Concurrent vs. immediate retrospective temporal sensory data collection: A case study on lemon-flavoured carbonated alcoholic drinks

Michel Visalli, Takahiro Wakihira, Pascal Schlich

Title

Concurrent vs. immediate retrospective temporal sensory data collection: A case study on lemon-flavoured carbonated alcoholic drinks

Auteurs

Michel Visalli[1,2*], Takahiro Wakihira[3], Pascal Schlich[1,2]

[1] Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, Université de Bourgogne Franche-Comté, F-21000 Dijon, France
[2] INRAE, PROBE research infrastructure, ChemoSens facility, F-21000 Dijon, France
[3] Brewing Science Laboratories, Asahi Quality & Innovations, Ltd., 1-21 Midori 1-Chome, Moriya, Ibaraki 302-0106, Japan

Keywords

Consumer study. Temporal perception. Home tasting. Method comparison. Repeated measures. Cognitive load.

Abstract

Most temporal sensory methods measure product perception concurrently with tasting. However, retrospective measurement could have the advantage of being easier to implement with consumers. To date, no study has explicitly compared these two modes of temporal data acquisition. The objective of this study was to compare the temporal description obtained from consumers who had drank and evaluated a full can of four lemon-flavoured carbonated alcoholic drinks at home and on different days. The consumers were separated into two panels and asked to select from eight attributes on a check-all-that-apply list that were applicable during three periods of perception of a sip— "in mouth before swallowing", "immediately after swallowing" and "aftertaste"—for three sips of each can. The first panel (RET, 97 consumers) did the evaluation retrospectively immediately after the tasting, while the second (SIM, 96 consumers) did it concurrently with the tasting. Data were analysed using the multiple-response correspondence analysis (mrCA) framework applied at different levels: product, period and sip. The data from RET and SIM captured the differences between products and periods, with the differences between the products being larger than the differences between the periods. In both methods, no sip effect were observed. Perception of the products were identical in sips 1, 4 and 7. The consumers of RET and SIM agreed overall on the relative differences between products, although the level of discrimination was stronger for the consumers of RET. The consumers of RET and SIM only partially agreed on the between-period differences by product. The results suggest that the retrospective temporal evaluation could better discriminate the products and that the temporality patterns may be different between the two methods; therefore, there is a need for more research about the resolution of temporal data.

1. Introduction

Temporal perception has been investigated in sensory science for the last 60 years using different methodologies (Kemp et al., 2017). The first developed methods measured the evolution of the intensity of the attributes over time. Some did it in a continuous way: time-intensity (TI, Lee & Pangborn, 1986), dual-attribute time-intensity (DATI, Duizer et al., 1997) or multi-attribute time-intensity (MATI, Kuesten et al., 2013). Others did it at fixed discrete time points: discrete time-intensity (Clark & Lawless, 1994), progressive profiling (Jack et al., 1994) or sequential profiling (Methven et al., 2010). More recent methods, which did not rely on intensities, measured the evolution of dominant attributes (Temporal Dominance of Sensations - TDS, Pineau et al., 2009) or applicable attributes (Temporal Check

All That Apply - TCATA, Castura et al., 2016) in continuous time. The common point between all these methods is that the data collection is carried out simultaneously to the tasting.

Few studies have applied retrospective temporal measurements in sensory analysis. Visalli et al. (2020) proposed the attack-evolution-finish (AEF) method to study the temporal perception of chocolates. They compared the results obtained with AEF with those obtained with TDS and concluded on very similar product discrimination. Mahieu, Visalli, Thomas, et al. (2020) extended the AEF concept by using Free-Comment, avoiding the issues inherent in the use of a predefined list of descriptors. While losing temporal resolution, the two methods present the advantage of being easy to implement with consumers, as the task does not require specific training or familiarization, contrary to TDS (Boinbaser et al., 2015; Hutchings et al., 2014; Kantono et al., 2018; Rodrigues et al., 2016; Arnaud Thomas et al., 2015; Velázquez et al., 2020) or TCATA (Gastón Ares et al., 2015; Jaeger et al., 2017; Rizo et al., 2020; Weerawarna et al., 2021).

The lack of attention to retrospective measurements in sensory analysis can probably be explained by the tacit assumption that concurrent temporal sensory data collection is more precise and less biased than retrospective data collection. Some psychologists have judged real-time data capture preferable for measuring changes over time, using concurrent assessment to ask people to report on their current experience (Stone et al., 2007). Others claimed that concurrent data collection does not eliminate other potential sources of bias in self-reports (Schwarz, 2012). Aldrovandi et al. (2015) presented a series of studies showing that retrospective evaluation of event sequences could be biased by memory and notably by the tendency to retrieve the most readily available information that is itself related to the valence of the information. However, meta-analytic findings (Block et al., 2010) revealed that if the duration was less than 60 seconds, there was no difference observed between conclusions obtained with simultaneous and retrospective paradigm. Noncongruent results related to taste perception have been reported. Liang et al. (2018) showed that the stress of the memory load influenced sensory perception by decreasing taste sensibility, while Daniel and Katz (2018) showed that a higher accuracy in taste recognition was observed over a 30-s delay. Recently, Botha et al. (2021) compared TCATA data collection with a retrospective approach for phenotyping thermal taste. The authors judged that TCATA avoids memory effects and allows participants to fully focus on the new taste instead of relying on their working memory, but they also recognized that the cognitive aspects of TCATA have not been studied to date. Varela et al. (2018) also used an immediate retrospective think-aloud protocol to review TDS results to better understand the dominance concept and gain insight about the temporal perception.

Indeed, memory is not the only aspect to consider, as it is just one component of the cognitive load. The cognitive load is a complex and multidimensional construct consisting of causal factors related to the task and the subject and assessment factors related to mental load, mental effort, and performance (Paas & Van Merriënboer, 1994). In sensory analysis, TDS and TCATA have been reported to be potentially demanding tasks by some authors (Botha et al., 2021; Castura et al., 2016; Jaeger et al., 2017; Pineau et al., 2012). This could affect the outcomes of these methods, as Wal & Dillen (2013) suggested that an increased task load could reduce taste perception. This was partially confirmed by Velázquez et al. (2020), who showed that a TCATA task was feasible by children on a simple video stimulus, but when applied to more complex food stimuli, they used TCATA (and TDS) methods as static methods.

The performances of retrospective and concurrent measurements have been compared in other research fields, still leading to controversy. Church et al. (2019) compared humans' metacognitive performances (the ability to choose between two answers with and without feedback) in prospective and retrospective paradigms, showing that the metacognitive response was used more robustly and accurately retrospectively when it was not in direct competition with the primary perceptual responses. Van Den Haak et al. (2003) showed that concurrent and retrospective think-aloud protocols revealed comparable sets of usability problems, but in the concurrent protocols, the requirement to think aloud while working had a negative effect on the task performance, raising the question about the reactivity of concurrent think-aloud protocols in the case of high task complexity. Peute et al. (2015) showed that the concurrent think-aloud method was more efficient in assessing the usability of a data query tool but also that it did not outperform retrospective methods that additionally elucidated unique problems. Kuusela & Paul

(2000) reported that the concurrent protocol generally outperformed the retrospective method in verbal protocol analysis, but they also noticed that the retrospective method provided more statements about final choices. Similarly, Whyte et al. (2010) concluded that concurrent verbal reports provided by nurses during and after administering care in a simulated task environment provided a more complete representation of the cognitions of research participants, but additional unique data were exclusively present in the retrospective reports. Ryan & Haslegrave (2007) showed that concurrent reports only contained a proportion of information about workers' thoughts, while additional information could be obtained from retrospective reports, although these appeared to be vulnerable to bias and reordering of reported information due to the rationalization of thoughts.

With regard to this context, the objective of this study was to compare the sensory descriptions of commercial products collected from consumers using either a concurrent or a retrospective data collection method.

2. Materials and methods

2.1. Simultaneous and retrospective protocols

Two data collection protocols based on Check-all-that-apply (CATA, Ares & Jaeger, 2015) questions were used in this study, the first one simultaneous to the tasting (SIM), the second one retrospective to the tasting (RET). A detailed description of the two protocols can be found in the related data paper (Visalli et al., 2022).

2.1.1. Samples

The five products (P1, P2, P3, P4, and P5) were 350 ml cans of commercial lemon-flavoured carbonated drinks. P1 was a nonalcoholic drink that served as a warm-up. P2 to P5 were white liquor-based (Japanese Shochu or Vodka) alcoholic drinks, referred to as "Chu-hai" in Japanese. The cans were blinded by white-colored masking films and coded using three-digit labels, and they were presented according to a Williams Latin square, but P1 was served first to every consumer.

2.1.2. Consumers

Two hundred consumers aged 20 to 39, who were regular consumers of lemon-flavoured carbonated alcoholic beverages, were recruited through an online questionnaire from a panel of consumers belonging to a research agency in Japan. "Regular" referred to consumers drinking flavoured carbonated alcoholic beverages with a frequency of at least twice a week and lemon-flavoured carbonated alcoholic beverages at least once a month. The design of the test was explained to consumers in the online questionnaire. They were informed they would do the test on their smartphones, and they had to sign a consent form to participate in the study. They were financially compensated for their participation. The consumers were separated into two panels that were balanced in their composition (gender, age, frequency of consumption). The first panel (RET) had to evaluate the products retrospectively to the tasting just after they declared that they no longer perceived anything, while the second (SIM) evaluated them concurrently to the tasting.

2.1.3. Descriptors

The same list of descriptors was provided for both the RET and SIM panels: Alcohol, Bitter, Carbonated, Lemon, Refreshing, Sour, Sweet aroma and Sweet taste. (In Japan, sweet aroma is used when it is not possible to describe detailed quality of sweetness in terms of aroma. Sweet aroma can include different types of aromas, such as fruity, floral, caramel, vanilla, honey, etc.) The descriptors were presented as a check-all-that-apply (CATA) list in a random order on the screen, but this order was constant for each consumer across evaluations. No definitions of the descriptors were given to the consumers.

2.1.4. Experimental procedure

The consumers had to evaluate the products using the browser of their smartphones (TimeSens V2 web app). The procedures for the full consumption of the can is summarized in Figure 1. Only three sips were evaluated to limit the duration of the task and to avoid boredom (Thomas et al., 2018).

[INSERT FIGURE 1 HERE]

Figure 1. Description of the tasting of the drink over seven sips (common for RET and SIM). Only sips 1, 4 and 7 were evaluated.

[INSERT FIGURE 2 HERE]

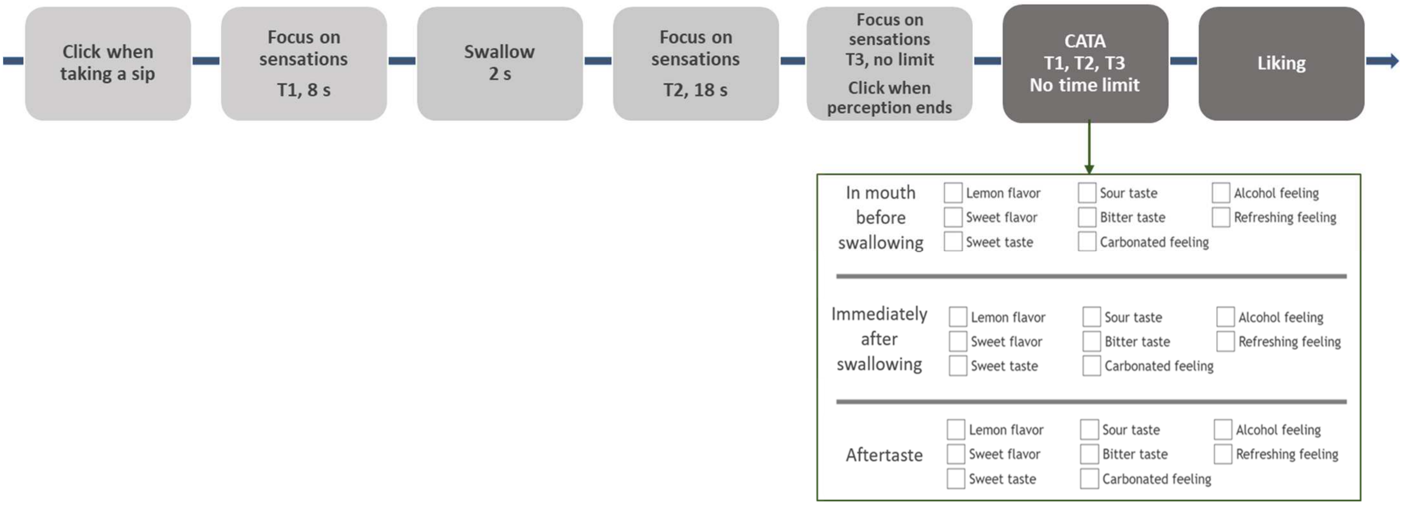

Figure 2. Evaluation of sips 1, 4, 7 for panel RET using retrospective CATA. Only one CATA measurement screen was displayed, after the tasting. "In mouth before swallowing", "Immediately after swallowing" and "Aftertaste" correspond to periods T1, T2 and T3.

Figure 2 describes the tasting of sips 1, 4, 7 for panel RET. During the tasting, several screens were displayed to guide the consumer. A first screen invited the consumer to take a sip without swallowing. Then, a screen was displayed during eight seconds (T1, "In mouth before swallowing") instructing the consumer not to swallow while focusing on perceived sensations and memorizing them. Then, a screen was displayed during two seconds, inviting the consumer to swallow. Then, a screen was displayed during eighteen seconds (T2, "Immediately after swallowing") instructing the consumer to focus on perceived sensations and memorize them. Then, a screen instructing the consumer to focus on perceived sensations and memorize them was displayed until the consumer declared he no longer perceived anything (T3, "Aftertaste"). Then, the CATA screen was displayed with no time limit, with the instruction: "What did you perceive during the different steps of the tasting? Please select all the sensations that apply during each period". Finally, the liking score was asked to the consumer (on a 0-10 continuous scale, 0 being labelled "not at all" and 10 "very much").

[INSERT FIGURE 3 HERE]

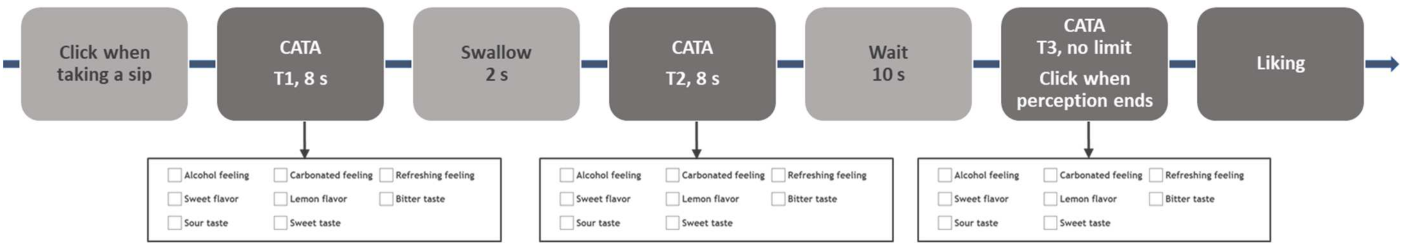

Figure 3. Evaluation of sips 1, 4, 7 for panel SIM using simultaneous CATA. The CATA measurement screen was displayed during the tasting within periods "In mouth before swallowing" (T1), "Immediately after swallowing" (T2) and "Aftertaste" (T3).

Figure 3 describes the tasting of sips 1, 4, 7 for panel SIM. During the tasting, several screens were displayed to guide the consumer. A first screen invited the consumer to take a sip without swallowing. Then, a CATA screen was displayed during eight seconds (T1, "In mouth before swallowing"), with the instruction: "What do you perceive now? Please select all the sensations that apply". Then, a screen was displayed during two seconds, inviting the consumer to swallow. Then, a screen was displayed during eight seconds (T2, "Immediately after swallowing") instructing the consumer to select all the sensations that apply (same instruction as in T1). Then, a screen inviting the consumer to wait was display during ten seconds. Then, the last CATA screen was displayed until the consumer declared he no longer perceived anything (T3, "Aftertaste"), instructing the consumer to select all the sensations that apply (same instruction as in T1). Finally, the liking score was asked to the consumer (on a 0-10 continuous scale, 0 being labelled "not at all" and 10 "very much").

2.2. Data analysis

This article focuses on methodological aspects related to concurrent and retrospective evaluation. Conclusions on products and liking will be published elsewhere.

The data for product P1 (the warm-up) were excluded. Six consumers had incomplete data in the SIM session and one in RET because they did not complete their sessions. Their data were kept because the statistical methods used can deal with incomplete datasets.

Individual consumer data were stored as a table with seven columns: "Panel" (RET, SIM), "Consumer", "Sip number" (1, 4 or 7), "Product" (P2, P3, P4 or P5), "Attribute" (Carbonated, SweetF, Refreshing, Bitter, Lemon, SweetT, Sour, Alcohol), "Period" (T1, T2, T3) and "Score" (1 applicable, 0 otherwise). Each consumer contributed to 288 rows (3 sips x 4 products x 8 descriptors x 3 periods).

Statistical analyses were performed using R 4.1.0 software (R Core team, 2020) and the package MultiResponseR (Mahieu, 2021).

2.2.1. Use of descriptors

For the RET and SIM panels, mean number of citations and descriptors were obtained in various ways. Citations were averaged: (i) across sips, products and consumers within each period (T1, T2, T3); (ii) across products and consumers by sip (1, 4 or 7) and overall. Descriptors were averaged by sip (1, 4 or 7) and overall: (i) across products and consumers; (ii) across consumers. The 95% confidence intervals were computed based on Poisson log-linear models for count data (R function glm, family = poisson). Two means were considered different when their confidence intervals did not intersect.

2.2.2. Product/period/sip comparisons

CATA results were aggregated into contingency tables, the column variables being the descriptors (see examples in supplementary materials). For studying product-by-sip sensory trajectories over periods, one contingency table was computed for each panel, the row variables being the combinations of "product x period x sip". For studying between-sip differences, twelve contingency tables (one for each "product x period" combination) were computed for each panel, the row variables being the sip. For studying between-product differences, four contingency tables (one for each period plus one with citations averaged over periods, with sips being pooled and considered to be independent observations) were computed for each panel, the row variables being the products. For studying between-period differences, five contingency tables (one for each product plus one with citations averaged over products, with sips being pooled and considered to be independent observations) were computed for each panel, the row variables being the periods.

As the consumers could check multiple attributes for a product during a period, the usual χ^2 framework was not well suited to analyse these multiple-response data. To overcome this limitation, the approach described in Mahieu et al. (2021) was used. The dimensionality of the dependence between row and column variables was tested using the dimensionality test (2000 simulations) based on multiple-response χ^2 framework. Then, if and only if at least one dimension was significant (alpha=0.05), the multiple-response correspondence analysis (mrCA) was computed on the contingency table. Outputs of mrCA were displayed using a standard biplot, and 95% confidence ellipses were computed with a total bootstrap procedure (Cadoret & Husson, 2013) with 2000 simulations, Procrustes rotations being performed on the significant dimensions. When the periods were analyzed, the sensory trajectory of each product was represented by a two-segment arrow from period 1 to 2 and then from period 2 to 3 as in Mahieu et al. (2020). For each pair of row variables of the contingency table, a total bootstrap test was performed on the significant dimensions for assessing the significance of difference. When significant (alpha=0.05), for each pair of row and column variables (cell), a multiple-response hypergeometric test (2000 simulations, alpha = 0.05) was performed to test if the descriptor was cited in a proportion that was significantly greater than the overall average citation proportion. No adjustments have been made for the multiple hypergeometric tests.

2.2.3. Feedback of the consumers about the task

The answers to the free-text question about the difficulty of the task (Q6, Visalli et al., 2022) were investigated. Q6 was labelled: "Please tell us what you felt about this tasting survey (if the tasting method was difficult, easy, etc.)". Similar answers about the difficulty of the task were manually grouped by the experimenter, then counted. No statistical analysis was performed with these data.

3. Results

One hundred and ninety-three consumers participated in the study: 97 consumers in panel RET, and 96 in panel SIM.

3.1. Use of descriptors by the consumers

[INSERT FIGURE 4 HERE]

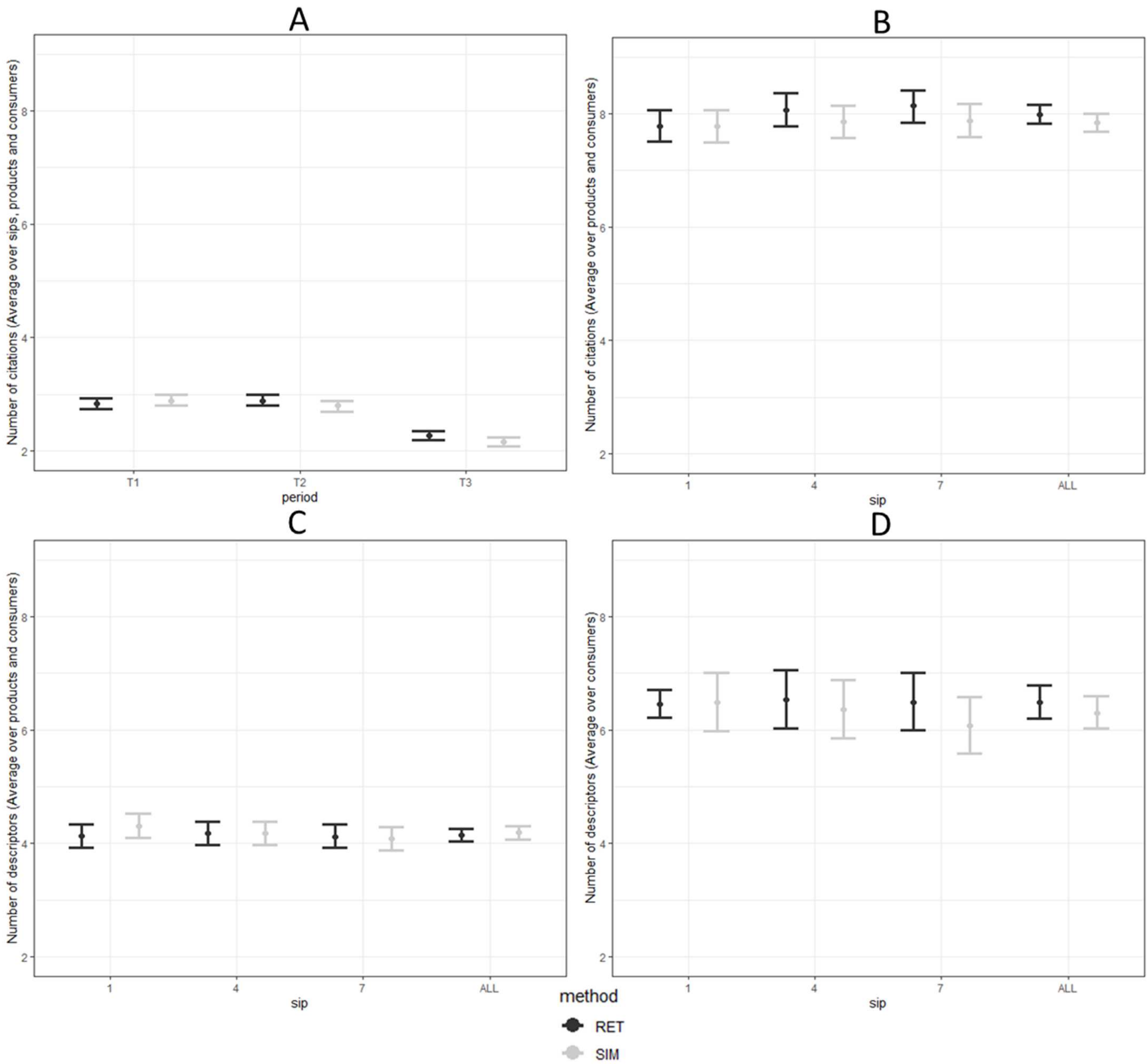

Figure 4: A – Number of citations averaged over sips, products and consumers, by period. B – Number of citations averaged over products and consumers, by sip and all sips pooled. C – Number of descriptors averaged over products and consumers, by sip and all sips pooled. D – Number of descriptors averaged over consumers, by sip and all sips pooled. For all figures, vertical bars represent 95% confidence intervals of the means based on Poisson log-linear models for count data.

Figure 4 plots how the descriptors have been used by the two panels. Figure 4A shows that the consumers of the two panels cited almost 3 descriptors by product in periods T1 and T2; then, this number significantly decreased to just over two in period T3. Figure 4B shows that the number of citations by product was approximately eight. These number of citations slightly increased over the three sips (not significantly). Figure 4C shows that the number of descriptors used by the product was approximately four for the two panels. The number of descriptors slightly decreased over the sips in panel SIM (not significantly). Figure 4D shows that the total number of descriptors used to characterize the four products was between six and seven (over the eight of the CATA list). The decrease observed in Figure 4D for panel SIM was more pronounced but still not significant.

3.2. Product comparison

3.2.1. Product-by-sip sensory trajectories over periods

[INSERT FIGURE 5 HERE]

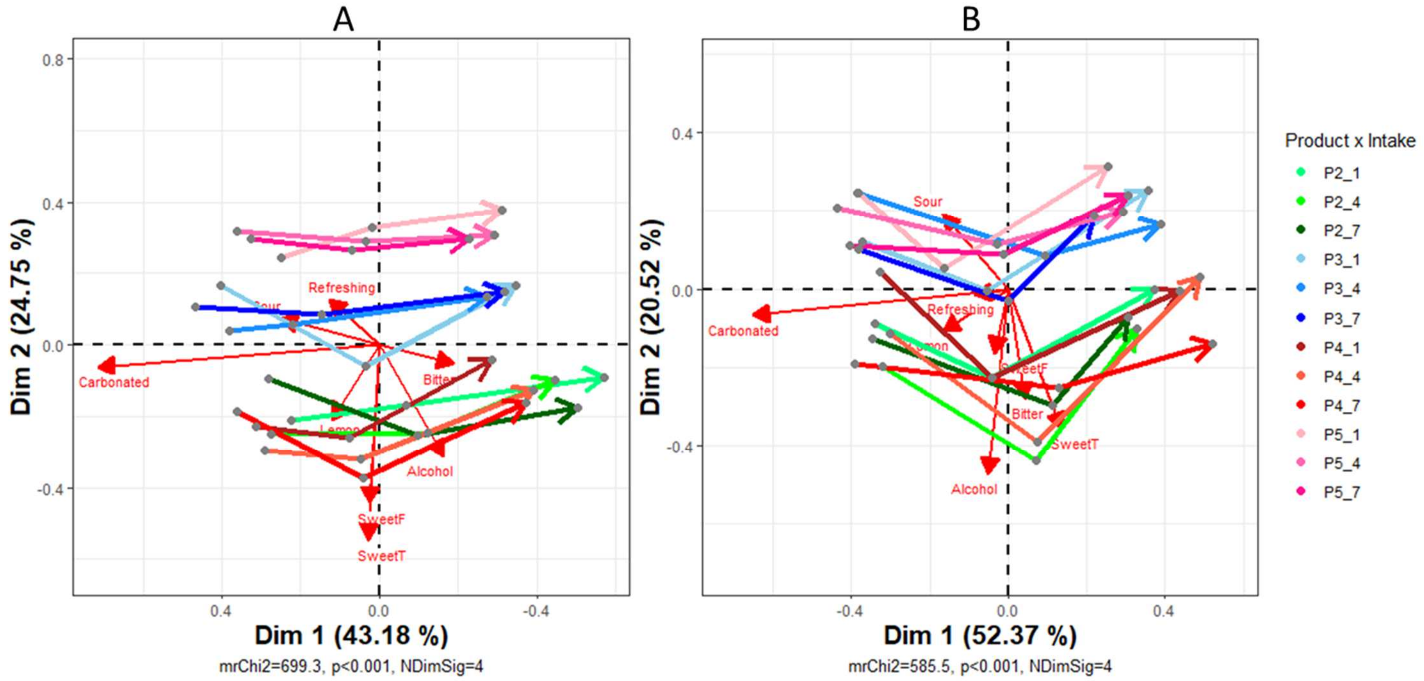

Figure 5: mrCA of 36 evaluations (product x sip x period) representing the product by sip sensory trajectories over periods on dimensions 1 and 2. Figure 5A (left) corresponds to the RET panel, and Figure 5B (right) corresponds to the SIM panel. "mrChi2" is the statistic of the mr χ^2 for the dimensionality test, "p" is the p-value of the dimension test, "NDimSig" is the number of significant dimensions. The sensory trajectory of each product was represented by a two-segment arrow from period 1 to 2 and then from period 2 to 3. Red arrows (with the triangle arrowheads) indicate the relative importance and correlation between the descriptors and the dimensions.

For the two panels, the first dimension depicted the evolution from *Carbonated* to *Bitter* over the three successive periods of perception, whereas the second dimension opposed the products mostly characterized by *Sour* and *Refreshing* to the products mostly characterized by *SweetT, SweetF* and *Alcohol*. *Bitter* was perceived more in period T3 in panel RET and in period T2 in panel SIM. The trajectories were straighter in the RET panel than in the SIM panel, meaning that the sips were not that different (for RET panel) in dimension two within the same product. The same number (4) of significant dimensions was observed, denoting a similar complexity of differences. However, discrimination was better with panel RET (mr χ^2 of 699.3 vs. 585.5 for SIM). The period trajectories exhibited a clear evolution of perception, meaning that the descriptors were not selected in the same way in periods T1 (more citations of *Carbonated*), T2 (more citations of *SweetF*), and T3 (less citation in average). However, these trajectories seem to be rather similar across products. Products P2 and P4 showed similar trajectories for both panels. Products P3 and P5 also showed similar trajectories for panel SIM but were more separated for panel RET.

There was almost no evolution of perception across sips regardless of the product, except for a tendency with product P4 in period 1 with the panel SIM (detailed results are presented in supplementary materials). Therefore, in the rest of this article, sips will subsequently be pooled and considered to be independent observations.

[INSERT FIGURE 6 HERE]

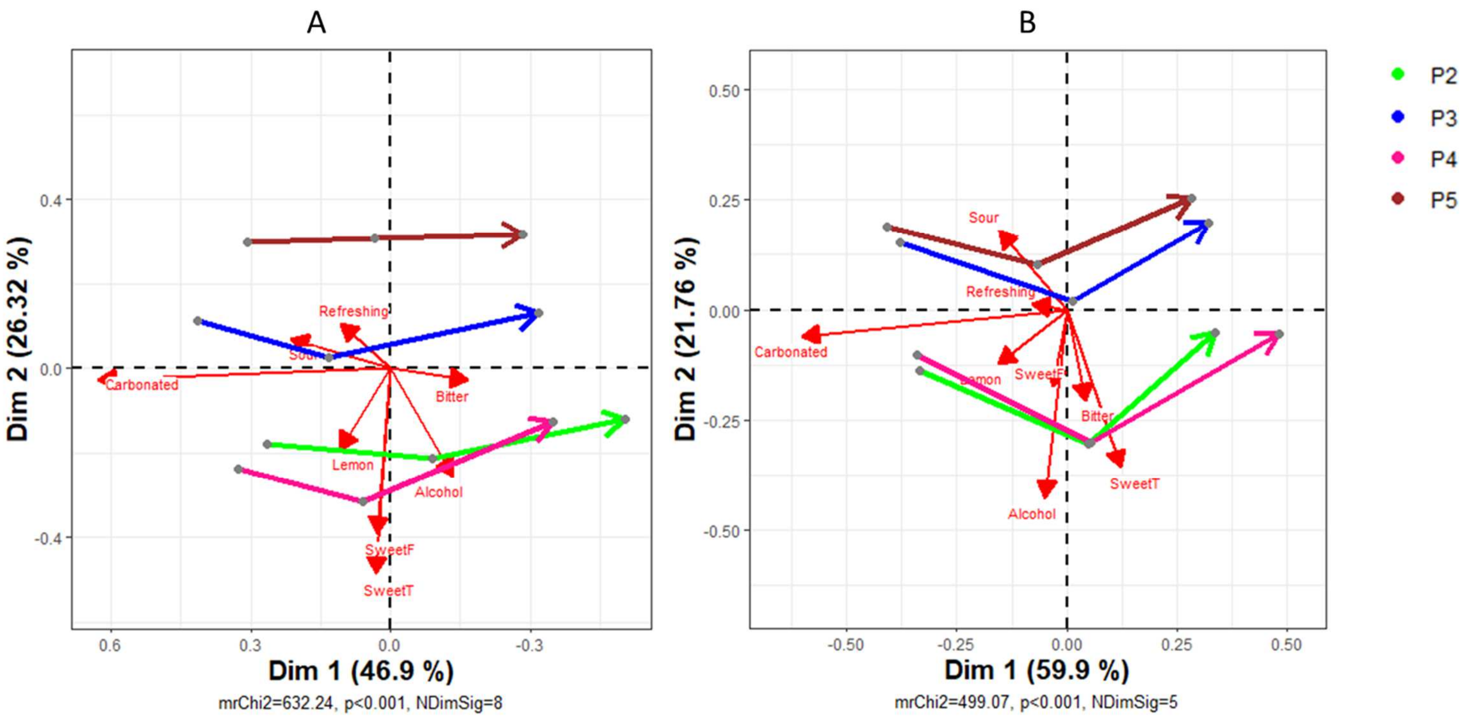

Figure 6: mrCA of 12 evaluations (product x period, sips as independent observations) representing the product sensory trajectories over periods on dimensions 1 and 2. Figure 6A (left) corresponds to the RET panel, and Figure 6B (right) corresponds to the SIM panel. "mrChi2" is the statistic of the mr χ^2 for the dimensionality test, "p" is the p-value of the dimension test, "NDimSig" is the number of significant dimensions. The sensory trajectory of each product was represented by a two-segment arrow from period 1 to 2 and then from period 2 to 3. Red arrows indicate the relative importance and correlation between the descriptors and the dimensions.

Figure 6 shows the same sensory interpretation as Figure 5 The RET panel was still more discriminative (632.24 vs. 499.07) and exhibited more complex differences (eight vs. five significant dimensions).

3.2.2. Comparison of products by period

[INSERT FIGURE 7 HERE]

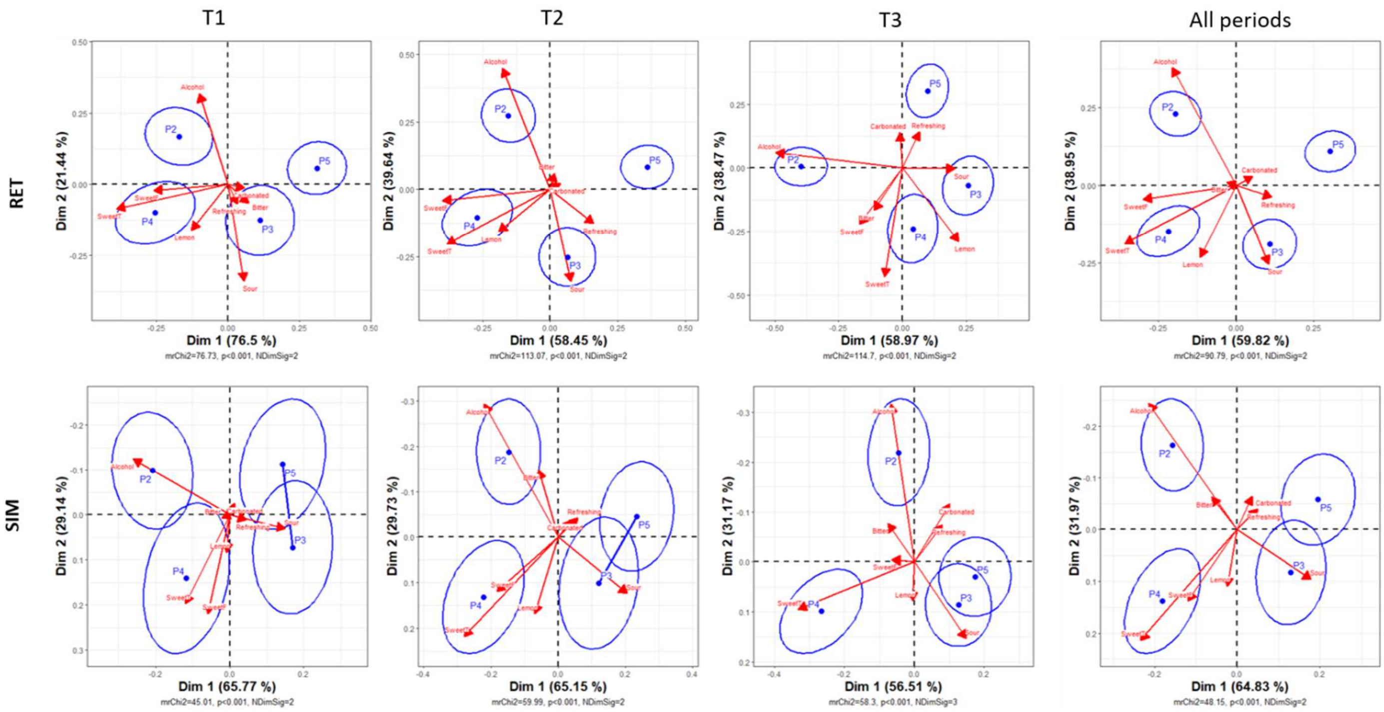

Figure 7: mrCA of the four products, by period and averaged over all periods on dimensions 1 and 2, for panels RET (first row) and SIM (second row). T1 (first column), T2 (second column), T3 (third column) and "All periods" (fourth column) correspond to the periods. "mrChi2" is the statistic of the mr χ^2 for the dimensionality test, "p" is the p-value of the dimension test, "NDimSig" is the number of significant dimensions. Red arrows indicate the relative importance and correlation between the descriptors and the dimensions. Blue lines connect not discriminated products (whatever the represented dimensions).

Figure 7 illustrates the differences between products within each period or all periods pooled, with citations averaged over periods. Both panels discriminated the four products (P3 and P5 on dimension 3 and only in T3 and overall, not represented in panel SIM). The mr χ^2 statistics show that panel RET was still more discriminative than SIM in every period. RET and SIM panels were more discriminative in period T1 than in periods T2 and T3.

	T1				T2				T3				All periods			
	P2	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**
Alcohol	33.3*	20.8	25.4	24.7	50.0*	24.3	35.8	33.3	46.9*	18.1	26.7	27.1	43.4*	21.2	29.5	28.5
Bitter	19.1	22.9	20.1	22.9	31.9	29.5	29.9	31.9	33.7*	28.5	33.0	24.7	28.1	27.1	27.8	26.7
Carbonated	67.4	70.8	67.0	72.2*	51.0	50.7	49.7	52.8	25.0	23.6	21.9	27.8*	47.9	48.3	46.2	50.7*
Lemon	59.0	60.8	67.4*	52.4	60.1	65.3	70.1*	51.7	43.1	62.5*	62.9*	45.8	54.2	62.9*	66.7*	50.0
Refreshing	30.6	34.0	32.3	33.3	30.2	39.9*	33.0	40.3*	26.0	28.8	25.7	32.3*	29.2	34.4	30.2	35.4*
Sour	32.6	47.2*	43.4	41.0	30.6	50.4*	42.7	41.3	24.0	36.1*	31.9	33.0	29.2	44.4*	39.6	38.2
SweetF	19.4*	13. 9	21.5*	9.7	17.4*	13.5	20.8*	5.6	17.7*	12.2	16.3	8.7	18.1*	13.2	19.8*	8.0
SweetT	21.5*	14.6	26.0*	6.9	18.4	19.8	26.4*	7.6	17.0	16.3	21.9*	6.9	18.8	16.7	24.7*	7.3

Table 1: Percentages of citations by attribute (rows), product and period (columns) for panel RET. Significant (alpha=0.05) multiple-response hypergeometric tests per cell for product comparison are indicated by an asterisk.

	T1				T2				T3				All periods			
	P2	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**	**P2**	**P3**	**P4**	**P5**
Alcohol	39.2*	19.9	30.4	24.7	44.0*	26.2	34.4	28.5	33.3*	21.3	22.3	20.4	38.5*	22.0	29.3	25.0
Bitter	23.1	22.0	23.1	21.8	35.9*	28.7	30.8	30.6	26.7	22.7	25.6	21.8	28.6	24.5	26.4	25.0
Carbonated	71.8	71.6	70.3	72.9	46.2	46.5	44.0	47.9	27.8	23.8	21.3	29.6*	48.7	47.5	45.1	49.7
Lemon	64.5	67.4	69.2	63.7	60.4	65.6	70.7*	59.9	49.1	55.3	52.8	49.3	57.9	61.7	64.5*	58.5
Refreshing	32.6	35.8	33.7	35.2	34.8	35.5	32.6	37.0	31.9	31.2	26.0	31.0	33.0	33.7	31.1	34.9
Sour	29.7	40.8*	33.3	39.1*	27.5	39.0*	30.8	39.8*	22.3	33.0*	23.8	29.6	26.0	36.9*	29.7	36.6*
SweetF	13.6	15.3	19.1*	10.9	14.3	13.1	18.3*	10.2	13.6	14.9	13.6	8.8	13.6	13.5	17.6*	10.9
SweetT	16.5	15.3	20.9*	11.3	21.6	20.2	30.4*	14.1	18.7	14.9	30.4*	14.4	19.1	17.4	26.7*	12.7

Table 2: Percentages of citations by attribute (rows), product and period (columns) for panel SIM. Significant (alpha=0.05) multiple-response hypergeometric tests per cell for product comparison are indicated by an asterisk.

Tables 1 and 2 show the percentages of citations by attribute, product and period for the two panels, focusing on between-product comparisons. It should be noted that the results are presented in a single table for convenience, but the multiple-response hypergeometric tests have been made by period. Thus, the multiple-response hypergeometric tests per cell (Tables 2 and 3) should be interpreted by block (T1, T2, T3 or all periods) because each block summarizes the result of one mrCA. For example, *Alcohol* was cited in P2 significantly more than in other products during period T1 in the two panels.

The first three blocks of the tables allowed us to compare each product to the others inside each period. P2 was more often described as *Alcohol* during periods T1, T2, T3 (RET, SIM), more *Bitter* in period T3 (RET) or T2 (SIM), more *SweetF* during periods T1, T2, T3 (RET), and more *SweetT* during period T1 (RET). P3 was more often described as *Lemon* during period T3 (RET), more *Refreshing* during period T2 (RET), and more *Sour* during periods T1, T2, T3 (RET, SIM). P4 was more often described as *Lemon* during periods T2 (RET, SIM), T1, T3 (RET), more *SweetF* in periods T1 and T2 (RET, SIM) and more *SweetT* in periods T1, T2, T3 (RET, SIM). P5 was more often described as *Carbonated* in periods T1 (RET) and T3 (RET, SIM), *Refreshing* in periods T2 and T3 (RET) and *Sour* in periods T1 and T2 (SIM).

The last block of the tables allowed us to compare each product to the others without considering temporality. P2 was more often described as *Alcohol* (RET, SIM) and *SweetF* (RET). P3 was more often described as *Sour* (RET, SIM) and *Lemon* (RET). P4 was more often described as *Lemon, SweetF* and *SweetT* (RET, SIM). P5 was more often described as *Carbonated* (RET), *Refreshing* (RET) and *Sour* (SIM). If we ignored periods, products were still discriminated, but some differences were visible only during specific periods (for example, *Bitter* for product P2 during T3 and *Refreshing* for P3 in T2 in panel RET, *Bitter* for product P2 during T2 or *Carbonated* for Product P5 during T3 in panel SIM) and would have been missed without considering a temporal approach.

The two panels were in overall agreement except for P5, which was perceived as more *Sour* only by panel SIM. Overall, panel RET highlighted more differences (34) between products than panel SIM (22). The citation rates for *Alcohol* seemed to follow the alcoholic contents of the products.

3.2.3. Comparison of periods by product

[INSERT FIGURE 8 HERE]

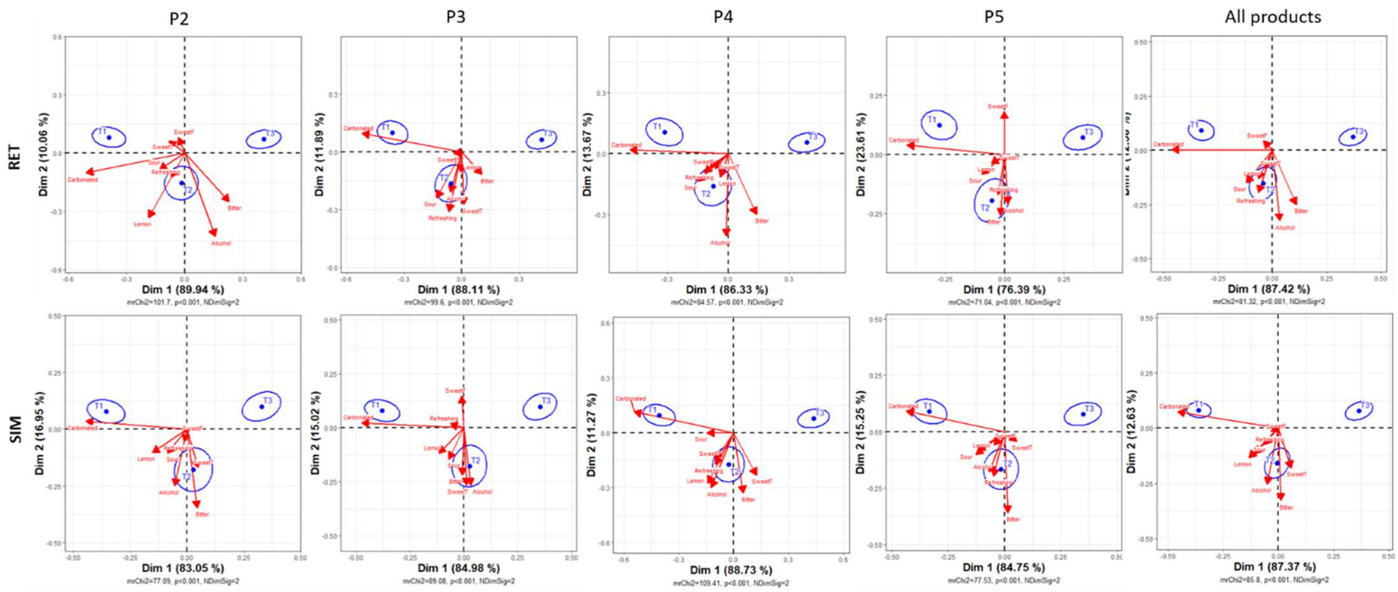

Figure 8: mrCA of the three periods, by product and averaged over all products on dimensions 1 and 2. RET (first row) corresponds to the retrospective panel, SIM (second row) to the simultaneous panel. P2 (first column), P3 (second column), P4 (third column), P5 (fourth column) and "All products" (fifth column) correspond to the products. "mrChi2" is the statistic of the mr χ² for the dimensionality test, "p" is the p-value of the dimension test, "NDimSig" is the number of significant dimensions. Red arrows indicate the relative importance and correlation between the descriptors and the dimensions.

Figure 8 shows that all periods were separated, denoting a clear temporality, but this temporality was globally the same regardless of the product. Overall, there is no difference in discrimination between the two panels, RET being more discriminative for P2 and P3, and SIM being more discriminative for P4 and P5 and for "All products". The period T1 was always associated with *Carbonated*, while T3 was the period having the less multidimensional differences.

	P2			P3			P4			P5			All products		
	T1	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**
Alcohol	33.3	50.0*	46.9*	21.2	24.3*	18.1	25.4	35.8*	26.7	24.7	33.3*	27.1	26.0	35.8*	29.9
Bitter	19.1	31.3	34.7*	22.6	28.6	30.6*	20.5	31.3*	31.3*	23.3	32.6*	23.6	21.5	30.9*	29.9*
Carbonated	68.40*	50.4	24.3	72.6*	49.7	21.9	65.6*	50.7*	23.6	70.8*	53.5	28.8	69.4*	51.0	24.7
Lemon	60.42*	61.5*	43.4	63.2	67.0	63.2	65.3	68.4*	62.2	50.7	50.7	45.5	59.7	61.8*	53.5
Refreshing	31.3	31.6	26.7	35.1	41.7*	30.6	31.6	31.3	24.0	32.6	39.2*	31.6	32.6	36.1*	28 .1
Sour	32.3*	29.9	23.6	46.5	49.3*	35.4	44.1*	44.1*	32.6	41.3	42.0*	33.3	41.0*	41.3*	31.3
SweetF	19.4	17.4	17.7	14.2	13.5	12.2	21.2	20.8	16.3	9.4	5.6	8.7	16.0	14.2	13.5
SweetT	21.2*	18.1	16. 7	13.9	19.4*	15.3	26.4	26.7	22.6	7.3	8.0	7.3	17.4	18.1	15.6

Table 3: Percentages of citations by attribute (rows), product and period (columns) for panel RET. Significant (alpha=0.05) multiple-response hypergeometric tests per cell for period comparison are indicated by an asterisk.

	P2			P3			P4			P5			All products		
	T1	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**
Alcohol	38.8	44.3*	33.3	19.5	27.3*	21.3	30.8	33.7*	22.3	25.0	27.5*	20.4	28.1	32.6*	24.2
Bitter	24.2	35.9*	26.7	23.8	29.1*	22.7	21.6	30.4*	25.6	20.4	30.3*	21.8	22.8	31.6*	23.9
Carbonated	71.1*	45.4	27.8	70.9*	44.0	23.8	71.1*	45.4	21.3	73.6*	49.7	29.6	70.9*	46.0	25.3
Lemon	65.9*	60.8	49.1	69.5*	66. 7	55.3	67.4*	70.0*	52.8	61.6*	59.2	49.3	66.0*	63.5*	50.9
Refreshing	33.0	34.1	31.9	36.5	33.0	31.2	33.3	34.1	26.0	34.9	39.1*	31.0	35.1	34.7	29.8
Sour	28.9	28.2	22.3	39.7	40.8	33.0	34.4*	29.7	23.8	40.1*	38.4	29.6	35.1*	34.0*	26.7
SweetF	13.6	13.9	13.6	15.3	12.1	14.9	19.1	19.1	13.6	10.9	10.9	8.8	14.4	13.7	12.3
SweetT	16.1	21.6*	18.7	14.5	20.6*	14.9	21.3	30.4*	30.4*	11.6	13.7	14.4	15.4	21.1*	19.3

Table 4: Percentages of citations by attribute (rows), product and period (columns) for panel SIM. Significant (alpha=0.05) multiple-response hypergeometric tests per cell for period comparison are indicated by an asterisk.

Tables 3 and 4 show the percentages of citations by attribute, product and period for the two panels, focusing on between-period comparisons. The fifth block illustrated the overall product temporality. The interest of investigating temporal differences was confirmed, as products were overall more often described *Alcohol* in period T2 (RET, SIM), more *Bitter* in period T2 (RET, SIM) and T3 (RET), more *Carbonated* in T1 (RET, SIM), more *Lemon* in T1 (SIM) and T2 (RET, SIM), more *Refreshing* in T2 (RET), more *Sour* in T1 and T2 (RET, SIM), more *SweetT* in T2 (SIM). Only the citation rates of *SweetF* were constant over the periods.

The four first blocks of the tables allowed us to compare each period to the other ones by product. For P2, period T1 was more often described as *Carbonated*, *Lemon* (RET, SIM), *Sour* and *Sweet* (RET); period T2 was more often described as *Alcohol* (RET, SIM), *Lemon* (RET), *Bitter* and *SweetT* (SIM); and period T3 was more often described as *Alcohol* (RET). For P3, period T1 was more often described as *Carbonated* (RET, SIM) and *Lemon* (SIM); period T2 was more often described as *Alcohol* and *SweetT* (RET, SIM) and *Refreshing* and *Sour* (RET); and period T3 was more often described as *Bitter* (RET). For P4, period T1 was more often described as *Carbonated* and *Sour* (RET, SIM), *Lemon* (SIM); period T2 was more often described as *Alcohol, Bitter* and *Lemon* (RET, SIM), *Carbonated* and *Sour* (RET), *SweetT* (SIM); period T3 was more often described as *Bitter* (RET), *SweetT* (SIM). For P5, period T1 was more often described as *Carbonated* (RET, SIM), *Lemon* and *Sour* (SIM); period T2 was more often described as *Alcohol*, *Bitter* and *Refreshing* (RET, SIM), *Sour* (RET). These results showed that despite similar evolutions, some attribute temporal changes were product dependent (*SweetT, Refreshing, Sour*).

Panel RET highlighted about the same number of differences (34) as panel SIM (32), but some conclusions were different depending on the panel, some significant attributes switching by a period between RET and SIM.

3.2.4. Feedback of the consumers about the task

The analysis of the feedback on the simultaneous task showed that 29 consumers found the task easy, and seven found it difficult. However, some consumers complained about "the pace of rating being too fast", "the response time being too short", and "the time to select an impression being too long", and some found it "a little difficult to drink while being aware of the difference in change" or were "distracted by timing", "pressed for time", and "wanted slightly more time". Other expressed limitations found in panel SIM included "too fast," "I'm in a hurry," "a little quick," and "difficult timing".

The analysis of the feedback on the retrospective task showed that 28 consumers found the task easy, and eight found it difficult. Only two consumers reported verbatim statements related to memory: "forgot the taste" and "difficult to match the timing".

4. Discussion

4.1. Statistical methodology

The sensory trajectories, between-sip, between-product and between-period comparisons have been performed using the mrCA paradigm, proving its versatility. The approach proposed here is based on a step-by-step strategy of comparison as in Mahieu, Visalli, Thomas, et al. (2020). It has the advantage of clearly identifying the sources of differences. In this study, the differences between the products were larger than the differences between the periods, and there was no difference between the sips. Several limitations can be noticed. As there was no observed difference between the sips, the combination "consumer x sip" was considered as an observation, artificially increasing the degrees of freedom. It is fairly common when drawing TDS or TCATA curves, but possible alternatives would have been to sum or average the citations over the sips, however, in this case the Tables 1-4 could not have been interpreted in terms of citation rates. The mrCA is by nature multidimensional; thus, it takes dependence between the attributes into consideration, but it does not allow us to statistically compare the magnitudes of the citation rates of the attributes between them. It does not allow us to test whether a citation rate is large enough to be considered important or applicable at the panel level. Using a generalised linear model such as in Visalli et al. (2020) or Weerawarna et al. (2021) could have also been considered as a univariate alternative.

4.2. Retrospective vs. concurrent measurement

The two panels (RET and SIM) were able to temporally discriminate the products. As in Visalli et al. (2020), retrospective and concurrent temporal measures allowed us to draw relatively similar conclusions. This was congruent with the conclusions of Van Den Haak et al. (2003) but contrary to the observation of Liang et al. (2018) about the decrease in taste sensitivity due to the memory load that was not observed with panel RET. However, this decrease was observed for intensities, not for citation rates.

However, some differences have been observed. The RET panel highlighted more differences between the products, as observed in different scientific contexts by Kuusela & Paul (2000); Ryan & Haslegrave (2007); Whyte et al. (2010). Notably, the differences in sweetness between products seem to have been better captured with panel RET. Indeed, the second dimension of Figure 6A captured the Brix measure gradient (Degrees Brix measures the sugar content of an aqueous solution): P4 (6.33) > P2 (6.07) > P3 (4.82) > P5 (2.63). However, as the "sensory reality" is unknown, it is not possible to conclude the superiority of retrospective measures. Indeed, the differences were small, and no repeatability measurements were performed to quantify the level of noise. Thus, it was not possible to conclude if these differences were due to randomness or to the task, and only hypotheses can be formulated. No definition of the descriptors was given to the consumers. It is therefore possible that some sensory terms have been interpreted differently depending on the context. Indeed, differences were observed with *Refreshing* in Table 1 (RET) but not in Table 2 (SIM). The consumers in panel RET could have interpreted the descriptor as "thirst-quenching", whereas, as it was evaluated earlier by the consumers in panel SIM, it could have been interpreted as something related to the temperature or the dryness of the drink.

The SIM panel had limited duration and allowed us to check the applicable attributes simultaneously to the tasting; it could have increased the difficulty of the task. The experimenters pre-tested the evaluation time of eight seconds for periods T1 and T2 and found it well adapted to this type of carbonated beverage. Keeping the beverage in mouth for eight seconds was even found quite long, but the experimenters wanted the in-mouth and after-swallowing evaluation times to be the same. As there was no difference in the number of citations between the two panels, it can be supposed that consumers from SIM panel had sufficient time to check the attributes. In addition, for panel SIM, the number of different descriptors used tends to decrease over sips. It is thus possible that the cognitive load was increased by the simultaneous task and influenced perception, as reported by Wal & Dillen (2013). To reduce the cognitive load and/or the boredom, the consumers could for example have reported dominant sensations instead of applicable ones.

Most of the discriminant attributes were the same regardless of the panel, but some appeared more important at different periods. As SIM was expected to be more temporally accurate than RET, it can be supposed that the switches of periods between the two panels could be due to temporal inaccuracy due to the memorization process involved in the retrospective choice of the period. This could corroborate the conclusions about the memory effect identified by Daniel and Katz (2018) and Aldrovandi et al. (2015). However, the switches of descriptors between panels and periods went both ways, e.g. *SweetT* for P2 from T1 in RET to T2 in SIM, and *Sour* for P5 from T2 in RET to T1 in SIM, suggesting that the changes were related to variability rather than to the method. Moreover, if the consumers of panel RET had randomly chosen the periods for the descriptors, it is unlikely that a better discrimination would have been observed. One of the most notable differences was observed with *Bitter*, which was most cited for products P2 and P3 during period T3 by panel RET and in period T2 by panel SIM. For this descriptor, RET seems to be in agreement with previous studies. Dietz et al. (2022) showed that bitterness was more important at the end of the tasting, and Higgins et al. (2021) demonstrated that the bitterness peak was after 20 seconds. This observation remains to be confirmed because the products (beers) were not flavoured in the mentioned studies.

The analysis of consumers' feedback did not allow to conclude about the difficulty of the task. However, it should be noted that in this study, the evaluation times were relatively short. Retrospective evaluation could be unsuitable for products having lingering tastes or flavours for longer durations, which could cause difficulty in memorizing those sensory perceptions.

Regarding these results, it can be assumed that concurrent and retrospective measurement could align with Kahneman's theory (O'Brien, 2012) about two systems of thinking: the first one being fast, intuitive and emotional; the second one being slower and more deliberative; each one having its own advantages and drawbacks for developing consumer methodologies. The most appropriate measure probably depends on the objective. The concurrent measure was the most spontaneous and likely the most suitable for studying physiological processes involved in temporal perception. The retrospective measure was presumably the most cognitive and analytical, but also the closest to the remembered experience and thus could be more important to explain later consumer choices. Outside of food science, the superiority of remembered experience in predicting choices was observed by Kahneman et al. (1993) and Wirtz et al. (2003). If no similar study has been conducted with food choices, the importance of cognitive processes for determining answers to food cues was largely studied as demonstrated in the review of Higgs (2016). However, in this study, the retrospective measure was made just after the tasting, and it could be interested to redo the experiment with different delays after the tasting to study what is really recalled of the products and how it impacts consumers' choices.

4.3. Periods and temporality

Retrospective measurement imposes temporal measurements by period. In Visalli et al. (2020), the chosen periods were quite subjective, as "attack", "evolution" and "finish" were not precisely defined and could evoke different moments between the panellists. Here, this potential subjectivity was limited using specific and meaningful tasting points to delimitate the periods: "in mouth before swallowing," "immediately after swallowing" and "aftertaste." This study, as well as those of Visalli et al. (2020) and Mahieu, Visalli, Thomas, et al. (2020), proved that measurements by period captured temporal differences within the product. Indeed, the two protocols used in this study could be seen as "discrete time TCATA" with the advantage of no need for fading (Ares et al., 2016) to help the subjects to unselect the applicable attributes and with the possibility to include a larger number of descriptors (only with the retrospective protocol). Using periods, the temporal precision was expected to be lower than that of a continuous time measurement. However, the temporal precision of TDS and TCATA has not been extensively studied. In this way, it should be interesting to test the variability of the continuous time-dependent conclusions, for example, by using replicates or bootstrap, and to see if continuous time measures truly brought additional usable information compared to periods. In any case, if the temporal precision is not of crucial importance or is not the main source of difference between the studied samples, the discrete time TCATA methods (both concurrent and retrospective, but particularly retrospective) should be considered as useful temporal measurements in consumer studies.

5. Conclusion

This study compared the temporal description obtained from consumers at home in two conditions: concurrently with the tasting (panel SIM) and retrospectively with the tasting (panel RET). Consumers were asked about the applicable attributes for four full portions of lemon-flavoured carbonated alcoholic drinks at three specific time periods: "in mouth before swallowing", "immediately after swallowing", and aftertaste. Data were analysed using multiple-response correspondence analysis framework applied at different levels: product, period and sip. RET and SIM captured the differences between products and periods, with the differences between the products being larger than the differences between the periods. In both methods, no sip effect were observed. Perception of the products were identical in sips 1, 4 and 7. Overall, the consumers of RET and SIM agreed on the relative between-product differences, with RET showing more differences than SIM. The consumers of RET and SIM only partially agreed on the within-product differences. The results suggest that the retrospective temporal evaluation could better discriminate the products and point to the need for more research about the temporal precision of the data.

Highlights

Descriptive temporal data were collected from consumers at home on smartphones.

Retrospective and concurrent CATA descriptions over 3 periods were compared.

The big picture was the same, but the retrospective panel was more discriminative.

Retrospective measure should be considered in consumer studies.

CRediT author statement

Pascal Schlich: Conceptualization, Methodology, Software, Supervision, Funding acquisition, Writing - Review & Editing.

Michel Visalli: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing - Original Draft, Visualization.

Takahiro Wakihira: Conceptualization, Methodology, Validation, Data Curation, Resources, Project administration, Investigation, Writing - Review & Editing.

Acknowledgments

We would like to thank and acknowledge Seiichi Higuchi and Yasushi Nagatomi at Asahi Quality & Innovations, Ltd. for funding acquisition and endorsement of this project.

References

Aldrovandi, S., Poirier, M., Kusev, P., & Ayton, P. (2015). Retrospective evaluations of sequences: Testing the predictions of a memory-based analysis. *Experimental Psychology*, *62*(5), 320–334. https://doi.org/10.1027/1618-3169/A000301

Ares, G., & Jaeger, S. R. (2015). Check-all-that-apply (CATA) questions with consumers in practice: experimental considerations and impact on outcome. *Rapid Sensory Profiling Techniques and Related Methods: Applications in New Product Development and Consumer Research*, 227–245. https://doi.org/10.1533/9781782422587.2.227

Ares, Gastón, 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.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*(3), 330–343. https://doi.org/10.1016/J.ACTPSY.2010.03.006

Boinbaser, L., Parente, M. E., Castura, J. C., & Ares, G. (2015). Dynamic sensory characterization of cosmetic creams during application using Temporal Check-All-That-Apply (TCATA) questions. *Food Quality and Preference*, *45*, 33–40. https://doi.org/10.1016/J.FOODQUAL.2015.05.003

Botha, J. J., Cannon, P., & Hort, J. (2021). Comparing a new rapid combined method (RapCoTT) with traditional approaches for phenotyping thermal taste. *Physiology & Behavior*, *238*, 113482. https://doi.org/10.1016/J.PHYSBEH.2021.113482

Cadoret, M., & Husson, F. (2013). Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality and Preference*, *28*(1), 106–115. https://doi.org/10.1016/J.FOODQUAL.2012.09.005

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.

Church, B. A., Jackson, B. N., Beran, M. J., & Smith, J. D. (2019). Simultaneous versus prospective/retrospective uncertainty monitoring: The effect of response competition across cognitive levels. *Journal of Experimental Psychology: Animal Learning and Cognition*, *45*(3), 311–321. https://doi.org/10.1037/XAN0000207

Clark, C. C., & Lawless, H. T. (1994). Limiting response alternatives in time-intensity scaling: an examination of the halo-dumping effect. *Chemical Senses*, *19*(6), 583–594. https://doi.org/10.1093/CHEMSE/19.6.583

Daniel, T. A., & Katz, J. S. (2018). Primacy and recency effects for taste. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(3), 399–405. https://doi.org/10.1037/XLM0000437

Dietz, C., Cook, D., Yang, Q., Wilson, C., & Ford, R. (2022). A TCATA by modality approach to study the multisensory temporal profile of hop bitter and flavour products applied in lager. *Food Quality and Preference*, *97*, 104470. https://doi.org/10.1016/J.FOODQUAL.2021.104470

Duizer, L. M., Bloom, K., & Findlay, C. J. (1997). Dual-attribute time-intensity sensory evaluation: A new method for temporal measurement of sensory perceptions. *Food Quality and Preference*, *8*(4), 261–269. https://doi.org/10.1016/S0950-3293(96)00052-3

Higgins, M. J., Gipple, J. T., & Hayes, J. E. (2021). Common bitter stimuli show differences in their temporal profiles before and after swallowing. *Food Quality and Preference*, *87*, 104041. https://doi.org/10.1016/J.FOODQUAL.2020.104041

Higgs, S. (2016). Cognitive processing of food rewards. *Appetite*, *104*, 10–17. https://doi.org/10.1016/J.APPET.2015.10.003

Hutchings, S. C., Foster, K. D., Grigor, J. M. V, Bronlund, J. E., & Morgenstern, M. P. (2014). Temporal dominance of sensations: A comparison between younger and older subjects for the perception of food texture. *Food Quality and Preference*, *31*, 106–115.

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*(3), 539–543.

Jaeger, S. R., Beresford, M. K., Hunter, D. C., Alcaire, F., Castura, J. C., & Ares, G. (2017). Does a familiarization step influence results from a TCATA task? *Food Quality and Preference*, *55*, 91–97.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, *4*(6), 401–405. https://doi.org/10.1111/J.1467-9280.1993.TB00589.X

Kantono, K., Hamid, N., Shepherd, D., Lin, Y. H. T., Brard, C., Grazioli, G., & Thomas Carr, B. (2018). The effect of music on gelato perception in different eating contexts. *Food Research International*, *113*, 43–56. https://doi.org/10.1016/J.FOODRES.2018.06.030

Kemp, S. E., Hort, J., & Hollowood, T. (2017). Time-Dependent Measures of Perception in Sensory Evaluation. In J. Hort, S. E. Kemp, & T. Hollowood (Eds.), *Time-Dependent Measures of Perception in Sensory Evaluation*. John Wiley & Sons Ltd.

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*(2), 260–273. https://doi.org/10.1016/J.FOODQUAL.2013.06.007

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology*, *113*(3), 387–404. https://doi.org/10.2307/1423365

Lee, W. E. I., & Pangborn, M. (1986). Time–intensity: The temporal aspects of sensory perception. *Food Technology*, *40*, 78–82.

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.

Liang, P., Jiang, J., Ding, Q., Tang, X., & Roy, S. (2018). Memory Load Influences Taste Sensitivities. *Frontiers in Psychology*, *0*(DEC), 2533. https://doi.org/10.3389/FPSYG.2018.02533

Mahieu, B. (2021). *MultiResponseR*. https://github.com/MahieuB/MultiResponseR

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*, 104256. https://doi.org/10.1016/J.FOODQUAL.2021.104256

Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020). Using Free-Comment with consumers to obtain temporal sensory descriptions of products. *Food Quality and Preference*, *86*, 104008. https://doi.org/10.1016/J.FOODQUAL.2020.104008

Methven, L., Rahelu, K., Economou, N., Kinneavy, L., Ladbrooke-Davis, L., Kennedy, O. B., Mottram, D. S., & Gosney, M. A. (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*(8), 948–955.

O'Brien, D. T. (2012). Thinking, fast and slow by Daniel Kahneman. *Journal of Social, Evolutionary, and Cultural Psychology*, *6*(2), 253. https://doi.org/10.1037/H0099210

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review 1994 6:4*, *6*(4), 351–371. https://doi.org/10.1007/BF02213420

Peute, L. W. P., de Keizer, N. F., & Jaspers, M. W. M. (2015). The value of Retrospective and Concurrent Think Aloud in formative usability testing of a physician data query tool. *Journal of Biomedical Informatics*, *55*, 1–10. https://doi.org/10.1016/J.JBI.2015.02.006

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:

Rizo, A., Vidák, K., Fiszman, S., & Tarrega, A. (2020). Influence of fading duration on TCATA evaluation. *Food Quality and Preference*, *79*, 103619.

Rodrigues, J. F., de Souza, V. R., Lima, R. R., Carneiro, J. de D. S., Nunes, C. A., & Pinheiro, A. C. M. (2016). Temporal dominance of sensations (TDS) panel behavior: A preliminary study with chocolate. *Food Quality and Preference*, *54*, 51–57.

Ryan, B., & Haslegrave, C. M. (2007). Use of concurrent and retrospective verbal protocols to investigate workers' thoughts during a manual-handling task. *Applied Ergonomics*, *38*(2), 177–190. https://doi.org/10.1016/J.APERGO.2006.03.003

Schwarz, N. (2012). Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture. In Mehl & Connor (Eds.), *Handbook of Research Methods for Studying Daily Life.* Guilford.

Stone, A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). *The Science of Real-Time Data Capture: Self-Reports in Health Research (Google eBook)*. 416. https://books.google.com/books/about/The_Science_of_Real_Time_Data_Capture.html?hl=fr&id=LmlnDAAAQBAJ

Thomas, A, van der Stelt, A. J., Schlich, P., & Lawlor, J. Ben. (2018). Temporal drivers of liking for oral nutritional supplements for older adults throughout the day with monitoring of hunger and thirst status. *Food Quality and Preference*, *70*, 40–48.

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

Van Den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour and Information Technology*, *22*(5), 339–351. https://doi.org/10.1080/0044929031000

Varela, P., Antúnez, L., Carlehög, M., Alcaire, F., Castura, J. C., Berget, I., Giménez, A., Næs, T., & Ares, G. (2018). What is dominance? An exploration of the concept in TDS tests with trained assessors and consumers. *Food Quality and Preference*, *64*, 72–81.

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*, 104002. https://doi.org/10.1016/J.FOODQUAL.2020.104002

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*, 103956. https://doi.org/10.1016/J.FOODQUAL.2020.103956

Visalli, Michel, Wakihira, T., & Schlich, P. (2022). A dataset on concurrent and retrospective methods for measuring sensory perception and preferences of lemon-flavoured carbonated alcoholic drinks. *Data in Brief*.

Wal, R. C. van der, & Dillen, L. F. van. (2013). Leaving a Flat Taste in Your Mouth: Task Load Reduces Taste Perception. *Https://Doi.Org/10.1177/0956797612471953*, *24*(7), 1277–1284. https://doi.org/10.1177/0956797612471953

Weerawarna, M. N. R. P., Godfrey, A. J. R., Ellis, A., & Hort, J. (2021). Comparing temporal sensory product profile data obtained from expert and consumer panels and evaluating the value of a multiple sip TCATA approach. *Food Quality and Preference*, *89*, 104141. https://doi.org/10.1016/J.FOODQUAL.2020.104141

Whyte IV, J., Cormier, E., & Pickett-Hauber, R. (2010). Cognitions associated with nurse performance: A comparison of concurrent and retrospective verbal reports of nurse performance in a simulated task environment. *International Journal of Nursing Studies*, *47*(4), 446–451. https://doi.org/10.1016/J.IJNURSTU.2009.09.001

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on Spring Break? The Role of Predicted, on-Line, and Remembered Experience in Future Choice. *Psychological Science*, *14*(5), 520–524. https://doi.org/10.1111/1467-9280.03455