Understanding quality judgements of red wines by experts: effect of evaluation condition

María-Pilar Sáenz-Navajas[a,b*], José Miguel Avizcuri[c], José Federico Echávarri[d], Vicente Ferreira[a,c], Purificación Fernández-Zurbano[c,d] and Dominique Valentin[b,e]

[a] *Laboratorio de Análisis del Aroma y Enología (LAAE), Department of Analytical Chemistry, Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2) (UNIZAR-CITA), Associate unit to Instituto de las Ciencias de la Vid y el Vino (ICVV) (UR-CSIC-GR), c/ Pedro Cerbuna 12, 50009 Zaragoza, Spain*

[b] *Centre des Sciences du Goût et de l´Alimentation, UMR6265 CNRS – INRA-UB, 9E Boulevard Jeanne d´Arc, 21000 Dijon, France*

[c] *Instituto de las Ciencias de la Vid y el Vino (ICVV) (Universidad de La Rioja-CSIC-Gobierno de La Rioja). Carretera de Burgos, km.6, Finca de la Grajera, 26007 Logroño, Spain*

[d] *Department of Chemistry, Universidad de La Rioja, c/ Madre de Dios 51, 26006 Logroño, La Rioja, Spain*

[e] *AgroSup Dijon, 1 Esplanade Erasme, 21000 Dijon, France*

**corresponding author: Maria-Pilar Sáenz-Navajas (maria@saenz.info)*

Abstract

The effect of evaluation condition on quality judgements of wine experts was evaluated. Quality perceived by wine experts was investigated under the assumption that this construct is built from multimodal sensory inputs. Twenty-one wine experts from Rioja (Spain) scored the intrinsic quality of 16 Spanish red wines under four conditions: (i) visual stimulation only, (ii) orthonasal olfaction alone, (iii) in-mouth sensations only (wearing a nose clip) and (iv) global tasting. Agreement among judges and the effect of evaluation condition were evaluated by principal component analysis (PCA) and ANOVA, respectively. In parallel, a trained panel described aroma, taste and in-mouth sensory properties such as astringency, global intensity and persistence. CIELab colour coordinates were also obtained. These descriptive data were submitted to regression analyses to explore their relationship with quality scores derived from the four evaluation conditions. Common mental representations of wine quality under visual, olfactory and global conditions were confirmed, while there was not a clear quality construct based exclusively on taste and mouthfeel properties. Wine taste and mouth-feel quality concept is suggested to be built only in combination with aroma and/or colour stimuli, and thus within a wine context.

Global quality judgement integrated information provided by visual and olfactory cues, even if olfactory stimuli were suggested to have more importance on the construction of the global quality concept of wine experts. Significant interactions between wine and evaluation condition revealed significant differences in quality scores dependent on the stimuli received during tasting and on the wine judged. Sensory cues driving quality, especially visual and in-mouth properties varied depending on the evaluation condition, which suggested that global wine quality concept would be the result of the integration of perceptual and cognitive information rather than a collection of independent stimuli.

Key words: evaluation condition; wine; quality perception; experts

1. Introduction

Quality is generally defined as the judgment of a products´ overall excellence or superiority (Zeithaml, 1988). Understanding the mechanisms underlying food quality perception is important as it is involved in the decision-making process of consumers at purchase situations (Marin & Durham, 2007). Wine is a particular case study within the general food and beverage domain as the opinion of wine experts, especially of the so-called wine gurus, exerts an important influence on wine market. It is thus important to understand sensory drivers of experts´ quality perception as their judgements tend to generate quality prototypes among wine consumers. Despite the known relevance of understanding quality perception for the wine industry, this concept is not yet fully understood in part because it is a multidimensional concept, which makes it difficult to define.

1.1. Multidimensionality of quality

The multidimensional character of perceived quality is related to factors such as the properties of the product itself, and the characteristics of consumers.

Quality perception is influenced by the characteristics of the product which have been mainly classified into intrinsic and extrinsic factors (Charters & Pettigrew, 2007). Intrinsic cues are those related to the product itself (physical part of it) and its organoleptic properties such as aroma, in-mouth properties or colour. Extrinsic cues refer to properties which are not physically part of the product such as package design or region of origin. For the specific case of wine, intrinsic cues of previously experienced wines are determinant in repurchase situations (Mueller, Osidacz, Francis, & Lockshin, 2010). The importance of extrinsic properties lies on the fact that at wine purchase the consumer is rarely able to taste wine and thus has to rely on extrinsic cues to infer wine quality.

Quality cannot be understood unless the characteristics of the consumer judging the product are considered. This is particularly important for wine since consumers' perceptions are quite heterogeneous and is highly influenced by consumer's level of expertise and different from that of experts (Ballester, Patris, Symoneaux, & Valentin, 2008). Experts seem to have common memorised wine prototypes, especially within the same production area (Hopfer & Heymann, 2014; Torri et al., 2013), contrary to less experienced consumers (Urdapilleta, Parr, Dacremont, & Green, 2011). The fact that quality assessment is based on technical winemaking processes for experts and on individual experiences for consumers results in a misalignment in the quality concept between wine professionals and low-experienced consumers (Lattey, Bramley, & Francis, 2010; Sáenz-Navajas, Ballester, Pêcher, Peyron, & Valentin, 2013).

1.2. Flavour: an integrated percept

Food flavour has been defined as the combination of stimuli perceived in the oral cavity combining taste, olfactory as well as trigeminal somatosensory and thermal perception. Prescott (2012b) suggested that during food experiences rather than the perception of individual discrete sensations, products are perceived as an integration of these signals. Discrete physiological sensory systems (taste, odours or tactile sensations) are anatomically separated, but they are functionally connected (Gibson, 1966). They are integrated into a single perception (flavour). Perceptions are constructed from a combination of both perceptual and cognitive signals, these lasts including the sensory properties of the object that are encoded in the memory (Small & Prescott, 2005).

In the context of wine flavour, Castriota-Scanderbeg et al. (2005) showed that the pattern of brain activations was different in wine consumers with different levels of expertise (experts vs naïve consumers). Experts showed activation of areas implicated in gustatory/olfactory integration in primates and involved higher cognitive functions such as memory. They showed higher sensitivity to combined olfactory and taste perception and thus the ability of integrating several sensory modalities, which would result in flavour representation (Pazart, Comte, Magnin, Millot, & Moulin, 2014). Differently, naïve consumers showed activations in the primary gustatory cortex and brain areas related to a more emotional and global experience when drinking a wine (Castriota-Scanderbeg et al., 2005). Less-experienced consumers seem to have recourse to more analytical approaches than experts, thus a complex stimulus seems to be perceived as the individual elements rather than integrated as a flavour.

1.3. Wine quality evaluated by experts

Wine quality is usually judged by wine professionals. For this purpose, either analytical (based on descriptive analysis) (Etaio et al., 2010) or integrated (holistic) (Goldwyn & Lawless, 1991) methodologies are described in the bibliography. Concerning analytical methods, it is widely extended in the wine sector that groups of experts from a same region carry out the sensory quality control, especially in Protected Designations of Origins (PDO) contexts such as that accredited and described by Etaio et al. (2010) for young red wines from Rioja. Usually, a panel of around five-seven experts carries out a descriptive task by scoring the intensity of individual parameters linked to visual, aroma and in-mouth properties and/or selecting positive attributes or defects from a previously established list. The parameters included in the score card are previously selected by a group of experts during the method development. These attributes have to be specific of the wine category object of evaluation and to influence its sensory quality. An overall quality score is calculated by applying a weighting factor to each parameter of the scorecard. The contribution of each parameter to the overall sensory quality is defined by consensus among experts during method development. For example, Etaio et al. (2010) attributed weighting factors of 10%, 30% and 60% to parameters evaluated in the presence of exclusively visual, aroma and all perceived in-mouth (aroma, taste and trigeminal sensations) cues, respectively. Accordingly, in-mouth and visual properties were suggested to be more and less important, respectively, for the overall sensory quality.

Integrated quality assessments consist in the direct evaluation of quality based on a holistic approach (Goldwyn & Lawless, 1991; Hopfer & Heymann, 2014). Experts are asked to score quality as a single multidimensional attribute of wine. This approach considers both the common mental representation of wine quality among wine experts from the same production area, and their heterogeneity, as mental concepts are based on individual experiences (e.g. past tastings), ideas and expectations. This methodology considers quality as an integrated percept (flavour) rather than the summation of individual discrete sensations (taste and mouth-feel, aroma, colour) in contrast to analytical approaches.

Most popular score cards for wine tasting combine both, analytical and holistic approaches. Therefore, in the first step of wine evaluation, quality of wine is scored based on exclusively visual stimuli. Then, judges evaluate wine quality based on olfactory cues and the last step involves the scoring of overall wine quality with access to all sensory stimuli: visual, olfactory and gustatory. Even if this wine tasting protocol is widely extended, there is a lack of scientific work exploring the relationship between global quality perception (with access to all stimuli) and quality scored in the presence of isolated sensory stimuli (e.g. visual or olfactory). In the present work, quality perception was evaluated in these three conditions: with visual stimulation only (Qv), with orthonasal olfaction alone (Qo), and global tasting (Qg: with visual, olfactory, taste and trigeminal stimuli) together with a fourth perception mode in the presence of in-mouth sensations only (Qm: wearing a nose clip). Even if wearing nose clips could be rather disturbing, they have been employed as a means of closing participants' nostrils in previous studies (Labbe, Damevin, Vaccher, Morgenegg, & Martin, 2006; Lawless et al., 2004; Parr et al., 2015) and are considered a suitable method to prevent olfactory perception. This permitted us to study the contribution of exclusively in-mouth stimuli (taste and trigeminal sensations) to the overall wine quality perception. Together with visual cues, orthonasal olfaction, in-mouth properties (taste, and trigeminal stimulation), retronasal olfaction is also involved in the perception of wines. However, the direct evaluation of this chemosensory process deems difficult, since in the oral cavity retronasal aroma stimuli and taste/mouthfeel properties are perceived simultaneously. Taking into account that procedures for the direct measure of retronasal aroma would be rather onerous for experts, even more than wearing nose clips, direct quality evaluation of wines based on exclusively retronasal aroma was not considered in the present study.

In this context, the present research aimed at exploring: 1) the presence of shared mental representations for quality in the presence of different sensory stimuli (visual, olfactory, in-mouth and global), 2) the effect of evaluation condition on perceived quality of red wines by experts, and 3) associations between quality perception and wine intrinsic cues (colour coordinates, aroma and in-mouth properties such as taste, astringency, global intensity and persistence).

2. Material and methods

2.1. Wines

Sixteen Spanish red wines from different wine making areas, varieties, vintages and with different ageing periods in both bottle and oak barrels were selected to cover a wide range of sensory properties. The detailed list of samples, including wine information and basic compositional oenological parameters, is shown in Table 1.

2.2. Quality evaluation by wine experts

2.2.1. Judges

The panel of judges was composed of 21 established winemakers from DOCa Rioja (Spain), twelve females and nine males ranging from 28 to 57 years of age (median = 35). Wine tasting and quality judging was part of their everyday professional tasks as they mainly base their winemaking and commercial decisions on tasting outcomes.

2.2.2. Evaluation protocol

Each judge completed four sessions (ca. 20 min each) in individual booths within the same day. In the first session each judge evaluated the quality of each of the 16 wines in dark glasses (to avoid visual influence) attending exclusively to orthonasal aroma properties (Quality olfaction-Qo-). In the second session, judges scored the quality based on exclusively visual stimuli (Quality visual-Qv-). In the third session, judges had to taste the wines in dark glasses while wearing a nose clip to avoid aroma and visual interactions and to score quality based on perceived in-mouth properties: taste and trigeminal sensations (Quality in-mouth-Qm). In the last session, wines were served in clear glasses and judges had access to all stimuli: visual, olfaction, retro-olfaction, taste and trigeminal sensations (Quality global-Qg-) of wines, as in conventional tastings. A break of 10 min was enforced after each session.

Just after judges had scored wine quality in the visual, olfactory and in-mouth conditions, they were asked to freely elicit visual, olfactory or in-mouth terms, respectively linked to high and low quality wines according to their own criteria.

Twenty-mL wine samples were presented randomly in coded dark (for Qo and Qm) or clear (for Qv and Qg) approved wine glasses (ISO 3591, 1977) at room temperature and covered with a Petri dish. The three-digit code assigned to each wine was different in each of the four sessions. Presentation order was randomised across judges within and across sessions. Water and unsalted crackers were available so that participants could cleanse their palate between wines. Judges were encouraged to expectorate wine samples.

Judges had to evaluate the samples once in the proposed order, in order to minimize any bias introduced by the sample presentation order. Afterwards, they could examine the samples as many times as they wanted and in any order. Unstructured 10-cm-long scales anchored with ‘‘very low quality” at the right-end and ‘‘very high quality” at the left-end were used to score quality in the four sessions (Hopfer & Heymann, 2014).

Participants were advised that they would taste and score quality of twenty wines in four sessions. They were not given any other information about the study.

2.3. Aroma and in-mouth characterisation of wines by a trained panel

2.3.1. Panellists

Panellists were recruited via email from Universidad de La Rioja affiliates, including students and staff, and gave oral consent to participate in the study. A total of 52 panellists were recruited on the basis of their interest and their availability during five months. They were not paid for their participation. For attendance reasons and based on panellist's individual performance evaluated using the reproducibility index developed by Campo et al. (2008), the responses of forty-one panellists (17 males and 24 females from 21 to 57 years old, median = 28) were considered for data analyses.

2.3.2. Panel training

The panellists were trained during eighteen sessions (ca. one hour per session) over a period of five months. This training period included two phases: a general (10 sessions) and a product specific (8 sessions) training phase. The wines selected for the general training phase presented intense and easily recognizable aroma, taste and astringency properties and included red, white and rosé wines of diverse grape varieties and origins. The objectives of the specific training sessions were for panellists to gain familiarity with the type of wines selected for the study. During a typical training session panellists became familiar with the specific vocabulary of an initial list of 110 aroma descriptors (Sáenz-Navajas, Fernandez-Zurbano, Martin-Lopez, & Ferreira, 2011) and with the rating of six attributes evaluated in-mouth: sweetness, acidity, bitterness, astringency, global intensity and persistence. In each session reference standards were presented as described elsewhere (Sáenz-Navajas, Fernandez-Zurbano, et al., 2011) to illustrate the aroma and in-mouth attributes. Then, panellists evaluated three to five different wines by describing their aroma properties (orthonasally) by choosing up to five descriptors from the list (Campo et al., 2008) and by rating tastes and astringency on a 10-point scale (0 = "absence", 1 = "very low" and 9 = "very high"), global intensity on a 9-point scale (1 = "very low" and 9 = "very high") and global persistence on a nine-point scale (1 = "very short" and 9 = "very long"). The session ended with a discussion during which the panel leader compared the aroma descriptors and the taste intensity scores given by panellists to describe each wine. During training, the panellists modified the initial list of terms by eliminating those terms they considered irrelevant, ambiguous or redundant and by adding additional attributes they considered pertinent. At the end of the training, the list included 113 terms.

2.3.2. Formal descriptive sessions

Trained panellists described wines following the procedure described in Sáenz-Navajas et al. (2011). Twenty-mL wine samples were presented in dark approved wine glasses (ISO 3591, 1977) labelled with 3-digit random codes and covered with plastic Petri dishes according to a random arrangement and monadic sequential presentation. Each panellist completed two sessions (ca. 45 min each) for the analysis of 20 samples (16 samples + 4 replicates of the same wine for evaluating individual and panel repeatability within sessions and reproducibility between sessions) involving ten samples per session. Panellists were asked to smell each wine, describe their odour by choosing a maximum of five attributes from the list of 113 according to the citation frequency method (Campo et al., 2008). Then, they were asked to taste the wine and rate sweetness, sourness, bitterness, astringency, global intensity, and global persistence of the samples using the above mentioned structured scales for each wine. Trained panellists rated samples using the sip and spit protocol described by Colonna, Adams, and Noble (2004). Therefore, ten seconds after wine was sipped, it was expectorated. Ten seconds later, apple pectin solution (1 g/L) was sipped, which was spat out after another 10 s. Between wine–rinse combinations, subjects rinsed twice with de-ionised water for 20s.

All wines were served at room temperature and were evaluated in individual booths. Panellists were not informed about the nature of the samples to be evaluated.

2.4. Visual characterisation of wines by CIELab coordinates

The CIELab coordinates of wines were calculated in order to have a complete characterisation of the colour of samples. Therefore, the transmittance spectra of this set of wines were measured. Measurements were carried out in Agilent 8453 UV-Vis spectrophotometer with photodiode array, using 0.2 cm path-length quartz cuvettes. Measurements were taken every 1 nm between 380 and 780 nm. Wine samples were previously clarified by centrifuging and passing wine through 0.45 µm filters. From the spectra, the colour coordinates were calculated using the CIE method, with the CIE 1964 10º standard observer and the illuminant D65, according to the OIV rules (Resolution Oeno

1/2006). The values correspond to the degree of wine lightness (L_{10}*) and the degree of red (when a_{10}* > 0), green (when a_{10}* < 0), yellow (when b_{10}* > 0), and blue (when b_{10}* < 0) colour (Ayala, Echavarri, & Negueruela, 1997).

2.5. Data analysis

2.5.1. Expert's agreement in quality evaluation

Quality scores were calculated by measuring the distance between the origin of the scale and the mark indicated by the participants, ranging from 0 to 10. Principal Component Analysis (PCA) was run on individual quality scores (judges in columns and wines in rows) derived from assessments under the four evaluation conditions (Qo, Qm, Qv, and Qg) in order to evaluate inter-individual consistency and thus judges' agreement. For that, a table with the wines in rows and the judges in columns was compiled for each condition (Ballester, Dacremont, Le Fur, & Etievant, 2005). Simple linear regression coefficients between the average (of the 21 judges) quality scores for a given condition and the individual score of each participant were calculated to evaluate panel agreement.

For the in-mouth condition (Qm) no agreement among judges was observed, thus quality scores grouped in a wine-by-participant matrix were submitted to hierarchical cluster analysis (HCA) with the Ward criteria in order to identify groups of participants scoring wines similarly. Accordingly, two groups of experts (clusters 1 and 3) and a judge (J3-cluster 2) evaluated in-mouth quality differently. Further PCA was conducted with the average quality scores for each cluster to evaluate their inter-relationship.

2.5.2 Correlation between evaluation conditions

A PCA was run on the quality scores averaged across judges in the visual, olfactory, in-mouth and global evaluation conditions to evaluate correlations between conditions.

2.5.3. Effect of evaluation condition on quality assessments

A three-way ANOVA, with judge as random factor and wine and evaluation condition as fix factors considering all main effects and interactions was calculated on the quality ratings. When a wine by evaluation condition effect was observed a two-way ANOVA (judges as random factor and evaluation condition as fix factor) was performed to evaluate the effect of evaluation condition on the quality scores of each wine sample. Bonferroni correction was applied to adjust for the effects of multiple testing. When a significant effect of evaluation condition was observed, pairwise comparisons were carried out using a Bonferroni pairwise comparison post-hoc test.

2.5.4. Sensory descriptive analysis

Evaluation of panel performance. For evaluating the individual performance of panellists in the orthonasal aroma description, average repeatability and reproducibility indexes (Ri) were calculated for each of the panellists from duplicate assessments of one wine, within the same session and between sessions. The minimum average Ri required to keep a judge response was set at 0.20 (Campo et al., 2008). The median of the average of Ri index (which varies from 0 to 1) was 0.58 and all were above 0.2, thus all subjects were considered in further analysis.

A contingency table, in which rows were the wines (including the replicates) and columns were the terms, was submitted to Correspondence Analysis (CA) to explore the global repeatability and reproducibility of the panel by evaluating the projection of wine replicates on the two-dimensional CA map. Replicates were close to each other on the map; thus the panel was considered globally repeatable and reproducible.

A PCA was run for each of the six attributes evaluated in mouth in order to assess judges' agreement. For that, a table with the wines in rows and the judges in columns was employed. Judges' projections were grouped in the loading plot for sourness, bitterness, astringency, global intensity and persistence. Thus, the panel agreed in the interpretation of these terms. On the contrary, for sweetness, judges were spread over the loading plot, which suggested that either the assessors do not interpret similarly these attributes or the sensory differences among wines for this attribute were marginal. Hence, sweetness was not further considered in subsequent analyses.

Selection of significant aroma terms. Chi-square tests were applied to the 113 aroma attributes to select the attributes with frequencies of citation (FC) higher than those expected by chance as described elsewhere (Sáenz-Navajas, Gonzalez-Hernandez, Campo, Fernández-Zurbano, & Ferreira, 2012). Twenty-eight individual attributes were discriminant. Among these discriminant attributes, those belonging to the same sensory category were then combined in order to obtain more general families/categories reaching higher FCs and larger magnitudes of variation. Accordingly, it was possible to establish that 10 aroma categories were relevant for the characterization of the sensory properties of the 16 wine samples. The final list of terms is presented in Table 2.

Multivariate analysis. A CA was performed on the wine by general terms contingency table. Only dimensions with an eigenvalue higher than the mean eigenvalue (Kaiser law) were retained. Quality scores obtained in the olfactory (Qo) and global (Qg) condition were projected as illustrative variables on the CA plot.

2.5.5. Relationship between quality scores and descriptive variables

The relationship between quality scores and descriptive variables was studied by multiple linear regressions (MLR) (Freedman, 2009) with cross-validation. Therefore, all factors derived from the CA calculated with combined aroma terms, in-mouth variables and colour coordinates were considered. As sensory descriptive scores and quality scores are not necessarily linearly related, linear and power correlations were also considered.

2.5.6. Classification of wines based on global quality perception

In order to identify groups of wines according to global quality, a first cluster analysis (HCA) was performed on all the PCs derived from the PCA calculated for global quality scores. With the three clusters identified, a two-way ANOVA analysis was performed with judges (random) and clusters (fix) as factors. Fischer post-hoc pairwise comparisons (95%) were calculated for significant effects.

To evaluate the presence of significant differences among the three clusters, one-way ANOVA (with cluster as fix factor) for colour coordinates, two-way ANOVA (with judges and cluster as random and fix factors, respectively) for in-mouth attributes and Chi-square ($\chi 2$) test for aroma attributes were performed.

3. Results

3.1. Experts´ agreement in quality evaluation based on different sensory stimuli

Figure 1 shows the loading of the judges onto the first two principal components (PC) derived from the quality scores in the four evaluation conditions: visual (Qv, Figure 1a), orthonasal olfaction (Qo, Figure 1b), in-mouth sensations with nose clips (Qm, Figure 1c) and global perception (Qg, Figure 1d). Figure 1a shows that in the visual condition, judges´ loadings are grouped on the positive side of the first PC (explaining almost 60% of the original variance), indicating a good inter-judge agreement. Figure 1b shows that in the olfactory condition, twenty out of 21 judges loaded on the positive side of the first PC (explaining 30% of variance). One judge (J17) loaded negatively on the first PC and positively on the second one, suggesting a strong opposition with quality scores of most judges. Figure 1c shows that in the in-mouth condition (with nose clip) judges' loadings are spread out over the PCA, suggesting disagreement among judges. Further cluster analysis calculated on individual scores allowed the identification of three groups of judges using similar quality criteria under this condition. The most numerous group was cluster 1, which was composed of 71% of judges, followed by cluster 3 (24%) and cluster 2 (5%). Cluster 2 was formed by exclusively one judge: J3, nevertheless their records were studied to further understanding in-mouth quality scores provided by the whole panel of experts. Scores of this judge were independent from the other two clusters as it can be observed in the PCA plot shown in Figure 2. The first PC, explaining 43% of the total variance, revealed a clear opposition between quality scores of cluster 3 (negative values for PC1 and plotted on the left part of Figure 2) versus judge 3 (cluster 2), which acquired positive values of PC1 (plotted on the right part). Thus, samples SO_C07 and CT_B07, related to quality perceived by cluster 1, were opposed to samples projected on the right part of the plot. Samples MG_V05 and CZ_D08 were especially related to quality perceived by judge 3, which were confronted to the youngest wines of the study (projected on the top-left part of the plot). The second

PC, explaining almost 40% of the original variance, is driven by quality scores of cluster 3 and thus related to wines with higher values of PC2 such as the young wines BE_R10 and RM_R10.

Figure 1d shows that in the global condition, most judges loaded positively on the first PC (explaining 30% of variance). As an exception, judge J16 loaded mostly on the sixth component (r=-0.61) of the PCA, suggesting that his or her judgement was different from that of most judges.

For each condition, average simple linear correlation coefficients (r) calculated between the average quality scores of the panel of experts and individual scores (given by each judge) showed that the highest average correlation coefficient was obtained for the visual condition (average r=0.73, ranging from +0.14 to +0.95), followed by the olfaction (average r=0.50, ranging from +0.09 to +0.74) and global condition (average r=0.48, ranging from +0.00 to +0.80). The lowest average correlation coefficient was observed for the in-mouth condition (average r=0.28, ranging from -0.29 to +0.68). These data evidence the presence of a relatively homogeneous concept of quality among judges under visual, followed by olfaction and global conditions, while there is a more heterogeneous non-consensual quality construct in the in-mouth condition.

3.2. Correlation between evaluation conditions

Figure 3 shows the projection of wines and quality scores in the four evaluation conditions onto the first two PCs of the PCA. The quality scores obtained in the four conditions are positively correlated with PC1 (r > 0.72), which explained almost 70% of the original variance. This suggested that there is a certain congruency in quality judgements of wines regardless the evaluation condition. Wines projected on the right side of the plot (GC_B10, BO_B10, RM_R10 and CT_B07) were perceived higher in quality (score > 1 on PC1) in the four conditions. On the contrary, wines AY_C05, CZ_D08 and SO_C07 (score < 1 on PC1) were perceived as lower quality exemplars.

Besides the commonalities observed on PC1, differences among the olfaction and visual evaluation conditions are shown on PC2 which explains about 19% of original variance. Olfaction and visual qualities were negatively (r=-0.60) and positively (r=+0.64) correlated with this PC, respectively.

Simple linear regressions calculated between the average quality scores for the global condition and the other three evaluation conditions suggested that judges could globally rely to a greater degree on olfactory ($r=0.77$; $P<0.05$) than on visual ($r=0.66$; $P<0.05$) information when judging global quality. Even if average global in-mouth quality scores were significantly correlated ($r=0.63$; $P<0.05$) this result has to be interpreted with caution given the high disagreement observed among judges in this condition.

3.3. Effect of evaluation condition on quality scores

Three-way ANOVAs calculated on quality scores (judges as random factor and condition and wine as fixed factors) showed significant effects for both main factors: condition ($F=7.3$, $P<0.001$) and wine ($F=15.2$, $P<0.001$) as well as their interaction ($F=3.6$, $P<0.001$). Thus, even if a global effect of the evaluation condition on quality scores was observed, this effect seemed to be dependent on the wine evaluated. This dependency could be further confirmed by calculating two-way ANOVAs (judges and evaluation condition as random and fix factors, respectively) for each wine on quality scores. Results showed significant main effects of the evaluation condition ($P<0.05$) for 38% of samples (RM_R10, SO_C07, GC_B10, CH_R10, CZ_D08, CD_C10), and no significant effect for the remaining wines. Among these six wines, four (SO_C07, GC_B10, CH_R10, CZ_D08) did not present significant differences between global and olfactory quality scores. Global and in-mouth quality scores did not significantly differed for four wines (RM_R10, CH_R10, CZ_D08, CD_C10) and two wines (RM_R10, GC_B10) showed no significant difference between global and visual quality scores.

3.4. Terms associated with low and high quality

Table 3 shows visual, aroma and in-mouth (taste and mouthfeel) terms associated with high and low quality. These terms were freely cited by judges after scoring wine quality in the visual, olfactory or in-mouth conditions. Visual attributes such as limpidity/clarity, depth (intense in colour), and red-purple colour were related to high quality, on the contrary, oxidised-brown colour, turbidity and light in colour to low quality.

The most elicited aroma attributes related to high quality were fruit, integrated wood, intensity, complexity and varietal aroma, while terms such as oxidation, reduction, dirty aroma, low intensity, *brettanomyces*, excessive old wood, faulty or green/vegetal aromas were linked to low quality.

Terms associated with high in-mouth quality were balance, volume/body, persistency, round/smooth tannins or fatty mouthfeel; in opposition to excessive astringency and sourness, unbalance, light/short, green sensation, bitterness or coarse tannins for low quality.

These results indicated that there were robust associations of visual, aroma and in-mouth terms to quality. It was interesting to note that even if judges showed no agreement in the concept of in-mouth quality (based exclusively on taste and mouthfeel sensations) when scoring quality of the studied sample set, there was a global agreement in associating in-mouth sensory terms to quality. Among these terms, together with classical terms such as astringency, balance or sourness, terms linked to more specific mouthfeel sensations such as round/smooth tannins, volume/body, fatty or green mouthfeels were cited (Table 3).

3.5. Linkage between quality scores and sensory variables

3.5.1. Linkage between quality scores and visual properties

A highly significant model was obtained (P<0.001) in the prediction of visual quality (Qv) from colour coordinates (Table 4). The b_{10}^* and L_{10}^* coordinates appeared to be significant negative predictors of visual quality: more yellow (and less blue: higher b_{10}*) and light-coloured (higher L_{10}*) wines were perceived lower in quality in the visual condition.

A second regression was calculated to evaluate the role played by the visual cues (colour coordinates) on global quality perception. Results showed a less significant model (P<0.05; R^2=0.36), involving the a_{10}^* coordinate as significant variable and suggesting that the red colour was the main visual cue driving global quality.

3.5.2. Linkage between quality scores and aroma properties

Ten dimensions of the CA retained 100% of the original variance. These 10 dimensions were used as predictors in multiple regression analysis of olfactory and global quality scores. The first two dimensions were the only significant dimensions in the model. So only these two dimensions will be presented in what follows. Figure 4 shows the projection of wines and terms into these dimensions together with the quality scores (projected as illustrative variables) in the olfaction (Qo) and global (Qg) conditions. The first dimension, which explained almost 35% of variance, was driven primarily by the terms herbal, lactic and roasted (positively) and by the term vegetables (negatively). For the sake of simplicity in the presentation of results, dimension 1 will be denoted as roasted/lactic/herbal aroma factor onwards. The second dimension, retaining more than 28% of the original variance, was driven primarily by the terms vegetables and red fruits (positively) and woody (negatively). This dimension will be denoted vegetables/red fruit aroma factor onwards. According to Figure 4, higher perceived qualities (evaluated in the olfaction and global condition) were linked to wines located on the bottom-right quadrant of the plot, while lower quality wines were located on the opposite side (top-left of the plot). Thus, wines mainly characterised by the term roasted (composed by the individual terms toasted bread, caramel and coffee) were linked to higher quality samples, while vegetal aromas and to a lesser extent animal were negatively correlated with perceived quality in both conditions.

In agreement with this observation, the regression models were significant in both olfactory (Qo) and global (Qg) evaluation conditions ($P<0.001$) but the regression coefficient was higher for Qo than for Qg ($R^2=0.60$ vs 0.50). Both models involved factors 1 (roasted/lactic/herbal) and 2 (vegetables/red fruits) (Table 5), but their role in the models was slightly different. On the one hand, Qo was linearly correlated with the roasted/lactic/herbal aroma factor (higher values for this factor resulted in higher Qo scores); while a quadratic relationship was observed for the vegetables/red fruits vector. This quadratic relationship suggested that when judges had exclusively access to olfactory information, the contribution of vegetables/red fruit aroma to the formation of the quality concept was more important in wines with higher intensity for this aroma factor, while it was less relevant for wines with lower values for this factor. Thus, for wines with negative values for factor 2 (plotted on the bottom part of Figure 4) the role of the vegetal/red fruit aroma factor was not as important as for wines plotted on the top part of Figure 4 (positive values for factor 2), for which higher vegetables/red fruit aroma resulted in lower quality scores. On the other hand, Qg was linearly correlated with the vegetable/red fruit aroma factor, while a quadratic relationship was observed for the roasted/lactic/herbal vector. These results indicate that when judges had access to olfactory, in-mouth and visual information (as in regular wine tastings), wines with higher vegetal-like aroma were scored lower in quality according to the simple negative correlation between quality and F2. Moreover, the negative quadratic correlation between quality and F1, suggested that for wines with lower intensity for factor 1 (roasted/lactic/herbal) the negative role played by the roasted/lactic/herbal aroma on quality perception was more important than for wines with higher intensity for this aroma.

3.5.3. Linkage between quality scores and in-mouth properties

A significant quadratic regression model ($P<0.05$) could be built for cluster 1, in which the sour taste was the sole significant variable (Table 6). Among wines with the lowest sour taste (<2.6), the lower this taste was, the higher in-mouth quality was perceived. However, for sourer wines (>2.6), the contribution of this taste to in-mouth quality judgements was limited. However, the relationship between quality and sourness should be considered with caution as a low variation in the sour taste of the studied wines was perceived (ranging from 2.2 to 3.3).

For judge 3 (J3), called cluster 2, a highly significant quadratic model ($P<0.01$) was obtained involving exclusively the astringent perception (Table 6) as it can be observed in Figure 5. This quadratic relationship suggested that the judge relied more on the tactile sensation in wines presenting higher astringency.

For the third cluster of judges, formed by 24% of participants, in-mouth properties considered for scoring in-mouth quality were less clear. No significant model could be built regressing in-mouth properties on global quality scores. Only a weak significant ($P<0.1$; $R^2=0.15$) simple positive linear correlation was observed between quality and sourness (Table 5). This result suggested that in-mouth quality for these judges was driven by other in-mouth sensory dimensions (different from taste, astringency, global intensity or persistence) that have not been described by the trained panel.

3.6. Linkage between global quality scores and sensory variables

A significant linear model was obtained ($P<0.001$; $R^2 = 0.85$) in the prediction of global quality from aroma, visual and in-mouth descriptors. The model is shown in equation 1.

$$\mathbf{Qg}=3.4+1.2\mathbf{*roasted/lactic/herbal}-0.87\mathbf{*vegetables/redfruits}+2.3*\mathbf{(vegetables/redfruits)}^2+0.002*\mathbf{a_{10}}^2-0.13*\mathbf{astringency}^2 \quad \text{(equation 1)}$$

The regression model showed that olfactory (roasted/lactic/herbal and vegetables/red fruits aroma vectors), visual (a_{10}^* coordinate) and in-mouth properties (astringency) were involved in global quality judgements. All the terms contributed significantly to the model ($P<0.05$ in all cases).

For further understanding wine quality judgements based on global evaluation, a PCA followed by cluster analysis was carried out with the individual quality scores. Three main clusters of wines were identified (Figure 6). With these clusters, a two-way ANOVA (judges as random and clusters as fix factors) followed by Fischer post-hoc pairwise comparisons (95%) were calculated. A significant effect of cluster was obtained ($F=37.1$, $P<0.0001$), which indicated that quality scores were significantly different among the three clusters. The cluster of wines with higher average quality scores (5.8±2.2) was composed of five samples: GC_R10, RM_R10, BO_B10, CT_B07 and CD_C10. Wines scored lower in quality (3.1±2.2) were CZ_D08, AY_C05, SO_C07, while the remaining eight wines belonged to the medium quality category (4.4±2.1).

The three wines with lower quality (CZ_D08, AY_C05, SO_C07) presented the highest frequency of citations for the terms vegetables and for two of them (CZ_D08, SO_C07) for animal aroma. These attributes were negatively correlated with perceived quality (Figure 4). This cluster presented significantly (chi-square=3.99; $P<0.05$) higher frequency of citations in comparison with the remaining 13 wines for the term vegetable (13.7 *vs* 4.5), while lower for roasted (3.7 *vs* 10.2; chi-square=6.3; $P<0.05$).

Leaving aside these three wines with negative aroma and thus low quality, the drivers responsible for differences between average and high quality wines were investigated. Results show that higher quality exemplars presented significantly higher values (F=11.6, P<0.01) for the a_{10}^* coordinate (50 *vs* 40) and significantly higher frequency of citations (chi-square=3.13; P<0.1) for the spicy attribute (13 *vs* 6). None of the in-mouth terms described by the trained panel presented a significant difference among high and average quality wines. This could be explained because the relationship between wine quality and astringency was not linear but quadratic as indicated in equation 1. A second potential explanation would be the fact that the set of in-mouth sensory descriptors scored by the trained panel was limited and experts would rely on other mouthfeel properties such as those cited in the declarative task (e.g., balance, volume/body, fatty mouthfeel, coarse, round or smooth tannins).

4. Discussion and Conclusions

4.1. Quality concept under different evaluation conditions

The lowest variability among the panel of experts when judging quality was observed when participants had access to visual stimulation (Qv) exclusively, followed by both orthonasal olfaction only (Qo) and conjoint visual, olfaction, taste and trigeminal (Qg) stimulations. These results indicated that there was a global agreement among judges when evaluating wine quality, which supports the notion of agreed mental representations for wine quality under these three evaluation conditions. This fact was further confirmed by the fact that judges exhibited robust verbal associations between sensory terms and quality evaluated under visual and olfactory conditions. This collective wine quality image was previously observed for constructs such as potential for aging (Langlois, Ballester, Campo, Dacremont, & Peyron, 2010) and typicality (Ballester et al., 2008). Wine experts are used to attending formal wine tasting sessions, in which they often have information about the wines they taste, which leads to lower variability and higher consistency in responses compared to novices (Urdapilleta et al., 2011). This higher consistency is attributed to the building of shared semantic sensory memory representations of wine knowledge through exposure, especially for experts belonging to the same wine region (Ballester et al., 2008; Langlois et al., 2010), even if groups of experts from different regions (Rioja in Spain vs Côtes du Rhône in France) have also been reported to present such commonalities (Sáenz-Navajas et al., 2013). Thus, when tasting a wine, experts compare its sensory properties with idiotypic recollections generated during previous experience to perform their quality judgement (Hughson & Boakes, 2002).

Concerning in-mouth evaluation of quality, there was an apparent consensus among judges from declarative data as terms such as balance, volume/body, persistency, round/smooth tannins or fatty mouthfeel were positively linked to wine quality, while excessive astringency or sourness, unbalance, light/short sensation, green mouthfeel, bitterness or coarse tannins were linked to low quality. However, this was not confirmed from a behavioural point of view as judges showed a generalised disagreement. A first potential cause for this disagreement could be linked to the fact of wearing nose clips, which may have disoriented them. This disagreement could also be explained in terms of absence of a shared mental representation and thus heterogeneity among participants in the in-mouth quality construct (access exclusively to taste and mouthfeel stimuli) of in-mouth quality concept among judges. This last possibility could be explained in terms of flavour integration and memory patterns. Experts process wine sensory information by similitude with wine flavours that they have stored in memory during previous experiences to try to recognise all characteristics of wine (Pazart et al., 2014). Binding and joint encoding of odours after pairing with tastes and tactile sensations has been described to be automatic (Prescott, 2012a). However, in the in-mouth condition, the stimuli they received did not seem to be familiar to them, as they usually evaluate taste and mouth-feel sensations in a context, in the presence of olfactory and/or visual cues simultaneously. Thus, the absence of mental prototypes of quality based exclusively on taste and trigeminal sensation stored in their memory could generate this disagreement among participants. This result suggested that the evaluation of wine quality based on taste and trigeminal sensation should be evaluated within a context, in which at least aroma should be present.

4.2. Linkage between global quality judgements and quality evaluated under isolated stimuli

Significant correlation coefficients between average global quality and quality scores evaluated with access to exclusive visual or olfactory sensory cues suggested that global quality judgement integrated information provided by visual and olfactory clues. These commonalities were stronger between global and olfactory quality scores, which would indicate the higher importance of olfactory, followed by visual cues olfaction cues on global perceived quality. Concerning in-mouth quality evaluation, the average scores were also significantly correlated with the average global quality score, which would suggest that judges also rely on in-mouth cues when evaluating overall quality. However, this result has to be interpreted with caution given the high disagreement among judges in the in-mouth condition (wearing nose clips). Even if judges seemed to rely on aroma as well as on visual and probably on in-mouth stimuli, a significant interaction of the evaluation condition and wine was observed, which suggested that the effect of evaluation condition was wine dependent. This result supported that global quality perception of wine was not a collection of independent stimuli but an integration of information from physiologically distinct sensory modalities leading to a new construct as stated by Small and Prescott (2005).

In this context, it would be important to consider whether a simple holistic and integrated approach, evaluating global quality impressions of wine experts similar to that employed in the present work and also proposed by Goldwyn and Lawless (1991) or Hopfer and Heymann (2014), would be more suitable for obtaining an overall quality judgement of wines than traditional quality evaluation schemes, which propose analytic approaches (individual flavour stimuli are evaluated separately) to generate an overall quality score calculated from the records of individual parameters. As already stated Lawless (1995), both analytical and integrated approaches have their advantages and disadvantages. The formers guaranty more reliable sensory descriptions derived from trained panels, easier to implement in quality control programs, while holistic methodologies take into consideration an integrated perception (closer to consumers´ experiences) and inter-judge diversity, which seems to better guaranty adaptation to changes in quality representations.

4.3. Drivers of quality judgements

Quality perceived under the four evaluation conditions were driven by different sensory attributes. Experts seemed to rely on both yellow colour (measured by b_{10}^* coordinate) and wine lightness (measured by L_{10}^* coordinate) when judging wine quality based on exclusively visual cues. Thus, more yellow and light-coloured wines were linked to low quality. Yellow nuances appear in prematurely aged red wines as a result of a deficient management of oxygen during wine making (Sanchez-Iglesias, Luisa Gonzalez-Sanjose, Perez-Magarino, Ortega-Heras, & Gonzalez-Huerta, 2009). This would explain why experts, which base their quality judgements mainly on technical variables such as oenological processes and viticulture variables (Parr, Mouret, Blackmore, Pelquest-Hunt, & Urdapilleta, 2011), associated yellow colour in wine with low quality. Concerning wine lightness, the role played by this variable in quality judgements would be more oriented in terms of wine prototypes stored in the memory of experts and related to specific wine regions. Thus, in the Spanish Rioja region, darker wines have been linked to higher quality samples (Sáenz-Navajas, Echavarri, Ferreira, & Fernandez-Zurbano, 2011). This could be linked to the fact that quality wines elaborated with Tempranillo (most cultivated variety in the region) are aimed at reaching high colour intensity. Notwithstanding, it could be hypothesised that for wines from regions elaborated with varieties yielding light-coloured wines such as Pinot noir in Burgundy, wine colour intensity (measured by L_{10}^*) would be differently linked to visual quality evaluated by experts in that production area.

Concerning olfactory quality, both declarative and behavioural data, suggested that the first driver of quality was the absence of defective aromas related to vegetal and animal nuances. From declarative data mainly fruity and integrated woody aromas were linked to high quality, while the behavioural task revealed that judges relied on roasted aroma when judging olfactory quality. This was well in accordance with literature dealing with assessments carried out by experts from different countries or highly-involved wine consumers in Australia (Lattey et al., 2010; Mueller, Osidacz, Francis, & Lockshin, 2010), Spain (Sáenz-Navajas, Fernandez-Zurbano, et al., 2011;

Sáenz-Navajas et al., 2012), France (Sáenz-Navajas et al., 2013) or Uruguay (Varela & Gambaro, 2006).

Regarding in-mouth quality judgments, three groups of judges showing different quality concepts were obtained. On the one hand, a certain linkage between sourness and quality was suggested for cluster 1 and Judge 3 (cluster 1 and 3), which was consistent with previous works carried out with Spanish wines evaluated by experts (Sáenz-Navajas, Fernandez-Zurbano, et al., 2011). However, this result should be interpreted with caution firstly because the range of intensity of sourness in the studied wines was low and secondary because the relationship between sourness and quality was not strong enough. On the other hand, the cluster formed by a sole judge relied on astringency when evaluating in-mouth quality as reported in the literature (Sáenz-Navajas, Fernandez-Zurbano, et al., 2011; Varela & Gambaro, 2006). However, the original variance explained was in all cases low (<50%).

Globally, these results indicated that there were not strong relationships between quality perceived in mouth scored by judges and in-mouth attributes evaluated by the trained panel. This fact together with the results derived from the declarative task, where several terms related to mouthfeel properties were cited, suggested that attributes traditionally measured by trained panels (such as taste or astringency) are insufficient for understanding in-mouth quality. Thus, further work should be carried out to develop an operational tool describing a wider range of in-mouth sensations as suggested by Gawel, Iland, and Francis (2001).

Intrinsic sensory cues driving global quality involved colour (red colour), aroma (defective and roasted aroma) and in-mouth (astringency) properties. It is interesting to note that visual and in-mouth sensory cues differed depending on the information that experts had access to when judging wine. Red colour of wines was a significant parameter taken into account (together with other sensory parameters) when evaluating the global quality of wines, but when judges had access to exclusively visual cues the sensory drivers considered in their judgements differed and were related to yellow nuances and wine lightness. For the in-mouth condition, no strong relationships between quality and studied in-mouth attributes could be found, while when they had access to all stimuli, astringency appeared to drive quality assessments. Concerning, aroma drivers, even if the role played was different to a certain degree, similar aroma terms were involved in both olfactory and global conditions. This reinforced the result related to the fact that olfactory cues had more importance on global quality judgements than visual or in-mouth drivers.

Acknowledgements

This work was funded by the Spanish Ministry of Economy and Competitiveness (MEC/FEDER) with the AGL2010-22355-C02 and AGL2014-59840 projects. J.M.A. acknowledges the Government of Navarra for his predoctoral (FPI) fellowship and M.P.S.N. the Spanish Ministry of Economy and Competitiveness (MINECO) for her postdoctoral fellowship (Formación Posdoctoral 2013). LAAE acknowledges the continuous support of Diputación General de Aragón (T53) and European Social Fund.

References

Ayala, F., Echavarri, J. F., & Negueruela, A. I. (1997). A new simplified method for measuring the color of wines .1. Red and rose wines. *American Journal of Enology and Viticulture, 48*(3), 357-363.

Ballester, J., Dacremont, C., Le Fur, Y., & Etievant, P. (2005). The role of olfaction in the elaboration and use of the Chardonnay wine concept. Food Quality and Preference, 16(4), 351-359. doi: 10.1016/j.foodqual.2004.06.001

Ballester, J., Patris, B., Symoneaux, R., & Valentin, D. (2008). Conceptual vs. perceptual wine spaces: Does expertise matter? *Food Quality and Preference, 19*(3), 267-276. doi: 10.1016/j.foodqual.2007.08.001

Campo, E., Do, B. V., Ferreira, V., & Valentin, D. (2008). Aroma properties of young Spanish monovarietal white wines: a study using sorting task, list of terms and frequency of citation. *Australian Journal of Grape and Wine Research, 14*(2), 104-115. doi: 10.1111/j.1755-0238.2008.00010.x

Castriota-Scanderbeg, A., Hagberg, G. E., Cerasa, A., Committeri, G., Galati, G., Patria, F., . . . Frackowlak, R. (2005). The appreciation of wine by sommeliers: a functional magnetic resonance study of sensory integration. *Neuroimage, 25*(2), 570-578. doi: 10.1016/j.neuroimage.2004.11.045

Charters, S., & Pettigrew, S. (2007). The dimensions of wine quality. *Food Quality and Preference, 18*(7), 997-1007. doi: DOI 10.1016/j.foodqual.2007.04.003

Colonna, A. E., Adams, D. O., & Noble, A. C. (2004). Comparison of procedures for reducing astringency carry-over effects in evaluation of red wines. *Australian Journal of Grape and Wine Research, 10*(1), 26-31.

Elortondo, F. J. P., Etaio, I., Albisu, M., Gaston, E., Ojeda, M., & Schlich, P. (2008). Development of a quantitative sensory method for the description of young red wines from Rioja Alavesa. *Journal of Sensory Studies, 23*(5), 631-655.

Etaio, I., Albisu, M., Ojeda, M., Gil, P. F., Salmeron, J., & Elortondo, F. J. P., (2010). Sensory quality control for food certification: A case study on wine. Method development. *Food Control, 21*(4), 533-541. doi: DOI 10.1016/j.foodcont.2009.08.013

Freedman, D.A. (2009). Statistical Models: Theory and Practice. Cambridge University Press.

Gawel, R., Iland, P. G., & Francis, I. L. (2001). Characterizing the astringency of red wine: a case study. *Food Quality and Preference, 12*(1), 83-94. doi: 10.1016/s0950-3293(00)00033-1

Gibson J.J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin Company

Goldwyn, C., & Lawless, H. (1991). How to taste wine. *ASTM Standardization News, 19*(3), 32-37.

Hopfer, H., & Heymann, H. (2014). Judging wine quality: Do we need experts, consumers or trained panelists? *Food Quality and Preference, 32*, 221-233. doi: 10.1016/j.foodqual.2013.10.004

Hughson, A. L., & Boakes, R. A. (2002). The knowing nose: the role of knowledge in wine expertise. *Food Quality and Preference, 13*(7-8), 463-472. doi: 10.1016/s0950-3293(02)00051-4

ISO 3591:1977. Sensory analysis - Apparatus - Wine-tasting glass.

Labbe, D., Damevin, L., Vaccher, C., Morgenegg, C., & Martin, N. (2006). Modulation of perceived taste by olfaction in familiar and unfamiliar beverages. *Food Quality and Preference, 17*, 582–589.

Langlois, J., Ballester, J., Campo, E., Dacremont, C., & Peyron, D. (2010). Combining Olfactory and Gustatory Clues in the Judgment of Aging Potential of Red Wine by Wine Professionals. *American Journal of Enology and Viticulture, 61*(1), 15-22.

Lattey, K. A., Bramley, B. R., & Francis, I. L. (2010). Consumer acceptability, sensory properties and expert quality judgements of Australian Cabernet Sauvignon and Shiraz wines. *Australian Journal of Grape and Wine Research, 16*(1), 189-202. doi: 10.1111/j.1755-0238.2009.00069.x

Lawless, H. (1995). Dimensions of sensory quality- A critique. *Food Quality and Preference, 6*(3), 191-199. doi: 10.1016/0950-3293(94)00023-o

Lawless, H, & Heymann, H. (2010). Quality Control and Shelf-Life (Stability) Testing. In In *Sensory Evaluation of Food Principles and Practices*: Boca Raton (FL): CRC Press.

Lawless, H. T., Schlake, S., Smythe, J., Lim, J., Yang, H., Chapman, K., et al. (2004). Metallic taste and retronasal smell. *Chemical Senses, 29*, 25–33.

Marin, A. B., & Durham, C. A. (2007). Effects of wine bottle closure type on consumer purchase intent and price expectation. *American Journal of Enology and Viticulture, 58*(2), 192-201.

Mueller, S., Osidacz, P., Francis, I. L., & Lockshin, L. (2010). Combining discrete choice and informed sensory testing in a two-stage process: Can it predict wine market share? *Food Quality and Preference, 21*(7), 741-754. doi: 10.1016/j.foodqual.2010.06.008

Resolution Oeno 1/2006. Determination of chromatic characteristics according to CIELab. Method OIV-MA-AS2-11.

Parr, W. V., Mouret, M., Blackmore, S., Pelquest-Hunt, T., & Urdapilleta, I. (2011). Representation of complexity in wine: Influence of expertise. *Food Quality and Preference, 22*(7), 647-660. doi: 10.1016/j.foodqual.2011.04.005

Pazart, L., Comte, A., Magnin, E., Millot, J.-L., & Moulin, T. (2014). An fMRI study on the influence of sommeliers' expertise on the integration of flavor. *Frontiers in Behavioral Neuroscience, 8*. doi: 10.3389/fnbeh.2014.00358

Prescott, J. (2012a). Chemosensory learning and flavour: Perception, preference and intake. *Physiology & Behavior, 107*(4), 553-559. doi: 10.1016/j.physbeh.2012.04.008

Prescott, J. (2012b). Multimodal chemosensory interactions and perception of flavor. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*: Boca Raton (FL): CRC Press.

Sáenz-Navajas, M. P., Ballester, J., Pêcher, C., Peyron, D., & Valentin, D. (2013). Sensory drivers of intrinsic quality of red wines. Effect of culture and level of expertise. *Food Research International, 54*(2), 1506-1518.

Sáenz-Navajas, M. P., Echavarri, F., Ferreira, V., & Fernandez-Zurbano, P. (2011). Pigment composition and color parameters of commercial Spanish red wine samples: linkage to quality perception. *European Food Research and Technology, 232*(5), 877-887. doi: 10.1007/s00217-011-1456-2

Sáenz-Navajas, M. P., Fernandez-Zurbano, P., Martin-Lopez, C., & Ferreira, V. (2011). Sensory properties of premium Spanish red wines and their implication in wine quality perception. *Australian Journal of Grape and Wine Research, 17*(1), 9-19. doi: DOI 10.1111/j.1755-0238.2010.00115.x

Sáenz-Navajas, M. P., Gonzalez-Hernandez, M., Campo, E., Fernández-Zurbano, P., & Ferreira, V. (2012). Orthonasal aroma characteristics of Spanish red wines from different price categories and their relationship to expert quality judgements. *Australian Journal of Grape and Wine Research, 18*(3), 268-279. doi: 10.1111/j.1755-0238.2012.00195.x

Sanchez-Iglesias, M., Luisa Gonzalez-Sanjose, M., Perez-Magarino, S., Ortega-Heras, M., & Gonzalez-Huerta, C. (2009). Effect of Micro-Oxygenation and Wood Type on the Phenolic Composition and Color of an Aged Red Wine. *Journal of Agricultural and Food Chemistry, 57*(24), 11498-11509. doi: 10.1021/jf902173d

Small, D. M., & Prescott, J. (2005). Odor/taste integration and the perception of flavor. *Experimental Brain Research, 166*(3-4), 345-357. doi: 10.1007/s00221-005-2376-9

Torri, L., Dinnella, C., Recchia, A., Naes, T., Tuorila, H., & Monteleone, E. (2013). Projective Mapping for interpreting wine aroma differences as perceived by naïve and experienced assessors. *Food Quality and Preference, 29*(1), 6-15

Urdapilleta, I., Parr, W., Dacremont, C., & Green, J. (2011). Semantic and perceptive organisation of Sauvignon blanc wine characteristics Influence of expertise. *Food Quality and Preference, 22*(1), 119-128. doi: 10.1016/j.foodqual.2010.08.005

Varela, P., & Gambaro, A. (2006). Sensory descriptive analysis of Uruguayan Tannat wine: Correlation to quality assessment. *Journal of Sensory Studies, 21*(2), 203-217.

Zeithaml, V.A. (1988). Consumer perceptions of price, quality and value – a means-end model and synthesis of evidence. *Journal of Marketing, 52*, 2-2

Figure Captions

Figure 1. PCA plots on dimensions 1 and 2 calculated on the individual quality scores given by judges based on: a) exclusively visual stimuli (Qv), b) exclusively olfactory stimuli (Qo), c) exclusively in-mouth stimuli (Qm) and d) global cues (Qg). The arrows represent the judges.

Figure 2. PCA plot on dimension 1 and 2 calculated on the average in-mouth quality scores of cluster 1, cluster 2 (formed by exclusively one judge: J3) and cluster 3

Figure 3. Projection of wines and quality scores in the four evaluation conditions on dimensions 1 and 2 of the PCA.

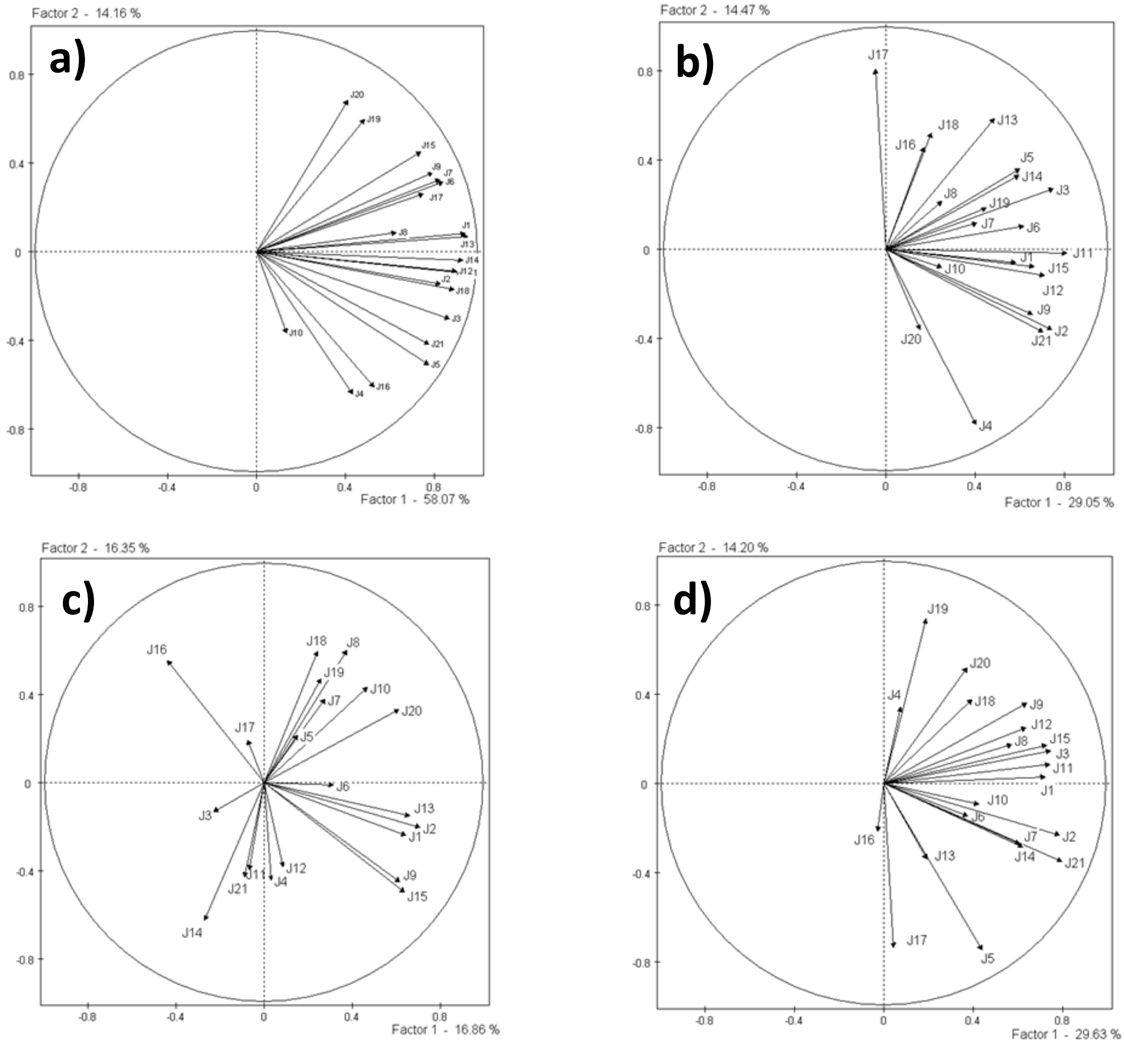
Figure 4. Projection of aroma descriptors and wines on the correspondence analysis space (dimensions 1 and 2). The arrows (illustrative variables) represent the average quality scores given by judges under the olfaction (Q olfaction, Qo) and global (Qg) conditions.

Figure 5. Second order-potential relationship between in-mouth quality scores (Qm) given by judge 3-cluster 2 (5% of the panel)- and astringent score derived from the trained panel.

Figure 6. Mean quality scores obtained for the 16 studied wines under the global condition (Qg: with access to visual, olfactory and in-mouth stimuli). Error bars are calculated as $s/n^{1/2}$; s: standard deviation, n: number of panellists. The three clusters of wines (High, Medium and Low quality) derived from the HCA are represented with different bar colours.

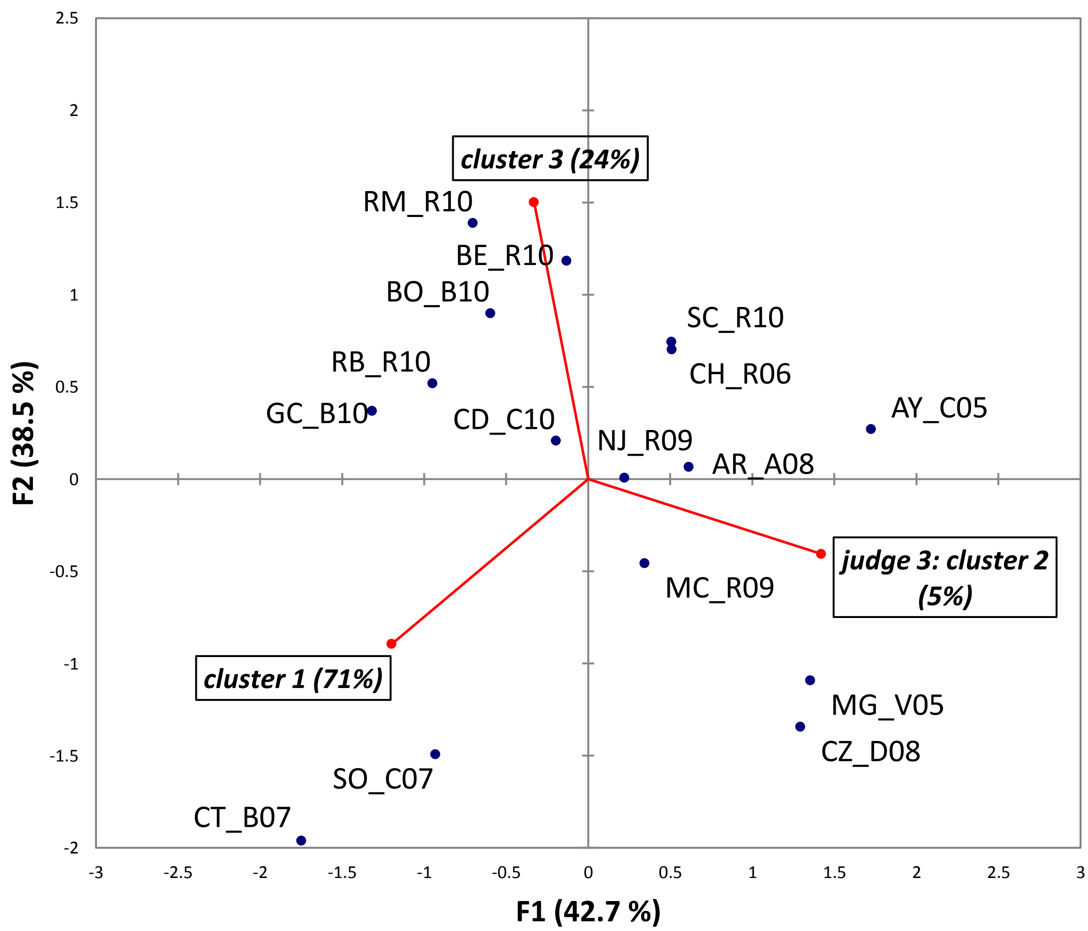
Figure

Figure 1.



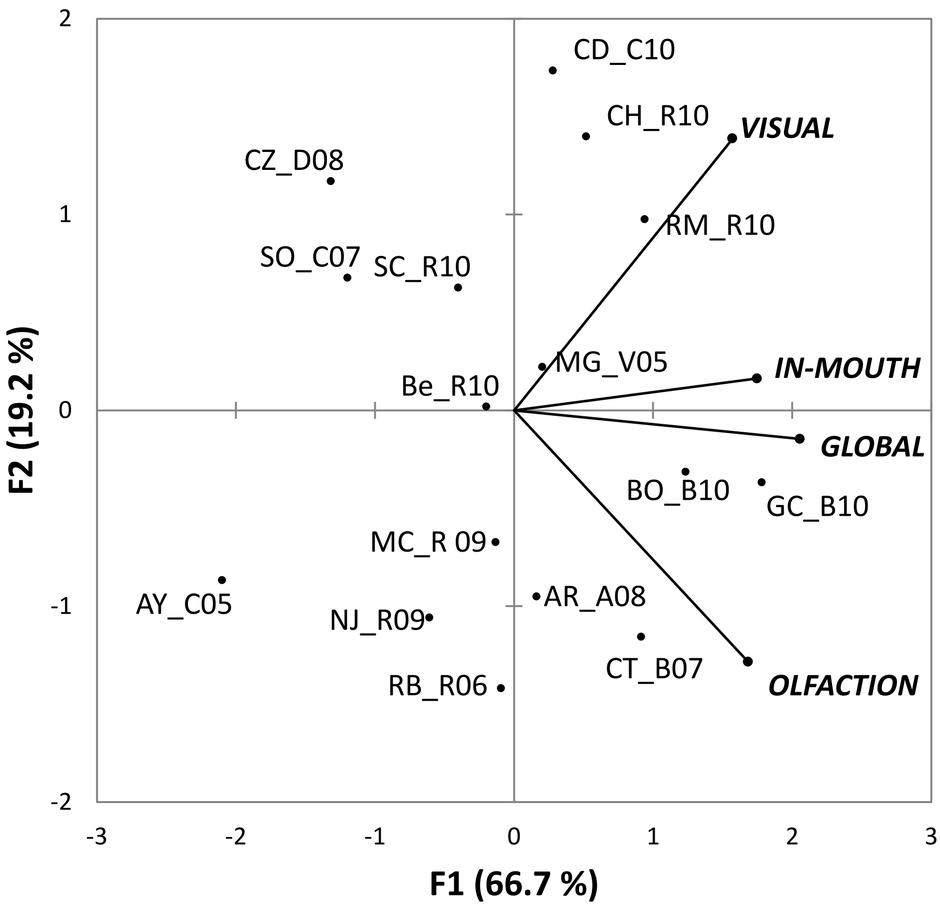
Figure

Figure 2



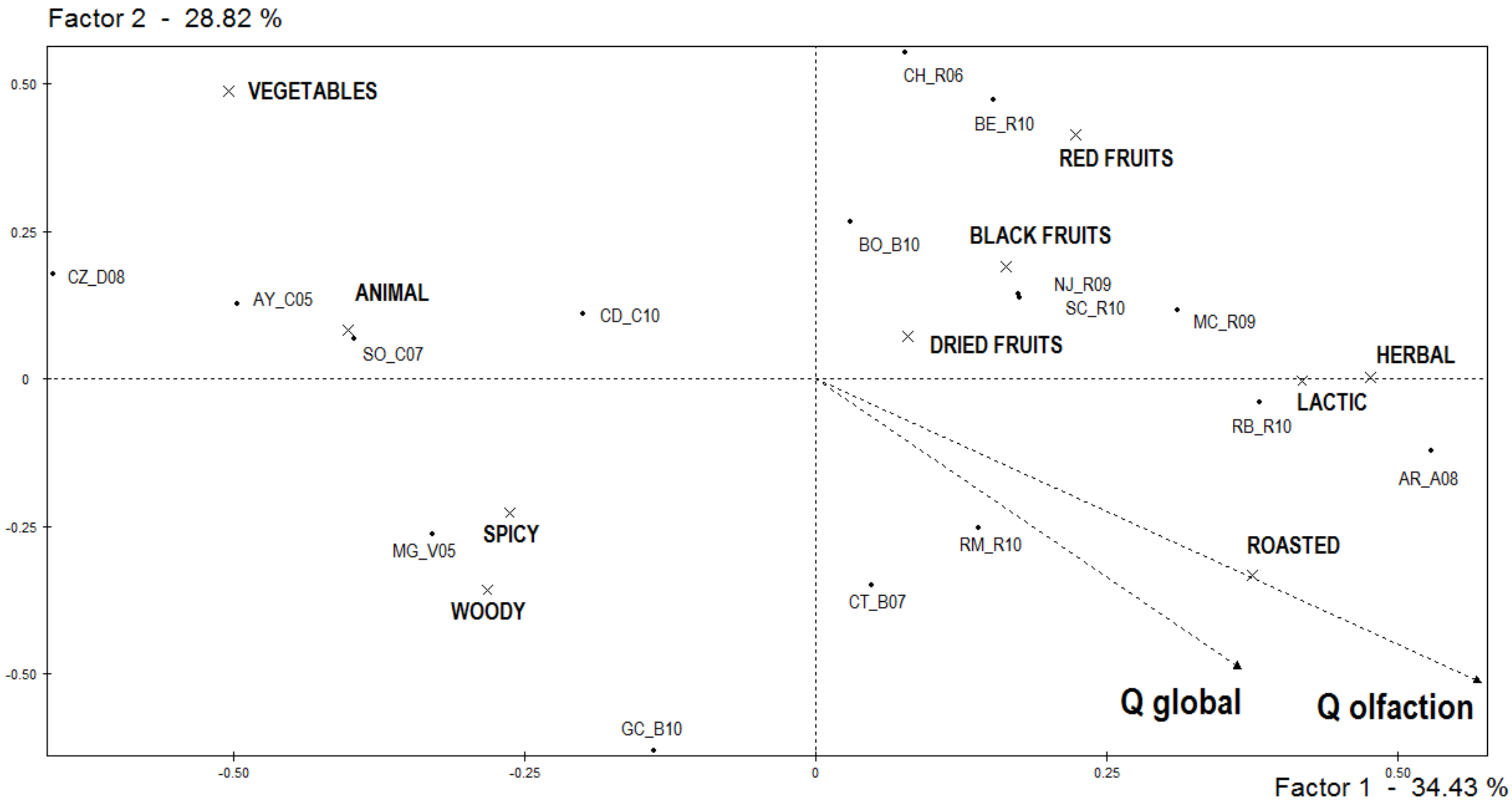
Figure

Figure 3.



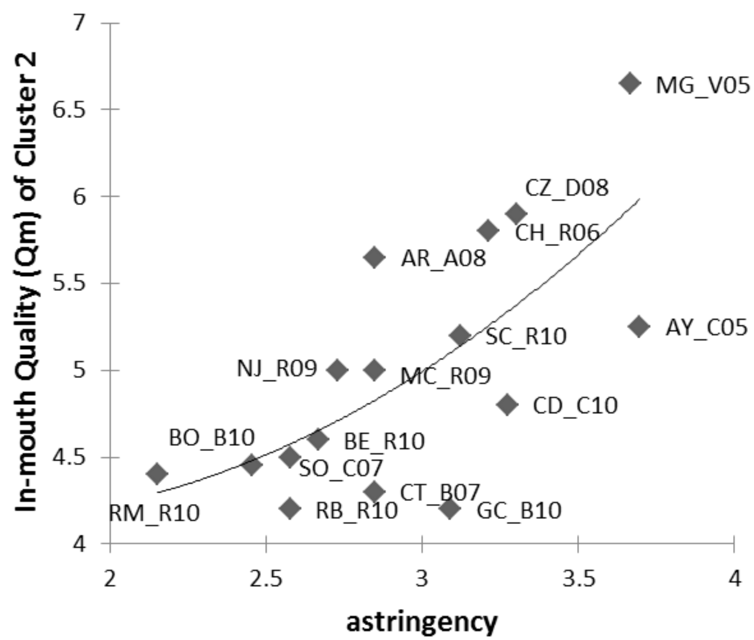
Figure

Figure 4.



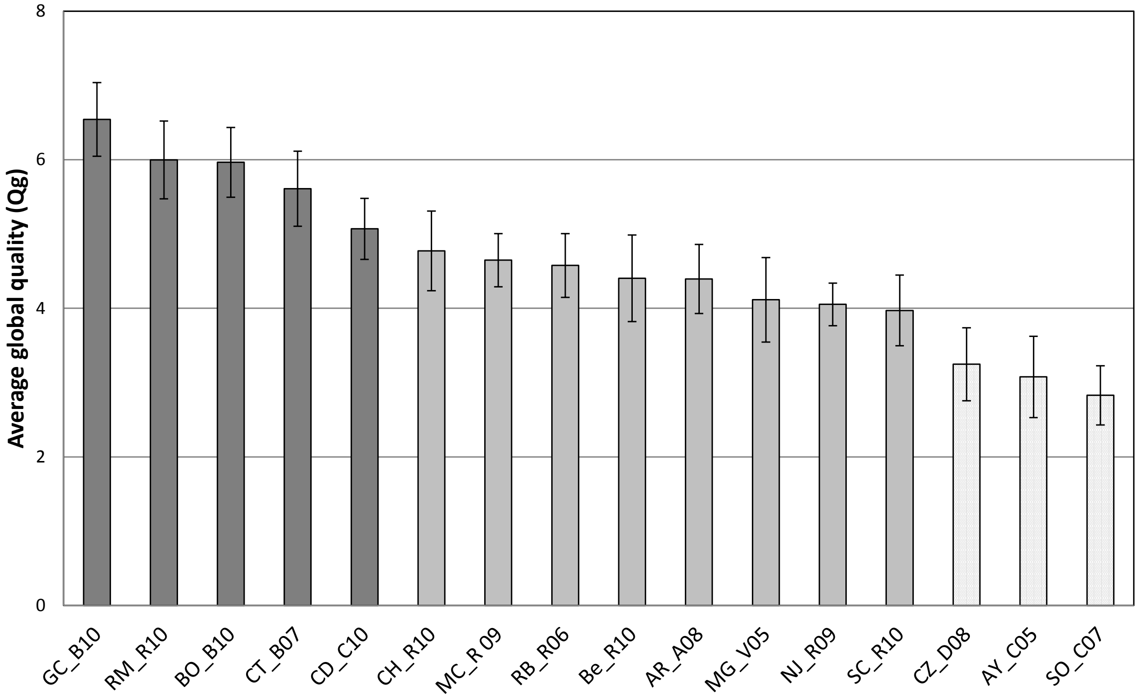
Figure

Figure 5



Figure

Figure 6



Table

Table 1. The sixteen studied commercial wines and their original oenological parameters.

wine code	origin	vintage	grape variety	oak aging	TPI[a]	pH	TA[b]	AV[c]	RS[d]	MA[e]	LA[f]	alcohol (% v/v)
MG_V05	DO Dominio de Valdepusa	2005	cabernet sauvignon	12	83.4	3.65	4.91	0.56	4.35	0.29	0.77	15.2
AY_C05	DO Cariñena	2005	merlot, tempranillo, cabernet sauvignon	10	74.3	3.52	5.86	0.69	3.39	0.33	1.00	14.3
GC_B10	DO Borja	2010	garnacha	4	71.4	3.43	6.14	0.42	3.61	0.25	0.68	14.7
RM_R10	DOCa Rioja	2010	graciano	8	66.4	3.57	5.80	0.41	2.31	0.19	1.45	14.8
CD_C10	DO Cariñena	2010	garnacha, tempranillo, cabernet sauvignon	0	66.4	3.63	5.30	0.53	2.57	0.24	0.90	13.5
CZ_D08	DO Duero	2008	tempranillo	18	62.0	3.65	5.33	0.57	1.71	0.35	2.47	13.4
BO_B10	DO Borja	2010	garnacha, syrah, tempranillo	0	61.0	3.66	5.04	0.47	2.68	0.17	1.07	14.8
CH_R06	DOCa Rioja	2006	tempranillo, viura	0	60.3	3.88	4.45	0.62	1.77	0.20	3.30	14.1
CT_B07	DO Borja	2007	garnacha	15	59.1	3.47	5.66	0.51	4.34	0.30	0.75	13.9
SC_R10	DOCa Rioja	2010	tempranillo, garnacha	0	57.8	3.72	4.84	0.48	2.32	0.18	2.52	13.4
SO_C07	DO Cariñena	2007	garnacha, tempranillo, cabernet sauvignon	18	54.9	3.53	5.66	0.75	3.81	0.18	1.21	13.8
AR_A08	DO Arlanza	2008	tempranillo	12	53.0	3.73	5.57	0.63	1.98	0.24	2.79	13.6
MC_R09	DOCa Rioja	2009	tempranillo,graciano, mazuelo	12	52.3	3.64	4.92	0.52	2.09	0.21	2.11	13.7
NJ_R09	DOCa Rioja	2009	tempranillo, garnacha	18	49.7	3.65	5.35	0.66	1.67	0.18	2.14	13.6
RB_R10	DOCa Rioja	2010	tempranillo, garnacha	18	49.4	3.49	5.37	0.57	2.23	0.23	1.45	14.3
BE_R10	DOCa Rioja	2010	tempranillo, garnacha	0	45.4	3.61	5.09	0.25	1.52	0.18	1.86	13.9

[a]Total Polyphenol Index. Absorbance at 280nm measured in 10-cm cuvettes
[b]Total titratable acidity expressed in g L^{-1} of tartaric acid
[c]Volatile acidity expressed in g L^{-1} of acetic acid
[d]Reducing sugars expressed in g L^{-1}
[e]Malic acid expressed in g L^{-1}
[f]Lactic acid expressed in g L^{-1}

Table 2. Combined terms (Cx) formed by individual attributes with their significance (*P* value) according to the χ^2 distribution. Data is expressed as percentage of frequency of citation (% FC).

Combined terms	**Red fruits (C3)**	**Black fruits (C2)**	**Dried fruits (C2)**	**Roasted (C3)**	**Woody (C2)**	**Spicy (C4)**	**Vegetables (C4)**	**Herbal (C3)**	**Animal (C2)**	**Lactic (C2)**
Individual terms	Red fruits	Black fruits	Dried fruits	Toasted bread	Wood	Spicy	Vegetal	Fresh tobacco	Animal	Butter
	Strawberry	Blackberry	Prune	Caramel	New wood	Liquorice	Vegetables	Thyme	Leather	Lactic
	Cherry			Coffee	Wood smoke	Black pepper	Olive	Menthol/ fresh		
						Vanilla	Backed potato			
Significance	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)	(P<0.001)
Maximum (% FC)	39%	41%	22%	51%	73%	49%	37%	37%	32%	12%
Samples for Max.	BE_R10	MC_R09	NJ_R09	AR_A08	GC_B10	CT_B07	AY_05	AR_A08	CZ_D08	AR_A08
Minimum (% FC)	2%	12%	2%	2%	7%	2%	2%	2%	2%	0%
Samples for Min.	GC_B10	CZ_D08	BE_R10	CH_R06	BE_R10	CH_R06	RB_R10 GC_B10 CT_B07	CH_R06	NJ_R09 CT_B07	SC_R10 SO_C07 AY_C05
Range (% FC)	37%	29%	20%	49%	66%	46%	34%	34%	29%	12%
Average (% FC)	20%	24%	13%	22%	28%	21%	15%	13%	12%	6%

Table 3. Visual, aroma and in-mouth (taste and mouthfeel) terms linked to high and low quality perception. Terms cited by less than 15% of experts have been omitted for clarity. Numbers in brackets are the frequency of citation for a term expressed in %.

	High quality	**Low quality**
Visual terms	Limpidity/clarity (81%), high depth-intensity (71%), red-purple colour (43%),	Oxidized-brown colour (81%), turbidity (67%), low colour intensity (57%)
Aroma terms	Fruit (71%), integrated wood (71%), intense aroma (43%), complex aroma (29%), varietal aroma(24%)	Oxidation (57%), reduction (52%), dirt (48%), low intensity (48%), *brett* (43%), excessive old wood (33%), fault (33%), green/vegetal (24%), mould (19%)
Taste and mouthfeel terms	Balance (67%), volume/body (48%), round/smooth tannins (43%), persistency (24%), fatty mouthfeel (19%)	Excessive astringency (67%), excessive sourness (52%), unbalance (48%), light/short (33%), green (29%), bitterness (29%), coarse tannins (19%)

Table 4. Regression models predicting visual quality (Qv) and global quality (Qg) from visual variables (a_{10}^*-red colour-, b_{10}^*-yellow colour-, L_{10}^*-lightness-), R-squared value, F-ratio and significance: *P<0.1, **P<0.05, ***P<0.01, ****P<0.001.

	equation	R^2	F	P
Qv	$13.4 - 0.12 \times \mathbf{b}_{10}^* - 0.13 \times \mathbf{L}_{10}^*$	0.92	88.7	****
Qg	$0.20 + 0.10 \times \mathbf{a}_{10}^*$	0.38	8.67	**

Table 5. Regression models predicting olfactory quality (Qo) and global quality (Qg) from aroma factors derived from CA analysis (F1: contributed mostly by roasted/lactic/herbal, F2: vegetables/red fruit), R-squared value, F-ratio and significance: *P<0.1, **P<0.05, ***P<0.01, ****P<0.001.

	equation	R^2	F	P
Qo	$4.5 + 1.9*\mathbf{F1} - 1.5*\mathbf{F2} + 2.8*\mathbf{F2}^2$	0.60	8.34	***
Qg	$5.2 - 1.4*\mathbf{F2} - 5.4*\mathbf{F1}^2$	0.50	8.66	***

Table 6. Regression models predicting in-mouth quality (Qm) perceived by three clusters of experts (cluster 1 formed by 71% of participants, cluster 2 by 5% and cluster 3 by 24%) from in-mouth attributes, R-squared value, F-ratio and significance: *P<0.1, **P<0.05, ***P<0.01, ****P<0.001.

	equation	**R^2**	**F**	**P**
Qm (Cluster 1) 71%*	29.8 – 17.1***sourness** + 2.9 ***sourness**^{2}	0.50	8.07	***
Qm (Cluster 2) 5%	3.3 + 0.2***astringency**^{2}	0.44	12.7	***
Qm (Cluster 3) 24%	0.3 + 1.5***sourness**	0.15	3.54	*

*for this model, AY_C05 was an outlier