DOI 10.5162/16dss2022/P49

From Whisky to Aroma
Investigating mixture data for odor prediction

S. Singh[1], S. Saloman[2*], T. Scholz[2], H. Haug[1,3], T. Gorges[2,4], A. Strube[1], B. Saha[2], A. Grasskamp[1†], S. Hettenkofer[2†]*

[1]Sensory Analytics and Technology, Fraunhofer Institute for Process Engineering and Packaging IVV, Freising, Germany
[2]Digital Sensory Perception, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
[3]Chair of Aroma and Smell Research, Department of Chemistry and Pharmacy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[4]Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
Contact: andreas.grasskamp@ivv.fraunhofer.de & sebastian.hettenkofer@iis.fraunhofer.de

Introduction

Hedonic impression of food products largely depends on consistency, taste, and smell. While process control over other factors is quite feasible, perception of smell is not something easily predictable or measurable as volatile organic compounds (VOCs) are immensely diverse in odorant quality and detection threshold. Furthermore, investigating key aroma compounds in food products is costly in time and effort. Therefore, it is desirable to develop a system for the efficient and reliable interpretation of the decisive aroma components that influence consumer satisfaction. Recent years have seen considerable progress in computer aided analysis of molecules for different purposes. To avail ourselves of this progress, we utilized machine learning methods to predict the aroma qualities of whisky spirits, which are classically valued for their diverse aroma profiles depending on process parameters like aging duration, cask origin and blending.

Materials and Methods

Description of sensory and analytical data

As described in Haug et al. [1] 16 different commercial whiskies were chosen, including 9 Scotch and 7 American whiskies. Analytical and sensory data obtained by the studies described in Haug et al. were used for the present research. In short, the following procedures were conducted [1]: The alcohol content of these samples was adjusted to 40% vol., additionally whiskies were diluted to 20% vol. to gain insight into their aroma profiles. The 16 samples at 40% vol. were chemically analyzed using Stir-Bar-Sorptive-Extraction [2] followed by thermal desorption gas chromatography-mass spectrometry (GC-MS) and then an in-house semi-automated software processed the GC-MS data to create a list of molecules identified in each sample along with their SMILES [3] strings and the confidence of the match [4], [5]. It must be noted that the data on identified analytes used for the present study slightly differs from the data used in Haug et al. [1]. The reason for this is the use of an older version of the database for analyte identification as well as a stricter rejection threshold of 0.9. For these whisky samples, 818 compounds were detected in total, 102 of which were unique.

Furthermore, samples were analyzed sensorially [1]. For this purpose, 11 members of the IVV sensory panel were provided with said samples in either 20% vol. or 40% vol. (in total 32 samples relevant for the present study; however, whiskies with an original alcohol content higher than 40% vol. were also investigated, but are not taken into account) and asked to rate the 5 most applicable descriptors from the

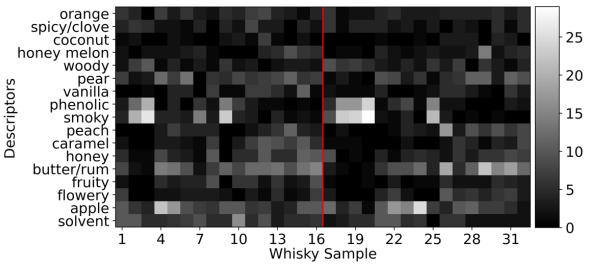

Fig. 1: Distribution of descriptors across all whisky samples. The sample on the left of the red line contain 40% vol. alcohol, and on the right are samples diluted to 20% vol.

given descriptors by Rate All That Apply (RATA [6], [7]) in the range of 1-3 in terms of their applicability through orthonasal assessment . These RATA values were later summed up for each descriptor across different panelists and an aggregated RATA label set was created where the applicability ranged from a low of 0 to the highest applicability of 29. The distribution of the descriptors across different whisky samples is shown in Figure 1. Upon visual inspection, descriptors such as '*smoky, 'phenolic'* and *'apple*' occur more often than others and descriptors '*smoky'* and *'phenolic'* seem to be coexistent.

Encoding of mixture proportions

In a related work a neural network for odor prediction based on mono-molecular data was trained with multiple datasets and validated on the DREAM Olfaction Prediction Challenge data [8]. Its feature space, capturing the relation

between odors and molecular properties was added to the whisky data by adding all corresponding features to each molecule contained in a whisky. For each molecule contained in a whisky the features were weighted by the normalized peak area of the GC-MS measurements to encode the mixture proportions as shown in Figure 2.

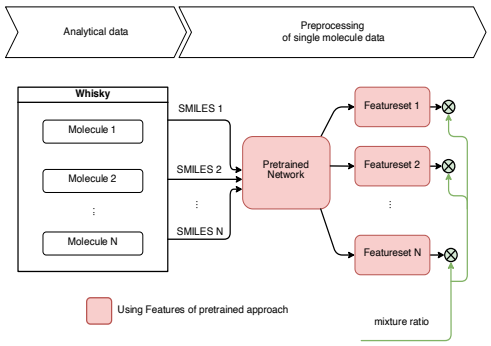

Fig. 2: Encoding of the single molecular properties with a pretrained network to generate meaningful feature sets

Classification Targets

For the classification of the aroma profile of the whiskies, a new dataset was created by binarizing the 4 highest rated scents per whisky. In case of a tie in the top 4, the scent which was rated more frequently was used.

Modeling a mixture problem

Two models were trained on all possible 0.8/0.2 train-test splits: a recurrent neural network (RNN) and a random forest (RF), shown in Figure 3. The RNN was used due to the structure of the problem, that each whiskey has a different number of molecules. The RNN architecture allows a variable input length thus no padding is needed. Additionally, the network was allowed to transport the whole feature sequence through the network for a deeper analysis. The number of layers was set to four, with three leading recurrent layers with 16 cells each and a trailing Dense layer to map the human sensory data. This keeps the number of total trainable parameter low, so that the network can train only with a few samples. As metric the Top-K accuracy is used with a k of 4 (Equation 1)

$$\text{Metric} = \sum_{i \in Topk} \delta\left(y_i^{pred}, y_i\right)/k \qquad (1)$$

where *Top-K* is the set of indices corresponding to the 4 highest rated scents per whisky and $\delta\left(y_i^{pred}, y_i\right) = 1$, if $y_i^{pred} = y_i$ and zero otherwise.

Data preparation and feature extraction

To vectorize the SMILES strings obtained after processing the GC-MS data for each sample, an initial literature search was performed to obtain a list of 390 substances commonly found in whisky samples [9–17]. The SMILES strings of these molecules served as a general dataset corpus. Features were extracted from these SMILES by performing a pairwise search to find the maximum common substructures. These extracted features consist of SMARTS [18] strings that denote the overlaps between substructures within a pair of molecules. These features were then vectorized by counting the occurrences of each of these substructure features in the 102 unique SMILES obtained from our whisky samples.

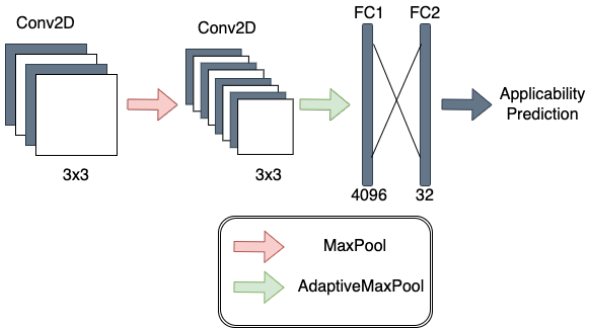

Fig. 4: Convolutional Neural Network architecture used to regress the descriptor applicability values

The final feature matrix obtained is of the shape (102, 3425) denoting 3425 extracted features for each of the 102 unique SMILES and served as the input for the neural network architecture chosen for regression.

Regression analysis

The regression problem tries to predict applicability ratings for reach descriptor associated with each whisky sample using a neural network architecture. For this purpose, a neural network architecture consisting of 2 convolutional

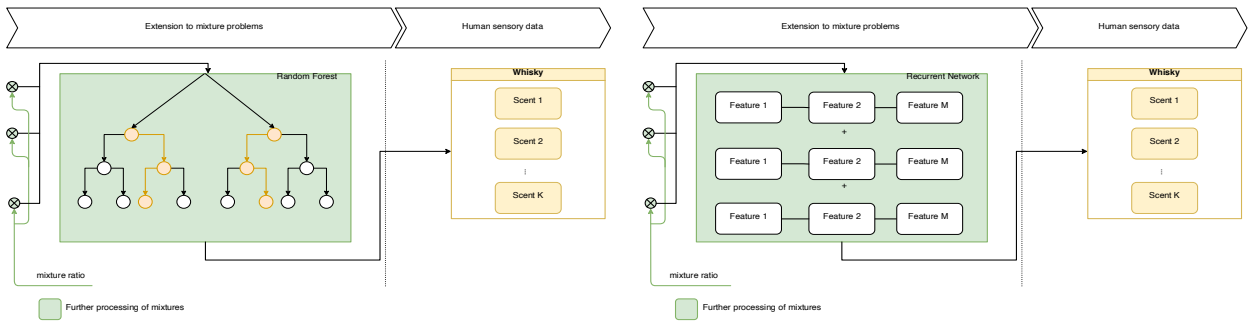

Fig. 3: Illustration of the mixed models

layers, adaptive pooling and two further fully connected layers was developed. The architecture is shown in Figure 4. Four-fold cross validation was used to train this network due to the limited amount of available data. For each fold, the training data was sampled from the feature matrix. For example, if sample 1 consisted of 50 identified substances, the input data was of the shape (50,3425) sampled from the global feature matrix and weighted by the normalized peak areas of those 50 substances.

The weighting of features using normalization of peak areas provided the neural network with the influence of the said substances to the overall aroma. A higher normalized peak area would indicate that a substance and its features would play a more important role than a substance present in scarcity.

Each fold was trained 100 epochs with a learning rate of 1e-3 and a rate decay of 0.1 after every 30 epochs. To reduce the effect of overfitting, a L2 penalty term of the order of 1e-6 was included in the Mean Square Error (MSE) loss function. The results of the training were evaluated using three metrics. These include the Mean Absolute Error (MAE) and R^2 score between the predicted and ground truth applicability, defined in Equation 2 and 3 respectively. Additionally, an overlap was calculated between the five highest predicted descriptors by the neural network and the five largest ground truth descriptors. These metrics were calculated across each fold of cross-validation and then averaged out.

$$MAE = \sum_i |y_i - x_i|/n \quad (2)$$

$$R^2 = 1 - \sum(y_i - \hat{y}_i)^2 \,/\, \sum(y_i - \bar{y}_i)^2 \quad (3)$$

Results and Discussion

Average rating of scents and whiskies

Analysis of the human sensory data shows that whiskies strongly differ regarding the perception of individual scents. Figure 5 shows the rating averaged over all whiskies and all panelists together with the standard deviation. For some whiskies, the standard deviation is very high (for example "smoky"), which indicates that the perception of this scent varies a lot within whiskies. Moreover, it can be seen that dilution plays an essential role in the perception of scents, however, no correlation can be observed.

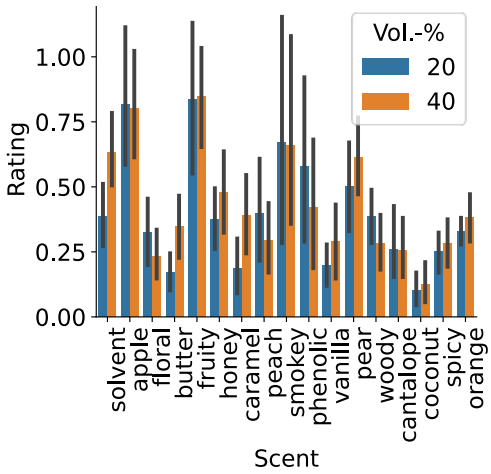

Fig. 5: Average Rating over all whiskies and panelists

Distinguishing between American whisky and scotch

Overall, aroma impressions such as caramel, vanilla or honey are more frequently observed in American whisky, a smoky scent is more frequently observed in Scotch (Figure 6). A close analysis of the molecular composition revealed that testing for 6-propyloxan-2-one (δ-octalactone) suffices to distinguish Scotches from American Whisky (Figure 7).

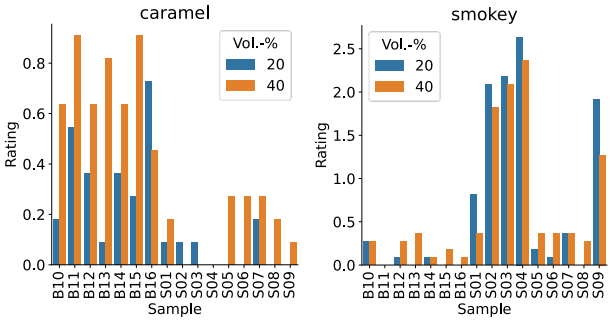

Fig. 6: Average rating of caramel and smoky over all panelists

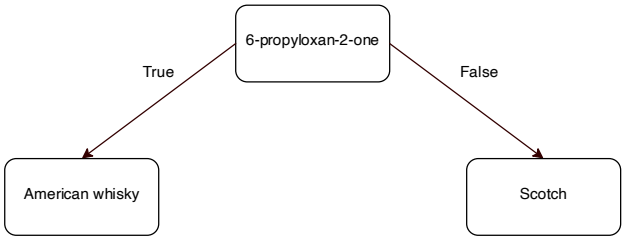

Fig 7: Decision tree distinguishing between American whisky and Scotch

Correlation between aroma profile and molecular composition

There are some whiskies in the data set that have an apparently dominant odor, which the majority of panelists rated as relevant. For further analysis, an odor was defined as "dominant" if it was rated by at least 60% of the testers (in at least one of the two dilutions). The resulting odors and whiskies are listed in Table 1.

The odors phenolic, smoky, fruity, apple, and vanilla have been identified as the dominant odor notes. Analysis of the whiskies exhibiting such a dominant odor revealed an overlap in the molecular composition: the apple-smelling whiskies have 67.8% of all molecules in common, the fruity ones 85.7%, the smoky ones 52.8% and the phenolic ones 69.2%.

Table 1: Mean and standard deviation of dominant odors

ID	Dilution	Scent	Mean	STD
B10	20	fruity	1.727	1.368
B13	20	fruity	2.000	1.296
B15	20	fruity	1.636	1.259
B15	40	vanilla	1.000	0.922
S02	20	smoky	2.090	0.950
S02	20	phenolic	1.545	1.286
S02	40	smoky	1.818	0.950
S03	20	smoky	2.181	1.037
S03	20	phenolic	1.545	1.221
S03	40	smoky	2.090	1.037
S04	20	smoky	2.636	0.912
S04	20	phenolic	2.181	1.069
S04	40	smoky	2.363	0.912
S04	40	phenolic	1.818	1.069
S05	40	apple	2.000	1.438
S06	20	apple	1.545	1.299
S06	40	apple	1.545	1.299
S07	20	apple	1.363	1.245
S08	20	apple	2.272	1.262
S09	20	smoky	1.909	1.181
S09	40	smoky	1.272	1.181

In addition, it was investigated whether whiskies with the same dominant odors had a similar composition in their molecules. This is shown in Figure 8 and Figure 9 demonstrating the examples of the scents fruity and apple respectively

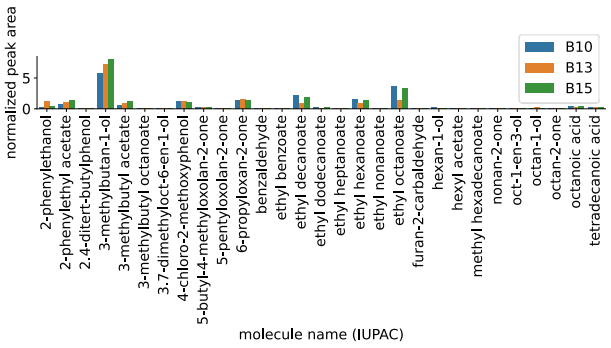

Fig. 8: The molecular composition for the Whiskies B10, B13 and B15 with the dominant scent fruity

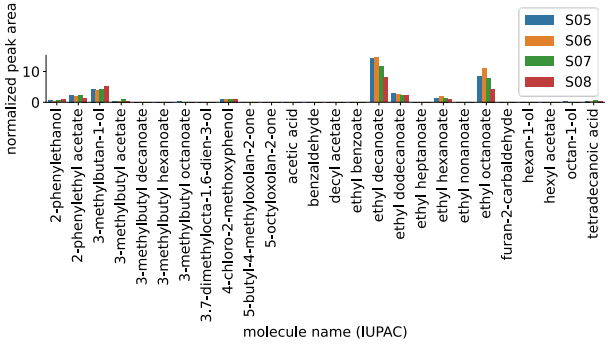

Fig. 9: The molecular composition for the Whiskies S05, S05, S07 and S08 with the dominant scent apple

Moreover, a vector was generated for each whisky containing the normalized peak area per molecule. These vectors were compared in pairs using cosine similarity and the minimum and maximum values were determined for each group (Table 2). The distance is overall high for each group (at least 0.9).

Table 2: The cosine similarity for different scents

Scent	Apple	fruity	smoky	phenolic
Max	0.995	0.978	0.989	0.984
Min	0.937	0.909	0.952	0.950

Classification results

The Results show that the train-test split strongly influences the model performance: Accuracies range between 25 and 68.75 % for the RNN and 12.5 and 68.75 for the RF (Figure 10). An analysis of the best and worst splits for each model was carried out as follows: For each whisky in the train set, its cosine similarity to each whisky in the test set was computed for both, the olfactory and molecular profile. Results are illustrated in Figure 11, where a light color indicates a high similarity and darker color a low similarity between the whiskies. It can be seen, that for the well-predicted whiskies in the test set, whiskies with a similar olfactory profile and molecular composition can be found in the training dataset. For the worst split, whiskies both similar in

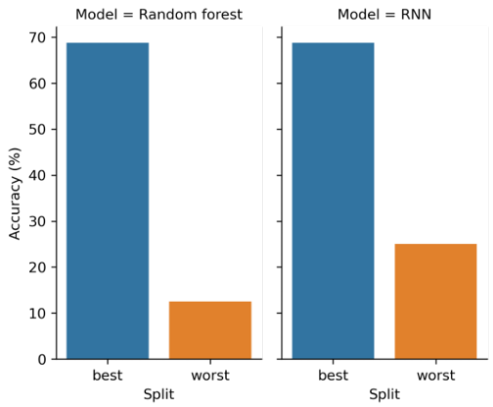

Fig. 10: Classification performance of recurrent neural network and random forest

scent and molecular composition were found to be either in the test or the training data.

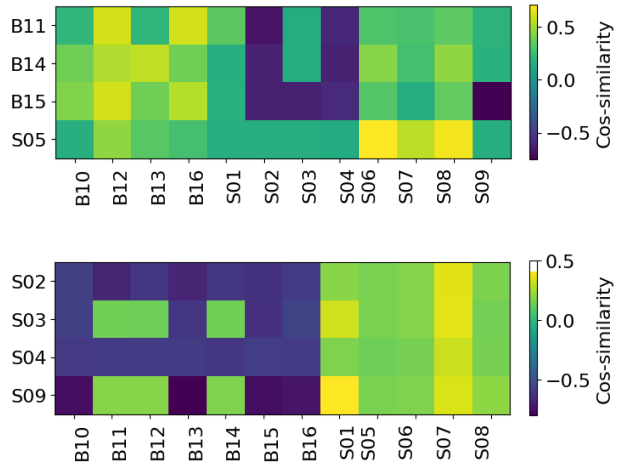

Fig. 11: Cosine similarity between the best (top) and worst (bottom) split for the random forest model

Regression results

The MAE and the R^2 metrics explained above were calculated for each fold and then averaged out. These are shown in Figure 12. As seen in the results for classification, the train-test split makes a significant influence on the ability of the neural network to regress the descriptor applicability

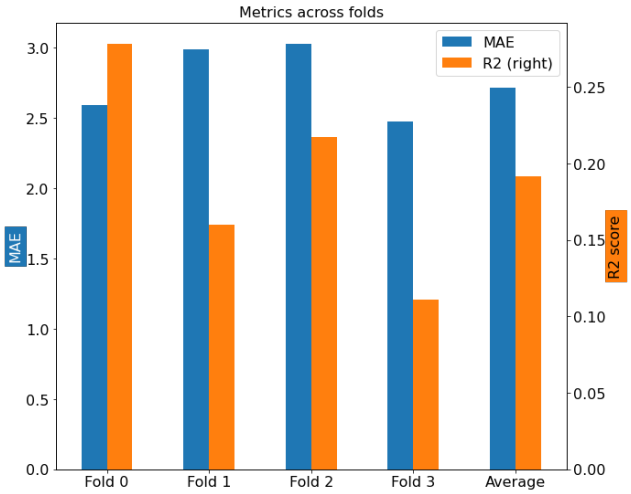

Fig. 12: Bar plot for metrics across folds

values. A 'favorable' split that consists of similarly dominant descriptors across the train and test set lead to better MAE and R^2 score, as seen in Folds 0 and Fold 2 compared to fold 1 and 3. On average, we achieve a R^2 score of 0.1917 and an average MAE of 2.7164.

Furthermore, for each whisky in each test set, an intersection overlap between the five largest actual descriptors and the five largest predicted descriptors was calculated to estimate the overall ability of the network to predict similar descriptor trends. For example, if a whisky is marked as having the five most dominant descriptors of *'smoky'*, *'phenolic'*, *'woody'*, *'apple'* and *'pear'* with predictions of *'phenolic'*, *'pear'*, *'woody'*, *'fruity'* and *'solvent'*, the intersection overlap would be of 3 descriptors that occur in both. The resulting overlaps are shown in Table 3. We achieve an overlap of 2.6875, i.e., we can, on average predict almost 3/5 of the largest descriptors correctly. This can be used to estimate how a hypothetical sensory panel would rate an unknown sample.

Table 3: Overlap of top 5 descriptors across different folds

Whisky	Overlap	Fold Index	Fold Avg.
Sample 1	4	0	3
Sample 2	2		
Sample 3	2		
Sample 4	4		
Sample 5	3	1	2.75
Sample 6	2		
Sample 7	2		
Sample 8	4		
Sample 9	4	2	2.75
Sample 10	4		
Sample 11	1		
Sample 12	2		
Sample 13	4	3	2.25
Sample 14	2		
Sample 15	1		
Sample 16	2		
Average	2.6875		

In conclusion, it can be shown that using the limited data that is available, neural network algorithms can be used to classify and regress meaningful information related to the aroma of a whisky sample. This can be in the form of a Top-K descriptor classification or to predict the intensity of the descriptors associated with such an unknown sample. While the whisky samples currently consist of not only a significant overlap in their composition but also an imbalance towards particular descriptors such as *'apple'*, *'smoky'* and *'phenolic'*, we hypothesize that this can be countered with a larger dataset of whisky samples that would allow capturing of more features that better distinguish the changes in perception of aroma across different dilutions and descriptors.

Acknowledgments

This work was funded as part of the Campus of the Senses by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (funding number: RMF-SG20-3410-2-14-3).

Literature

[1] Haug, Helen ; Grasskamp, Andreas T ; Singh, Satnam ; Strube, Andrea ; Sauerwald, Tilman: Quick insights into whisky – investigating rapid and efficient methods for sensory evaluation and chemical analysis (Submitted). In: *Analytical and Bioanalytical Chemistry* (2022)

[2] Baltussen, Erik ; Sandra, Pat ; David, Frank ; Cramers, Carel: Stir bar sorptive extraction (SBSE), a novel extraction technique for aqueous samples: Theory and principles. In: *Journal of Microcolumn Separations* Bd. 11, John Wiley & Sons, Ltd (1999), Nr. 10, S. 737–747

[3] Weininger, David: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. In: *Journal of Chemical Information and Computer Sciences* Bd. 28, American Chemical Society (1988), Nr. 1, S. 31–36

[4] Grasskamp, Andreas T ; Singh, Satnam ; Haug, Helen: Assisting the automated analysis of chemical-analytical measurements in spirits using validated algorithms and an intuitive user interface. In: *Sensoren und Messsysteme - 21. ITG/GMA-Fachtagung* (2022)

[5] Grasskamp, Andreas T ; Singh, Satnam ; Haug, Helen ; Sauerwald, Tilman: Assisting the automated analysis of chemical-analytical measurements in spirits using validated algorithms and an intuitive user interface (Submitted)

[6] Vidal, Leticia ; Ares, Gastón ; Hedderley, Duncan I. ; Meyners, Michael ; Jaeger, Sara R.: *Comparison of rate-all-that-apply (RATA) and check-all-that-apply (CATA) questions across seven consumer studies*. Bd. 67, 2018 — ISBN 5982922021

[7] Ares, Gastón ; Bruzzone, Fernanda ; Vidal, Leticia ; Cadena, Rafael Silva ; Giménez, Ana ; Pineau, Benedicte ; Hunter, Denise C ; Paisley, Amy G ; u. a.: Evaluation of a rating-based variant of check-all-that-apply questions: Rate-all-that-apply (RATA). In: *Food Quality and Preference* Bd. 36 (2014), S. 87–95

[8] Keller, Andreas ; Gerkin, Richard C. ; Guan, Yuanfang ; Dhurandhar, Amit ; Turu, Gabor ; Szalai, Bence ; Mainland, Joel D. ; Ihara, Yusuke ; u. a.: Predicting human olfactory perception from chemical features of odor molecules. In: *Science* Bd. 355 (2017), Nr. 6327, S. 820–826

[9] Jeleń, Henryk H ; Majcher, Małgorzata ; Szwengiel, Artur: Key odorants in peated malt whisky and its differentiation from other whisky types using profiling of flavor and volatile compounds. In: *LWT* Bd. 107 (2019), S. 56–63

[10] Demyttenaere, Jan ; Martínez, Jorge ; Verhé, Roland ; Sandra, Pat ; Kimpe, Norbert: Analysis of volatiles of malt whisky by solid-phase microextraction and stir bar sorptive extraction. In: *Journal of chromatography. A* Bd. 985 (2003), S. 221–232

[11] Poisson, Luigi ; Schieberle, Peter: Characterization of the Most Odor-Active Compounds in an American Bourbon Whisky by Application of the Aroma Extract Dilution Analysis. In: *Journal of Agricultural and Food Chemistry* Bd. 56, American Chemical Society (2008), Nr. 14, S. 5813–5819

[12] Daute, Martina ; Jack, Frances ; Baxter, Irene ; Harrison, Barry ; Grigor, John ; Walker, Graeme: Comparison of three approaches to assess the flavour characteristics of scotch whisky spirit. In: *Applied Sciences (Switzerland)* Bd. 11 (2021), Nr. 4, S. 1–27

[13] Câmara, J S ; Marques, J C ; Perestrelo, R M ; Rodrigues, F ; Oliveira, L ; Andrade, P ; Caldeira, M: Comparative study of the whisky aroma profile based on headspace solid phase microextraction using different fibre coatings. In: *Journal of Chromatography A* Bd. 1150 (2007), Nr. 1, S. 198–207

[14] Owens, Janel E. ; Zimmerman, Laura B. ; Gardner, Michael A. ; Lowe, Luis E.: Analysis of Whiskey by Dispersive Liquid–Liquid Microextraction Coupled with Gas Chromatography/Mass Spectrometry: An Upper Division Analytical Chemistry Experiment Guided by Green Chemistry. In: *Journal of Chemical Education* Bd. 93, American Chemical Society (2016), Nr. 1, S. 186–192

[15] Mac Namara, K., McGuigan, F., & Hoffmann, A.: Auf den Geschmack gekom- men Nr. 44. Gerstel Aktuell (2011), S. S. 12-14

[16] Nie, Yunyun ; Kleine-Benne, Eike: Determining Phenolic Compounds in Whisky using Direct Large Volume Injection and Stir Bar Sorptive Extraction Bd. I, S. 1–8

[17] Stupak, Michal ; Goodall, Ian ; Tomaniova, Monika ; Pulkrabova, Jana ; Hajslova, Jana: A novel approach to assess the quality and authenticity of Scotch Whisky based on gas chromatography coupled to high resolution mass spectrometry. In: *Analytica Chimica Acta* Bd. 1042 (2018), S. 60–70

[18] Daylight: Daylight Theory: SMARTS - A Language for Describing Molecular Patterns (2012)