**A new index for predicting differences in repeatability of Time-Intensity
curves: Time-Intensity Reliability Index (TI-RI)**

Carolina Chaya ^{a,*}, Celia Criado ^b, María Ángeles Pozo-Bayón ^b, Andrea
Echevarrías Marco ^a, Amanda Dupas de Matos ^{a,c}

^a Department of Agricultural Economics, Statistics and Business Management,
Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingeniería
Agronómica, Alimentaria y de Biosistemas, Ciudad Universitaria s/n, 28040
Madrid, Spain

^b Instituto de Investigación en Ciencias de la Alimentación (CIAL), (CSIC-UAM), C/
Nicolás de Cabrera, 9, 28049, Madrid, Spain

^c Visiting scholar at Universidad Politécnica de Madrid, Department of Agronomy,
Food, Natural Resources, Animals and Environment (DAFNAE), University of
Padova, Viale dell'Università, 16, 35020, Legnaro, Italy

* Corresponding author: Carolina Chaya, e-mail address: carolina.chaya@upm.es

Abstract

The Time-Intensity (TI) method evaluates the perceived intensity of a sensation over time, and it is a suitable method to measure perception changes during and/or after consumption. Compared to static methods, TI requires additional training and more replicates to obtain reliable results. In order to investigate inter-individual variation between panelists, this study proposes a new index called Time-Intensity Reliability Index (TI-RI). The TI-RI was tested with a dataset obtained from the sensory evaluation of four types of aromatized wine matrices by a trained panel. The proposed index is able to predict differences in panelists' repeatability of TI curves and it could be used to assess the precision of other sources of variation such as products, attributes and replicates.

Keywords: Time-Intensity method, Inter-individual variation, Repeatability, Reliability Index.

1. Introduction

Traditional sensory methods have treated food evaluation as a static phenomenon. Yet the dynamic nature of food sensory evaluation, however, arises from processes such as chewing, breathing, salivation, tongue movements, and swallowing (Dijksterhuis, 1996). This dynamic nature leads to changes of flavor and taste attributes over time, and for this reason, a methodology enabling the study of these changes over time is needed. Time-Intensity (TI) methodology evaluates the perceived intensity of a sensation over time and can be used to profile sensory attributes whose perception changes during and/or after consumption/use. However, compared to static methods, the TI method requires additional training and replicates in order to obtain reliable results (Peyvieux & Dijksterhuis, 2001; Chaya, 2017).

Most dynamic sensory methods of analysis try to reduce inter-individual differences and propose an average/aggregated curve (Liu & McFie, 1990; Overbosch et al., 1986). A “measure of the reliability” (T-IR) was proposed by Bloom, Duizer & Findlay (1995). In this case, T-IR was calculated as the absolute average of a set of standardized (centered and reduced) standard deviations (S_t), representing the variability at each sampling point (t) in a set of repeated time-intensity trials.

Furthermore, Peyvieux & Dijksterhuis (2001) proposed the use of Principal Component Analysis (PCA) to study the consistency of the panel and hence the efficiency of the training. The loading plots of the panelists and their replicates were used to identify the “odd” replicates. Echols et al. (2003) graphically assessed panel performance using Strip-plots of the parameter estimated for each sample through three replications by each panelist. In addition, Reinbach et al. (2007)

analyzed the ability of the assessors to replicate the assessments by calculating the Euclidean distance between each pair of replicates for the whole curves. Until now, most of the proposed methods have not been intended to account for panelist variation. For example, if one wishes to study the impact of individual oral physiology on sensory perception, the aggregated methods could help in the evaluation of differences among products, but they do not help in the investigation of inter-individual variation. Although TI-R could be the most suitable index to measure inter-individual variation, in its current stage computation, it is not sensitive enough.

Thus, the aim of this paper is to propose a new index, called Time-intensity Reliability Index (TI-RI), to evaluate the repeatability of TI data.

With respect to T-IR, this study proposes two main modifications: i) it does not reduce S_t (it does not divide by standard deviation) as a way of accounting for individual variations, and ii) it does not use a fixed and predetermined number of points, but instead, uses the total time (total duration of TI curve) with 1-second intervals. The proposed index will be tested with a dataset obtained from four types of spiked (aromatized) wine matrices. For this, the TI evaluations of the retronasal aroma intensity of four aroma descriptors associated to the chemical odorants added to the wines, were recorded by a trained panel.

In this paper, we will present the TI-RI values for all replicates by matrix, aroma and panelist, as well as a visual inspection of some TI replicate curves and their associated RI values. We will compare TI-RI with other measurements of repeatability, and analyze the effect of different factors (matrix, aroma and panelist) on TI-RI values. We will finally complete these results with the classical analysis of

the curves of the TI parameters. Conclusions about the possible uses of TI-RI will finally be outlined.

2. Material and Methods

2.1 Products

Four wine matrices were prepared using a commercial dealcoholized still rosé wine as the base. Addition of food grade ethanol (Panreac Química S.A., Barcelona, Spain) and procyanidin extract (Les Dèrives Resiniques & Terpéniques, S.A., France) was carried out at two different levels in order to have four different types of wine matrices (**Table 1**).

Four different aroma compounds (isoamyl acetate, ethyl butanoate, ethyl decanoate, ethyl hexanoate) (Sigma-Aldrich, Steinheim, Germany) all at the same concentration (60 mg/L) were individually added to each of the four wine matrices. Only one aroma compound was added to each type of wine matrix. In order to increase the sensitivity of the panel, and considering that the retronasal TI-evaluation of the wines was carried out immediately after mouth rinsing and spitting off the wine. The concentration of each aroma compound was much higher than its previously described odor threshold (Francis & Newton, 2005),

2.2 Training of panelists

Ten panelists (four male and six female) were recruited and the training was performed in three steps: 1) Recognition and evaluation of the aroma compounds at two different intensities (**Table 2**) defined in a preliminary test. The panel generated the terms described in **Table 2** for each of the four chemical odorants;

2) Introducing the assessment protocol to the panelists; 3) Familiarization of the panelists with the temporal sensory methodology and computer software. During training, one of the panelists (P8) gave up, and consequently their data could not be used in this study.

2.3 Assessment protocol

The panelists were instructed about the assessment protocol before each session.

In the procedure, 15 mL of the aromatized wine were placed into the oral cavity, performing a soft rinsing and then spitting out after 30 seconds. During rinsing, special care to keep the lips closed, not to swallow and not to open the velum-tongue border prior to expectoration was taken. Between samples, the panelists were asked to clean their mouths by rinsing with water, eating some crackers, and then finally drinking more water. All TI recordings were collected after spitting the samples out, producing an abrupt increase of recorded intensity at the very beginning of the TI curve, a comparison to other more conventional protocols of TI curves.

2.4 Time-Intensity (TI) evaluation

Time-Intensity (TI) evaluation of retronasal aroma intensity was assessed after spitting-out the wines recording up to a maximum of three minutes by means of Compusense software. Each wine sample was presented to the panelists twice in random order. In the same session (one day), the four aroma attributes were independently evaluated in the same wine matrix. Therefore, a total of eight time-intensity evaluation sessions were carried out (four wine matrices and the

corresponding duplicated sessions). A total of 288 TI curves were generated (nine panelists, four types of aromas, four wine matrices, and two replicates).

The TI parameters extracted from the Time-Intensity curve and analyzed in this study include the highest recorded intensity in the TI curve ($I_{\max}$), time at which the maximum intensity occurs ($T_{\max}$), total area under the curve (AUC), and time to finish the aroma perception (T_{end}).

2.5 Methods to explore reliability of TI measurements

2.5.1 Measure of the reliability (T-IR)

Few indicators of reliability of repetition are available in the literature. The only existing index is T-IR was proposed by Bloom et al. (1995). The T-IR is calculated as follows:

$$T-IR = \frac{1}{m} \sum_{t=1}^{t=m} |Z_{st}|$$

Where, m is the number of predetermined time points. This index reflects dispersion between curves, in which the absolute value ($|Z_{st}|$) of the differences ($S_t - \bar{S}$) is divided by their standard deviation (S_{st}) consequently reducing the computed variability among curves. ($\bar{S}$) is the mean of the standard deviation for all time points.

$$Z_{st} = \frac{S_t - \bar{S}}{S_{st}}$$

S_t is calculated as follows:

$$S_t = \sqrt{\frac{\sum_{j=1}^{j=r} (I_{tj} - \bar{I}_t)^2}{r - 1}}$$

Where, r is the number of replicates; I_{ij} is the intensity measured in the replicate j at the instant t , $\bar{I}_t$ is the mean of the intensity between replicates at time t .

2.5.2 Time-Intensity Reliability Index (TI-RI)

The Time-Intensity Reliability Index (TI-RI) is the sum of the absolute values of the difference between the standard deviation for each time point (S_t) and the mean of the standard deviation for all time points ($\bar{S}$), divided by the number of replicates minus one ($r-1$). The final value of the sum is divided by the total time evaluation or by the number of time points (t_{end}).

$$TI-RI = \frac{1}{t_{end}} \sum_{t=1}^{t_{end}} \frac{|S_t - \bar{S}|}{r - 1}$$

The lower the variation between curves (i.e., the higher the reliability) the lower the TI-RI index and vice versa, the higher the variation between curves, the higher the TI-RI index.

TI-RI has two modifications compared to the Measure of the reliability (T-IR): a) it does not divide by standard deviation as a way of accounting for individual variations and b) it does not use a fixed and predetermined number of points (m) but uses the total time with 1-second intervals.

2.5.3 Strip-plot of parameters

A graphical data technique known as Strip-plot is often used to summarize a univariate data set. Given the large volume of TI-curves, previous authors have

proposed the simultaneous visualization of the parameters using Strip-plots (Echols et al., 2003). The Strip-plot is an alternative to a histogram or a density plot, in which each horizontal line corresponds to one panelist, the vertical black lines are the values of the replicates, and the black cross is the mean. The Strip-plots provide a global insight of variation within and between individuals for a specified TI parameter. An example is shown in **Figure 2**.

2.6 Statistical analysis

Statistical analyses were performed using STATGRAPHICS Software (Statgraphics Centurion XVII version). Analysis of variance (ANOVA) followed by post hoc Tukey's Honestly Significant Differences (HSD) test was performed to study the effect of the panelist, the aroma type, and the wine matrix on TI-RI.

3. Results and Discussion

3.1 TI-RI results by panelist, aroma and matrix

The TI-RI value was calculated in order to evaluate the repeatability of the data according to panelist, aroma, and matrix (**Table 3**). **Table 3** shows the TI-RI values of nine panelists, four types of aromas and four wine matrices. Regarding the panelists, some differences in reliability for certain aromas and wine matrices were observed. For example, panelist 2 (P2) had better repeatability in the evaluation of almost all aromas, showing high reliability (low TI-RI values) for all wine matrices regarding four descriptors, with the exception of W10P for strawberry flavored candy and banana aromas.

If we look at the wine matrix, variation in reliability across the different aromas and panelists was found. For example, the W0C matrix yielded the lowest TI-RI (0.294), obtained from the banana aroma and P2. However, in the same W0C matrix, the highest TI-RI (1.819) was found in the strawberry flavored candy aroma, assessed by P9. **Figure 1 (a, b, respectively)** shows the TI curves related to these TI-RI results.

Figure 1. Time-Intensity curves obtained from W0C matrix for a) Panelist P2 and banana aroma and b) Panelist P10 and strawberry flavored candy aroma. Both examples illustrate a low (TI-RI=0.294) and a high (TI-RI= 1.819) reliability index, respectively.

Taking the aroma type into consideration, we can observe that pineapple aroma was evaluated with higher repeatability between panelists in comparison to the others aromas. The only case, in which the pineapple aroma showed a TI-RI value greater than one, occurred with the W10P matrix and P5. On the other hand,

banana aroma had higher TI-RI (10 cases for $TI-RI > 1$) values, revealing that the panelists did not coherently evaluate this descriptor in both sessions.

3.2 Comparing TI-RI to other measurements of TI curve variations,

Strip-plots provide a global insight of variation within and between individuals for a specified TI parameter. **Figure 2** shows an example of Strip-plot of Area Under the Curve (AUC) for the WOC matrix and dried plum aroma. Each horizontal line corresponds to one panelist, the vertical black lines are the values of the two replicates, and the black cross is the mean. For example, the variation of P5 is higher than P2 and P3, because the distance between both vertical black lines is higher. This is also consistent with a high TI-RI value of 1.053 for panelist P5.

Figure 2. Strip-plot of AUC for dried plum aroma evaluation in W0C matrix.
Computed TI-RI are shown in the same examples.

The differences between P2 and P3 (TI-RI values of 0.490 and 0.794, respectively) are not captured with Strip-plots of AUC. This is because Strip-plot of AUC is less sensitive than TI-RI. In addition, the results of other parameters such as $I_{\max}$ are even less sensitive.

An illustration of the better sensitivity of T-IR and TI-RI compared to Strip-plot is shown in **Figure 3 (a, b, c, respectively)**. **Figure 3** also presents the computed T-IR of the same curves. TI-R was seen to be less sensitive to variations than TI-RI.

Figure 3b represents a much lower repeatability than **Figure 3c**. This difference can be seen with TI-RI (0.794 and 0.490, respectively) but it is not with T-IR (0.535 and 0.555, respectively). In the case of a much lower repeatability (**Figure 3a**), an increase in T-IR (0.861) is observed but it is not as important for TI-RI (1.053).

A comparison between **Figure 2** with **Figure 3**, shows that TI-RI is more sensitive at capturing the differences between panelists' repeatability. Therefore, these examples show how T-IR could help in the identification of very low repeatability, but it is not however sensitive enough when subtler differences between curves are found. The remaining computed T-IR followed a similar trend (results not shown).

Figure 3. TI curves for W0C matrix and dried plum aroma recorded by panelists P5 (a), P3 (b) and P2 (c). Computed values of TI-RI and T-IR are also displayed in the plots.

3.3 Statistical analysis of panelist, matrix and aroma effects on TI-RI

The effect of panelist, wine matrix and aroma on TI-RI was analyzed with 3-way ANOVA. **Table 4** shows the ANOVA table, where it can be seen that panelist and aroma had a significant effect on TI-RI ($p\text{-value} < 0.05$).

Figures 4 and **5** respectively, show the effect of the panelist (at a level of significance of 0.05) and of aroma (at a level of significance of 0.10) on TI-RI. The plots show the effect of panelists (across all matrices and aromas) and the effect of aroma (across all matrices and panelists), respectively. **Figure 4** shows significantly higher TI-RI values of P10 compared to P2 and P6, which means that the reliability of panelist P10 was significantly worse than P2 and P6 reliability.

Figure 4. Study of the effect of panelists on TI-RI using Tukey's HSD test.

Figure 5 shows the effect of aroma on RI (across all panelists and matrices). Pineapple aroma presented a significant lower TI-RI than the strawberry flavored candy aroma, which means that the pineapple aroma was more easily identified by panelists and had better repeatability, i.e. the TI curves of pineapple aroma were closer between replicates. In contrast, the strawberry flavored candy aroma was not identified with the same accuracy between replicates.

Figure 5. Study of the effect of aroma on TI-RI using Tukey's HSD test.

3.4 Statistical analysis of matrix, panelist and aroma effects on TI parameters

Taking into consideration the most frequent TI parameters ($I_{\max}$, $t_{\max}$ and AUC), a classical analysis for each aroma on matrix and panelist with matrix x panelist interaction was carried out. The p-values of these analyses are presented in **Table 5**. The highest source of variation was the panelist, which showed significant

effects on all aromas and almost all parameters. This variation among panelists could be due to a real differences in the panelists' oral physiology (saliva composition and flow, breath flows, etc.) or to other differences, such as training efficiency, scale use, etc.,. Despite these reasons for such differences, the fact that the only significant effect of matrix was found in the pineapple aroma (see **Table 5** and **Figure 6**) could be related to the higher reliability found in this aroma and demonstrated by a significantly lower TI-RI (see **Figure 5**). It is well known that the lower the variation in the data, the higher the power of the test to show significant differences. This is why any method designed to improve the analysis of variation among replicates could positively contribute to TI evaluation.

Figure 6. Effect of wine matrix on $I_{\max}$ for pineapple aroma. See **Table 1** for matrix code explanation.

4. Conclusions

This work provides insight into the study of TI data reliability. The proposed method (TI-RI) to account for individual variations has shown its ability to predict differences in panelists' repeatability in TI curves. In addition, TI-RI has shown to be more sensitive than other indicators of TI data repeatability such as T-IR or Strip-plots of parameters. The slight modification introduced in its computation compared to T-IR is easy to include in the existing software.

Using TI-RI during panel training would be beneficial in identifying the most difficult attributes and/or the less reliable panelists. This could be very helpful not only in terms of training efficiency in TI method, but also in terms of discrimination of the whole panel when a specific panelist, attribute or product shows poor repeatability. If TI training is finished and replicates of TI evaluation have been obtained, TI-RI could positively contribute to find out about the panelists, attributes and/or products showing a better/worse accuracy. This information could guide the panel leader in making decisions, such as discarding or maintaining data from a panelist or repeating the evaluation of a specific attribute, based on TI-RI values. However, the robustness of the method should be verified with other product categories and TI protocols.

Given the protocol of the conducted experiment, this study did not allow the panelists to swallow the wines and the panelists had to wait to report TI data until they spat the samples out. Therefore, the TI curves have a particular shape (abrupt increase at starting time). This assessment protocol is quite specific to this data and could have an impact on TI-curves shapes. Therefore, more studies confirming that other TI-curves are sensitive to the proposed TI-RI should be carried out.

This study is based on two repetitions per panelist. A time period up to a maximum of 180 seconds recorded with 1-second intervals was used. The effect of more replicates and a longer or shorter duration should be considered in future studies. Finally, similarly to other measurements like T-IR (Bloom et al., 1995) and the computed distance between TI-curves (Reinbach et al., 2007), TI-RI, not quoted. It would be very difficult to set a limit above which the reliability is poor. However, as long as the replicates are obtained, it could be helpful in comparing the precision of different factors (panelists, products, attributes) within the same TI experiment,. And in the TI method, replicates are highly commended.

Acknowledgements

The authors thank the panelists for their participation. This work was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) (Project AGL2016-78936R).

5. References

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

Chaya, C. (2017). Continuous Time-Intensity. In: J. Hort, S. F. Kemp, & T. Hollowood (Eds.), *Time-Dependent Measures of Perception in Sensory Evaluation* (pp. 237-266). Chichester, United Kingdom: John Wiley & Sons.

336 Dijksterhuis, G. (1996). Time-Intensity methodology: review and preview.
337 Proceedings of the COST96 Meeting: Interaction of Food Matrix with Small
338 Ligands Influencing Flavour and Texture. Dijon, France, Nov. 20, 1995.
339
340 Echols, S., Lakshmanan, A., Mueller, S., Rossi, F., & Thomas, A. (2003).
341 Parametric modelling of time intensity data collected on product prototypes
342 generated from a fractional factorial experiment to quantify sources of texture
343 variability. *Food Quality and Preference*, 14, 527-536.
344
345 Francis, I. L., & Newton, J. L. (2005). Determining wine aroma from compositional
346 data. *Australian Journal of Grape and Wine Research*, 11(2), 114-126.
347
348 Liu, Y. H., & McFie, H. J. H. (1990). Methods for averaging Time-Intensity curves.
349 *Chemical senses*, 15, 471-484.
350
351 Overbosch, P., Enden, J. C, van den, J. C, & Keur, B. M. (1986). An improved
352 method for measuring perceived intensity/time relationships in human taste and
353 smell. *Chemical Senses*, 11, 331-338.
354
355 Peyvieux, C., & Dijksterhuis, G. B. (2001). Training a sensory panel for TI: a case
356 study. *Food Quality and Preference*, 12, 19-28.
357
358 Reinbach, H. C., Meinert, L., Ballabio, D., Aaslyng, M. D., Bredie, W.L.P., Olsen,
359 K., Møller, P. (2007). Interactions between oral burn, meat flavor and texture in chili

360 spiced pork petties evaluated by time-intensity. *Food Quality and Preference*,
361 18(6), 909-919.

Table captions

Table 1. Ethanol and procyanidin extract concentrations of wine matrices.

Wine matrix	Ethanol (% <i>, v/v</i>)	Procyanidin extract (mg/L)
W0C	0	0
W10C	10	0
W0P	0	200
W10P	10	200

Table 2. Aroma compounds used to aromatize the wines and concentrations used in the training of the panel.

Aroma		Concentration (mg/L)	
Compound	Descriptor	Low	High
Isoamyl acetate	Banana	30	90
Ethyl butanoate	Strawberry flavored candy	30	90
Ethyl decanoate	Dried plum	30	90
Ethyl hexanoate	Pineapple	30	90

369 **Table 3.** TI-RI values of panelists for each aroma compound in different wine
370 matrices.

	W0C				W0P				W10C				W10P			
Panelist	Banana	Strawberry	Dried plum	Pineapple	Banana	Strawberry	Dried plum	Pineapple	Banana	Strawberry	Dried plum	Pineapple	Banana	Strawberry	Dried plum	Pineapple
P01	0.555	-	0.541	0.405	0.628	1.011	1.247	0.750	0.805	0.507	0.976	0.665	0.364	0.748	0.345	0.687
P02	0.294	0.938	0.490	0.528	0.391	0.996	0.370	0.426	0.384	0.476	0.446	0.625	1.201	1.029	0.428	0.342
P03	1.266	0.981	0.794	0.515	0.592	0.644	0.375	0.585	1.463	0.946	1.018	0.817	0.402	0.387	0.587	0.955
P04	0.703	0.785	0.606	0.611	1.345	0.842	1.143	0.405	0.637	0.737	0.489	0.311	0.509	0.419	1.001	0.558
P05	0.858	0.848	1.053	0.541	0.676	0.966	0.669	0.558	0.964	0.704	0.479	0.623	0.702	0.628	0.674	1.104
P06	0.494	0.887	0.646	0.450	1.090	0.460	0.389	0.422	0.432	1.199	0.474	0.532	1.036	0.347	0.497	0.515
P07	1.030	-	-	0.631	0.449	0.541	0.374	1.124	0.421	0.644	0.674	0.440	0.405	0.345	0.385	0.950
P09	1.238	0.869	0.405	0.594	0.955	1.688	0.602	0.799	0.831	0.775	0.852	0.581	1.355	0.659	1.468	0.854
P10	0.795	1.819	-	0.621	0.885	0.885	0.679	0.464	0.851	1.209	1.765	0.774	1.199	0.674	0.728	0.943

371 See **Table 1** for matrix code explanation.

372 **Table 4.** ANOVA table of the effects of panelist, wine matrix and aroma on TI-RI.

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>p-Value</i>
MAIN EFFECTS					
A: PANELIST	2.0774	8	0.2597	3.03	0.0038
B: MATRIX	0.0287	3	0.0096	0.11	0.9530
C: AROMA	0.6923	3	0.2308	2.70	0.0488
RESIDUAL	10.6997	125	0.0856		
TOTAL (Corr)	13.5311	139			

373 Numbers in bold mean *p*-values < 0.05.

374 **Table 5.** Classical analysis of TI parameters ($I_{\max}$, $t_{\max}$, AUC) on panelist and wine
 375 matrix.

<i>Aroma</i>	<i>p-Value</i>			
	<i>Parameter</i>	<i>Panelist</i>	<i>Matrix</i>	<i>Matrix*Panelist</i>
Banana	$I_{\max}$	0.0001	0.8082	0.2843
	$t_{\max}$	0.0169	0.1340	0.3652
	AUC	0.0001	<i>0.0952</i>	0.504
Strawberry	$I_{\max}$	0.0001	0.6449	0.5947
	$t_{\max}$	0.3433	0.5600	0.9641
	AUC	0.0001	<i>0.0730</i>	0.4726
Dried plum	$I_{\max}$	0.0001	0.6514	0.4229
	$t_{\max}$	0.2703	<i>0.0718</i>	0.5359
	AUC	0.0001	0.7938	0.2163
Pineapple	$I_{\max}$	0.0001	0.0252	<i>0.0902</i>
	$t_{\max}$	0.0002	0.9681	0.6067
	AUC	0.0001	0.1045	0.1636

376 Numbers in bold mean p-values < 0.05, in italic mean p-value < 0.1.