples were coded using three-digit blinding codes.

2.2.2. Yogurt assessors – trained panel

The sensory panel was originally comprised of 12 assessors, aged 22–48 years old, selected according to the guidelines of the ISO 8586:2012 standard (ISO, 2012). All assessors had been involved with prior descriptive sensory evaluation of yogurts and dairy products. The yogurt panel had no experience with evaluating strawberry yogurts using temporal methods. In the first session, assessors were presented with five commercial strawberry yogurt samples, representing a wide range of sensory characteristics. Assessors were asked to try the yogurts and to individually think of sensory attributes that changed over the time of the evaluation period. Then, through open discussion with the panel leader, assessors agreed on the best attributes to fully describe the dynamics of the sensory characteristics of samples. Attributes were defined and attribute references selected in open discussion (Table 1). During eight 10-minute training sessions, assessors were familiarized with the TCATA task and the software used for data collection. The importance of checking attributes based on their applicability for describing the sample, and unchecking attributes that no longer applied for describing the samples was stressed. The number of training sessions was pre-established tentatively based on training times used in some previous studies (e.g. Zorn et al., 2014), and not extended due to positive assessor feedback on attribute and task familiarity. One assessor attended only 1 of the 3 data collection sessions, and was subsequently dropped from the panel.

2.2.3. Yogurt study – experimental design and data collection

Six samples were presented to each assessor in sequential monadic format. To balance bias associated with sample order and first-order carryover, samples were presented according to a Williams Latin square design (Williams, 1949). There were 8 sensory attributes, selected by the trained assessors and the panel leader: Ácido (*Sourness*), Dulce (*Sweetness*), Sabor a frutilla (*Strawberry flavour*), Sabor extraño (*Off Flavour*), Cremoso (*Creamy texture*), Recubrimiento graso (*Fatty Mouthfeel*), Sabor artificial (*Artificial flavour*), and Sabor a crema (*Cream flavour*). Attributes were arranged in a three-column layout. Different attribute list orders were obtained from a Williams Latin Square design, with list orders allocated such that each assessor received a different attribute list order in each session, but within a session used the same attribute list order to evaluate all samples.

Assessors were instructed to commence the evaluation by pressing a Start button simultaneous to putting a tablespoon (15 mL) of yogurt in the mouth. TCATA data collection commenced upon clicking Start, at which time it became possible to select and deselect attributes, corresponding to the sensory terms that described the sample characteristics at each time. No precise instructions were given regarding the moment at which assessors should swallow the sample. The total duration of the test (30 s) was determined with trained assessors in the first training sessions in order to evaluate sensory characteristics after swallowing.

Results were collected for a single-spoon evaluation. Data from the study were collected on laptops using Compusense at-hand 5.6 (Compusense Inc., Guelph, Ontario, Canada), which was modified to permit collection of TCATA data. The trained panel evaluated each product in triplicate in 3 sessions. Data were collected in 3 sessions on separate days within a 12-day time frame in February-March 2014 at the Universidad de la República in Montevideo, Uruguay.

2.3. Data analysis

In this paper, data analysis focused on exploratory techniques, which were conducted using R 3.1.2 (R Core Team, 2014).

A unique combination of a particular assessor in a particular session will be referred to as a "TCATA run". Raw TCATA data can be denoted as having $a = 1, \ldots, A$ attributes, $k = 1, \ldots, K$ products, $r = 1, \ldots, R$ TCATA runs, and $t = 1, \ldots, T$ time slices. TCATA data create a multivariate binary time series; the response $x_{arkt} = 1$ if a was checked by r for k at t; otherwise, $x_{arkt} = 0$.

2.3.1. Visualization of a raw TCATA data

Data from each TCATA evaluation were arranged into matrices, $\boldsymbol{X}_{a,t}^{(rk)}$, i.e., one matrix per product per run, with attributes in rows and time slices in columns. Indicator charts were prepared for each evaluation from these matrices, which permit the visualization of the attribute selections $1, \ldots, A$ used in the run to characterize the sensory properties of product k at $1, \ldots, T$. Indicator charts permit inspection of elicitation onset and extinction times, elicitation durations, elicitation sequences, and elicitation co-occurrences. The plots are useful in understanding the characteristics for

Table 1
Definition and references of the attributes evaluated by the trained TCATA assessors.

Attribute	Definition	Reference
Sourness	Sour taste	Commercial sample of unsweetened stirred yogurt
Sweetness	Sweet taste	Plain yogurt formulated by the authors, containing 8% sugar and 0.015% sucralose
Strawberry flavour	Flavour characteristic of fresh and natural strawberry	Stirred yogurt formulated by the authors, containing strawberry puree
Off flavour	Non-characteristic flavour of strawberry yogurt	Commercial sample of plain stirred yogurt with probiotics, which was described mainly as chemical or plastic; the same off flavour appeared in one of the samples which also contained the same probiotic strain
Creamy texture	Sensation related to a product of smooth texture, homogeneous, with intermediate thickness and moderate melting rate	Plain yogurt formulated by the authors containing 3% milk fat and 1% modified starch
Fatty mouthfeel	Sensation caused by the film of fat remaining in the mouth after swallowing the sample	Milk with 20% added milk cream
Artificial flavour	Non-natural strawberry flavour	Commercial sample of strawberry-flavoured drinkable yogurt
Cream flavour	Flavour characteristic of milk cream	Milk with 20% added milk cream

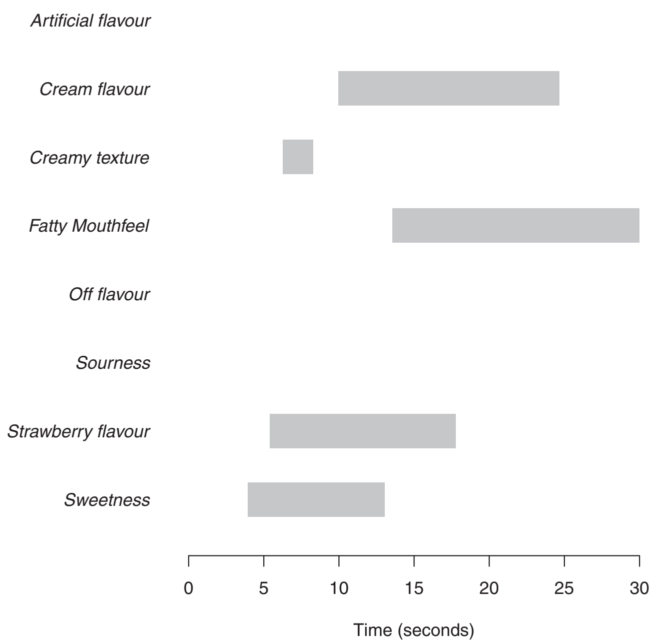

Fig. 1. TCATA indicator plot for Trained Assessor 1 for Yogurt Sample 1.

TCATA data, including the possibility for concurrent selection of attributes.

Fig. 1 provides the indicator chart based on raw data for a trained assessor who evaluated Yogurt 1.

2.3.2. Visualization of product data

2.3.2.1. TCATA curves. The number of citations for each product k for each attribute a and time slice t was obtained, and expressed as a proportion of the number of runs (R) to obtain the citation proportion p_{atk}. These citation proportions are calculated independently for each time slice. In TCATA, it is possible for assessors to check attributes concurrently, thus the sum of proportions can exceed 1. TCATA curves were created using these proportions based on the procedure illustrated in Fig. 2. Note the similarities between the procedure for aggregation for TCATA data and the procedure for aggregation of TDS data (cf. Pineau et al., 2009, Fig. 1). Smoothing is performed via the `smoothing.spline` function in R. (Smoothing is optional; any smoothing algorithm that fits the data might be considered to conduct this step.) Note that assessors are free to check any attribute that is not selected, and to uncheck any attribute that is selected; in Fig. 2, Assessor 2 has checked then unchecked *Strawberry flavour* twice in Session 2.

To provide context for the evolving citation proportions of attributes for product k relative to the other products in the study, reference lines are calculated, based on the citation proportions of pooled data from the other products in, $\{1,..,K\} \setminus k$ henceforth denoted $!k$. More specifically, the reference line is a line corresponding to k on, a and obtained by summing the number of citations for a for $!k$ at each t, and dividing by the number of evaluations $R(K-1)$ to obtain the citation proportion $p_{at(!k)}$. Again, these citation proportions are calculated independently for each time slice, and construction follows the procedure in Fig. 2.

To improve the legibility of the plots, we prefer to suppress the reference line a at t if there is evidence suggesting that the difference in citation proportions for product k and $!k$ is due to chance.

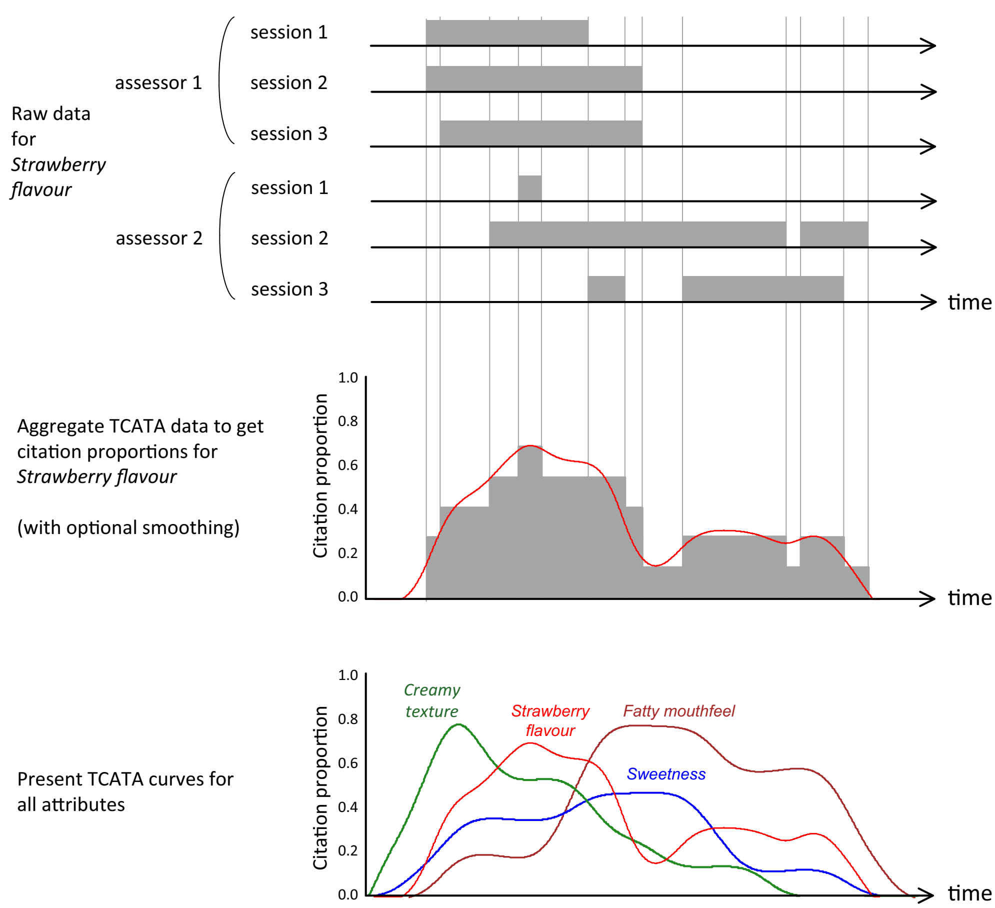

Fig. 2. The procedure for aggregation of TCATA data is illustrated using data for *Strawberry Flavour* for two assessors and 3 sessions.

Specifically, for each product k for attribute a at time t, we form the 2×2 table $\mathbf{C}$. Row 1 contains observations related to k, and row 2 contains observations related to $!k$. Data under columns 1 and 2 indicate that the attribute was checked and not checked, respectively. Note that sums for rows 1 and 2 are unequal but fixed by the experiment, i.e., R and $R(K-1)$, respectively. We use the two-sided Fisher–Irwin test (Fisher, 1935; Irwin, 1935) using Irwin's rule to test the hypothesis H_0 : no association of rows and columns versus H_1 : rows and columns associated. To conduct the test, all possible tables with the observed row and column sums are obtained. The p-value indicates the proportion of those tables with values as extreme or more extreme than the observed values under the marginal constraints. H_0 is rejected in favour of H_1 at level α if the citation proportions p_{atk} and $p_{at(!k)}$ are not homogeneous. In this case, the reference line is considered potentially informative, and is displayed. To further improve legibility of the plot, the weight of the TCATA curves is heavier than the reference curves, and TCATA curves are highlighted wherever the corresponding reference curves are displayed (e.g., Figs. 3 and 4).

2.3.2.2. TCATA difference curves. Pairwise product differences are often of interest to product developers. For consistency with the TCATA curves, differences in TCATA profiles of two products are obtained by taking the difference in TCATA citation proportions for each of the attributes at each of the time slices. Optionally, smoothing can be performed. The superimposition of all attributes onto a single TCATA difference plot permits inspection of the product differences over time. The line showing the difference in citation frequencies is suppressed for attribute a at time t if evidence suggests that the difference in citation proportions for products g and h could arise by chance. As in Section 2.3.2.1, a 2×2 table $\mathbf{C}$ is formed, with rows 1 and 2 indicating the split of checked and not checked responses for products g and h respectively. Row sums for each product are equal (R), and the data are again treated as if arising from a comparative trial, and evaluated using the two-sided Fisher–Irwin test as in Section 2.3.2.1. Wherever H_0 is rejected at level α in favour of H_1, the line showing the difference in citation proportions for products g and h is displayed (e.g. Fig. 5).

2.3.3. TCATA product trajectories

Given a contingency table of CATA counts with products in rows and attributes in columns, it is possible to summarize relationships between products and attributes using correspondence analysis (CA), and to visualize these results in a biplot, by focusing on the cosine of the angles between products and attributes (cf. Meyners et al., 2013). TCATA data can be organized as many contingency tables of CATA counts with products in rows and attributes in columns, with one such contingency table per time slice. If CA is conducted on the TCATA data, the product at each time slice forms a trajectory that reveals the sensory evolution of that product over time. The points along this trajectory are interpreted as any static point would be in CA, by checking the cosine of the angle between this point along the trajectory and the attributes.

In this paper, the trajectories are created by aggregating TCATA data into a contingency table. Rows consisted of a compound variable, i.e., a product at a time slice, where times started at 3 s and onward until the end of the evaluation, at 1 s increments. Attributes were in columns. Cells were citation frequencies. CA was conducted using the χ^2 metric using code adapted from the R package `ca` (Greenacre, Nenadic, & Friendly, 2014; Nenadic & Greenacre, 2007). CA plots were constructed to visualize results. Asymmetric maps were given, with rows in principal coordinates, and columns in standard coordinates multiplied by the square root of each corresponding column mass, an approach for visualizing CA results recommended by Greenacre (2006). For background information on CA, the interested reader is directed to Greenacre (2007). Product trajectories for the sequence of times points were smoothed along each dimension independently using the R function `loess` (Cleveland, Grosse, & Shyu, 1992). This approach for smoothing product trajectories is analogous to the one used by Gower, Lubbe, and Le Roux (2011, p. 119) to obtain smoothed object trajectories in a principal component analysis biplot. Examples of product trajectories are shown in Figs. 6 and 7.

2.3.4. Changes in selection and concurrent selections

TCATA data were investigated to understand the manner in which attributes were checked and unchecked to characterize

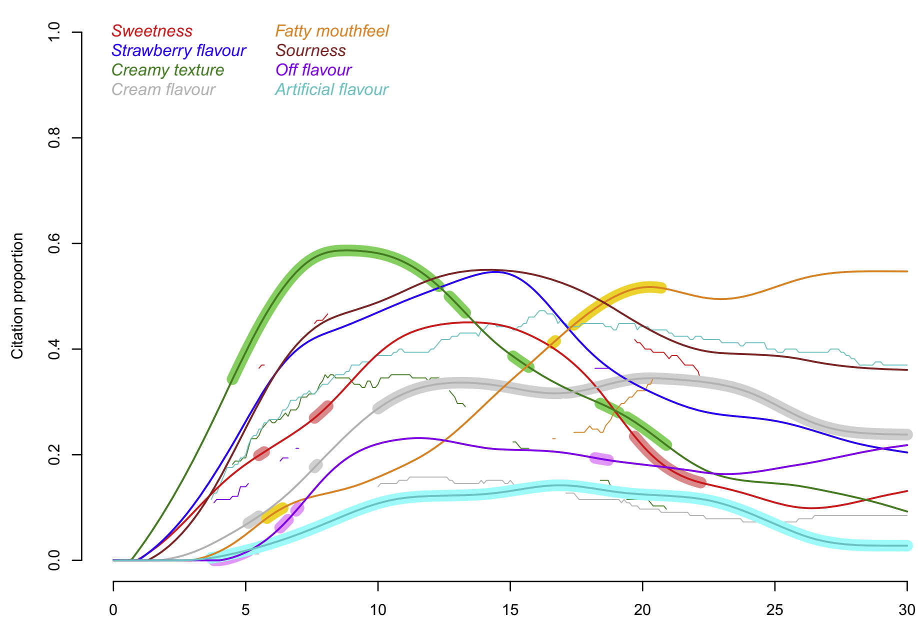

Fig. 3. TCATA curves for Yogurt 1. Solid lines indicate smoothed citation proportions for each of the attributes. Reference lines, which indicate citation proportions for data pooled from the other yogurts in the study, are shown in a lighter line weight only when Yogurt 1 is considered meaningfully different from other products. Where a reference line is shown, the corresponding product curve is highlighted.

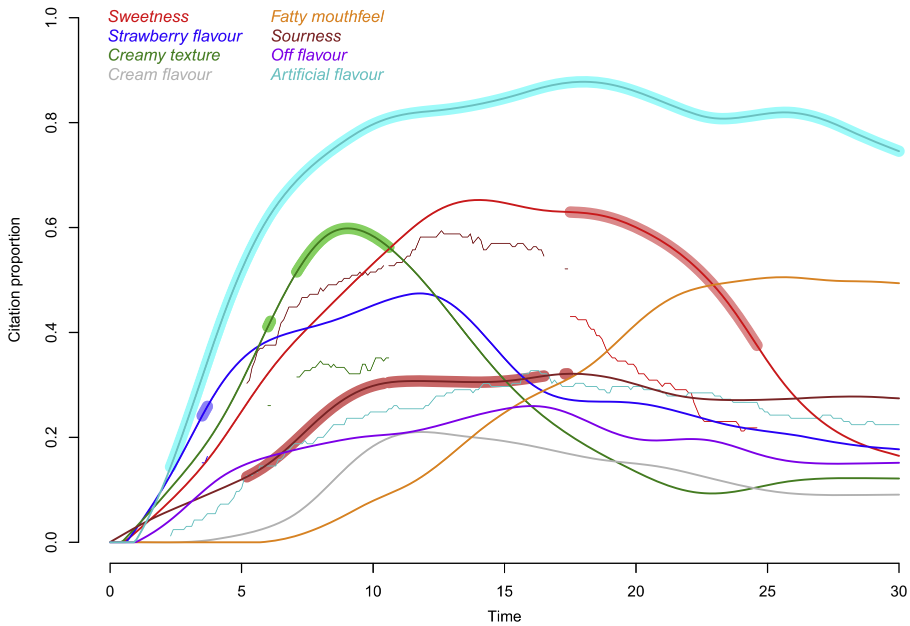

Fig. 4. TCATA curves for Yogurt 5. Solid lines indicate smoothed citation proportions for each of the attributes, and are highlighted when reference lines (which are shown in a lighter weight) are displayed.

samples in during the TCATA evaluations. We calculated the average number of citations per TCATA run. We also calculated the average number of attributes per evaluation that were checked then unchecked, as well as the average number of attributes per evaluation that were checked but not unchecked.

Concurrent attribute selection was also of interest. To investigate patterns of concurrent attribute selection in the data, data prior to the initial onset were discarded, providing data on an onset-trimmed timeline. I.e., if T is the total evaluation time, and O_{rk} is the end of the pre-onset time in the evaluation of product k in run r, trimming creates unequal trimmed evaluation times of duration $T - O_{rk}$, which were then time standardized (cf. Lenfant, Loret, Pineau, Hartmann, & Martin, 2009) to ensure that each run was given equal weight. Using the standardized trimmed data, we obtained a matrix indicating the number of concurrent attribute citations per run at each standardized time slice from 0 to 1, in 0.001 increments. The global average indicates the average number of attributes cited concurrently. Column averages of this matrix indicate that proportion of attributes selected concurrently along the average time-standardized TCATA evaluation.

2.3.5. Repeatability and agreement

Repeatability refers to the ability of an assessor to provide consistent data across multiple evaluations (cf. Rossi, 2001). Agreement refers to the extent to which an assessor provides data that are similar to those provided by other assessors. One approach for quantifying repeatability and agreement for TCATA data is proposed. If a response for a particular product on an attribute at a time slice is 0 or 1, and the same product, attribute, and time slice in a different session is 0 or 1, then the observations are either matched (1–1 or 0–0) or not matched (0–1 or 1–0). The proportion of matches to observations will indicate similarity. Alternatively expressed, the similarity coefficient $1 - d(\boldsymbol{x}, \boldsymbol{y})$ is calculated, where $d(\boldsymbol{x}, \boldsymbol{y})$ denotes the average city-block distance between two m-variate observations $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{m}\sum_{i=1}^{m}|x_i - y_i|. \quad (1)$$

The average city-block distance corresponds to the average Manhattan distance or the average L_1-distance. For binary data, it also corresponds to the average Hamming distance (Hamming, 1950). Note that $1 - d(\boldsymbol{x}, \boldsymbol{y})$ is a commonly used similarity coefficient (cf. Johnson & Wichern, 2007, p. 675), and will permit quantification of assessor repeatability (Section 2.3.5.1) and reproducibility (Section 2.3.5.2).

2.3.5.1. Repeatability.

To quantify the repeatability of assessor j over S sessions (where $S > 1$), it is possible to calculate the average city-block distance over all possible pairs of sessions under consideration. Let $p = 1, \ldots, P$, where $P = \binom{S}{2}$, and let x_p and y_p refer to the first and second session with the particular pair of sessions containing raw binary responses for each product, attributes, and time slice. The repeatability of assessor j is quantified

$$1 - d_{Rj}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{1}{P}\sum_{p=1}^{P} d(x_p, y_p). \quad (2)$$

The result $1 - d_{Rj}(\boldsymbol{x}, \boldsymbol{y})$ is a proportion, which quantifies the repeatability of assessor j. The repeatability index has a lower bound of 0 (no repeatability) and an upper bound of 1 (perfect repeatability).

The repeatability of a panel across sessions can also be determined using (2), where x_p and y_p are citation means for the panel across all assessors, rather than raw binary responses. The range of the repeatability index is again 0 to 1, with the same interpretation.

2.3.5.2. Agreement.

To quantify agreement between assessor j and the other assessors on the panel, the average city-block distance over all assessor pairs including assessor j is calculated. Let $\boldsymbol{q} = \{1, \ldots, J\} \setminus j$, where J is the total number of assessors, and assessor j is not in q. Let x_j be the average citation selection proportion for assessor j over all sessions, for each product on each attribute and time slice. Let y_q be the corresponding average citation

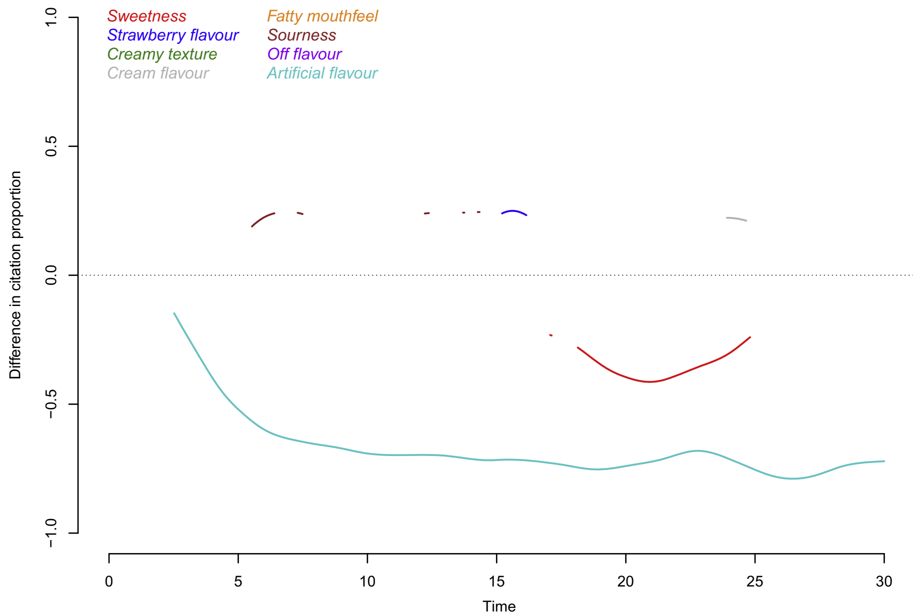

Fig. 5. TCATA difference curves between citation proportions for Yogurt 1 and the citation proportions for Yogurt 5.

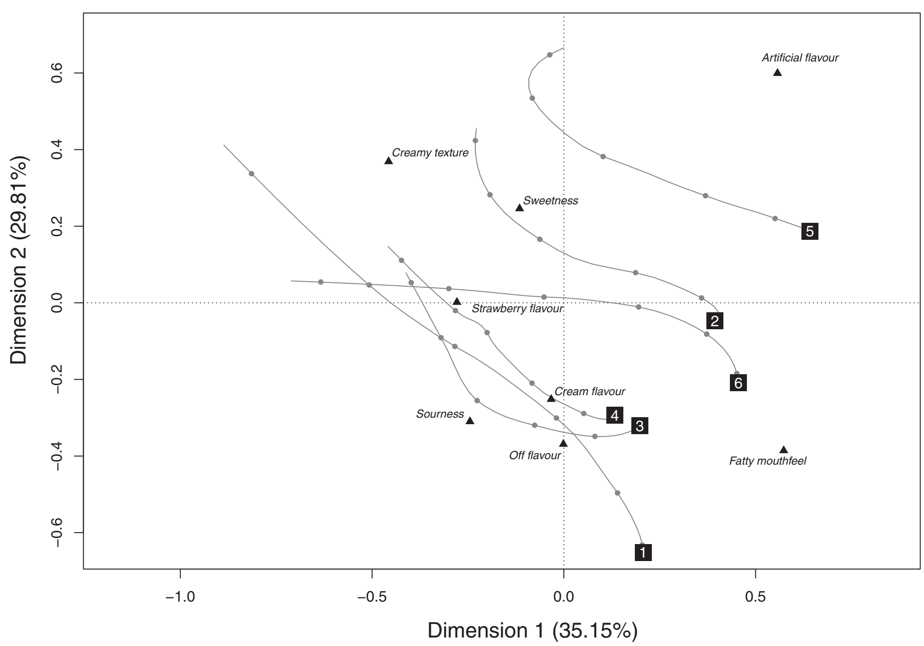

Fig. 6. Correspondence analysis for yogurt data, showing Dimensions 1 and 2. Product trajectories are smoothed, and terminate in a label indicating the sample number. Dots are shown at 5 s increments.

selection proportions of the other assessors in q. The agreement of assessor j with the other assessors can then be quantified

$$1 - d_{Aj}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{1}{J-1}\sum_{q=1}^{J} d(x_j, y_q). \quad (3)$$

Thus $1 - d_{Aj}(\boldsymbol{x}, \boldsymbol{y})$ quantifies agreement, with possible values between 0 (no agreement) to 1 (complete agreement).

3. Results

3.1. Results – single evaluation: indicator charts

Indicator charts were obtained using raw data, and used for inspection purposes. Inspection of indicator charts showed that the yogurt trained assessors used simultaneous selection of attributes to describe the samples being evaluated. Fig. 1 shows that the elicitation onset sequence for Yogurt 1 by Trained Assessor 1 is *Sweetness*, *Strawberry flavour*, *Creamy texture*, *Creamy flavour*, and *Fatty mouthfeel*. Onset and extinction times for each attribute are shown, as well as elicitation durations. For most of the evaluation period, two or three attributes are selected concurrently; exceptions are *Sweetness* (3.94–5.41 s) and *Fatty mouthfeel* (24.67–30.00 s) near the start and end of the evaluation, respectively, where these two attributes were the only ones checked. *Fatty mouthfeel* remained checked at the end of the evaluation. Three attributes (*Artificial Flavour*, *Off flavour*, *Sourness*) were not checked at all. Four of the five attributes that were checked were subsequently unchecked, but none of these attributes were subsequently checked a second time. Generally aggregated results will

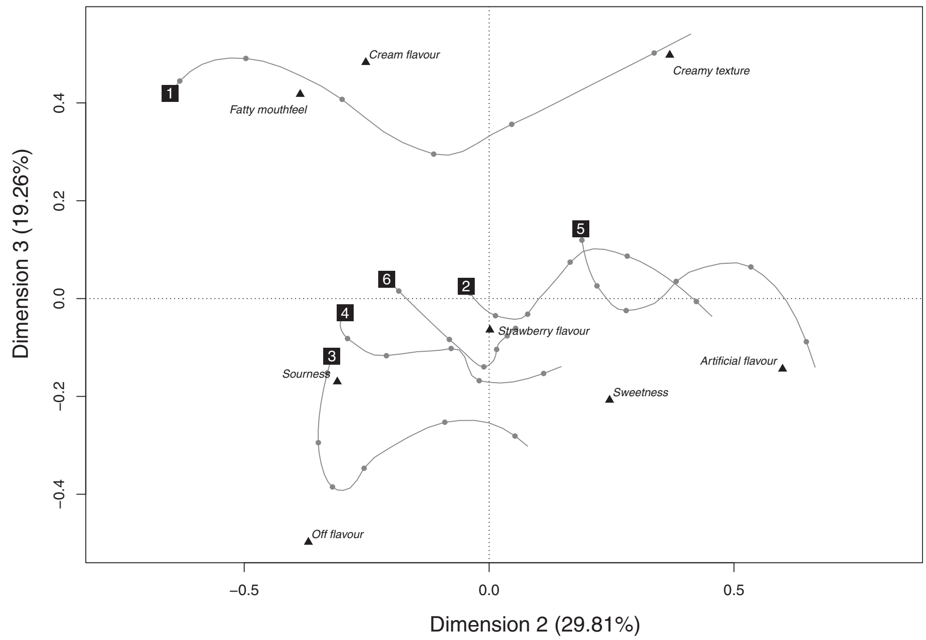

Fig. 7. Correspondence analysis for yogurt data, showing Dimensions 2 and 3.

be more useful, except when assessor performance is of interest. For brevity this will be the only indicator chart discussed in this manuscript.

3.2. Results – aggregated evaluations

TCATA curves are given for Yogurt 1 (Fig. 3) in thick solid lines. As described in Section 2.3.2.1, the thinner reference lines are shown conditionally, and indicate the attribute citation proportions based on pooled results for Yogurts 2–6 at times when the difference between citation rates for Yogurt 1 and the other yogurts is significantly different from zero. Highlighting is added to the corresponding segments of the TCATA curves to draw attention to these differences.

If Fig. 3 were interpreted using attribute citation frequencies alone, what would stand out are the high citation proportions for *Creamy texture* up to 12 s, followed by high citation proportions for *Sourness* and *Strawberry flavour* from 12 s to 18 s, and *Fatty mouthfeel* from 18 s to the end of the evaluation.

Reference lines provide additional guidance. *Creamy texture* is found to have high citation rates relative not only to other attributes in the early evaluation, but also relative to other products at that time, especially in the first 13 s. *Cream flavour* is cited with high citation rates relative to the other products, whereas the high citation rates for *Sourness* and *Strawberry flavour* in mid-evaluation are characteristics not only of Yogurt 1 but also of other products. The *Fatty mouthfeel* reference line indicates that its citation proportion increases for other yogurts near the end of the evaluation as well, although there is some suggestion that perhaps the increase is slightly more than average for Yogurt 1. *Artificial flavour* is cited with low citation proportions for Yogurt 1 almost throughout the evaluation period, and *Sweetness* is cited at lower rates for Yogurt 1 than other yogurts, especially between 20 s and 22 s.

TCATA curves are also presented for Yogurt 5 (Fig. 4), with reference lines and curve highlighting determined as described in Section 2.3.2.1. If Fig. 4 was to be interpreted using only attribute citation rates, the high citation proportions for *Artificial flavour* would stand out. The progression seen in other secondary attributes—especially *Creamy Texture*, *Strawberry flavour*, *Sweetness*, and *Fatty mouthfeel*—might also be noted. The addition of the reference lines suggests that citation proportions for *Fatty mouthfeel* are not particular to this product, but rather somewhat similar to the average of the other yogurts. There is a short-duration surge in citation rates for *Creamy Texture* between 5 s and 10 s in Yogurt 5, and *Sweetness* is used to describe Yogurt 5 somewhat more often than the other yogurts between 17 s and 25 s. The reference lines also highlight the relatively low *Sourness* citation proportion for Yogurt 5 relative to other yogurts through the first half of the evaluation period.

The differences between Yogurts 1 and 5 are investigated further via TCATA difference curves (Fig. 5). As expected, Yogurt 1 receives fewer citations for *Artificial flavour* throughout the evaluation period, and has fewer *Sweetness* citations between 18 s and 25 s. There is some suggestion of Yogurt 1 having slightly higher citation rates for *Sourness*, *Strawberry flavour*, and *Cream flavour* than Yogurt 5 in early-, mid-, and late-evaluation, respectively; these short-lived differences are best treated as spurious, but could be investigated further if they were of interest.

3.3. Results – TCATA product trajectories

Yogurt trained panel data were prepared as described in Section 2.3.3. The χ^2 test provides evidence that rows and columns of the matrix are dependent ($X^2 = 3019$ on 1169 degrees of freedom), justifying further investigation via CA. Dimensions 1 and 2 (Fig. 6) illustrate a temporal pattern of attribute citations; the upper left quadrant is associated with attributes that tend to be elicited early in the evaluation (*Creamy texture*, *Sweetness*, *Strawberry flavour*), whereas the lower right quadrant is associated with attributes that tend to be elicited later in the evaluation (*Fatty mouthfeel*, *Cream flavour*, *Off flavour*). The first dimension is mainly defined by *Fatty mouthfeel* and *Artificial flavour* in opposition to *Creamy texture*, *Strawberry flavour*, and *Sourness*. The second dimension is mainly defined by *Sourness*, *Off flavour*, and *Fatty mouthfeel* and their opposition to *Artificial flavour*, *Creamy texture*, and *Sweetness*. All yogurts follow a similar evolution, for example, by *Creamy texture* at the start of the evaluation and by *Fatty mouthfeel* towards the end, as revealed by the somewhat parallel paths in the sensory space.

Yogurts 1 and 5 have very low and higher citation frequencies, respectively, for *Artificial flavour* than would be expected if products and attributes were independent, consistent with Figs. 4 and 5. Both yogurts have more citations for *Creamy texture*, *Cream flavour*, and *Fatty mouthfeel* in early-, mid-, and late-evaluation respectively than would be expected if attributes and products were independent. Yogurt 5 is more aligned with *Sweetness* and less aligned with *Sourness*, while Yogurt 1 has the reverse of these associations. Other products can be interpreted in likewise manner, guided by angles between product and attribute at different time slices along each product trajectory for interpretation.

CA provides an effective multivariate summary based on the χ^2 metric, but a disadvantage is that a summary view might be misinterpreted. Reliance only on Fig. 6 for the interpretation of Yogurt 1 might suggest that this product receives many citations for *Off flavour*; however, this is not at all true, and it just happens that Yogurt 1 has similarities to two other products, Yogurts 3 and 4, which receive high citation frequencies for *Off flavour*. The relationships between Yogurt 1 and *Off flavour* is clarified in Dimension 3 (Fig. 7), in which high citations for *Cream flavour* is shown to be a characteristic of this yogurt.

3.4. Results – changes in selection and concurrent selections

There were 4.35 attributes checked over the course of the average single-bite TCATA evaluation of a yogurt sample. In an average run, 2.53 of checked attributes (58%) were subsequently unchecked before the evaluation ended at 30 s, and 1.82 checked attributes (42%) remained checked when the evaluation ended at 30 s. More than 95% of TCATA runs had one or more attributes still checked when the maximum time was reached.

Times were standardized as described in Section 2.3.4, putting each TCATA evaluation on a 0 to 1 timeline. The average number of attributes selected concurrently within a TCATA run was determined at each standardized time slice. The number of concurrent selections at standardized time 0 corresponds to the first onset attribute for each TCATA run, at which, by definition, the average number of attributes selected is 1. At its peak, there were an average 3.08 attributes selected concurrently, which occurred between standardized times 0.35 and 0.37. Standardized time 1 corresponded to the end of the evaluation, at which there were 1.86 attributes selected concurrently.

The slight discrepancy between the number of attributes that remained checked at the end of the evaluation on the raw vs. standardized timeline occurs because in 9 of 198 TCATA runs all attributes were set to a not-checked state prior to reaching the maximum time (30 s) after which no further attributes were checked.

3.5. Results – repeatability and agreement

Assessor repeatability was quantified as described in Section 2.3.5.1. Trained Assessor 1 was observed to have the best repeatability index (0.84). Trained Assessors 9 and 5 have the lowest repeatability indices (0.69 and 0.72). The panel's session-to-session repeatability was 0.77.

Assessor agreement was quantified as described in Section 2.3.5.2. The average agreement index amongst the assessors was 0.72. Trained Assessor 1 was observed to have the highest agreement index (0.74), indicating the best agreement with the other assessors. Again, Trained Assessors 5 and 9 had the lowest agreement with the other assessors, based on their agreement indices (0.69 and 0.70).

4. Discussion

TCATA is a novel method for dynamic sensory characterization. It is based on continuous selection of sensory attributes that are considered applicable to describe samples. In the present work the method was used by trained assessors to evaluate strawberry yogurts. The approaches proposed for data analysis allowed exploring different aspects of the dynamics of sensory perception.

Results presented here show that trained assessors selected attributes concurrently, and both checked and unchecked attributes to track dynamic sensory changes in products, suggesting that they understood and could perform the task.

Indicator charts (cf. Fig. 1) allow elicitation onset and extinction times to be reviewed visually, along with elicitation durations, elicitation sequences, and elicitation co-occurrences. Cross-referencing an indicator chart for a single assessor's sample evaluation with other results provides insight into whether the assessor is providing a response that is similar to, or different from, the panel. However, such a procedure would be unwieldy if there are many TCATA runs.

TCATA curves (Figs. 3 and 4) permit visualization of aggregated TCATA data, and enable identification of the main sensory characteristics of the products at each moment of the evaluation. Although these plots are analogous to TDS curves (cf. Pineau et al., 2009), the following differences are highlighted:

(i) TCATA is not based on the concept of dominance: attributes are selected according to whether they are applicable to describe the sensory characteristics of the samples.
(ii) The TCATA task permits concurrent selections, thus the sum of line heights may exceed 1.
(iii) There are no chance and significance lines provided for the hypothesis test H_0 : all attributes cited with equal proportion versus H_0 : one or more attribute citation proportions unequal.
(iv) Rather, for each attribute and each product, the TCATA citation proportion for product k can be compared to the citation proportions for the other products (earlier, denoted $!k$). A reference line can be shown at time slices where differences in citation proportions are significantly different from zero.

TCATA difference curves (Fig. 5) permit visualization of differences in TCATA attribute citation proportions in a manner analogous to TDS difference curves (cf. Pineau et al., 2009), with line heights in the interval $(-1, 1)$. As with TDS difference curves, the differences in the TCATA citation proportions are only shown at time slices where significantly different from zero.

Note that the significance in both TCATA curves and difference curves is determined heuristically. The objective approach is based on the two-tailed Fisher–Irwin test. However, it must be acknowledged that this test is technically flawed for such applications because the binomial proportions being compared are not independent (cf. Meyners & Pineau, 2010). Thus the Fisher–Irwin test is offered only to provide descriptive guidance. Development of randomization tests for making statistically valid inferences on TCATA data for curves and difference curves is an area requiring additional research.

TCATA provided valuable insights on the differences in the dynamic sensory profiles of the yogurts. Strawberry flavour and the taste of sweetness are congruent sensations, and influence the perception of one another (Frank & Byram, 1988). In this study, strawberry flavour was evaluated using two attributes: *Strawberry Flavour* and *Artificial Flavour*, defined as natural and non-natural, respectively, by both definition and reference (Table 1). The strawberry character in Yogurts 1, 3, and 4 was more frequently characterized as natural than other yogurts, while Yogurts 2, 5, and 6 were more frequently characterized as non-natural than other yogurts (cf. Fig. 6). Generally the high citation rates for *Sweetness* tended to be associated with high citation rates for strawberry flavour, regardless of its natural or artificial quality; however, artificial strawberry flavour tended to maintain higher citation rates for a longer duration (cf. Fig. 4) than natural strawberry flavour, which was shorter lived. Thus, high (low) *Sweetness* citation rates (as defined by reference lines) tended to co-occur with high (low) *Artificial Flavour* citation rates (cf. Figs. 6 and 7).

Fat-reduced yogurts are often observed to have lower intensities of sweetness, butteriness, and creaminess, and potentially higher intensities of sourness, bitterness, and astringency (Majchrzak, Lahm, & Dürrschmid, 2010). Results from this TCATA study, which are based on citation frequencies, are in accordance with these observations regarding the flavours that tend to co-occur. Significantly lower citation rates for *Sourness* than other yogurts were concurrent with higher citation rates for *Creamy texture* and/or *Fatty mouthfeel* (e.g. Fig. 4), and vice versa. Yogurt 1 had higher citation rates for *Cream flavour* than the other yogurts, and also had high citation rates for *Creamy Texture* and *Fatty mouthfeel* (see Figs. 3 and 7). When *Off Flavour* was cited at rates higher than other products, it tended to co-occur with low citation rates of *Creamy texture* and *Fatty mouthfeel* (cf. Fig. 7). In general, yogurts without significant off-flavours in this study were perceived either as having a creamy texture and/or a fatty mouthfeel (e.g. Fig. 3), or as having an artificial strawberry flavour and as sweet at higher rates for a longer duration (e.g. Fig. 4).

CA biplots (Figs. 6 and 7) provide a multivariate summary of product characteristics and changes. TCATA data were kept in original units, trimming data prior to 3 s due to low citation rates. The product trajectories are obtained from a double row structure rather than modelling product and time separately, but the interpretations fit well with the individual TCATA curves.

One methodological issue with this study is the lack of instructions to assessors regarding the time of swallowing. It might be expected that particular sensations occur prior to, concurrent with, or following a swallowing event. Thus, a potential methodological refinement to the method presented is to either provide direction to assessors for swallowing (or, if permitted, expectorating) at a specific time, or to permit assessors to indicate when the sample is swallowed or expectorated.

Assessor repeatability, quantified in Section 3.5, straddles a wide range. Note that repeatability can be high even in the absence of sensory discrimination, so it requires additional information to be actionable. Assessor agreement, also quantified in Section 3.5, has a narrower range. No assessor stands out as providing responses that are unusually dissimilar from the other assessors, but it deserves mention that Trained Assessors 5 and 9 had the lowest indices for both repeatability and agreement.

Refinement of the proposed coefficients for measuring repeatability and agreement is possible. It would also be possible to investigate repeatability and agreement for individual attributes, and at each time slice, to determine whether any assessors deviate from the others mainly on particular attributes or at particular time intervals. Other similarity metrics (cf. Johnson & Wichern, 2007, p. 675) might also be considered, as well as hypothesis tests to investigate whether assessor repeatability or agreement differs significantly from levels that would be expected by chance. Either these indices, or more refined versions of these indices, could be used to facilitate panel monitoring.

In many cases attributes remained checked at the end of the evaluation period. This phenomenon is noted in Section 3.4, and is evident in Figs. 3 and 4, where all citation proportions do not return to zero. One possibility is that TCATA assessors might have found the task too demanding, and they were unable to deselect attributes efficiently as a result. This possibility cannot be refuted with the observed data. Another possibility that seems more plausible is that the evaluation period was not long enough to follow all sensations to extinction. A published TDS study on yogurt and yogurt-like products (Bouteille et al., 2013) often had samples that did not reach extinction of sensations until nearly 60 s after starting the evaluation, which is more than twice as long as the time provided to assessors in the TCATA yogurt study presented here. Our emphasis for this study was on the important up-front attributes, but in retrospect a longer evaluation time should have been provided to permit investigation into this aspect of the TCATA methodology, which a future study will evaluate in greater detail.

Along similar lines, a potential refinement to the TCATA method is to make attribute selections more ephemeral in nature, where the appearance of a selected attribute gradually fades until it becomes obvious to the assessor that the attribute is no longer selected. The assessor can re-endorse the same attribute if it remains applicable. This strategy was recently introduced for TDS (Thomas et al., 2015), where the dominant attribute selection fades to a deselected state within 3 s. A similar modification to the TCATA method could be investigated. One of the characteristics of TCATA is that it permits concurrent attribute selection, which may permit assessors to indicate sensations that are perceived simultaneously (cf. Zwillinger & Halpern, 1991). In this study, multiple attributes were selected concurrently in many TCATA runs and for a substantial duration, but attributes were also unchecked, as instructed by the task. If, after having made a refinement whereby TCATA attributes are selected ephemerally, and TCATA assessors routinely re-endorse the same attributes repeatedly, it would add stronger experimental evidence to indicate that assessors are capable of both perceiving and expressing multiple attributes concurrently.

5. Conclusions

In this paper, Temporal Check-All-That-Apply (TCATA) is introduced as a new dynamic sensory method. TCATA provides a continuous time extension of CATA data collection, and builds on the Time-Quality Tracking method (Halpern, 1991; Zwillinger & Halpern, 1991) and the TDS method (Pineau et al., 2009) by permitting concurrent selections. TCATA was used by trained assessors to evaluate yogurt. Results in this paper describe the TCATA method and illustrate some techniques for exploratory data analysis, providing insights into the temporal properties of the evaluated samples, and identifying differences amongst products in their dynamic sensory profiles. Results show that the method has potential for characterizing products based on their sensory temporal properties.

TCATA is an extension of CATA, which is often used with consumers without attribute orientation. In the study reported here, TCATA is used with a trained panel, and only brief attribute-specific training was conducted. The authors are actively investigating the TCATA method. Active projects include a study to apply TCATA with consumers, to evaluate differences between TCATA results and TDS results for the same products, as well as between TCATA results and CATA results for the same products, and to investigate its potential to characterize the temporal evolution of sensations in a range of food and non-food products. It is expected that contrasting results from different methods will illuminate advantages and limitations of each methodology, and increase understanding of the data that each method provides, such that the most appropriate methodology can be selected when designing a study.

Methodological investigations are required to understand and mitigate potential biases. Appropriate methods for data preprocessing, based on the initial onset times, could be investigated, and may depend on experimental protocols and also on the product category. It is possible that assessors will have different thresholds for checking a not-selected attribute vs. unchecking a selected attribute, and further research could help to understand such factors. For example, refinements of the TCATA method, such as by making each attribute selection ephemeral, requiring assessors to re-endorse attributes, could be investigated to determine how such modifications to the task affect TCATA data.

Various elaborations on the TCATA method are possible. Whereas instructions yielded applicability data in this study, the instructions could be altered to focus on presence/absence of particular sensory attributes, or intensity above a baseline provided by a memorized reference sample, especially with a trained panel, or even dominance or noticeability. It would be possible to organize TCATA studies in which assessor chew rates and swallowing times are regulated or recorded, or in which multiple exposures (e.g. multiple bites or sips) to products are tracked. Complex stimuli, including heterogeneous products, could also be investigated using TCATA. Improved indices could be developed to further investigate repeatability and agreement of TCATA evaluations. Collection of TCATA data from consumers could be supplemented with collection of liking or other consumer responses, including emotions or purchase intention.

Acknowledgements

Trained assessors and panel leaders, particularly Leticia Vidal at the Universidad de la República, are acknowledged for their efforts in this study. James Castura, Alan Leonard, and other members of the Compusense Development team are acknowledged for their efforts in implementing the TCATA method to enable data collection. The authors are grateful to Comisión Sectorial de Investigación Científica (Universidad de la República, Uruguay) for financial support. We are grateful to two anonymous reviewers who provided constructive feedback on an earlier draft of this manuscript, as well as Thierry Worch, who provided helpful comments.

References

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check all-that-apply response compared to traditional scaling of attributes for salty snacks. In *7th Pangborn Sensory Science Symposium*, 12–16 August, Minneapolis, MN, USA.

Ares, G., Etchemendy, E., Antúnez, L., Vidal, L., Giménez, A., & Jaeger, S. R. (2014). Visual attention by consumers to check-all-that-apply questions: Insights to support methodological development. *Food Quality and Preference, 32*, 210–220.

Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference, 28*, 141–153.

Ares, G., Jaeger, S. R., Bava, C. M., Chheang, S. L., Jin, D., Giménez, A., et al. (2013). CATA questions for sensory product characterization: Raising awareness of biases. *Food Quality and Preference, 30*, 114–127.

ASTM International. (2013). *E 1909-13 Standard guide for time-intensity evaluation of sensory attributes (E1909-13)*. West Conshohocken, PA, USA: ASTM International.

Bouteille, R., Cordelle, S., Laval, C., Tournier, C., Lecanu, B., This, H., et al. (2013). Sensory exploration of the freshness sensation in plain yoghurts and yoghurt-like products. *Food Quality and Preference, 30*(2), 282–292.

Bruzzone, F., Ares, G., & Giménez, A. (2012). Consumers' texture perception of milk desserts. II – Comparison with trained assessors' data. *Journal of Texture Studies, 43*, 214–226.

Caul, J. F. (1957). The profile method of flavor analysis. In E. M. Mrak & G. F. Stewart (Eds.), *Advances in food research* (Vol. 7, pp. 1–40). New York: Academic Press.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S*. Wadsworth & Brooks/Cole.

Duizer, L. M., Bloom, K., & Findlay, C. J. (1996). Dual-attribute time–intensity measurement of sweetness and peppermint perception of chewing-gum. *Journal of Food Science, 61*, 636–638.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, 98*, 39–54.

Frank, R. A., & Byram, J. (1988). Taste–smell interactions are tastant and odorant dependent. *Chemical Senses, 13*, 445–455.

Gordin, H. H. (1987). Intensity variation descriptive methodology: Development and application of a new sensory evaluation technique. *Journal of Sensory Studies, 2*, 187–198.

Gower, J. C., Lubbe, S. G., & Le Roux, N. J. (2011). *Understanding biplots*. Singapore: John Wiley & Sons Ltd.

Greenacre, M. (2006). Tying up the loose ends in simple correspondence analysis. Economics Working Papers 940, Department of Economics and Business, Universitat Pompeu Fabra.

Greenacre, M. (2007). *Correspondence analysis in practice* (2nd ed.). Boca Raton, FL, USA: Chapman and Hall/CRC.

Greenacre, M., Nenadic, O., & Friendly, M. (2014). ca: Simple, Multiple and Joint Correspondence Analysis. R package version 0.55. http://CRAN.R-project.org/package=ca.

Halpern, B. P. (1991). More than meets the tongue: temporal characteristics of taste intensity and quality. In H. T. Lawless & B. P. Klein (Eds.), *Sensory science theory and applications in foods* (pp. 37–105). New York: Dekker.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal, 29*, 147–160.

Irwin, J. O. (1935). Tests of significance for differences between percentages based on small numbers. *Metron, 12*, 83–94.

ISO (2012). *Sensory analysis—General guidance for the selection, training, and monitoring of selected assessors and expert sensory assessors. ISO Standard 8586:2012*. Geneva, Switzerland: International Organization for Standardization.

Jack, F. R., Piggott, J. R., & Paterson, A. (1994). Analysis of textural changes in hard cheese during mastication by progressive profiling. *Journal of Food Science, 59*, 539–543.

Jaeger, S. R., Cadena, R. S., Torres-Moreno, M., Antúnez, L., Vidal, L., Giménez, A., et al. (2014). Comparison of check-all-that-apply and forced-choice Yes/No question formats for sensory characterisation. *Food Quality and Preference, 35*, 32–40.

Jaeger, S. R., Giacalone, D., Roigard, C. M., Pineau, B., Vidal, L., Giménez, A., et al. (2013). Investigation of bias of hedonic scores when co-eliciting product attribute information using CATA questions. *Food Quality and Preference, 30*, 242–249.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Kostyra, E., Baryłko-Pikielna, N., Dąbrowska, U., & Wasiak-Zys, G. (2008). Possibility of simultaneous measurement of pungency and leading flavour attributes in natural food matrices – temporal aspects. In *3rd European conference on sensory and consumer research*, 7–10 September, Hamburg, Germany.

Kuesten, C., Bi, J., & Feng, Y. (2013). Exploring taffy product consumption experiences using a multi-attribute time–intensity (MATI) method. *Food Quality and Preference, 30*, 260–273.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: Principles and practices* (2nd ed.). New York: Springer.

Lenfant, F., Loret, C., Pineau, N., Hartmann, C., & Martin, N. (2009). Perception of oral food breakdown. The concept of sensory trajectory. *Appetite, 52*, 659–667.

Majchrzak, D., Lahm, B., & Dürrschmid, K. (2010). Conventional and probiotic yogurts differ in sensory properties but not in consumers preferences. *Journal of Sensory Studies, 25*, 431–446.

Methven, L., Rahelu, K., Economou, N., Kinneavy, L., Ladbrooke-Davis, L., Kennedy, O. B., et al. (2010). The effect of consumption volume on profile and liking of oral nutritional supplements of varied sweetness: Sequential profiling and boredom tests. *Food Quality and Preference, 21*, 948–955.

Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference, 30*, 309–319.

Meyners, M., & Castura, J. C. (2014). Check-all-that apply questions. In P. Varela & G. Ares (Eds.), *Novel techniques in sensory characterization and consumer profiling*. Boca Raton, FL: CRC Press.

Meyners, M., & Pineau, N. (2010). Statistical inference for temporal dominance of sensations data using randomization tests. *Food Quality and Preference, 21*, 805–814.

Nenadic, O., & Greenacre, M. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. *Journal of Statistical Software, 20*(3), 1–13.

Pecore, S., Rathjen-Nowak, C., & Tamminen, T. (2011). Temporal Order of Sensations. In *9th Pangborn Sensory Science Symposium*, 4–8 September, Toronto, ON, Canada.

Pineau, N., Cordelle, S., & Schlich, P. (2003). Temporal dominance of sensations: A new technique to record several sensory attributes simultaneously over time. In *5th Pangborn sensory science symposium* (p. 121), July 20–24, Boston, MA, USA.

Pineau, N., Goupil de Bouillé, A., Lepage, M., Lenfant, F., Schlich, P., Martin, N., et al. (2012). Temporal dominance of sensations: What is a good attribute list? *Food Quality and Preference, 26*, 159–165.

Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., et al. (2009). Temporal dominance of sensations: Construction of the TDS curves and comparison with time–intensity *Food Quality and Preference, 20*, 450–455.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reinbach, H. C., Giacalone, D., Ribeiro, L. M., Bredie, W. L. P., & Frøst, M. B. (2014). Comparison of three sensory profiling methods based on consumer perception: CATA, CATA with intensity and Napping®. *Food Quality and Preference, 32*, 160–166.

Rossi, F. (2001). Assessing sensory panelist performance using repeatability and reproducibility measures. *Food Quality and Preference, 12*, 467–479.

Seo, H.-S., Lee, M., Jung, Y.-J., & Hwang, I. (2009). A novel method of descriptive analysis on hot brewed coffee: Time scanning descriptive analysis. *European Food Research and Technology, 228*, 931–938.

Thomas, A., Visalli, M., Cordelle, S., & Schlich, P. (2015). Temporal drivers of liking. *Food Quality and Preference, 40*, 365–375.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, 2*, 149–168.

Worch, T., & Piqueras-Fiszman, B. (2015). Contributions to assess the reproducibility and the agreement of respondents in CATA tasks. *Food Quality and Preference, 40*, 137–146.

Zorn, S., Alcaire, F., Vidal, L., Giménez, A., & Ares, G. (2014). Application of multiple-sip Temporal Dominance of Sensations to the evaluation of sweeteners. *Food Quality and Preference, 36*, 135–143.

Zwillinger, S. A., & Halpern, B. P. (1991). Time-quality tracking of monosodium glutamate, sodium saccharin and a citric acid-saccharin mixture. *Physiology & Behavior, 49*, 855–862.