Analysis and visualization of comprehensive gastronomic data

Combining recipes, impressions and aroma measurements

Marta Bevilacqua[1], Barbara Santos Silva[2], Michael Bom Frøst[3], Benedict Reade[4], Kristen Rasmussen de Vasquez[5], Andra Tolbus[6], Mikael Agerlin Petersen[1], Rasmus Bro*[1]

1. Chemometrics and Analytical Technology, Department of Food Science, University of Copenhagen, Denmark,

2. Department of Chemical Engineering and Biotechnology, Université de Sherbrooke, Quebec, Canada,

3. Nordic Food Lab, Design and Consumer Behaviour, Department of Food Science, University of Copenhagen, Denmark

4. Edinburgh Food Studio, Edinburgh, Scotland, United Kingdom

5. Nutritional Sciences and Toxicology, University of California Berkeley, USA

6. Copenhagen NV, Denmark

There is an increasing abundance of data in science in general and also in flavor research. As an example, analytical chemical instrumentation makes it possible to gather comprehensive information of, e.g., food samples in an affordable manner. With more information available, there is an urgent need for methodologies to combine and analyze such data.

We show how we can analyze, visualize, and explore the fundamental differences and commonalities in a set of aromatic blends (mixes of spices, herbs, and other aromatic ingredients). We combine sensory analysis with gas chromatography for aroma profiling, as well as meta-information including ingredients and geographical and cultural aspects.

TABLE OF CONTENTS

BACKGROUND

Combining culinary experience with scientific understanding is a complex task that has received considerable attention in recent years [1, 2]. Aroma, colour, texture, temperature, and sound are just a few of the factors that play a major role in food perception. When we eat we enjoy food with all of our senses, which contributes to an overall sensory perception. Changing one contribution can affect the others. However, palatability of a food is determined by aroma, taste, and texture [3]. In addition to that, chemesthesis plays a role, and the combined impact of taste, smell, and chemesthesis is referred to as flavour [4]. Consider, for example, spices: they are typically only added in smaller amounts to foods and dishes; consequently, they cannot supply significant amounts of macronutrients. Their main function is to add flavour to the foods [5].

In light of the fact that we enjoy food with all of our senses, it is of interest to be able to combine information from different sources into *one* overall perspective. For example, quantitative objective information such as chemical profiling of flavor or aroma compounds can be combined with more subjective hedonic and sensory characteristics. In addition, it is necessary to couple such information as a whole to more generic metadata (data that provides additional information about factors such as the origin of the samples).

All of these different sources of information should be looked at in a combined manner. *Data fusion* is the scientific field devoted to integration of multiple sets of data and knowledge [6, 7], but there is no simple pre-defined data fusion approach that allows the combination of so many and so diverse sources of information.

Here we develop and approach for handling this problem using as an example, a set of spice blends and pastes designed for adding flavor to food.

THE SPICE MIXES

Twenty-nine aromatic blends were used, primarily constituting a set of traditional blends from around the world in addition to some more experimental blends. The blends can be considered the defining features of a certain cuisine in the same manner as characteristic seasonings can [8]. The aroma blend samples can be categorized as aqueous based, oil based, dry based, dairy based, or fermented. This is an adapted version of what Rozin calls “flavor systems” [9] – a type of structure that comprises the variety of seasoning substances (Table 1).

Table 1. Overview of the spice mixes used.

Type	Name	Main ingredients
Aqueous Based	BBQ Chipotle Blend	chipotle, panela, apple vinegar
	Jerk Paste	onion, vinegar, scotch bonnet pepper, allspice,
	Juniper and Ant paste	juniper berry, thyme oil, ant, verbena, lemon
	Tomato/ Epazote	tomato, epazote, piquin chili
	Pickling Blend	coriander seed, vinegar, juniper berry, bay leaf
	Vihna d'alhos Base	white wine, onion, paprika, garlic
Oil Based	Afro Bahian Base	coconut milk, cassava, green pepper, cilantro
	Aji Panca Adobo	aji panca, beer, grape seed oil, soy sauce
	Ligurian Pesto	parmesan, pine nut, French beans, basil, potato,
	Massaman	shallot, lemongrass, galangal, garlic, red chile,
	Mole Negro	guajillo chilli, mulato pasilla chilli, chile chipotle
	Pipian	pumpkin seed, tomato, corn, achiote, epazote,
	Recado	achiote, masa, cumin, pepper
	Salsa verde	capers, flat leaf parsley, olive oil, anchovy, lemon
Dry Based	Berbere mix	coriander, clove, fenugreek, black pepper,
	Dukkha	hazelnut, cumin, sesame, coriander seed, black
	Quatre Epice	black pepper, cinnamon, clove, ginger, nutmeg
	Za'atar	oregano, thyme, sumac, sesame seed, cumin
	Shichimi Togarashi	Szechuan peppercorn, buckwheat, koji, orange,
	Chinese Five-Spice	star anise, Szechuan peppercorn, fennel seed,
	Panch Puran	fenugreek seed, nigella, cumin seed, black
	Garam Masala	white pepper, cinnamon, clove, cardamom,
Dairy	Fresh dill Marinade	crème fraiche, Dijon mustard, dill, honey, fennel

Based	Kadi	yogurt (cow's milk), onion, graham flour, ginger,
	Aji Escabeche/Peanuts	aji escabeche, huacatay, peanut, milk, crème
	Tikka masala	yogurt (cow's milk), onion, ginger, garlic, paprika,
Fermented	NFL fermented bean	black soy bean inoculated with *Aspergillus* oryzae,
	Lacto blueberry	blueberry, sea salt
	Peaso	yellow peas, buckwheat koji, sea salt

The aromatic blends span a broad range of cuisines and cultures (Appendix 3) as well as a broad range of flavors and textures.

SENSORY ANALYSIS

All samples were kept frozen before being dispensed and brought to room temperature 45 minutes prior to sensory testing to ensure consistency between trials. Sensory analysis was performed using a projective mapping technique [10], with some minor modifications [11]. A panel consisting of 26 subjects including chefs and other food professionals participated in the sensory profiling. We chose food knowledgeable individuals due to the large sample size, which is uncommon in this kind of sensory analysis and which can easily lead to fatigue in novice tasters. Panelists performed a holistic assessment of each blend and placed the blends on a rectangular sheet of paper. Samples were positioned according to similarities: the closer the samples, the more similar they were with respect to criteria the assessor deemed important. By imposing a coordinate system on the sheet, the final scoring is therefore the x-y coordinates (two numbers) for each blend. During the sensory session, each assessor was also encouraged to use Ultra-Flash profiling methods, or freely associate subjective descriptors to each aromatic blend. Thus, in addition to the coordinates, the sensory analysis also provides a list of descriptors associated to each blend.

During this study, all 29 different aromatic blends were presented simultaneously to the testers, which is quite extensive. To reduce the likelihood of fatigue, we made subject placement of samples and data quantification easier by increasing the size and including a grid on the Napping®

sheet. This projective mapping method is developed by Nordic Food Lab, and coined as "Big Grid" Napping. For this method, an A0-sized plastic coated sheet was used and a grid of 60 by 90 cm with two cm squares was printed on the sheet. The study was conducted at Nordic Food Lab, under ambient temperature and daylight (Figure 1).

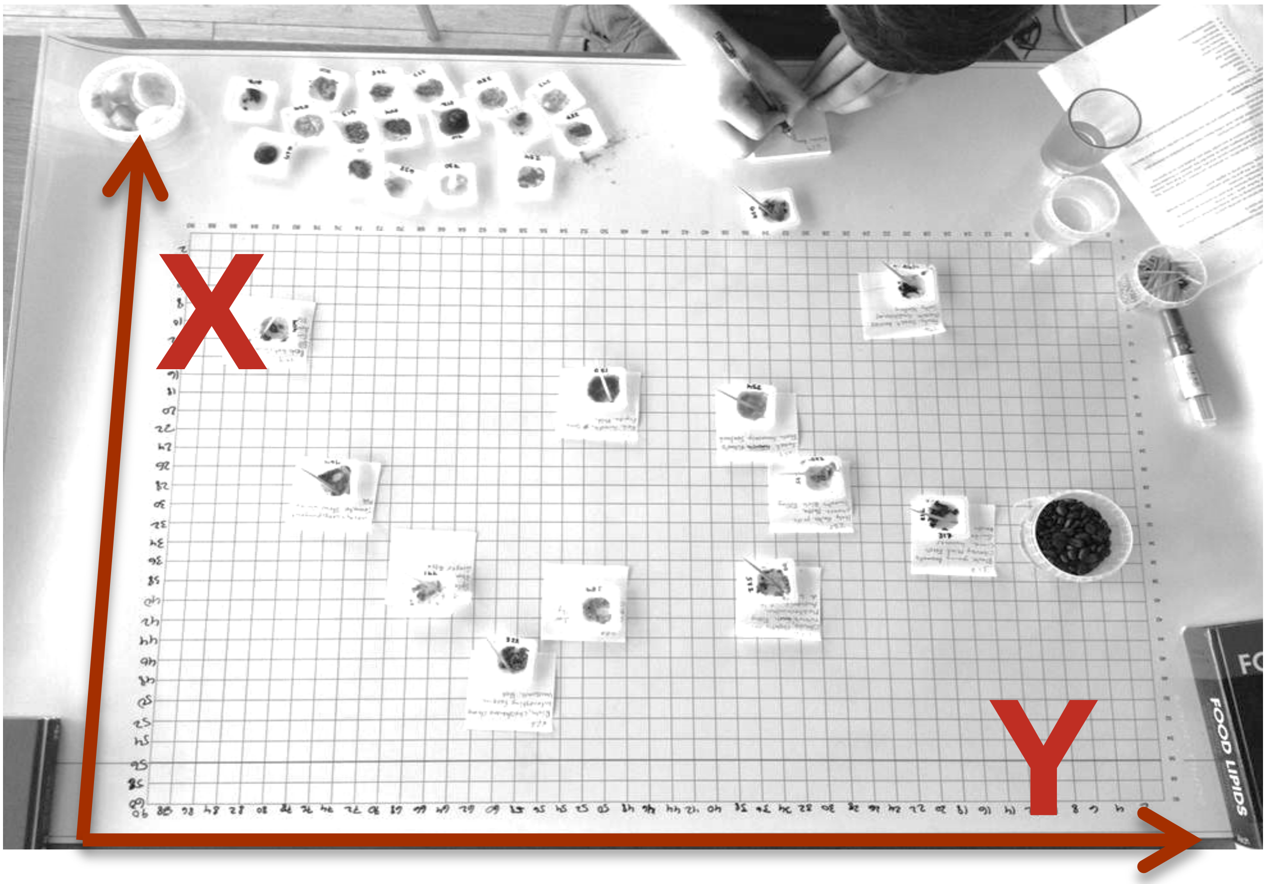

Figure 1. One assessor performing the projective mapping experiment. Coordinate axes are illustrated (note they are inverted due to perspective. Photo credit Nordic Food Lab.

For logistic reasons, respondents completed the study one at a time. Before the experiment, a quick five to ten minute instruction session was delivered. Participants were not provided with any details about potential word suggestions to describe samples, but were asked to determine their own criteria for placing and describing samples. Respondents did not discuss their evaluations with others during the assessment. To counteract sensory adaptation, respondents were provided with roasted coffee beans to smell when their noses adapted to the aromas of the blended spices/pastes and cucumbers to cleanse their palate. There was no coffee or cucumbers in any of the samples.

Sample descriptions from respondents were highly varied and, in addition to the 26 sets of scoring coordinates, a total of 545 descriptors were generated by the panelists. The descriptors were collected and analyzed for redundancy. It should be noted that the descriptors can be very idiosyncratic, and may differ as a consequence of the cultural and/or professional background of the respondent. A number of rules were applied in the analysis of the body of text. The first rule was stemming, where words that have the same stem are grouped (e.g. cumin and cuminy; chili, chile and chilli). Similarly, words and phrases that have similar meaning are grouped (e.g. smooth and silky; bitter and light bitter; fermented dairy, acidified milk product, and yogurt) to reduce the number of descriptors. This way, a consistent set of descriptive words was formulated and summed over respondents to be used for further statistical analysis. After stemming and grouping by meaning, 334 distinct descriptors remained. In total, 221 descriptors were used five times or less and those descriptors were removed leading to a total of 113 distinct and common descriptors, which are listed in Appendix 1. In the following, the descriptors are simply referred to as *Words* (see Figure 2).

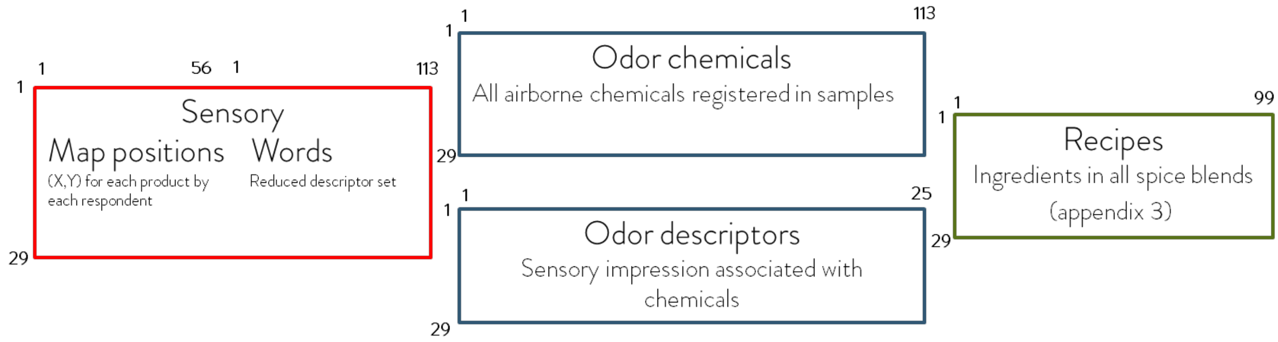

Figure 2. Overview of all data blocks available.

AROMA PROFILING

An amount of 0.2 g of sample was weighed into a gas washing bottle and mixed with 20 mL water and 1 mL internal standard solution (5 ppm 4-methyl-1-pentanol in water). The sample was purged with nitrogen (50 mL min^{-1}) for 20 min at 37°C. Volatile compounds were collected on Tenax-TA traps. The traps contained 250 mg of Tenax-TA with mesh size 60/80 and a density of 0.37 g mL-1 (Buchem bv, Apeldoorn, The Netherlands). After purging, water was removed from the traps using a flow of dry nitrogen (50 mL min^{-1} for 10 min).

The trapped volatiles were desorbed using an automatic thermal desorption unit (ATD 400, Perkin Elmer, Norwalk, USA). Primary desorption was carried out by heating the trap to 250°C with a flow (60 mL min^{-1}) of carrier gas (H2) for 15.0 min. The stripped volatiles were trapped in a Tenax TA cold trap (30 mg held at 5°C), which was subsequently heated at 300°C for 4 min (secondary desorption, outlet split 1:10). This allowed for rapid transfer of volatiles to a gas chromatograph-mass spectrometer (GC-MS, 7890A GC-system interfaced with a 5975C VL MSD with Triple-Axis detector from Agilent Technologies, Palo Alto, California) through a heated (225°C) transfer line.

Separation of volatiles was carried out on a DB-Wax capillary column 30 m long x 0.25 mm internal diameter, 0.50 µm film thickness. The column pressure was held constant at 2.4 psi resulting in an initial flow rate of approximately 1.2 mL min^{-1} using hydrogen as carrier gas. The column temperature program was: 10 min at 30°C, from 30°C to 240°C at 8°C min^{-1}, and finally 5 min at 240°C. The mass spectrometer was operating in the electron ionization mode at 70 eV. Mass-to-charge ratios between 15 and 300 were scanned. Volatile compounds were identified by probability based matching of their mass spectra with those of a commercial database

(Wiley275.L, HP product no. G1035A). The software program, MSDChemstation (Version E.02.00, Agilent Technologies, Palo Alto, California), was used for data analysis. At the end of the identification process, a total of 122 aroma compounds have been identified. Each sample was run in duplicate and peak areas averaged upon integration, thus obtaining a matrix with dimensions 29x122 called *Odor chemicals*. Moreover, an online database (www.thegoodscentscompany.com, Feb. 2016) was used to investigate the information about the sensory properties of the different volatile compounds. For 25 of the 122 chemical compounds, we were able to find the sensory properties associated with them. E.g. *Aroma/Odor* compound 66 is Camphene and has sensory properties best described as Woody/herbal/needle. The reduced set of peak areas (~relative concentrations) of the 25 aroma compounds has been collected into a matrix named *Odor descriptors*.

DATA AVAILABLE

For each of the 29 aromatic blends we thus have the following information available (Figure 2):

- Sensory map – Using the so-called projective mapping to characterize the samples. These measurements are collected in a matrix of dimension 29 x 52 (26 x 2 coordinate axes).
- Sensory descriptors (*Words*) - In addition to the normal projective mapping, there are also a number of descriptors/words, that sensory evaluators have attached to the samples. After grouping and stemming of the original set of descriptors, this set of data consist of a matrix with 29 rows and 113 columns.
- Aroma profiling - Through headspace sampling, GC-MS analysis was used to obtain an aroma profile for each spice mix. The aroma profile consisted of the relative integrated peak area (relative concentration) of 122 compounds and is called *Odor chemicals*

(Appendix 2). The compounds found with available sensory descriptors are collected in a smaller matrix (29 x 25 named *Odor descriptors*).

- Meta-data - name, *recipe* (or rather the fraction of each ingredient, but called recipe here), provides an approximate assessment of cultural identity of each recipe and base (oil, aqueous, dry, fermented and dairy). The details of these are given in Appendix 3.

MODELING EVERYTHING

Projective mapping and aroma profiling provide two complementary means to quantify the differences between blends, and can be used to perform a combined analysis that shows to what extent sensory results are already contained in the aroma analysis and vice versa.

Traditionally, projective mapping data is decomposed using MFA, Multiple Factor Analysis [12, 13]. Essentially MFA is performed by combining the data from each assessor (two coordinates). The matrix of data from each assessor is scaled according to variance and then assembled into a single matrix with as many variables as the number of assessors times two. A PCA model of this combined matrix provides the MFA model of the data.

MFA allows the identification of the common structure and provides a low-rank subspace (a score matrix) representing the main common variation in the 26 sets of scorings. In this case, it is desirable to use the aroma data to extract more information than is normally obtained from the projective mapping data. This is particularly interesting for this data set as it contains both an unusually high number of assessors and samples. In order to analyze the data, the following approach is adopted: the projective mapping data is held in a 29×52 matrix. Normally, each assessor matrix is scaled by its first eigenvalue but due to the similar magnitude of the individual data matrices, this is not needed nor done here. Instead, the whole matrix is scaled to unit variance and centered. Likewise, the data from aroma profiling is scaled to unit variance and centered but only after a logarithmic transformation of the original data (with the value two added throughout to avoid negative numbers). The logarithmic transform helps provide a more even distribution of aroma compounds. Without the logarithmic transform, the model would only reflect the few high intensity compounds. Compounds with low concentration may still have high importance from a sensory point of view.

The two blocks of projective mapping and aroma chemical data are combined into one matrix. Some (23) of the aroma variables are so badly described by a combined model that they are excluded during the data analysis process – as they are not important from nor reflected in, the model. Hence, the final model has 52 projective mapping and 99 flavor variables giving a matrix of dimension 29x151. The fused matrix can be properly analyzed by Principal Component Analysis (PCA) [14] and ten components are found to describe most of the variation (77%).

In Figure 3, the bar plot illustrates how most of the variables from both sensory and aroma profiling are explained by the PCA model of the combined data (29x151). On average, around 80% variance is explained of each variable and both blocks of data are general equally well explained. Hence, the model is really presenting the *fusion* of the two blocks and the components can be assessed from both an aroma and a sensory point of view.

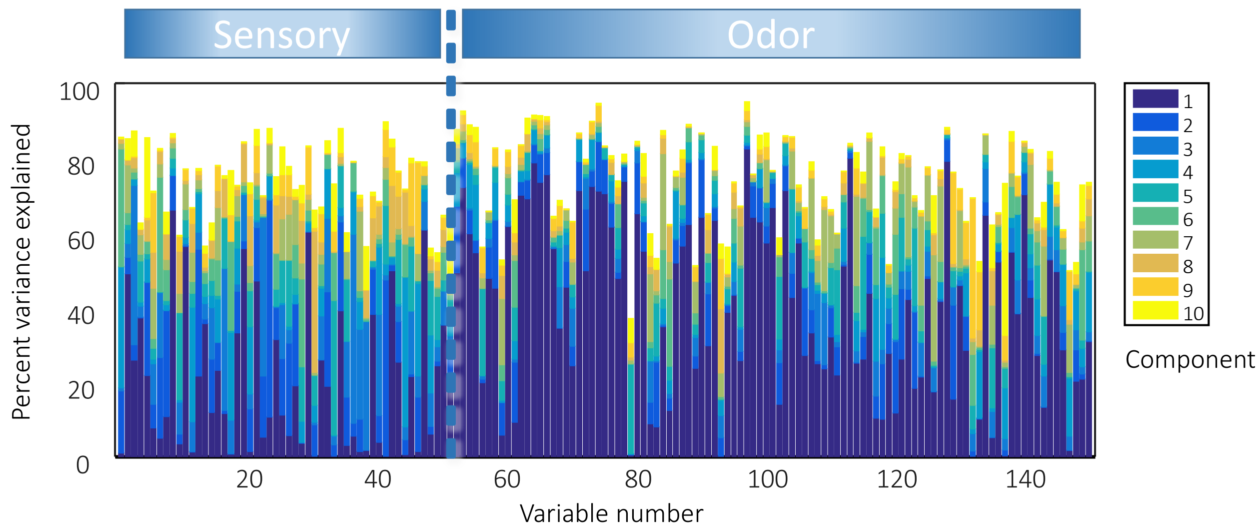

Figure 3. Percentage of variance explained of each of the 151 variables in the fused data. The amount of variance explained is given as function of the component number. The total height is the cumulative variation explained (in percentages) for each variable.

All the additional metadata available can be visualized in the PCA model using the data as *supplementary material,* as it is referred to in the MFA related literature [12]. It is possible to combine the scores with the *other* data sets such as *Recipe* and see what loadings the data matrix *Recipe* would generate in this case.

Hence, the scores of the PCA model

$$\mathbf{X}^{(sensory/aroma)} = \mathbf{T}\mathbf{P}^{T(sensory/aroma)}$$

are used as fixed latent variables to explore how these are manifested in e.g. the *Recipe* data (properly preprocessed):

$$\mathbf{X}^{(recipe)} = \mathbf{T}\mathbf{P}^{T(recipe)}.$$

The score matrix **T** is red to indicate it comes from the PCA model above and is not estimated using the recipe data. We call these loadings *pseudo-loadings*. The complete procedure for analyzing the data is visualized in Figure 4.

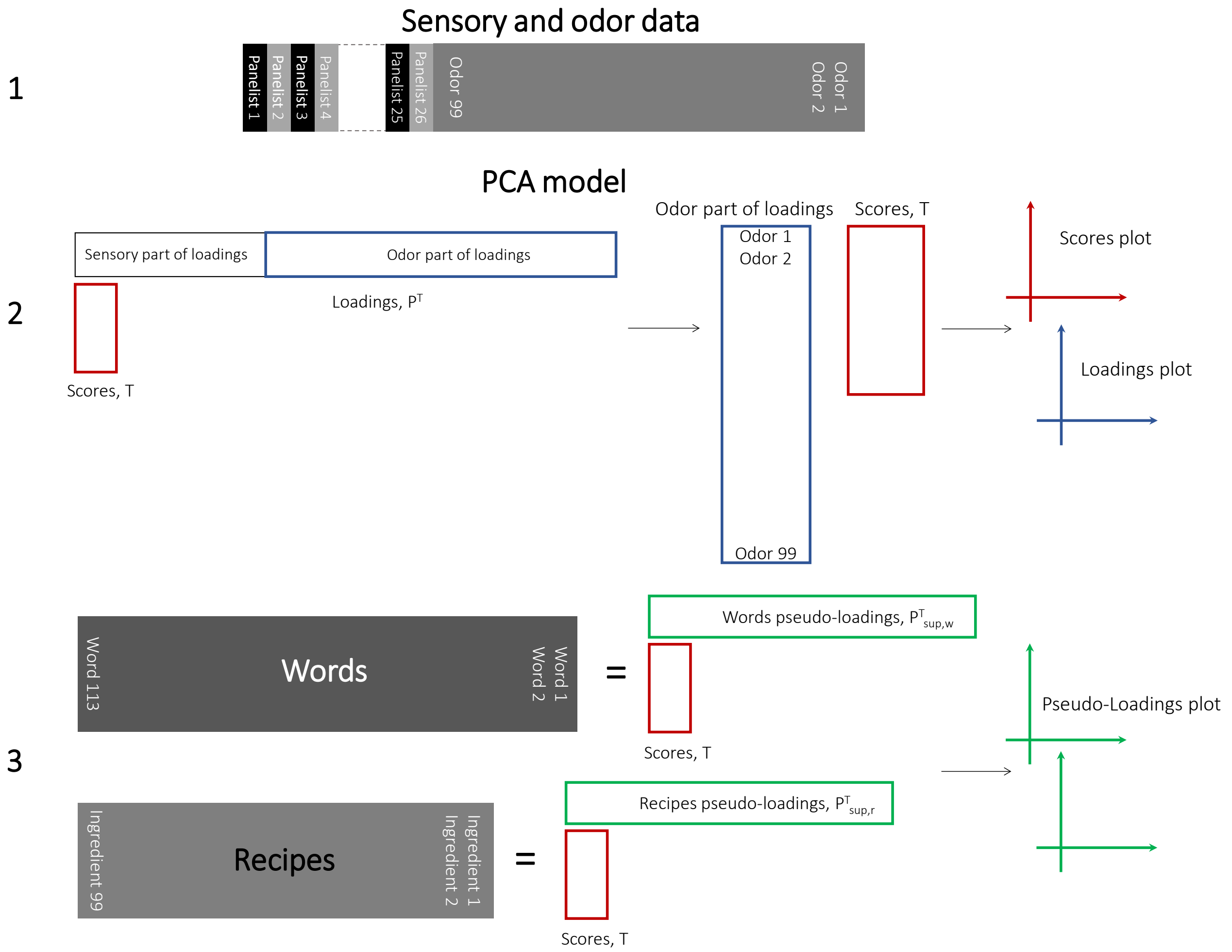

Figure 4. The data structure of the data fusion model.

COMBINING IT ALL

In order to understand the model, it is of course possible to do ordinary score and loading plots. However, we are particularly interested in understanding the relation between aromatic blends, aroma compounds, and the words used to describe the blends (sensory descriptors). Loading plots of the PCA are produced in a standard fashion, but only the part corresponding to aroma compounds is shown. The loadings corresponding to assessor coordinates are not shown as they are only of importance for steering the PCA model, and not for understanding the gastronomic underlying features. Instead, the loadings are shown for aroma compounds both in terms of actual common or IUPAC names and, where available, as most dominant sensory property (e.g. woody). Hence using either the matrix of odor chemicals or odor descriptors.

As already mentioned, all the external information, the words/attributes given by individual assessors and the lists of ingredients used in the recipes, is also visualized in "pseudo-loading" plots. The interpretation of these pseudo-loadings is similar to normal loadings. Summarizing, in total, there are then three sets of row-spaces that are visualized for each set of components: odor (chemicals or descriptors), words and recipes (ingredients).

The loading plot becomes too cluttered when there are very many variables. To better visualize the data, each loading element is reported on the figure on a grey scale, with size and color intensity proportional to its relative importance in the data set. It follows that the most important loadings will have not only higher values, but also higher visibility due to the bigger and darker font.

RESULTS

The results of the data analysis described in the previous paragraphs are summarized in comprehensive plots as presented in Figure 5 and Figure 7. In Figure 5, the first two and most important components of the PCA model are shown. These components explain most of the variability of the model (50%). In particular, the score plot with all of the samples colored by origin (geographical or developed by), are plotted on the top left corner of the figure. The upper right plot shows the words that the assessors have used for describing the same samples.

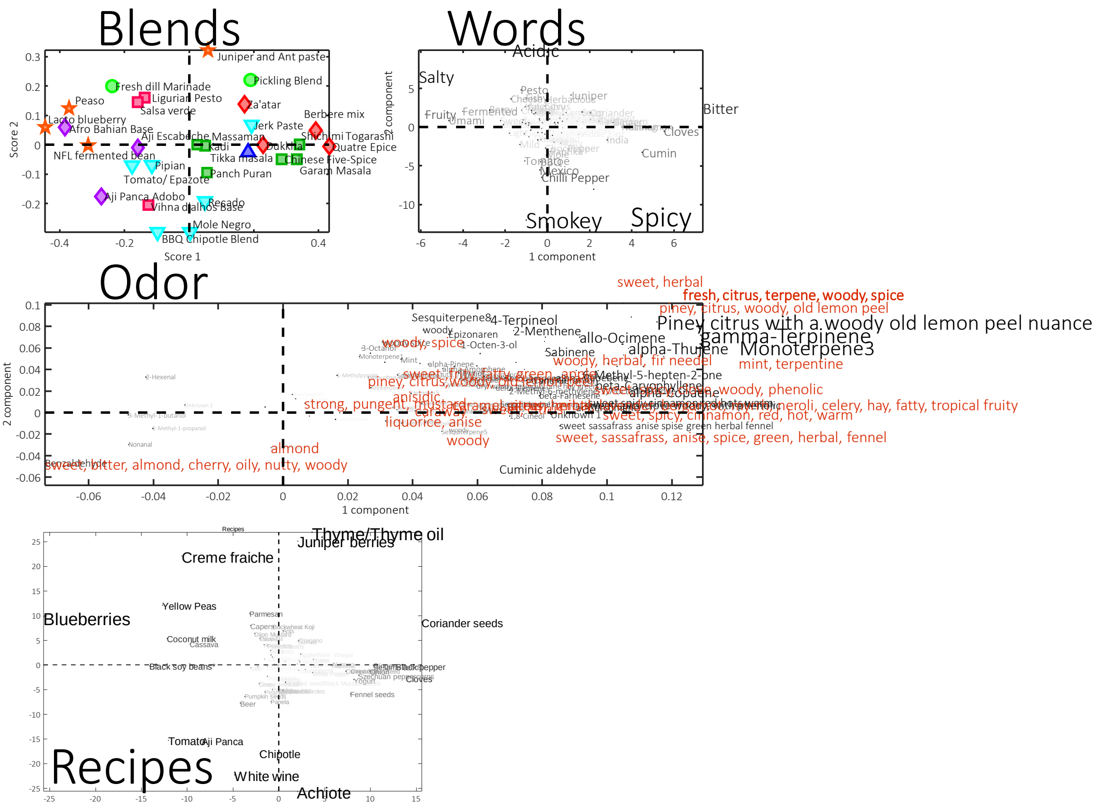

Figure 5. An overview of the first two components in the fused model. Legend: sample originated from Africa; Asia; Britain; Central America; Nordic Food Lab or noma Scandinavia; South America; Southern Europe.

Comparison of the two upper plots shows that the descriptors used by the assessors for the blends are in good agreement with their aroma compositions and that this information is well captured by the model. Indeed, the lowest part of the word-loading plot is characterized by attributes like spicy and smoky and, to some extent (going towards the origin of the plot), by the words chili, Mexico, BBQ, tomato, and mole. In the corresponding region of the score plot, we can observe the blends BBQ chipotle and mole negro that perfectly match with the previous description.

Similarly, the leftmost part of the word-loading plot is described by attributes such as fruity, berry, umami and fermented. Among the blends located in the leftmost part of the scores plot, the most extreme in this direction is lacto blueberry blend, constituted by fermented blueberry with sea salt. In the same direction we can find also the other two fermented blends (peaso and NFL fermented bean).

In the upper part of the scores plot we have the blends ligurian pesto, salsa verde, and fresh dill marinade on one side, and juniper and ant paste together with pickling blend on the other side. The direction of the first group is characterized by words such as salty, pesto, and cheese. However, these blends are also characterized by higher value of the score of the second principal component, with respect to all the others, so that they are also represented by the words in the top of the plot, in particular acidic. Among the ingredients of these blends we can find lemon and ingredients with pungent characteristics, such as mustard. The second group of samples is characterized by an acidic taste together with a juniper nuance, which is in perfect agreement with the composition of these blends, involving in both cases the presence of juniper berries and with an acidic ingredient such as lemon thyme or vinegar.

In the rightmost part of the score plot a cluster of five blends, shichimi togarashi, quatre epice, berbere mix, garam masala and Chinese five-spice are present. Although the geographical origin of these blends is very diverse, they are all mixes of dried spices. In the corresponding part of the word loading plot we can observe that they are associated to the spices cloves and cumin and to the attribute bitter.

The presence in this study of the aroma analysis using gas chromatography allows us to dig further into the data. In particular, observation of the middle plot shows a loading plot of the volatile odor compounds. On the same plot on top of the odor chemicals, the odor descriptors are shown. The plot shows how most of the odors are gathered in the upper-right part of the plot. This can be interpreted as a general trend of increased intensity of aroma, going from the samples in the bottom-left part of the scores plot towards those in the upper-right part. One can observe at first, all the fermented pastes, followed by the dairy, oil, and liquid samples and finally by the dried samples (Figure 6). A possible explanation of this trend along the first two components could be associated with the type of base of the blends that allows a slightly different release of volatiles during headspace sampling. This could sound like an analytical problem – but it is not. Release of volatiles is also what happens when eating, and the headspace setup mimics these conditions. Therefore, we are not measuring what is in the food, but what is released when it is consumed. The dry base consists mainly of mixes of pure spices that have been ground up, which contain a high concentration of many highly volatile compounds, e.g. terpenes. The fermented mixes will undoubtedly have a considerable content of acids and umami compounds which are not captured by the GC analysis, but will contribute to the taste of the mixes. The oil-based mixes appear to contain a medium level of aroma compounds, but this may be due to higher aroma retention in the oil under the conditions for trapping the aroma compounds. The conditions were set to mimic in-mouth release of volatile compounds, where the majority of highly fat soluble aroma compounds will not be released from the food matrix [15].

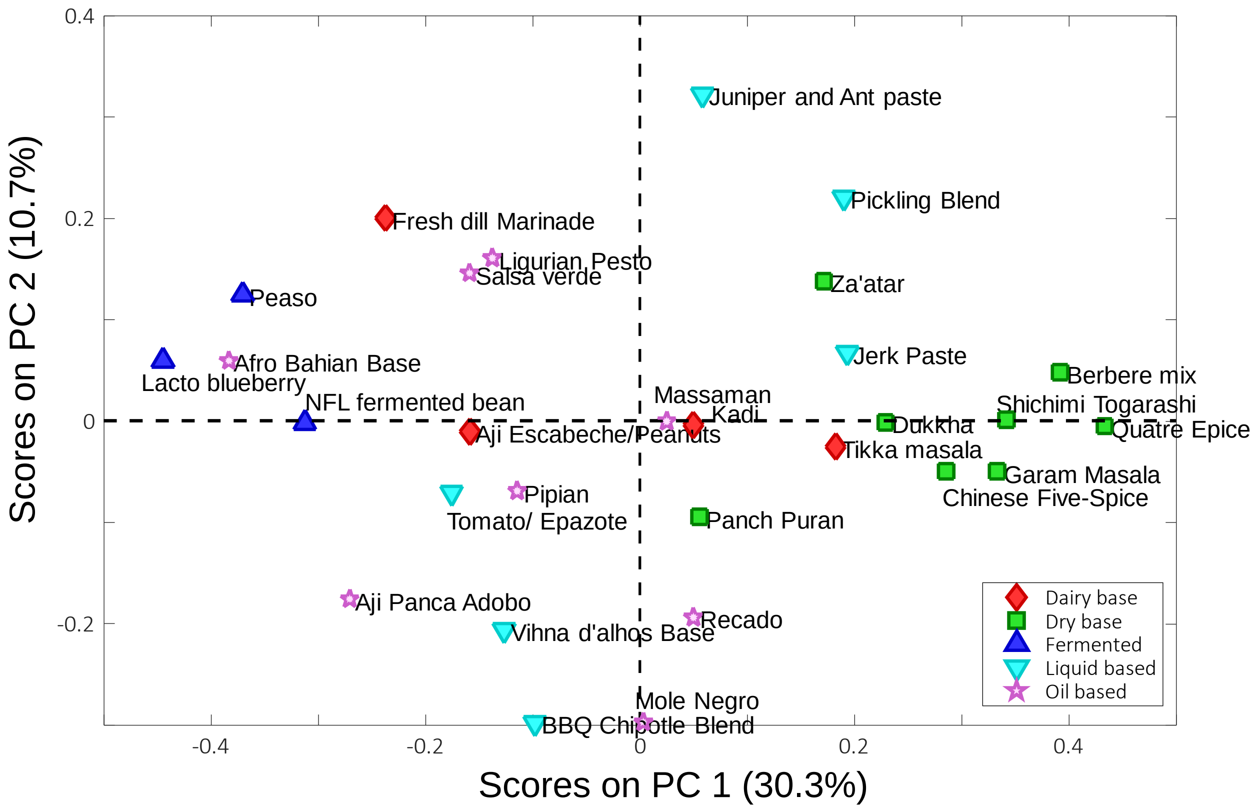

Figure 6. Score plot of the first two components. Similar to upper left plot in Figure 5, but here the samples are colored by their flavor systems ('base mix characteristics').

Finally, the recipe plot shows how the paste grouping can be interpreted on the basis of their ingredients. One can see how the ingredients naturally reflect the already mentioned groupings such as blueberry in the left part, juniper in the upper, and chipotle in the lower part.

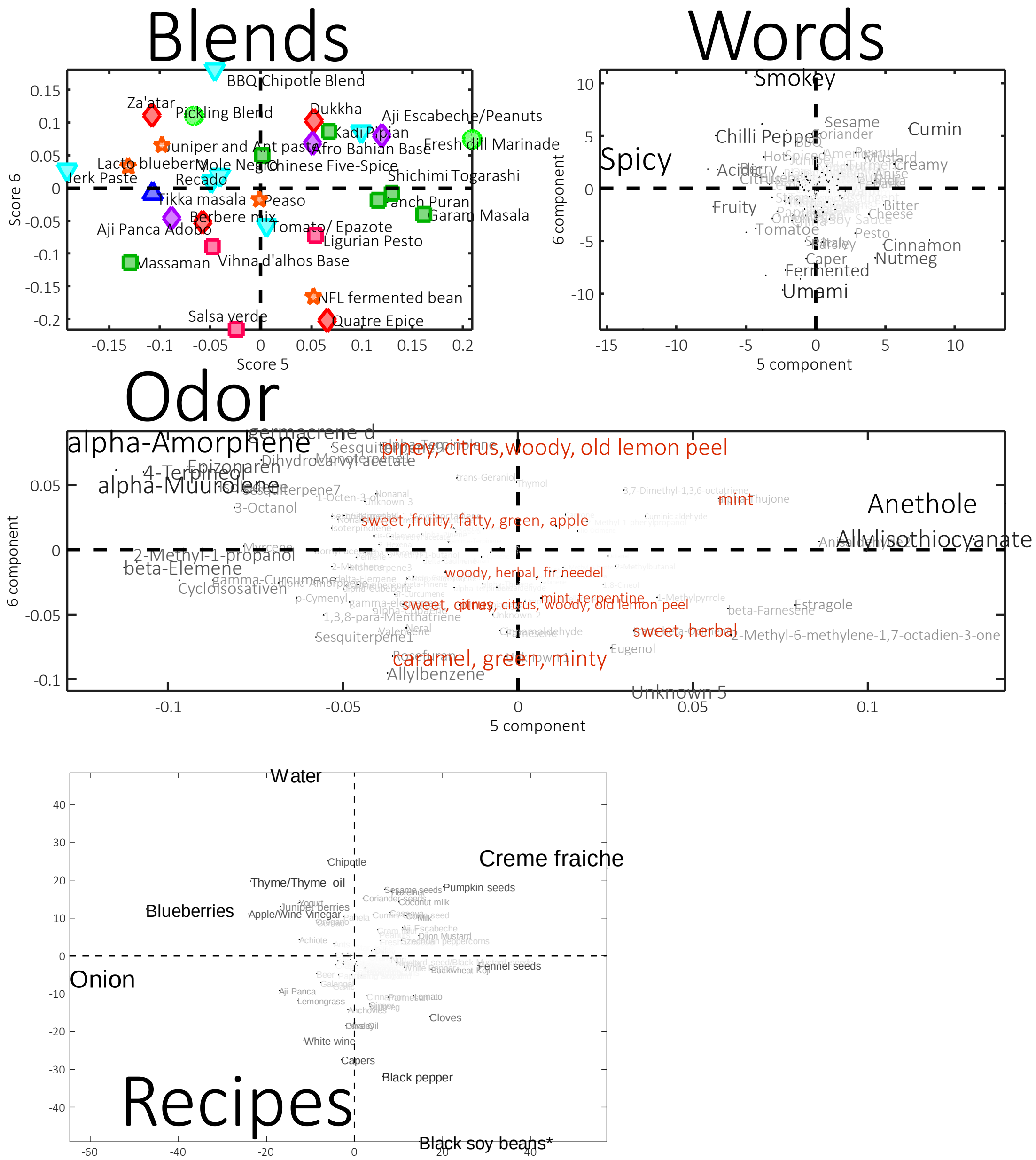

Figure 7. An overview of the component five and six in the fused model. Legend: Africa; Asia; Britain; Central America; Nordic Food Lab Scandinavia; South America; Southern Europe.

Since we are dealing with a complex set of data, there are several components of the PCA model and, as an example, we show here one of the other two-dimensional plots generated from the

PCA model where the contribution of the different original sets of data is more evident. Figure 7 shows the plots related to component five and six. It is shown how the perceived properties of the blends are divided quite clearly along the fifth component. The words characterizing the right side of the pseudo-loadings plot are mostly related to specific spices like cloves, cumin, nutmeg, cinnamon, and sesame, together with some attributes such as creamy and above all nutty. This is in perfect agreement with the recipe pseudo-loading plot where, in the upper-right side, we find ingredients such as pumpkin seeds, hazel nuts, sesame seeds and coriander seeds. The position of both the descriptors and ingredients agrees with the corresponding score plot of the blends, where the right most side is mostly characterized by the presence of dry blends such as shichimi togarashi, garam masala and panch puran, all coming from Asia and among whose main ingredients are the ones mentioned. In the same region of the plot we can find fresh dill marinade and aji escabeche/peanuts coming from completely different geographical origin and characterized by a dairy base. The presence in these blends of cheese and crème fraiche together with peanuts (in the case of aji escabeche/peanuts) could be the reason for the presence of words like creamy and nutty.

The leftmost part of the word loading plot is characterized by taste descriptors such as spicy, salty, and chili pepper. In the corresponding region of the score plot one can find the hot blends such as jerk paste, aji panca adobo, massaman and BBQ chipotle blend (notwithstanding the fact that this last one has been described very often as smokey and spicy and for this reason is located in a slightly different region of the space, having a higher value of score on the sixth component).

In the same region, more delicate blends are also located, such as lacto blueberry and tikka masala. These are constituted by fermented blueberry and yogurt, garlic and tomato. This can explain the presence of word like fruity, sour, acidic, garlic and tomato in the left part of the pseudo-loadings plot. In the same part of the plot we find two blends that we already discussed in

Figure 5, juniper and ant paste and pickling blend. They are indeed closely related for the presence of juniper and the perceived characteristics of pine and resinous taste. This last characteristic is not visible in the word loadings plot, but the presence of further information given by the odor compounds compensates for this. Indeed, looking at the odor plot in this figure one can see, in the same direction of the two blends, the presence of some chemical compounds like germacrene D and different kind of sesquiterpenes that are closely linked to the aroma and the odor of woody, spice, piney. In the same region of the plot, we can find an important contribution given by alpha-amorphene (known as woody aroma), present in a series of different leaves such as bay, present in some of the blends in this corner of the plot. In the same way, in the extreme right of this third plot we find in particular the presence of two compounds, allyl-isothicyanate and anethole. The first one in particular is related to a strong and pungent taste typical of mustard, and probably causes the presence of fresh dill marinade in the extreme part of the score plot on the right. Among the descriptors used by assessors for this specific blend, we find acidic, nordic, dill, creamy, mustard and horseradish that share a strong presence of allyl-isothicyanate with mustard. The second compound mentioned is typical of the taste of licorice and anise, and is present in some of the spices used for preparing the blend in the Asiatic group (e.g., Chinese five-spice and panch puran) – star anise, and fennel seeds.

CONCLUSION

A comprehensive way to fuse sensory projective mapping data, gastronomic information (recipes), and aroma profiling (gas chromatographic data) has been developed here. The method also allows for inclusion of additional data if available. More efforts are needed to help minimize the barriers that result from this highly cross-disciplinary experiment. There is a need for the ability to communicate across many fields of expertise such as chemistry, flavor research, gastronomy, mathematics, and sensory science. Finding a common language that allows open and creative communication is of paramount importance for the advancement of the field.

An effort for improving the visualization of the problem has been executed through this research, enhancing the readability of the most important loadings in all the loadings plots. However, work is still required and in the future we intend to expand on this, taking advantage of modern tools that computer science offers, such as interactive visualizations that allow a deeper exploration of complex data for which static plots are sometimes inadequate.

Reference List

1. Blumenthal, H. Weird but wonderful. The Guardian . 2002. 4-5-2002.

2. Barham P., Skibsted LH, Bredie WLP, Frøst MB, Møller P, Risbo J, Snitkjær P, Mortensen LM. Molecular gastronomy: a new emerging scientific discipline. Chemical Reviews 2010; 110

3. Yeomans MR. Taste, palatability and the control of appetite. Proceedings of the Nutrition Society 1998; 57:609-615

4. Shepherd GM. Smell images and the flavour system in the human brain. Nature 2006; 444:316-321

5. Rozin E, Rozin P. Culinary Themes and Variations. Natural History 1981; 90:6-14

6. Forshed J, Idborg H, Jacobsson SP. Evaluation of different techniques for data fusion of LC/MS and H-1-NMR. Chemom.Intell.Lab.Syst. 2007; 85:102-109

7. Acar E, Rasmussen MA, Savorani F, Næs T, Bro R. Understanding data fusion within the framework of coupled matrix and tensor factorizations. Chemom.Intell.Lab.Syst. 2013; 129:53-63

8. Ahn Y, Ahnert SE, Bagrow JP, Barabási A. Flavor network and the principles of food pairing. Scientific Reports 2011; 1:1-7

9. Rozin E. The Role of Flavor in the Meal and the Culture. In: Meiselman HL ed. Dimensions of the meal : The science, culture, business, and art of eating. Aspen Publishers, Inc. 2000: 134-142

10. Risvik E, McEwan JA, Colwill JS, Rogers R, Lyon DH. Projective mapping: a tool for sensory analysis and consumer research. Food Quality and Preference 1994; 5:263-269

11. Frøst MB, Giacalone D, Rasmussen KK. Alternative methods of sensory testing: working with chefs, culinary professionals and brew masters. In: Delarue J, Ben Lawlor J, Rogeaux M eds. Rapid Sensory Profiling Techniques and Related Methods - Applications in New Product Development and Consumer Research. Cambridge: Woodhead Publishing. 2015: 363-382

12. Escofier B, Pages J. Multiple factor analysis (AFMULT package). Computational Statistics and Data Analysis 1994; 18:121-140

13. Delholm C, Brockhoff PB, Meinert L, Aaslyng MD, Bredie WLP. Rapid descriptive sensory methods – Comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. Food Quality and Preference 2012; 26:267-277

14. Bro R, Smilde AK. Principal Component Analysis. Analytical Methods 2014; 6:2812-2831

15. Taylor AJ, Linforth RST. Flavour release in the mouth. Trends in Food Science & Technology 1996; 7:444-448

Appendix 1

Sensory descriptor groups:

Single Ingredient (1)	Single Ingredient (2)	Geography/culture
Cumin	Carrots	India
Chilli Pepper + Hot + picado	Corn	Mexico
Cloves	Caraway	Italy
Ginger	Cloves	Asia
Garlic	Ants	Middle East
Pepper+black pepper+white pepper	Mussels	Thailand
Coriander	Mushroom	Southeast Asia
Tomato	Cauliflower	Japan
Sesame	Rose	China
Lemon	Star Anise	American
Cinnamon	Mayo	France
Nutmeg	Tahini	South America
Turmeric	Apple	Mediterranean
Cardamom	Cayenne	Africa
Anise	Lavender	Spain
Paprika	Dill	Morocco + monoco
Juniper	Allspice	Korea
Peanut	Vanilla	Arabic
Onion	Potato	Ethiopia
Bean	Wasabi	Israel
Mint	Lettuce	Colombia
Mustard	Young Pandan	Caribbean
Blueberry	Sumac	Korean
Caper	Rapeseedy	Dutch
Oregano	Carob	Muslim
Parsley	Tobacco	Malaysia
Basil	Olive Oil + Olive	Singapore
Cucumber	Mace	East Asia
Lemongrass	Fenugreek	Southeast USA
Licorice	Spinach	Nepal
Fennel	Black Rasberry	Central America
Tamarind	Weed	Istanbul
Thyme	Lingonberry Sauce	Russia
Coconut	Zucchini	**Taste descriptor**
Pine	Rhubarb	Acidic + sour + sharp
Horse Raddish	Cherry	Sweet
Egg	Celery	Umami
Pineapple	Cilantro	Astringent
Almond	Pumpkin Puree + squash	Greasy
Peas	Caramel	Soap
Turkey	Asparagus	Bittersweet
Coconut Milk	Watermelon	Alcoholic
Rice	Hickory	Metallic
Pasta	Bamboo	Pungent + punchy
Leek	Tarragon	Spicy + spiced
Barley + Grain	Black currant	Salty
Lamb		Bitter
Legume		

Complex Compound ingredients (1)	Complex Compound ingredients (2)	Other Descriptors (2)
Anchovy + fishy	bruschetta	Murky
Vinegar	Cake	Breakfast
Chocolate + cocoa	Crispy	Puke flavor
Seaweed	Ginger Ale	Hospital + medicinal + dental clinic
Miso + Soy Sauce + Soy	Toothpaste	Sauna
Zatar	Enchilada	Tasty + Yummy + nice
Taco +Tortilla	Goulash	Anesthetic
Tea	Stirfry	Vibrant
Wine	Popcorn	Rich
Falafel	Nigella	Warm
Pizza	Hay	Unfamiliar
Laksa	Cookie	Creamy
Oyster Sauce	Rubber	Floral
Black Bean Sauce	Remoulade	Roasted + toasted
Yeast + bready	Campbell Soup	Balanced + round
Hummus	Sour Cream	Complex
Sea + Sushi	Citrus + lime + zesty	Bright
Satay	Curry	Cool + refreshing
Empanada	Cheese	Soft
Roti	Pesto	Healthy
Paratha	Berry	Mellow
Chanar	Meaty	Artificial
Chutney + spicy pickle	Mole	Familiar
Hamburger	Salsa	Low/No aroma + muted
French Fries	Butter	Smooth
Noodles	**Other Descriptors (1)**	Plain and Flat
Chilequiles	Fresh + clean + nordic	Seasoning
Koji	Strong	Desserts
Nectar	Herbacious	Farmhousy + animal
Cherry Bread	Fermented	Fire
Chorizo	Bark	Baking
Nougat	Smokey	Mineral
Blackberry Cobbler	mild + light + Low heat	Resin
Gazpacho	Earthy	Winter + Christmasy
Fish Sauce	Canned	Long + Lingering + long lasting
Muesli	Frosting	Delicate
Broth	Industry	Pickled
Meat Vegetable Stew	Toxic	Interesting
Fruity	Urinal Cakes	Paint
Greens + vegetal	Rain	Pet Store
Lactic + dairy + fermented dairy	Perfumed	Unpleasant
BBQ	Limestone	Surprising
Insect	Market	Heavy
Finger Food + picnic	Stale + Not Fresh + old	Simmery
Color Pepper + red spice	Classy	
Jammy	Crumbly	
Humita	Chalky + powdery + dusty	
Romesco	Subtle	
Nutty	Husked	
Rice Stew	Mature	
Seedy	Chewy	
Candy		

Appendix 2

List of identified compounds in aroma profiling with GC-MS

Carbondisulfide	3-Octanol	alpha-Terpinolene
Tetrahydrofuran	Rosefuran	Dihydrocarvyl acetate
Ethyl acetate	Allylbenzene	1,5-Dimethyl-1,5-cyclooctadiene
2-Methylbutanal	Monoterpene2	germacrene d
3-Methylbutanal	Perillene	Epizonaren
Methyl 2-methylbutanoate	1,3,8-para-Menthatriene	Zingiberene
alpha-Pinene	Nonanal	alpha-Muurolene
alpha-Thujene	p-Cymenyl	Valencene
Ethyl butanoate	Acetic acid	beta-Farnesene
Camphene	Unknownsulfur1	gamma-Muurolene
Hexanal	1-Octen-3-ol	Carvone
2-Butenenitrile	alpha-Cubebene	Farnesene
beta-Pinene	Furfural	Unknown_sulfur1
2-Methyl-1-propanol	delta-Elemene	delta-Cadinene
Sabinene	Sesquiterpene1	Sesquiterpene8
2-Menthene	Sesquiterpene2	ar-Curcumene
1-Methylpyrrole	Cycloisosativen	Cuminic aldehyde
3-Carene	alpha-Copaene	Sesquiterpene9
2-Butenenitrile	camphor	Nerol
Monoterpene1	Benzaldehyde	2-Methyl-1-phenylpropanol
l-Phellandrene	Sesquiterpene3	cis-Calamene
Myrcene	Monoterpene3	Anethole
alpha-terpinene	Camphene	trans-Geraniol
Heptanal	Isoledene	Carvacryl acetate
Limonene	Bornyl acetate	Unknown 1
beta-Phellandrene	beta-Elemene	Unknown 2
1,8-Cineol	beta-Caryophyllene	alpha-Calacorene
3-Methyl-1-butanol	4-Terpineol	Calamene
3,7-Dimethyl-1,3,6-octatriene	Sesquiterpene4	Unknown 3
2-Hexenal	gamma-elemene	Anisaldehyde2
gamma-Terpinene	Alloaromadendrene	Cinnamaldehyde
trans-beta-Ocimene	Sesquiterpene5	Eugenol
p-Cymene	Sesquiterpene6	Thymol
Methylallyldisulfide	alpha-Amorphene	Unknown 4
alpha-terpinolene	Sesquiterpene7	Carvacrol
Isoterpinolene	Estragole	Unknown 5

Octanal	alpha-Humulene	Unknown 6
2-Methyl-6-methylene-1,7-octadien-3-one	Unknownbenzene	Unknown 7
4-Methyl-1-pentanol ISTD	Neral	
6-Methyl-5-hepten-2-one	alpha-Amorphene	
Allylisothiocyanate	gamma-Curcumene	

Appendix 3. Additional metadata.

BLEND	BASE	GEOGRAPHICAL ORIGIN	INGREDIENTS	
BBQ Chipotle Blend	Liquid based	Southern US	Water	37
			Chipotle	36
			Panela	15
			Apple Vinegar	12
Jerk Paste	Liquid based	Jamaica	Onion	58
			Apple Vinegar	29
			Scothbonnet	6
			Allspice	3
			Black pepper	2
			Cinnamon	1
			Nutmeg	1
Juniper and Ant paste	Liquid based	Noma	Juniper berries	43.1
			Thyme oil	40.9
			Ants	12.9
			Verbena	2.1
			Lemon Thyme*	0.6
			Woodruff*	0.2
Tomato/ Epazote	Liquid based	Guatemala-Mexico	Tomato	97
			Epazote	2
			Piquin Chili	1
Pickling Blend	Liquid based	Scandinavia	Water	53
			Corriander seeds	27
			White wine vinegar	13.4
			Juniper berries	5.3
			Bay leaf	0.4
Vihna d'alhos Base	Liquid based	Portugal	White wine	65
			Onion	18
			Paprika	14
			Garlic	3
Afro Bahian Base	Oil based	Brazil	Coconut milk	54
			Cassava	43
			Green peppers	2
			Cilantro	1
Aji Panca Adobo	Oil based	Ecuador- Peru	Aji Panca	52.6
			Beer	26.3
			Grape seed oil	13.1
			Soy sauce	7.9

BLEND	BASE	GEOGRAPHICAL ORIGIN	INGREDIENTS	
Ligurian Pesto	Oil based	Italy	Parmesan	39
			Pine nuts	15
			French beans	15
			Basil	11
			Potatoes	11
			Garlic	5
Massaman	Oil based	Thailand	Onion	27
			Lemongrass	27
			Galangal	16
			Garlic	11
			Red chilies	5
			Shrimp paste	5
			Coriander seeds	3
			Cumin	2
			Peppercorns	2
			Cloves	1
			Cardamom	1
Mole Negro	Oil based	Mexico	Guajillo Chili	10.5
			Mulato	10.5
			Pasilla chili	10.5
			Chile Chipotle Mora	2
			Sesame seeds	10.5
			Peanuts	10.5
			Almonds	10.5
			Walnuts	5.5
			Pecans	5.5
			Raisins	10.5
			Semi sweet roles	10.5
			Cinnamon	0.9
			Black Pepper	0.3
			Cloves	0.3
			Cumin	0.3
			Thyme	0.3
			Mexican oregano	0.3
			Bay leaf	0.3
Pipian	Oil based	Mexico	Onion	27
			Lemongrass	27
			Galangal	16
			Garlic	11
			Red chilies	5

			Shrimp paste	5
			Coriander seeds	3
			Cumin	2
			Peppercorns	2
			Cloves	1
			Cardamom	1

BLEND	**BASE**	**GEOGRAPHICAL ORIGIN**	**INGREDIENTS**	
Recado	Oil based	Central America	Achiote (Annato)	78
			Masa	16
			Cumin	4
			Black Pepper	2
Salsa verde	Oil based	Spain	Capers	33
			Flat leaf parsley	22
			Olive Oil	22
			Anchovies	17
			Lemon and rind	6
Berbere mix	Dry base	Ethiopia	Corriander seeds	17
			Cloves	11
			Fenugrek	11
			Black Pepper	9
			Cayenne	9
			Ginger	9
			Allspice	8
			Cumin	8
			Cardamom	6
			Cinnamon	6
			Nutmeg	6
Dukkha	Dry base	Egypt	Hazelnuts	43
			Cumin	22
			Sesame	22
			Coriander seeds	22
			Black pepper	2
Quatre Epice	Dry base	France, North Africa	Black pepper	40
			Cinnamon	15
			Clove	15
			Ginger	15
			Nutmeg	15
Za'atar	Dry base	North African, Middle East	Oregano	22
			Thyme	22
			Sumac	21
			Sesame seeds	21
			Cumin	14
Fresh dill Marinade	Dairy base	Scandinavia	Creme fraiche	67
			Dijon Mustard	19
			Dill	6

			Honey	5
			Fennel seeds	3

BLEND	BASE	GEOGRAPHICAL ORIGIN	INGREDIENTS	
Kadi	Dairy base	India	Cow's milk yogurt	49
			Onion	23
			Gram flour	21
			Ginger	5
			Cumin	1
			Mustard seeds	1
			turmeric powder	0.9
Aji Escabeche/Peanuts	Dairy base	Peru	Aji Escabeche	24.6
			Huacatay	0.1
			Peanuts	16.4
			Milk	32.8
			Creme fraiche	13.1
			Fresh cheese	13.1
Tikka masala	Dairy base	Britain	Cow's milk yogurt	63
			Onion	34
			Ginger	15
			Garlic	11
			Paprika	3
			Coriander seeds	2
			Cumin	2
			turmeric powder	2
			Cayenne	2
			Tomato	2
NFL fermented bean	Fermented	Nordic Food Lab	Black soy beans*	76.7
			Icing sugar	7.67
			Food Molasses	2.3
			White wine vinegar	9.4
			Black Garlic	3.8
Lacto blueberry	Fermented	Noma	Blueberries	98
			Sea Salt	2
Peaso	Fermented	Nordic Food Lab	Yellow Peas	58.2
			Buckwheat Koji	36.9
			Sea salt	4.8
Shichimi Togarashi	Dry base	Japan	Szechuan peppercorns	21.9
			Buckwheat Koji	36.9
			Orange	7

			Garlic	8.8
			Nori	17.5
			Red chili peppers	5.3
			Ginger	13.2
			Sesame seeds	26.3

BLEND	BASE	GEOGRAPHICAL ORIGIN	INGREDIENTS	
Chinese Five-Spice	Dry base	China	Star anise	12.8
			Szechuan peppercorns	25.6
			Fennel seed	12.8
			Cloves	12.8
			Coriander seed	12.8
			Cinnamon	23.1
Panch Puran	Dry base	East India	Fenugreek seeds	11.1
			Nigella	22.2
			Cumin seeds	22.2
			Black mustard seeds	22.2
			Fennel seeds	22.2
Garam Masala	Dry base	Pakistan	White pepper	18.9
			Cinnamon	2.1
			Clove	27.4
			Cardamom	3.2
			Nutmeg	3.2
			Fennel seeds	27.4
			Coriander seeds	16.8
			Bay leaf	1 leaf

LIST OF ABBREVIATIONS

BBQ	Barbecue
NFL	Nordic Food Lab
GC-MS	Gas chromatograph-mass spectrometer

Psi	Avoirdupois pound-force per Square Inch
MFA	Multiple Factor Analysis
PCA	Principal Component Analysis
IUPAC	International Union of Pure and Applied Chemistry

Acknowledgements: We kindly acknowledge the 26 persons that gave their valuable time to evaluate the spice mixes. In addition, we acknowledge restaurant noma's employee Lars Williams for the donation of Ants and juniper mix. We acknowledge Kiki Sontiyart**,** Ana Caballero and Guillemette Barthouil for their contribution to the creation of the spice mixes.