Product discrimination ability of temporal sensory evaluation methods used with consumers

Michel Visalli, Sylvie Cordelle, Noëlle Béno, Pascal Schlich

Food Quality and Preference 115 (2024) 105123

Contents lists available at ScienceDirect

Food Quality and Preference

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

Product discrimination ability of temporal sensory evaluation methods used with consumers

Michel Visalli [a,b,*], Sylvie Cordelle [a,b], Noëlle Béno [a,b], Pascal Schlich [a,b]

[a] Centre des Sciences du Goût et de l'Alimentation, Institut Agro Dijon, CNRS, INRAE, Université Bourgogne, F-21000 Dijon, France
[b] INRAE, PROBE Research Infrastructure, ChemoSens Facility, F-21000 Dijon, France

ARTICLE INFO

Keywords:
TDS
TCATA
AEF
Rapid methods
Retrospective vs. concurrent measures
Gustometer

ABSTRACT

The objective of this study was to determine which temporal sensory evaluation methods (TDS, TCATA, AEF-D, AEF-A), variables (citations, durations, times of citation) and statistical analyses are the most discriminative when measurements are collected from consumers. This study is based on data collected on model stimuli delivered by a gustometer (controlled temporal differences) and on four categories of commercial food products (crisps, guacamoles, chocolates, ice teas) presenting different levels of sensory complexity.

With controlled stimuli, analyses of citations and durations both revealed the same differences, suggesting that product discrimination is mainly brought by presence/absence of sensory attributes. Compared to TCATA, TDS highlighted slightly more significant differences (including expected and unexpected differences). Short differences in durations were not captured. AEF-A was less discriminative, but presented no validity nor repeatability issues.

With commercial food products, analyses of citations were more discriminative than analyses of durations and times of citations. However, the combined analysis of citations and time-related variables resulted in a small gain of discrimination. Applicability-based methods (TCATA and AEF-A) were overall more discriminative compared to dominance-based methods (TDS and AEF-D), but had more repeatability issue.

These results suggest that when data are collected from consumers, in most situations analysing durations and times of citations would not bring much more information compared to a time-independent analysis of citations. Performances of TCATA and TDS are almost similar when used with simple stimuli presenting clear temporal patterns. With complex food products, TCATA is the most discriminative, but also the less reliable, and AEF-A presents the better compromise between discrimination and repeatability. This study showed that there is probably no ultimate method, and that every gain in temporal resolution or discrimination is assumably at the cost of a loss of reliability.

1. Introduction

To measure the temporal sensory perception of panellists during the tasting of food products, essentially two methods are used in practice: Temporal Dominance of Sensations (TDS, Pineau et al., 2009) and Temporal Check-All-That-Apply (TCATA, Castura et al., 2016). TDS and TCATA are qualitative evaluation methods that enable panellists to record temporal sequences of dominant or applicable attributes while they are tasting products. Dominant attributes are most often defined as those that trigger the panellist's attention at a given time (only one attribute can be chosen at any time), while applicable attributes are all those that are perceived at a given time (one or several attributes can be chosen at any time). While the operational definitions between dominance and applicability are different, the impact of these definitions on the temporal profiles collected is not obvious. Indeed, it has been shown recently that the TDS profiles obtained from naïve consumers are independent of the definition of dominance provided to the consumer. Different definitions of dominance (i.e. "most attention-catching sensation", "most intense sensation", or "most changing sensation") resulted in only marginal differences in TDS profiles suggesting that the

Abbreviations: AEF-A, Attack-Evolution-Finish Applicability; AEF-D, Attack-Evolution-Finish Dominance; AEF-RATA, Attack-Evolution-Finish with rating of applicable attributes; CATA, Check-All-That-Apply; CVA, Canonical Variate Analysis; mrCA, multiple response Correspondence Analysis; TCATA, Temporal Check-All-That-Apply; TDS, Temporal Dominance of Sensations.

* Corresponding author at: Centre des Sciences du Goût et de l'Alimentation, Institut Agro Dijon, CNRS, INRAE, Université Bourgogne, F-21000 Dijon, France.
E-mail address: michel.visalli@inrae.fr (M. Visalli).

https://doi.org/10.1016/j.foodqual.2024.105123
Received 1 June 2023; Received in revised form 3 October 2023; Accepted 25 January 2024
Available online 30 January 2024

definition of dominance provided to the consumers does not matter (Hutchings et al., 2022). Similarly, findings from another recent study carried out on control temporal stimuli have suggested that concepts of dominance and applicability were interpreted similarly by most consumers (Visalli, Béno, et al., 2023).

With trained panels, TDS (Bord et al., 2019; Braghieri et al., 2016; Bruzzone et al., 2013; Charles et al., 2017; Devezeaux de Lavergne et al., 2015; Etaio et al., 2016; Frost et al., 2018; Labbe et al., 2009; Meillon et al., 2009; Ng et al., 2012; Paulsen et al., 2014) and TCATA (McMahon et al., 2017; Varela et al., 2021) have proven their ability to capture differences between products that cannot be measured by static descriptive evaluation methods. Even if the added value of temporal measurements seems established when they are collected from trained panels, there is no such clear evidence with consumer studies. Indeed, some authors have questioned the complementarity of static and temporal measurements collected from consumers. After having compared results obtained with CATA (Adams et al., 2007) and TCATA, Vidal et al. (2019) and Bommel et al. (2020) concluded that there is no guarantee that temporal measurements bring additional information compared to static measurements. Others authors are more nuanced, observing that additional insights can be obtained for some specific attributes (Alcaire et al., 2017).

Most of these conclusions were obtained based on the study of the discriminative ability of the evaluation methods. With trained or semi-trained panels, TCATA showed either a better (Ares et al., 2015; Berget et al., 2020; Delompré et al., 2020; Meyners, 2020; Nguyen et al., 2018) or similar discrimination ability relative to TDS (Sharma & Duizer, 2019; Wu et al., 2019). With consumer panels, TCATA showed either a better (Ares et al., 2015; Gonzalez-Estanol et al., 2022), worse (Nguyen & Wismer, 2022), or similar discrimination ability relative to TDS (Esmerino et al., 2017; Paglarini et al., 2020; Velázquez et al., 2020). However, the statistical tools used to study discrimination vary as several variables can be derived from TDS and TCATA data. Among all the studies mentioned above, most conclusions relied on evaluations of the evolution of citation rates over time using TDS (Pineau et al., 2009) or TCATA (Castura et al., 2016) curves or difference curves. Others relied on inferential statistics, including among others univariate and multivariate analyses of durations (Galmarini et al., 2017) or citations (Lenfant et al., 2009). Thus, it is difficult to know which variables better explain the discrimination between products and the differences between methods.

Recently, Attack-Evolution-Finish (AEF) has been introduced as an alternative temporal evaluation method, in between static and dynamic methods. AEF enables consumers to report periods of perception of dominant (AEF-D, Visalli et al., 2020) or applicable (AEF-A, Visalli et al., 2022) attributes retrospectively after the tasting. Unlike concurrent measurements collected with TDS and TCATA (continuous data collection), AEF measurements are of low temporal resolution: only three periods corresponding to the beginning, middle and end of perception are evaluated. Despite that, the discrimination ability of AEF-D was similar to that of TDS (Visalli et al., 2020) and when AEF-A was used to characterize lemon-flavored carbonated alcoholic drinks a similar discrimination pattern was obtained relative to a concurrent evaluation method (Visalli et al., 2022). These two studies have revived the debate on what is really measured with temporal methods depending on whether consumers evaluate products concurrently or retrospectively to the tasting. Thus, additional works have been conducted to document the temporal resolution (Visalli, Mahieu, et al., 2023) and the validity (Visalli, Béno, et al., 2023) of temporal measurements collected from consumers with TDS, TCATA and AEF methods. These works showed that the temporal resolution of TDS and TCATA was below most expectations, and that repeatability and heterogeneity between consumers could be an issue threatening the validity of the measurements.

The objective of this study was to determine which temporal sensory evaluation methods (TDS, TCATA, AEF-D, AEF-A), variables (citations, durations, times of citation) and statistical analyses are the most discriminative while ensuring valid differences when measurements are collected from consumers. For this, an in-depth analysis was performed on a published dataset including measurements carried out on controlled temporal taste stimuli delivered by a gustometer (Béno et al., 2023). Then an attempt at generalization was made based on an analysis of another published dataset including four categories of commercial food products (crisps, guacamoles, chocolates, ice teas) presenting different levels of sensory complexity (Visalli, Cordelle, et al., 2022).

2. Materials and methods

2.1. Datasets

Table 1 summarizes the main characteristics of the datasets. All consumers were preselected from a population registered in the ChemoSens Platform's PanelSens database (declared to the relevant authority, authorization number 1148039). They had to fill out an informed consent form, and were compensated for their participation with vouchers worth 10 euros per session. They were randomly assigned to one panel among A, B and C (Béno et al., 2023) or among D, E, F or G (Visalli, Cordelle, et al., 2022), respectively, with a constraint of balance in gender and age between panels. Each panel used a different temporal method to evaluate all the samples. The between-subjects design was implemented with the objective of comparing methods, not to gain insight into the products. As we wanted to study individual differences, the samples were presented in the same order for all the consumers as recommended (Bell, 2013; Mollon et al., 2017). In doing so, an order effect may have affected the product comparison, but in the same way for all methods.

Liquid solutions were delivered using a gustometer model Burghart GU002. The gustometer is an instrument that enables to program the delivery of predetermined volume and concentration of liquid compounds to the panellist's tongue over a specified period of time.

Fig. 1 describes the controlled temporal stimuli delivered using the gustometer. They include three to five compounds, either aromatic (basil hydrosol, citral) or sapid (citric acid, sodium chloride, saccharose), delivered at constant concentration and *iso*-intensity. A list of eight attributes was proposed to the panellists to describe the stimuli. This list included five attributes corresponding to the delivered compounds (*Basil*, *Lemon, Acid*, *Salty*, *Sweet*) and three distractors (*Bitter*, *Mint*, *Licorice*).

Table 2 details the times of begin and end of delivery of the compounds constituting the controlled temporal stimuli. Depending on the chemical composition of the stimuli, the differences in sensory perception between pairs of stimuli can be of three different natures: differences in attribute citations (corresponding to one attribute present in the first stimulus but not in the second, e.g. between S01 and S03 for *Acid*), differences in times of first citations (corresponding to different times of begin of delivery, e.g. between S02 and S03 for *Lemon*), and differences in durations of dominance/applicability (corresponding to different durations of delivery, e.g. between S01 and S02 for *Sweet*).

The chosen products were commercial products commonly consumed by a majority of consumers, easy to prepare for the experimenters, and varying in texture. Crisps varied in fat and salt; ice teas varied in sugar/sweetener, flavour and variety of tea; guacamoles varied in avocado content and fat; and dark chocolates varied in cocoa origin and content (Table 3). A questionnaire derived from that established by Schlich et al. (2015) was given to the subjects after they finished evaluating each product category. The questionnaire included nine items assessed using structured scales, namely: the level of familiarity with the taste of the samples, the number of sensations perceived, the difficulty of identifying sensations, the harmony between attributes, the balance between attributes, the persistence of sensations, the strength of sensations, the level of similarity between samples and the overall level of sensory complexity of the samples. The last two items were used to define the levels of complexity and sensory differences of the product

Table 1
Characterization of datasets. Letters A-G between parentheses after panel sizes denote different panels. Levels of complexity and sensory differences were established from responses to a questionnaire.

Type of product (data paper)	Number of different samples (number of replicated samples)	Number of attributes	Settings	Complexity	Differences	Number of consumers (panel)				Mean durations of evaluation (panel)	
						TDS	TCATA	AEF-A/ AEF-RATA	AEF-D	TDS	TCATA
Model solutions (Béno et al., 2023)	4 (2)	8	Lab	–	–	50 (A)	50 (B)	49 (C)	–	26.7	40.7
Guacamoles (Visalli, Cordelle, et al., 2022)	4 (1)	12	Lab	High	High	70 (D)	73 (E)	74 (F)	75 (G)	36.7	43.9
Ice teas (Visalli, Cordelle, et al., 2022)	4 (1)	8	Lab	Medium	High	70 (D)	73 (E)	74 (F)	75 (G)	30.9	41.8
Crisps (Visalli, Cordelle, et al., 2022)	4 (1)	9	Home	Low	Medium	70 (D)	72 (E)	74 (F)	75 (G)	28.1	35.5
Chocolates (Visalli, Cordelle, et al., 2022)	4 (1)	12	Home	High	Medium	68 (D)	72 (E)	71 (F)	75 (G)	46.2	55.0

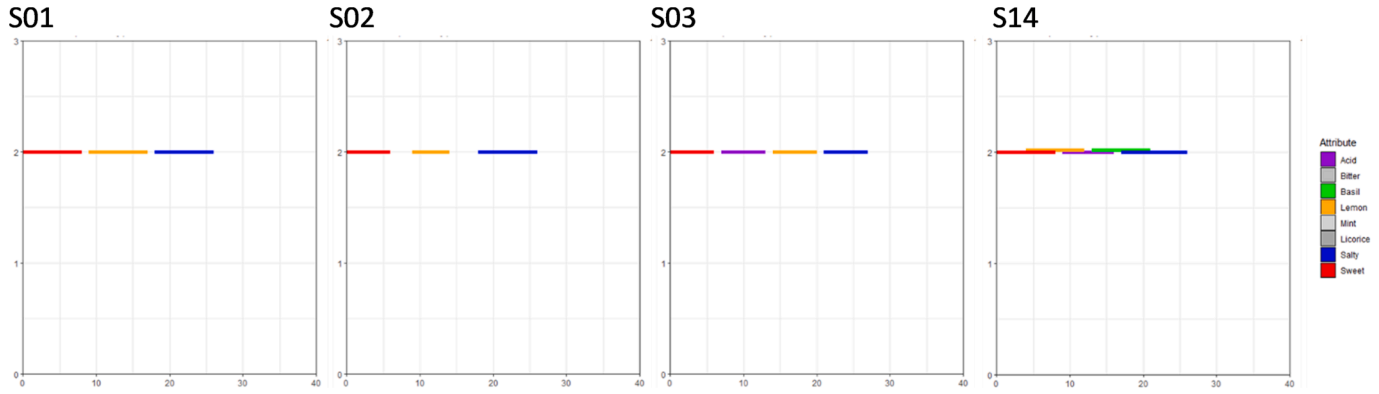

Fig. 1. Controlled temporal stimuli delivered using a gustometer. X-axis: time, y-axis: concentration of the compound. From left to right: stimuli S01, S02, S03, S14.

Table 2
Times of begin and end of delivery of the controlled temporal stimuli (in seconds).

	S01	S02/S02_2	S03	S14/S14_2
Acid			7–13	9–16
Basil				13–21
Lemon	9–17	9–14	14–20	4–12
Salty	18–26	18–26	21–27	18–26
Sweet	0–8	0–6	0–6	0–8

spaces reported in Table 1. More details can be found in Visalli, Cordelle, et al. (2022).

2.2. Data analysis

Statistical analyses were performed using R 4.1.0 software (R Core team, 2017). All tests were performed at the 5 % alpha-risk level.

For all the methods, to test whether discrimination was related to temporal aspects, TDS, TCATA and AEF data were transformed in CATA data. Attributes were set to 1 when they cited as dominant/applicable by a consumer at least once at any time during tasting, 0 otherwise.

To assess overall differences between samples, two multivariate analyses were run. Differences in presence/absence after CATA transformation were assessed using the multiple response Correspondence Analyses (mrCA, Mahieu et al., 2021) framework. Differences in durations of dominant/applicable attributes were assessed using Canonical

Table 3
Product categories, codes and composition (as reported on the packaging).

Product category	Codes	Composition
Crisp	C1, C1_rep (replicate of C1)	34 g fat, 1.3 g salt
Crisp	C2	23.9 g fat, 1.52 g salt
Crisp	C3	29 g fat, 1 g salt (sea salt)
Crisp	C4	34 g fat, 0.10 g salt
Guacamole	G1, G1_rep (replicate of G1)	92 % avocado, 16 g fat
Guacamole	G2	13 % avocado, 9.5 g fat
Guacamole	G3	90 % avocado, 14.6 g fat
Guacamole	G4	95 % avocado, 18 g fat
IceTea	IT1	4.7 g sugar, black tea, white peach
IceTea	IT2	4.5 g sugar, white tea, peach and rosemary
IceTea	IT3, IT3_rep (replicate of IT3)	4.3 g sugar, sweeteners, black tea, peach
IceTea	IT4	0 g sugar, sweeteners, black tea, peach
Chocolate	CH1	85 % cocoa, origin Madagascar
Chocolate	CH2	80 % cocoa, origin Equator
Chocolate	CH3	70 % cocoa, origin Peru
Chocolate	CH4, CH4_rep (replicate of CH4)	74 % cocoa, origin Côte d'Ivoire

Variate Analysis (CVA, Galmarini et al., 2017). Product maps were represented using biplots, and pairwise product comparisons were performed on the significant dimensions of mrCAs or CVAs. The two analyses were performed with and without including the replicated samples to quantify discrimination and non-repeatability, respectively.

To characterize differences between pairs of samples at the attribute level, the differences in citation rates at each time were plotted (referred to as "difference plots" in the rest of the manuscript). For TDS and TCATA, data were first discretized every 0.1 s and left standardized (as if the chronometer started with the first citation of an attribute) to reduce heterogeneity between consumers. The curves were smoothed using locally estimated scatterplot smoothing (LOESS, Cleveland, 1979). Significant differences were established based on proportion tests with Yates' continuity correction to prevent overestimation of statistical significance for expected count smaller than five (Yates, 1934). Significant differences were represented as bold lines (TDS and TCATA) or points (AEF-RATA, AEF-D, AEF-A) on the plots. Total durations of significant differences were computed and considered "significant" when they lasted more than 5 % of the total time (no statistical test performed). Absolute values of differences in citation rates were averaged over times/periods and attributes to quantify the overall discrimination for each pair of samples. Then, in the same idea as the F-ratio of an ANOVA, discriminating ratios were computed between the overall discrimination value of pairs of different samples (expected "high") and the overall discrimination value of pairs of replicated samples (expected "low"). Discriminating ratios higher than 1 were considered "significant" (no statistical test performed).

To characterize the nature of the differences between pairs of samples at the attribute level, three univariate analyses were run. For the assessment of differences in applicability, type-III ANOVA tables were computed by attribute after having adjusted the following generalized linear model with binomial error distribution: *citation (0 or 1)* = *sample* + *consumer* + *error* (Agresti, 2013). For the assessment of differences in durations, type-III ANOVA tables were computed by attribute after having adjusted either a linear model (for TDS and TCATA) or a generalized linear model with Poisson error distribution (for AEF): *duration* = *sample* + *consumer* + *error*. Pairwise differences between samples were estimated using least square means, and expressed as log of odd ratios (Sroka & Nagaraja, 2018) or difference in durations (Galmarini et al., 2017). Finally, as no statistical model fits for the analysis of attribute time of first citations in TDS and TCATA (because of unbalanced and non-normal data), the differences in median attribute times of first citations were assessed based on a bootstrap approach as described in Visalli, Mahieu, et al. (2023), with a threshold of 25 % for the overall citation rate for the selection of attributes. Significant differences between pairs of stimuli were qualified as:

- "incongruent" when differences were associated with attributes related to compounds not delivered by the gustometer (controlled stimuli only);
- "expected"/"unexpected" when real differences exist/not exist in citation, duration or time of first citation (depending on the variable and attribute analysed); significant differences between replicated pairs of stimuli were also qualified as "unexpected".
- "consensual" when differences were evidenced by at least three out of the four methods.

3. Results

3.1. Model solutions

Fig. 2 shows very similar product configurations regardless of the evaluation method and the analysed variable (citation or duration), but different interpretations regarding product characterization. The mrCAs (Fig. 2A, C and E) point out that the main differences observed (length of the arrows) were due to *Basil,* more cited for S14, and *Acid* more cited for S03 and S14. The arrows representing *Sweet*, *Lemon* and *Salty* (the three attributes corresponding to compounds present in all the stimuli) are very short, sometimes shorter than the arrows of the distractors. S01 and S02 (which differed only in durations) are not discriminated, nor are the replicated stimuli. These observations appear congruent with the chemical differences related to presence/absence of compounds in the stimuli (Fig. 1). When the replicated stimuli are included, the mrCAs highlight three significant dimensions in TDS and AEF-RATA, and two in TCATA. When the replicated stimuli are not included, the mrCAs highlight three significant dimensions for all methods. The $mrChi^2$ statistic (representative of the overall size of the differences in citations) is greater with TDS (205.7 with replicated stimuli, 160.3 without), then with AEF-RATA (155.5 and 107.9), then with TCATA (125.1 and 84.4).

Basil and *Acid* are also discriminative attributes with CVAs of durations (Fig. 2B–F). However, different from mrCA, with CVA the lengths of the arrows are not proportional to the differences in durations of delivery of the compounds. Indeed, the arrows representing the distractors have approximately the same lengths than the arrows representing the attributes corresponding to the compounds actually delivered. The replicated stimuli are not discriminated except for S14 with TDS (on axis 1–3, not represented on Fig. 2, see pairwise differences on Fig. 4). The small differences between S01 and S02 in durations of delivery of the compounds (*Sweet* and *Lemon*) are no more captured than with mrCA. When the replicated stimuli are included, three axes were tested significant for TDS and TCATA, and two for AEF-RATA. When the replicated stimuli are not included, four axes were tested significant for TDS, three for TCATA and two for AEF-RATA. The F-statistic (representative of the overall size of the differences in durations) is greater with TDS (12.8 with replicated stimuli, 15.6 without), then with TCATA (8.7 and 9), then with AEF-RATA (6.7 and 6.4).

These results show that on controlled stimuli, multivariate analyses of citations (mrCA) and durations (CVA) both capture differences in presence/absence. However, short differences in durations between stimuli are not necessarily captured by CVA.

Fig. 3 shows the evolution of differences in attribute citations rates over time. Ideally, as S02_2, S14_2 were replicates of S02 and S14, no significant differences should be evidenced by difference curves between S02 and S02_2 nor between S14 and S14_2. However, unexpected differences between the replicated samples are observed with TCATA (for *Acid* between S02 and S02_2, Fig. 3.A) and TDS (for *Basil* and *Acid* between S14 and 14_2, Fig. 3.B).

Controlled stimuli S01 and S02 differed on the durations of delivery for *Lemon* and *Sweet* (Table 2), but only TDS highlights significant differences for *Lemon* (Fig. 3.C).

Stimuli S01 and S02 differed from S03 on the times and durations of delivery for *Lemon, Salty*, on the durations of delivery for *Sweet* (S01 vs. S03 only) and on the applicability for *Acid* (Table 2). TDS highlights significant differences for all these attributes, while TCATA does not capture difference in citation rates for *Lemon*, and AEF-RATA only captures difference in *Acid* (Fig. 3.D). The duration of significant differences observed for *Sweet* is far longer than the difference in the duration of delivery (2 s). Fig. 3.F evidences large differences in *Sweet* between S03 and S03 for both TDS and TCATA, while there was no difference between the two stimuli, saccharose being delivered between 0 and 6 s. This result shows that the difference in citation rates for *Sweet* is not due to the difference in the duration of delivery.

Stimuli S01 and S02 differed from S14 on the applicability for *Acid* and *Basil,* on the time of delivery for *Lemon* and on the durations of delivery for *Sweet* (S02 vs. S14 only, Table 2). All the methods agree and highlight significant differences for *Acid*, *Basil* and *Lemon* between S01/S02 and S14, but with TDS and TCATA also for *Sweet* (Fig. 3.E and 3.G) and *Salty* (Fig. 3.G). The period of time during which significant differences were observed was higher for TDS than TCATA.

Stimuli S03 and S14 differed on the applicability for *Basil,* on the times durations of delivery for *Acid*, *Lemon* and *Salty*, and on the duration of delivery for *Sweet* (Table 2). This is the only pair of samples for

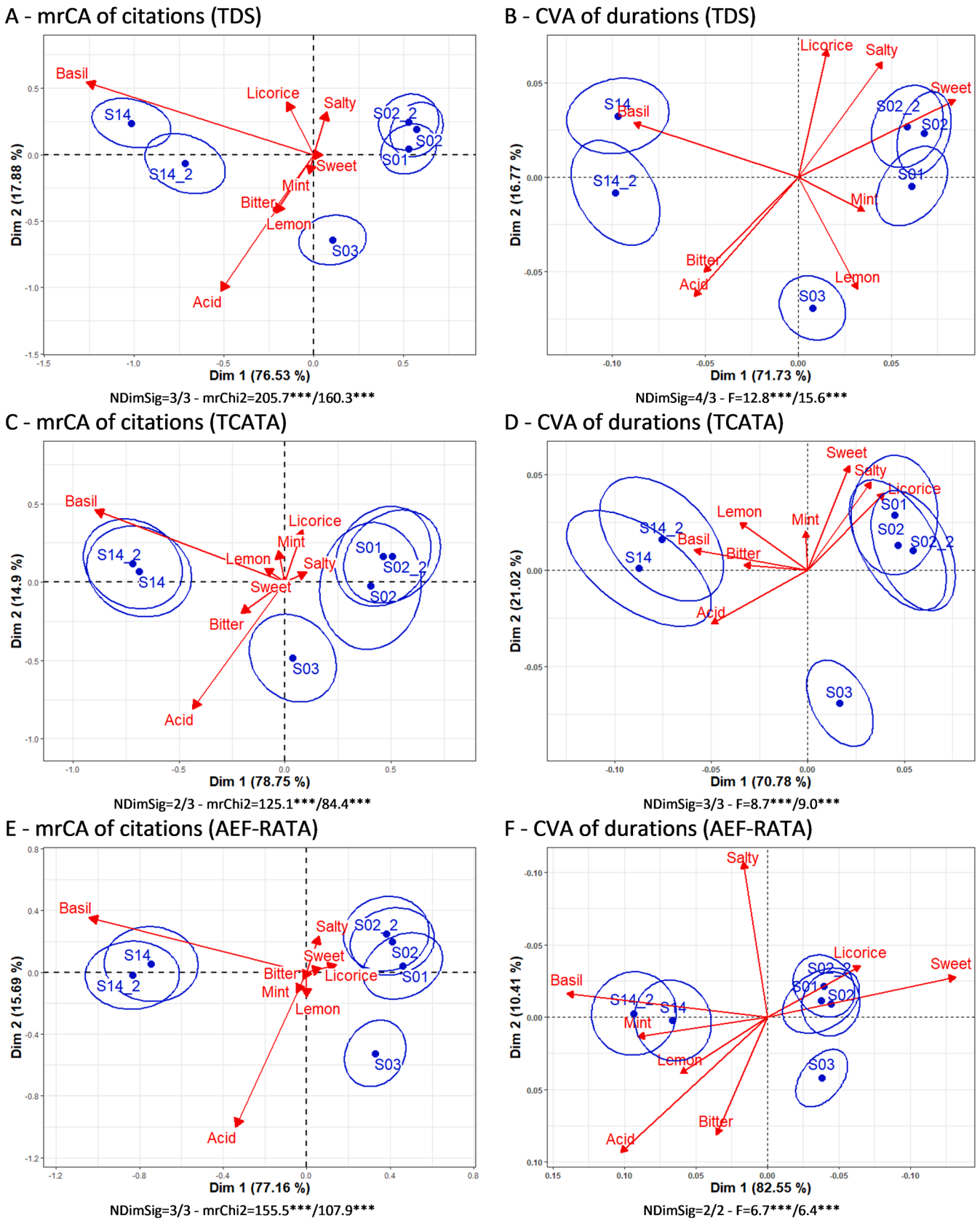

Fig. 2. Biplots of multivariate analyses (axes 1–2), with 95 % confidence ellipses. First row: TDS, second row: TCATA, third row: AEF-RATA. First column: mrCA of citations, second column: CVA of durations. NDimSig: number of significant dimensions (all stimuli/without replicates). mrChi2: mrChi2 statistic (all stimuli/without replicates). F: MANOVA statistic (all stimuli/without replicates). ***p-value < 0.001.

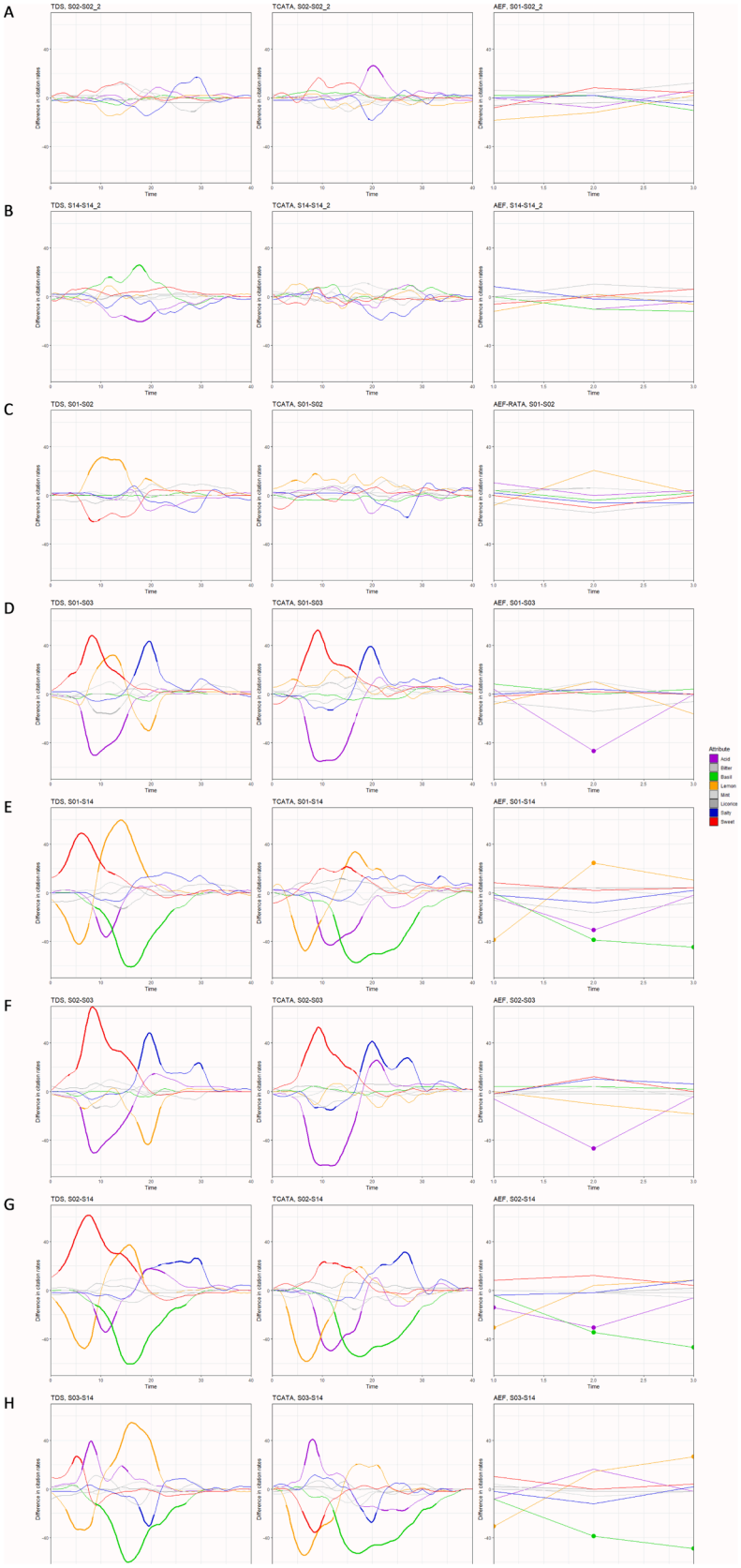

Fig. 3. Pairwise differences in citation rates (Y-axis) plotted over time (X-axis). A-H rows of plots each refers to one pair comparison: A = S02-S02_2, B = S14-S14_2, C = S01-S02, D = S01-S03, E = S01-S14, F = S02-S03, G = S02-S14, H = S03-S14. Within each row, first column refers to TDS, second to TCATA and third to AEF-RATA. Bold lines (for TDS and TCATA) and points (for AEF-RATA) correspond to significant differences in citation rates (alpha = 0.05).

Controlled stimuli		S02_S02_2			S14-S14_2			S01-S02			S01-S03			S01-S14			S02-S03			S02-S14			S03-S14		
		TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA	TDS	TCATA	AEF-RATA
Acid	Citation	0.6	2.1	0.1	-0.7	-0.2	-0.5	-0.8	-1.4	0.4	-4.0	-4.5	-2.3	-2.6	-3.2	-1.4	-3.3	-3.1	-2.7	-1.8	-1.9	-1.8	1.4	1.2	0.9
	Duration	0.6	1.5	-0.1	-2.8	-0.2	-0.2	-0.9	0.4	0.4	-3.5	-3.6	-0.7	-1.1	-3.7	0.6	-2.7	-4.0	-1.1	-0.3	-4.1	-1.0	2.4	-0.1	0.1
	Time	0.0	0.4	NT	0.2	-0.4	NT	-1.2	-1.1	NT	9.2	9.1	NT	6.8	7.1	NT	10.5	10.4	NT	8.0	8.2	NT	-2.4	-2.0	NT
	Citation rate over time	0.0	3.6	0.0	6.8	0.0	0.0	0.0	0.0	0.0	9.6	10.6	1.0	5.2	9.4	1.0	11.6	14.2	1.0	9.6	9.8	2.0	4.8	7.4	0.0
Basil	Citation	-0.9	0.3	-0.5	2.5	-0.3	-0.6	0.5	-0.3	-0.6	-0.7	-0.8	0.8	-7.8	-4.2	-6.0	-1.2	-0.5	1.5	-8.3	-3.9	-5.4	-7.0	-3.4	-6.8
	Duration	-0.3	0.7	-0.3	2.2	1.2	-0.2	0.1	-1.0	0.1	-0.2	-0.3	0.8	-6.5	-8.7	-1.6	-0.3	0.6	0.7	-6.6	-7.7	-1.6	-6.3	-8.3	-2.3
	Time	NT	NT	NT	-0.2	-0.4	NT	NT	NT	NT	NT	NT	NT	2.9	5.8	NT	NT	NT	NT	2.0	-1.8	NT	-0.2	1.8	NT
	Citation rate over time	0.0	0.0	0.0	3.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.0	17.4	2.0	0.0	0.0	0.0	18.4	17.2	2.0	18.4	20.8	2.0
Lemon	Citation	-1.0	-0.5	-1.4	0.4	0.0	-0.2	2.9	1.7	2.1	-0.7	0.8	0.2	-0.2	0.0	0.9	-3.5	-0.9	-1.9	-3.1	-1.7	-1.2	0.5	-0.8	0.7
	Duration	-1.1	-1.3	-0.3	0.5	0.1	-0.1	3.3	3.8	0.1	0.4	2.9	-0.1	2.5	0.3	0.0	-2.9	-0.9	-0.3	-0.8	-3.4	-0.2	2.1	-2.5	0.1
	Time	0.4	-0.3	NT	0.2	-0.7	NT	-1.0	-0.8	NT	-1.2	-0.8	NT	4.6	4.4	NT	0.0	0.0	NT	5.4	5.3	NT	5.6	5.2	NT
	Citation rate over time	0.0	0.0	0.0	0.0	0.0	0.0	9.4	2.6	0.0	8.4	1.5	0.0	17.0	12.4	2.0	6.0	0.0	0.0	15.0	9.4	1.0	15.8	11.4	2.0
Salty	Citation	0.2	-0.6	-0.4	0.1	-1.2	0.0	-0.8	-0.2	-0.4	0.9	0.4	0.7	0.1	1.5	0.4	1.7	0.7	1.1	1.0	1.7	0.7	-0.7	1.0	-0.3
	Duration	-0.3	-0.7	0.0	-1.4	-1.7	0.0	-0.8	0.2	-0.1	2.5	2.7	0.0	1.6	2.4	-0.1	3.3	2.5	0.1	2.3	2.2	0.0	-0.9	-0.3	-0.1
	Time	0.8	0.2	NT	1.0	0.0	NT	-0.4	-0.4	NT	-2.8	-2.8	NT	-1.0	0.2	NT	-2.4	-2.4	NT	-0.5	0.3	NT	1.9	2.7	NT
	Citation rate over time	0.4	0.2	0.0	0.0	0.2	0.0	0.0	0.6	0.0	4.8	6.0	0.0	0.2	0.8	0.0	6.8	12.6	0.0	8.4	5.6	0.0	3.2	3.0	0.0
Sweet	Citation	1.1	0.0	-0.4	1.0	-1.0	-1.3	0.9	-1.7	0	1.5	0.0	0.3	2.0	-0.7	1.6	0.6	1.7	0.3	1.1	1.0	1.6	0.5	-0.7	1.3
	Duration	1.0	1.2	0	1.3	-0.7	0.0	-1.4	0.2	-0.1	4.1	4.6	0	4.1	2.6	0.1	5.5	4.3	0.1	5.5	2.3	0.2	0.0	-2.0	0.1
	Time	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT	0.0	0.0	NT
	Citation rate over time	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	11.0	9.5	0.0	11.0	5.8	0.0	13.0	11.0	0.0	15.8	6.0	0.0	2.4	4.6	0.0
Bitter*	Citation	-0.2	-1.5	0.8	-0.2	0.5	1.1	0.7	1.5	-1.5	-0.9	-0.9	-1.2	-0.6	-1.0	-1.5	-1.6	-2.3	0.3	-1.3	-2.5	0.0	0.3	-0.2	-0.3
	Duration	-0.4	-1.2	0.5	0.0	-0.3	0.3	1.3	1.2	-0.6	0.0	0.2	-0.6	0.1	-0.1	-0.6	-1.2	-1.0	0.0	-1.1	-1.3	-0.1	0.1	-0.3	-0.1
	Time	NT	NT	NT	1.4	1.2	NT	1.0	NT	NT	7.4	3.2	NT	3.3	1.6	NT	6.5	5.1	NT	2.6	3.1	NT	-4.4	-1.8	NT
	Citation rate over time	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.0	3.0	0.0	0.0	1.8	0.6	0.0	0.0	1.0	0.0	0.0	0.0	0.0
Licorice*	Citation	0.3	0.2	-0.9	0.8	-0.3	0.0	-1.4	0.5	0.9	0.0	1.6	0.6	-2.0	1.0	1.2	1.4	1.1	-0.3	-0.6	0.5	0.3	-2.0	-0.6	0.6
	Duration	-0.6	0.4	-0.6	0.4	-0.6	-0.3	-0.5	0.6	0.6	0.0	1.9	0.3	-0.6	2.0	0.6	0.5	1.3	-0.4	-0.1	1.4	0.0	-0.6	0.0	0.4
	Time	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT
	Citation rate over time	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mint*	Citation	1.2	-0.5	0.0	-0.6	0.2	0.3	0.3	0.7	1.3	0.3	0.7	0.6	0.6	0.0	0.3	0.0	0.0	-0.7	0.3	-0.7	-1.0	0.3	-0.7	-0.3
	Duration	0.9	-1.0	0.0	-0.4	1.6	-0.1	-0.1	2.0	0.7	0.4	1.7	0.4	0.7	0.3	0.1	0.5	-0.3	-0.3	0.8	-1.6	-0.6	0.3	-1.4	-0.3
	Time	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT
	Citation rate over time	1.8	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.4	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Multivariate	mrChi2	3.2	6.0	2.4	2.9	1.0	2.1	8.5	5.6	6.5	18.7	15.5	10.2	46.0	29.8	32.2	24.0	11.5	14.4	48.0	28.6	28.3	33.0	16.4	30.7
	F-MANOVA	0.9	2.0	0.8	2.2	1.2	0.6	2.7	1.8	1.5	6.9	5.7	2.9	18.5	9.3	5.5	9.5	7.2	3.2	18.0	9.0	7.6	17.5	4.9	10.6
	Discriminating ratio	0.9	0.9	1.1	1.1	1.1	0.9	1.1	1.3	1.1	1.8	2.3	1.3	2.5	2.7	2.3	2.0	2.0	1.2	2.7	2.6	2.0	2.1	2.3	2.1

Fig. 4. Summary of pairwise differences between controlled stimuli, by attribute (rows) and method (columns). Attributes suffixed by an asterisk are distractors. Citation: log of the odd ratio (ANOVA *citation* = *sample* + *consumer* + *error*). Duration: mean difference in duration (ANOVA *duration* = *sample* + *consumer* + *error*). Time: median time of first citation (bootstrap test). Citation rate over time: total duration of significant differences in citation rates (difference plot). mrChi2: mrChi2 statistic restricted to the pair of samples (mrCA of citations). F-MANOVA: F statistic of the MANOVA restricted to the pair of samples (CVA of durations). Discriminating ratio: average difference between the pair of samples divided by average difference between the replicated samples (citation rates over time). Red and orange cells: significant unexpected differences between replicated samples (red cells: differences not captured with overall citation rates, after CATA transformation). Yellow cells: significant incongruent differences (compared to the chemical composition of the stimuli). Green cells: significant expected differences (compared to the chemical composition of the stimuli). Blue cells: non-significant differences while expected (compared to the chemical composition of the stimuli). NT: discrimination not tested. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which conclusions are opposed between TDS and TCATA, for *Sweet* (higher citation rates in S03 with TDS and in S14 with TCATA, Fig. 3.H). The two methods agreed on other differences while AEF-RATA only show significant differences for *Basil* and *Lemon*.

These results suggest that, on controlled stimuli, citation rates capture qualitative changes of applicability over time, but with a relatively low precision. Short differences in durations of delivery in the stimuli are not systematically highlighted on the temporal curves. TDS highlights more significant differences than TCATA, and AEF-RATA less. However, whatever the method, some differences are not congruent with the chemical composition of the stimuli.

Fig. 4 shows that whatever the multivariate analyses used (last three rows), except with the pair S01/S02, all pairs of different samples are discriminated (green cells) by the analyses of citations (with mrChi2), durations (with MANOVA) and discriminating ratio (with difference plots). The products are not discriminated by the analyses of citations. This result is in line with expectations as S01 and S02 only differed in the duration of the period of time during which the stimuli were delivered. However, only TDS captures significant differences between S01 and S02 based on durations. All methods capture differences between S01 and S02 based on discriminating ratio, but this analysis is likely to be too sensitive. Indeed, the difference plots also evidence unexpected discrimination (red cells) between replicated samples (S02 with AEF-RATA and S14 with TDS and TCATA). Unexpected discrimination is also observed with CVA of durations (S14 with TDS).

The univariate analyses by attribute enable a more detailed investigation of the differences between the pairs of samples. Most of the times, the methods agree on the differences, but AEF-RATA is clearly less discriminative than TDS and TCATA. S02 and S14 are unexpectedly discriminated from their replicates. With TCATA, S02 was discriminated from its replicate (S02_2) for *Acid* (citations and durations of significant differences), meanwhile with TDS S14 significantly differed from it replicate (S14_2) for *Acid* (durations and durations of significant differences) and *Basil* (citations, durations and durations of significant differences). This result shows that some repeatability issues can be observed when consumers perform temporal measurements. Significant differences incongruent with the chemical composition of the stimuli are also observed for some attributes. These differences are mostly observed with TDS and TCATA, some being common to both methods (*Acid* for S01/S03, S01/S14, S02/S03; *Salty* for S01/S14 and S02/S14; *Sweet* for S01/S14, S02/S03; S02/S14; S03/S14).

Table 4 shows that all existing differences due to presence/absence of compounds are captured by the analyses of citations, except one for AEF-RATA. In return, AEF-RATA has no repeatability issue, different from TDS and TCATA, which point out more incongruent differences. Only half of the existing differences in attribute durations are captured by TDS and TCATA, and a quarter by AEF-RATA. Notably, no method captures the short-duration differences for *Lemon* between S01 and S03,

Table 4
Count of discriminated pairs of samples. Expected: observed/expected differences corresponding to the chemical composition of the stimuli. Incongruent: differences not corresponding to the chemical composition of the stimuli. Unexpected: differences between the replicated stimuli (non-repeatability).

		TDS	TCATA	AEF-RATA
Citation	Expected	7/7	7/7	6/7
	Incongruent	4	2	1
	Unexpected	1	1	0
Duration	Expected	10/20	11/20	5/20
	Incongruent	5	2	0
	Unexpected	2	0	0
Time	Expected	8/10	8/10	Not tested
	Incongruent	5	3	Not tested
	Unexpected	0	0	Not tested
Citation rate over time	Expected	19/19	17/19	10/19
	Incongruent	8	7	0
	Unexpected	2	1	0

for *Sweet* between S01 and S02 and between S03 and S14, and for *Salty* between S03 and S14. TDS tends to point out more incongruent differences and unexpected differences compared to TCATA, while AEF shows none due to its low discriminative ability for durations. TDS and TCATA capture eighty percent of difference in times of first citations, with no unexpected difference but still more incongruent differences with TDS. The same pattern can be observed with durations of significant differences: TDS is more discriminative (100 % of expected differences captured), but shows more incongruent and unexpected differences compared to TCATA. AEF-RATA is less discriminative, but presents no potential validity nor repeatability issue. These results show that contrarily to multivariate analyses (that mostly conclude similarly with citations and durations), unidimensional analyses of durations and times of citations (time-dependent variables) are complementary to unidimensional analyses of citations (time-independent variable).

3.2. Food products

Multivariate analyses (CVA and mrCA maps), univariate analyses and difference plots are available in Supplementary material. Different from controlled stimuli, no hypotheses were formulated for expected and incongruent differences for commercial food products.

With multivariate analyses, all pairs of different guacamoles (six) and ice teas (six) are discriminated based on the analysis of citations after CATA transformation by the four methods. Only TCATA discriminates the six pairs of crisps, and only TDS the six pairs of chocolates, the other methods discriminating five pairs of samples. With respect to pairs of replicated samples, the guacamoles are unexpectedly discriminated by AEF-D and AEF-A; the ice teas by TCATA, AEF-D and AEF-A; the crisps by TDS and AEF-A; and the chocolates by no method.

The analysis of durations generally leads to similar conclusions, except for TCATA (ice teas IT3/IT4 and crisps C1/C2 not discriminated by durations) and TDS (crisps C1/C3 and chocolates CH2/CH3 not discriminated by durations). Repeatability issues previously observed with the analyses of citations of crisps with TDS and AEF-A are not observed with the analyses of durations. However, new repeatability issues are pointed out with the analyses of durations of guacamoles for TCATA. The discriminating ratios are most of the time greater than one, except for pairs of guacamoles G1-G2 (TDS), ice teas IT3-IT4 (AEF-D), crisps C1-C2 (all methods), and chocolates CH2-CH4 (TCATA). These results show that whatever the method, with a few exceptions, the conclusions about differences between pairs of samples are similar based on multivariate analysis of citations or durations.

Regarding the unidimensional analyses, the most consensual differences captured by the methods are observed for guacamoles with *Avocado*, *Onion/Shallot*, *Smooth/Fat* and *Spicy*; for ice teas with *Artificial*, *Sweet* and *Watery/Diluted;* for crisps with *Bland*, *Crackly/Hard* and *Salty;* and for chocolates with *Bitter*, *Fruity* and *Sweet*. These attributes structure the product configurations that lead to consensual multivariate differences.

Most differences are captured by several variables. However, some analyses of pairwise differences involved attributes few cited (by less than 25 % of consumers, as denoted by "NT" with TDS and TCATA). This can be observed with all product categories: for chocolates with Fluid, for ice teas with *Bitter* and *Watery/Diluted*; for crisps with *Bland, Melting,* and *Sticky/Pasty*; for chocolates with *Artificial*, *Floral* and *Sour*.

Some differences are method-dependent while others are variable-dependent. AEF-D and AEF-A detect more differences of *Lemon* and *Sweet* in Guacamoles; *Cocoa* and *Astringent* in Chocolates. TDS and TCATA capture more differences of *Artificial* in Ice teas. TDS never points out differences of *Bitter* in Guacamoles. Differences between C1 and C2 are only detected by TCATA and AEF-D. Differences of *Tomato* in Guacamoles are mainly shown with difference plots, differences of *Peach* in Ice teas or *Salty* in Crisps with analysis of time-dependent variables.

Table 5 summarizes the number of significant differences observed on the food products with the unidimensional analyses of the different variables for each method. It shows that except for AEF-A, more pairs of samples are discriminated by citations than by durations. This means that analysing only durations results in a loss of discrimination, with only a marginal reduction of error rate. The single analysis of times of first citations leads to a number of discriminated pairs of samples four to five times lower than with the analysis of citations, with an error rate above 5 %. However, the combined analysis of citation and durations (except for AEF-D) results in a gain of discrimination between 7 and 15 %. The gain is roughly the same with the combined analysis of citations and times of first citations of TDS and TCATA, but the error rate is higher. The only inspection of difference plots also results in a loss of discrimination for dominance-based methods, with a higher error rate. For applicability-based methods, this analysis is the most discriminative. However, the combined analysis of citations and difference plots leads to a gain of discrimination between 15 and 42 % (larger for applicability-based methods). Applicability-based methods are overall more discriminative compared to dominance-based methods, but they have also higher error rates, notably with TCATA. For this method, whatever the variable analysed, the error rate is always larger than 5 %.

Table 5
Count of discriminated pairs of samples, by variable, all categories combined. Discrimination: number of significant differences between pairs of different samples; between brackets: gain of discrimination when combined with analysis of citations. Non-repeatability: number of significant differences between pairs of replicated samples; between brackets: error rate = non-repeatability / (discrimination + non-repeatability). NT: not tested.

		TDS	TCATA	AEF-D	AEF-A
Citation	Discrimination	88	87	98	84
	Non-repeatability	2 (2.2 %)	5 (5.4 %)	2 (2 %)	3 (3.4 %)
Duration	Discrimination	55	78	67	86
	Non-repeatability	1 (1.8 %)	4 (4.9 %)	1 (1.5 %)	2 (2.3 %)
Time	Discrimination	18	22	NT	NT
	Non-repeatability	1 (5.3 %)	3 (12 %)	NT	NT
Citation rate over time	Discrimination	75	119	98	115
	Non-repeatability	3 (3.8 %)	10 (7.7 %)	5 (4.8 %)	5 (4.2 %)
Citation + Duration	Discrimination	94 (+6.8 %)	100 (+14.9 %)	99 (+0.1 %)	95 (+13.1 %)
	Non-repeatability	3 (3.1 %)	6 (6 %)	2 (2 %)	3 (3.1 %)
Citation + Time	Discrimination	93 (+5.7 %)	101 (+16.1 %)	NT	NT
	Non-repeatability	3 (3.1 %)	8 (7.3 %)	NT	NT
Citation + Citation rate over time	Discrimination	104 (+18.2 %)	124 (+42.5 %)	113 (+15.3 %)	117 (+39.3 %)
	Non-repeatability	3 (2.8 %)	10 (8.1 %)	4 (3.4 %)	5 (4.1 %)

4. Discussion

Qualitative temporal sensory evaluation methods record sequences of citations (0/1, i.e. perception/no perception) and their occurrence in time. Citations can be analysed with regard to temporal aspects (i.e. the sensory attribute is perceived from time t_1 to time t_2, t_1 being the time of

first citation and (t_2-t_1) the duration of dominance/perception), or not (i.e. whether the sensory attribute is perceived or not, at any time). In this study, the multivariate analyses of citations (with mrCAs) and durations (with CVAs) lead to very similar product discrimination whatever the sensory evaluation method. This result was expected because the two variables are strongly correlated (Visalli, Béno, et al., 2023). Thus, the differences captured by multivariate analyses are carried by the main characteristics of the samples which are common according to citation and durations. This makes the multivariate analyses of durations uninformative about the nature of the differences (durations or presence/absence) between the samples. Moreover, different from mrCA, CVA gives too much weight to attributes few cited, which can lead to misleading interpretation of these differences, as it can be seen on Fig. 2 (with the distractors). This suggests that it might make more sense to use mrCA to multidimensionally summarize differences due to the presence/absence of attributes, then to complement with other analyses better accounting for the temporal aspects of the data.

Univariate analyses (ANOVAs) of citations and durations show that more product pairs are discriminated with citations than with durations, demonstrating that discrimination is more related to the presence/absence of sensory attributes than to their durations. This result concurs with the findings of Bommel et al. (2020) and Vidal et al. (2019) based on the comparison of CATA and TCATA. In this study, regarding product characterization, if most often the differences captured by citations include differences captured by durations or times of citations, not all the differences intersect, as observed by Alcaire et al. (2017). The results presented in Table 5 show that the combined analyses of citations with durations enable to characterize more precisely the nature of the differences between the samples compared to the sole analyses of citations or durations. This is even clearer on Fig. 4, with the controlled stimuli designed to present temporal differences: TDS and TCATA enable samples to be discriminated based on differences in times of citations and durations for some attributes perceived in both samples.

In order to be measurable, univariate temporal differences between samples must be “large enough”, i.e. larger than two seconds with simple controlled stimuli. This value (the temporal resolution of TDS and TCATA) corresponds to the minimum bucket size used by Derks et al. (2022) to discretize temporal data. With complex food products, the minimum measurable temporal differences are likely to be larger. Indeed, even with controlled stimuli, some differences observed in Table 4 are not congruent with the chemical composition of the solutions. Some of these incongruent differences are common to TDS and TCATA and thus can possibly be explained by sensory interactions between attributes. However, some are likely “false positive” in view of the number of unexpected differences also observed between replicated stimuli. Thus, the gain in discrimination capacity obtained with temporal measurements and notably with concurrent evaluation methods is made to the detriment of the reliability and potentially to the validity of the conclusions obtained.

Reliability issues should be considered when interpreting the results, and notably subtle differences. This is especially the case with difference plots, which represent the evolution and the significance of differences in citation rates between pairs of samples. This analysis integrates the three temporal variables: citations, times of citations and durations. This is overall the most discriminative analysis, but also the one presenting the most incongruent and unexpected differences (Tables 4 and 5), as well as the lowest resolution (Fig. 3). This is explained by the fact that different from univariate analyses performed once and at individual level, proportion tests are made at the panel level and thus assume the homogeneity in temporal perception between consumers at each time point (Schlich, 2017), which has been proven to be an assumption that is often not fulfilled (Visalli, Mahieu, et al., 2023). To limit over-interpretation of “false positive” differences, a minimum duration corresponding to 5 % of the total duration of the stimulus has been chosen in this study. However, the difference in gain of discrimination still seems overrated compared to other analyses. This could be due to large differences in times of citations or to specific perception related to clusters of consumers captured only by difference. Indeed, a difference of 10 % can be significant with a proportion test when an attribute is not cited in a sample while it is cited in the other only by a small percentage of consumers. This situation reflects a low confidence in the real perception of the attribute. To avoid overinterpretation, a minimum overall citation rate can be required before testing differences in citation rates (similar to the threshold required before testing the differences in times of citations based on bootstrap tests, set to 25 % in this study). This threshold is difficult to establish (Visalli, Mahieu, et al., 2023). A candidate could be the significance level of the TDS curves, however this threshold is contested (Meyners & Castura, 2019). Indeed, as observed, some attributes can be either specific to some samples or never used, meaning that for some products the threshold used in TDS curves could be underestimated.

The issue with low citation rates observed is not specific to TDS or dominance-based methods but common to all static and temporal qualitative analyses using citation rates to determine whether an attribute has been perceived or not. This topic deserves further investigation, especially since no recommendation exists about the appropriate number of consumers to include in studies involving temporal methods. Conversely, with applicability-based methods and notably TCATA with which citation rates are higher, it is difficult to separate what falls under the agreement and the acquiescence bias observed with CATA (Mahieu et al., 2020) and CATA-like protocols (Callegaro et al., 2015; Kim et al., 2017). Other works tend to confirm that applicability-based methods overestimate citation rates (Visalli, Béno, et al., 2023). This could make the analyses more sensitive and increase the number of false positives with complex food products as seen in this article (Table 5). The question of the reliability and validity of conclusions based on citation rates is not new and has been studied with CATA (Ares et al., 2014; Jaeger et al., 2013), but it remains actual with temporal methods (Velázquez et al., 2020; Visalli, Béno, et al., 2023).

The size of the difference among samples has previously been reported to impact reproducibility of the result (Ares et al., 2014). In this study, the opposite effect is observed as the product category presenting the less differences and the most complexity (Table 1) is also the one for which the less reliability issues are observed. However, the two characteristics were subjectively evaluated by the consumers, thus it would be preferable to estimate the size of differences based on collected data. Adding replicated samples allows to quantify and compare the size of differences between replicated and different pairs of samples. This is the idea underlying the discriminating ratio computed for the difference plots (Fig. 4). A ratio lower or equal to one clearly indicates potential reliability issues due to too much similarity between samples.

It is interesting to notice that on controlled stimuli (Table 4), with a low number of attributes to follow, TCATA has less reliability issues compared to TDS, while the opposite is observed on real food products (Table 5). This suggests that the reliability of TCATA decreases with the number of attributes available to consumers or the complexity of the product, probably due to the increased cognitive load caused by the need to focus and the difficulty to react on multiple events temporally close. Surprisingly, AEF-A does not suffer the same issue. If the method is clearly not appropriate for measuring small temporal differences, it presents a fairly good compromise between discrimination (better than TDS but worse than TCATA) and reliability (better than TCATA but worse than TDS) on real food products. This is probably explained by the fact that the more complex the product space is, the more the consumers focus on differences of applicability rather than on temporal differences. In this sense, AEF-A, as a retrospective method, probably collects data very comparable to those that would have been collected with CATA.

The temporal differences captured with AEF-A with durations or difference plots but not with citations rarely align with the temporal differences captured by TDS or TCATA. This result can either suggest that these differences are reliable but non-valid on real food products (no such issue is observed on controlled stimuli), or confirm that

retrospective measurements result from a different evaluation process than concurrent measurements (read the discussion about Kahneman's theory in Visalli, Wakihira, et al., 2022). More than discrimination, this question is probably the most interesting to investigate as it is relative to the ecological validity of the measurements which is sought after with consumer studies (Galiñanes Plaza et al., 2019). According to this criterion, TCATA and probably even more TDS have an undeniable advantage as they are closest to natural immediate perception.

5. Conclusion

As demonstrated with controlled temporal stimuli, when temporal differences are established, concurrent evaluation methods (TDS and TCATA) undoubtedly perform better for product characterization than retrospective evaluation methods (AEF-D and AEF-A). These latter can be used by consumers to determine the main differences between samples, but due to their low temporal resolution they are obviously not appropriate to capture small temporal differences. However, with the real food products tested presenting unknown temporal differences, AEF-A, in practice very similar to CATA, discriminated samples as well as TDS and TCATA. Indeed, the main differences between samples could be explained by the presence or absence of sensory attributes regardless of temporal aspects. This result raises the question of the real interest of temporal measurements with consumers for answering most product development related questions. When discrimination is sought, using static or low temporal resolution measurements is unlikely to make any difference. Yet, in some cases, analysing time-dependent variables (times of citations and durations) in complement of citations can improve discrimination and bring complementary insights about the nature of the differences. As there is no possibility to anticipate if temporal differences between samples exist before having testing for them, it could therefore be tempting to systematically use temporal methods to avoid missing potential additional information. Furthermore, as perception is a dynamic phenomenon, concurrent evaluation methods are inherently more ecologically valid than retrospective ones, as the latter require consumers to integrate the dynamic perceptual output in a more analytical way. Nonetheless, temporal sensory data are complex to analyse and interpret due to the richness of information, but also of the noise, that they contain. There is no ultimate method, and every gain in temporal resolution or discrimination is assumably at the cost of a loss of reliability. This study shows how difficult it is to generalize conclusions relating to sensory evaluation methods. In the regrettable absence of large open databases of sensory studies that would allow meta-analyses to be conducted, the determination of the "best" method is likely to remain an endless debate.

CRediT authorship contribution statement

Michel Visalli: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Sylvie Cordelle:** Investigation, Writing – review & editing. **Noëlle Béno:** Investigation, Writing – review & editing. **Pascal Schlich:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data are available in data papers on Data in brief.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2024.105123.

References

Adams, J., Williams, A., Lancaster, N., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. *7th Pangborn Sensory Science Symposium*.

Agresti, A. (2013). Categorical data analysis. *Wiley-Interscience*. https://www.wiley.com/en-us/Categorical+Data+Analysis%2C+3rd+Edition-p-9780470463635.

Alcaire, F., Antúnez, L., Vidal, L., Zorn, S., Giménez, A., Castura, J. C., & Ares, G. (2017). Comparison of static and dynamic sensory product characterizations based on check-all-that-apply questions with consumers. *Food Research International, 97*, 215–222. https://doi.org/10.1016/J.FOODRES.2017.04.012

Ares, G., Antúnez, L., Giménez, A., Roigard, C. M., Pineau, B., Hunter, D. C., & Jaeger, S. R. (2014). Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers. *Food Quality and Preference, 36*, 111–121. https://doi.org/10.1016/J.FOODQUAL.2014.03.010

Ares, G., Jaeger, S. R., Antúnez, L., Vidal, L., Giménez, A., Coste, B., Picallo, A., & Castura, J. C. (2015). Comparison of TCATA and TDS for dynamic sensory characterization of food products. *Food Research International, 78*, 148–158.

Bell, A. (2013). Randomized or fixed order for studies of behavioral syndromes? *Behavioral Ecology, 24*(1), 16. https://doi.org/10.1093/BEHECO/ARS148

Béno, N., Nicolle, L., & Visalli, M. (2023). A dataset of consumer perceptions of gustometer-controlled stimuli measured with three temporal sensory evaluation methods. *Data in Brief*. , Article 109271. https://doi.org/10.1016/J.DIB.2023.109271

Berget, I., Castura, J. C., Ares, G., Næs, T., & Varela, P. (2020). Exploring the common and unique variability in TDS and TCATA data – A comparison using canonical correlation and orthogonalization. *Food Quality and Preference, 79*, Article 103790. https://doi.org/10.1016/J.FOODQUAL.2019.103790

Bommel, R. van, Stieger, M. P. dr., Graaf, K. de P. dr., & Jager, G. D. C. (2020). *Time will tell : dynamic sensory characteristics, hedonic perceptions and food-evoked emotions from first to last bite*. Wageningen University.

Bord, C., Guerinon, D., & Lebecque, A. (2019). Application of two sensory methods to investigate the impact of heating on the flavor perception of a French blue cheese. *Journal of Sensory Studies, 34*(5). https://doi.org/10.1111/JOSS.12509

Braghieri, A., Piazzolla, N., Galgano, F., Condelli, N., De Rosa, G., & Napolitano, F. (2016). Effect of preservative addition on sensory and dynamic profile of Lucanian dry-sausages as assessed by quantitative descriptive analysis and temporal dominance of sensations. *Meat Science, 122*, 68–75. https://doi.org/10.1016/J.MEATSCI.2016.07.020

Bruzzone, F., Ares, G., & Giménez, A. (2013). Temporal aspects of yoghurt texture perception. *International Dairy Journal, 29*(2), 124–134. https://doi.org/10.1016/J.IDAIRYJ.2012.10.012

Callegaro, M., Murakami, M. H., Tepman, Z., & Henderson, V. (2015). Yes–no answers versus check-all in self-administered modes: A systematic review and analyses. Doi: 10.2501/IJMR-2015-014a, 57(2), 203–224. doi: 10.2501/IJMR-2015-014A.

Castura, J. C., Antúnez, L., Giménez, A., & Ares, G. (2016). Temporal Check-All-That-Apply (TCATA): A novel dynamic method for characterizing products. *Food Quality and Preference, 47*, 79–90.

Charles, M., Endrizzi, I., Aprea, E., Zambanini, J., Betta, E., & Gasperi, F. (2017). Dynamic and static sensory methods to study the role of aroma on taste and texture: A multisensory approach to apple perception. *Food Quality and Preference, 62*, 17–30. https://doi.org/10.1016/J.FOODQUAL.2017.06.014

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Delompré, T., Lenoir, L., Martin, C., Briand, L., & Salles, C. (2020). Characterizing the dynamic taste and retro-nasal aroma properties of oral nutritional supplements using temporal dominance of sensation and temporal check-all-that-apply methods. *Foods, 9*(10). https://doi.org/10.3390/FOODS9101456

Derks, E., Ramnarain, S., Zhang, T., van Doorn, R., Nijmeijer, M., & van den Berg, M. (2022). Visualizing dynamic (after)taste effects by means of time-discrete TCATA data analysis. *Journal of Sensory Studies, 37*(3), e12737.

Devezeaux de Lavergne, M., van Delft, M., van de Velde, F., van Boekel, M. A. J. S., & Stieger, M. (2015). Dynamic texture perception and oral processing of semi-solid food gels: Part 1: Comparison between QDA, progressive profiling and TDS. *Food Hydrocolloids, 43*, 207–217. https://doi.org/10.1016/J.FOODHYD.2014.05.020

Esmerino, E. A., Castura, J. C., Ferraz, J. P., Tavares Filho, E. R., Silva, R., Cruz, A. G., Freitas, M. Q., & Bolini, H. M. A. (2017). Dynamic profiling of different ready-to-drink fermented dairy products: A comparative study using Temporal Check-All-That-Apply (TCATA), Temporal Dominance of Sensations (TDS) and Progressive Profile (PP). *Food Research International, 101*, 249–258. https://doi.org/10.1016/J.FOODRES.2017.09.012

Etaio, I., Meillon, S., Pérez-Elortondo, F. J., & Schlich, P. (2016). Dynamic sensory description of Rioja Alavesa red wines made by different winemaking practices by using Temporal Dominance of Sensations. *Journal of the Science of Food and Agriculture, 96*(10), 3492–3499. https://doi.org/10.1002/JSFA.7533

Frost, S. C., Blackman, J. W., Ebeler, S. E., & Heymann, H. (2018). Analysis of temporal dominance of sensation data using correspondence analysis on Merlot wine with

differing maceration and cap management regimes. *Food Quality and Preference, 64*, 245–252.
Galiñanes Plaza, A., Delarue, J., & Saulais, L. (2019). The pursuit of ecological validity through contextual methodologies. *Food Quality and Preference, 73*, 226–247. https://doi.org/10.1016/J.FOODQUAL.2018.11.004
Galmarini, M. V., Visalli, M., & Schlich, P. (2017). Advances in representation and analysis of mono and multi-intake Temporal Dominance of Sensations data. *Food Quality and Preference, 56*, 247–255. https://doi.org/10.1016/j.foodqual.2016.01.011
Gonzalez-Estanol, K., Cliceri, D., Biasioli, F., & Stieger, M. (2022). Differences in dynamic sensory perception between reformulated hazelnut chocolate spreads decrease when spreads are consumed with breads and wafers. *Food Quality and Preference, 98*. https://doi.org/10.1016/J.FOODQUAL.2022.104532
Hutchings, S. C., Cha, W., Dunshea, F. R., Sharma, C., & Torrico, D. D. (2022). Understanding dominance: The effect of changing the definition of dominance when using TDS with consumers. *Journal of Sensory Studies, 37*(4), e12750.
Jaeger, S. R., Chheang, S. L., Yin, J., Bava, C. M., Gimenez, A., Vidal, L., & Ares, G. (2013). Check-all-that-apply (CATA) responses elicited by consumers: Within-assessor reproducibility and stability of sensory product characterizations. *Food Quality and Preference, 30*(1), 56–67. https://doi.org/10.1016/J.FOODQUAL.2013.04.009
Kim, I. A., Hopkinson, A., van Hout, D., & Lee, H. S. (2017). A novel two-step rating-based 'double-faced applicability' test. Part 1: Its performance in sample discrimination in comparison to simple one-step applicability rating. *Food Quality and Preference, 56*, 189–200. https://doi.org/10.1016/J.FOODQUAL.2016.10.010
Labbe, D., Schlich, P., Pineau, N., Gilbert, F., & Martin, N. (2009). Temporal dominance of sensations and sensory profiling: A comparative study. *Food Quality and Preference, 20*(3), 216–221. https://doi.org/10.1016/J.FOODQUAL.2008.10.001
Lenfant, F., Loret, C., Pineau, N., Hartmann, C., & Martin, N. (2009). Perception of oral food breakdown. The concept of sensory trajectory. *Appetite, 52*(3), 659–667.
Mahieu, B., Schlich, P., Visalli, M., & Cardot, H. (2021). A multiple-response chi-square framework for the analysis of Free-Comment and Check-All-That-Apply data. *Food Quality and Preference, 93*, Article 104256. https://doi.org/10.1016/J.FOODQUAL.2021.104256
Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020). Free-comment outperformed check-all-that-apply in the sensory characterisation of wines with consumers at home. *Food Quality and Preference, 84*, Article 103937. https://doi.org/10.1016/J.FOODQUAL.2020.103937
McMahon, K. M., Culver, C., Castura, J. C., & Ross, C. F. (2017). Perception of carbonation in sparkling wines using descriptive analysis (DA) and temporal check-all-that-apply (TCATA). *Food Quality and Preference, 59*, 14–26. https://doi.org/10.1016/J.FOODQUAL.2017.01.017
Meillon, S., Urbano, C., & Schlich, P. (2009). Contribution of the Temporal Dominance of Sensations (TDS) method to the sensory description of subtle differences in partially dealcoholized red wines. *Food Quality and Preference, 20*(7), 490–499. https://doi.org/10.1016/J.FOODQUAL.2009.04.006
Meyners, M. (2020). Temporal methods: Are we comparing apples and oranges? *Food Quality and Preference, 79*. https://doi.org/10.1016/j.foodqual.2018.11.022
Meyners, M., & Castura, J. C. (2019). Did assessors select attributes by chance alone in your TDS study, and how relevant is it to know? *Food Research International, 119*, 571–583. https://doi.org/10.1016/j.foodres.2018.10.035
Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research, 141*, 4–15. https://doi.org/10.1016/J.VISRES.2017.11.001
Ng, M., Lawlor, J. B., Chandra, S., Chaya, C., Hewson, L., & Hort, J. (2012). Using quantitative descriptive analysis and temporal dominance of sensations analysis as complementary methods for profiling commercial blackcurrant squashes. *Food Quality and Preference, 25*(2), 121–134. https://doi.org/10.1016/J.FOODQUAL.2012.02.004
Nguyen, H., & Wismer, W. V. (2022). Temporal sensory profiles of regular and sodium-reduced foods elicited by temporal dominance of sensations (TDS) and Temporal Check-All-That-Apply (TCATA). *Foods, 11*(3), 457. https://doi.org/10.3390/FOODS11030457/S1
Nguyen, Q. C., Næs, T., & Varela, P. (2018). When the choice of the temporal method does make a difference: TCATA, TDS and TDS by modality for characterizing semi-solid foods. *Food Quality and Preference, 66*, 95–106.
Paglarini, C. de S., Vidal, V. A. S., dos Santos, M., Coimbra, L. O., Esmerino, E. A., Cruz, A. G., & Pollonio, M. A. R. (2020). Using dynamic sensory techniques to determine drivers of liking in sodium and fat-reduced Bologna sausage containing functional emulsion gels. *Food Research International, 132*, 109066. doi: 10.1016/J.FOODRES.2020.109066.
Paulsen, M. T., Nys, A., Kvarberg, R., & Hersleth, M. (2014). Effects of NaCl substitution on the sensory properties of sausages: Temporal aspects. *Meat Science, 98*(2), 164–170. https://doi.org/10.1016/J.MEATSCI.2014.05.020
Pineau, N., Schlich, P., Cordelle, S., Mathonnière, C., Issanchou, S., Imbert, A., Rogeaux, M., Etiévant, P., & Köster, E. (2009). Temporal dominance of sensations: Construction of the TDS curves and comparison with time–intensity. *Food Quality and Preference, 20*(6), 450–455.
R Core team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. url.
Schlich, P., Medel Maraboli, M., Urbano, C., & Parr, W. V. (2015). Perceived complexity in Sauvignon blanc wines: Influence of domain-specific expertise. *Australian Journal of Grape and Wine Research, 21*(2), 168–178. https://doi.org/10.1111/AJGW.12129
Schlich, P. (2017). Temporal Dominance of Sensations (TDS): A new deal for temporal sensory analysis. *Current Opinion in Food Science, 15*, 38–42. https://doi.org/10.1016/J.COFS.2017.05.003
Sharma, M., & Duizer, L. (2019). Characterizing the dynamic textural properties of hydrocolloids in pureed foods - A comparison between TDS and TCATA. *Foods, 8*(6). https://doi.org/10.3390/FOODS8060184
Sroka, C. J., & Nagaraja, H. N. (2018). Odds ratios from logistic, geometric, Poisson, and negative binomial regression models. *BMC Medical Research Methodology, 18*(1), 1–11. https://doi.org/10.1186/S12874-018-0568-9/FIGURES/1
Varela, P., Mosca, A. C., Nguyen, Q. C., McEwan, J. A., & Berget, I. (2021). Individual differences underlying food intake and liking in semisolid foods. *Food Quality and Preference, 87*, Article 104023. https://doi.org/10.1016/J.FOODQUAL.2020.104023
Velázquez, A. L., Vidal, L., Varela, P., & Ares, G. (2020). Can children use temporal sensory methods to describe visual and food stimuli? *Food Quality and Preference, 86*, Article 104002. https://doi.org/10.1016/J.FOODQUAL.2020.104002
Vidal, L., Antúnez, L., Ares, G., Cuffia, F., Lee, P. Y., Le Blond, M., & Jaeger, S. R. (2019). Sensory product characterisations based on check-all-that-apply questions: Further insights on how the static (CATA) and dynamic (TCATA) approaches perform. *Food Research International, 125*, Article 108510. https://doi.org/10.1016/J.FOODRES.2019.108510
Visalli, M., Mahieu, B., Thomas, A., & Schlich, P. (2020). Concurrent vs. retrospective temporal data collection: Attack-evolution-finish as a simplification of Temporal Dominance of Sensations? *Food Quality and Preference, 85*, Article 103956. https://doi.org/10.1016/j.foodqual.2020.103956
Visalli, M., Béno, N., Nicolle, L., & Schlich, P. (2023). Assessment of the validity and reliability of temporal sensory evaluation methods used with consumers on controlled stimuli delivered by a gustometer. *Food Quality and Preference, 110*, Article 104942. https://doi.org/10.1016/J.FOODQUAL.2023.104942
Visalli, M., Cordelle, S., Mahieu, B., Pedron, C., Hoffarth, B., Praudel, M., Coutière, M., & Schlich, P. (2022). A dataset of sensory perception of chocolates, guacamoles, ice teas and crisps collected with consumers using six temporal methods. *Data in Brief, 45*, Article 108708. https://doi.org/10.1016/j.dib.2022.108708
Visalli, M., Mahieu, B., Peltier, C., Cordelle, S., & Schlich, P. (2023). Temporal precision and resolution of sensory evaluation methods used with consumers. *Food Quality and Preference (Submitted)*.
Visalli, M., Wakihira, T., & Schlich, P. (2022). Concurrent vs. immediate retrospective temporal sensory data collection: A case study on lemon-flavoured carbonated alcoholic drinks. *Food Quality and Preference, 101*, Article 104629. https://doi.org/10.1016/j.foodqual.2022.104629
Wu, A. Z., Lee, R. W., Calvé, B. L., & Cayeux, I. (2019). Temporal profiling of simplified lemonade using temporal dominance of sensations and temporal check-all-that-apply. *Journal of Sensory Studies, 34*(6), e12531.