Ch. Schönberger, M. Krottenthaler, W. Back

Formulation of an alternative tasting scheme for the sensory evaluation of beer

In terms of sensory evaluation beer often suffers compared to wine since consumers often refuse to concede a sophisticated sensory evaluation for a beer as necessary. In this research a tasting scheme was developed emanating from the " trueness-of-type" scheme, which allows a sensory evaluation not only of the characteristic volatile flavours of beer but also of taste qualities in addition to the main off-flavours. The interpretation may be done in two different ways, in the form of visual comparisons, or in the form of score relations. It was shown that this scheme settles statistical claims in terms of reproducibility. Using this scheme in addition to the widespread DLG-scheme as an instrument for internal quality management holds the possibility to get substantial sensory information of the beer sample.

1 Introduction

For the sake of research projects focussing on the sensory determination of the influence of varied process parameters it is not only necessary to have professional tasters but also a sensory system to rely on which is capable to point out the sensory difference between beer samples. There are plenty of tasting schemes for the sensory evaluation of beer. They vary depending on the aim of the scheme, if it is to find out about differences, or preferences or flavour intensities. The common descriptions for the flavour and off-flavour in beer are often misleading and perceived differently by the assessors. The flavours can be divided into volatile and non volatile flavours. Volatile flavour substances in beer can be alcohols, aldehydes, ketones, carbondioxide, sulphur-bound substances etc. Non-volatile flavours can be amino acids, carbohydrates, products of maillard-type reactions, proteins, minerals and hop acids. Some esters are famous for giving a characteristic flavour to a beer as isoamylacetate, phenylacete, ethylacetate etc. (1). Most commonly used for sensory evaluation in Germany is the DLG-Scheme (2). The criteria to be evaluated are aroma, taste, body, liveliness and bitterness, whereas the evaluation of aroma, taste and bitterness weights double. Results are given in a weighted score. The standard deviation of 48 beer samples produced with different process parameters turned out to be only 0,19, and the mean value was 4,11 within a range of 0 – 5, which shows that the DLG-Scheme and its range for evaluation is rarely used, and the criteria are not the appropriate criteria to gain sensory data of the samples.

2 Materials and methods

2.1 Assessor training

The tasting panel of the institute consisted of 10 – 15 assessors. For the sensory evaluations more than 7 tasters had to attend. The age of the tasters varied between 25 and 40. The sessions were held in an undisturbed ambiance, not immediately before or after breakfast, lunch etc. The assessors were not allowed to talk and only scarce information about the samples was given. Table 1 shows the concentrations in which the reference substances were given. The five taste qualities one can distinguish are: Sour, salty, sweet, bitter, and umami (taste sensation activated by monosodiumglutamate).

The following aroma substances were given to the assessors to find a standardised terminology for the perceived flavours (Table 2). Table 3 shows the reference substances for the off-flavours given to the assessors. The assessors had to memorise the given descriptions for each flavour. The flavour substances were administered in different concentrations to train the assessors' memories. The individual experience with each substance is, especially for off-flavours, an important hedonic factor, and revealed that some assessors turned out to respond utmost sensitive to certain flavours.

2.2 Sensory schemes

Figure 1 shows the "trueness-of-type" scheme. The trueness-of-type scheme consists of two parts. Part A is for the description of positive flavours, Part B for negative flavours. The assessors have to judge if a flavour is intense enough, less intense or too intense. For both, too intense or less intense flavours, the score is lower than for the correct intensity. The scores for each flavour are added, and the sum of the negative flavours is substracted from the sum of the positive flavours. The result is given in percentage of the highest score.

2.3 Standard deviation, variance, variance coefficient

Using statistics for sensory data one can judge the assessor's sensory capability. Depending on the intended tasting scheme statistical interpretations have to objectify the results (5). The standard deviation for an array of values is equivalent to the square root of the sum of the deviations divided by the number of terms. The deviation of a term in the array is equivalent to the mean of all the numbers in the array minus the individual number itself. The sum of the variances is the sum of this (6).

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}} \quad s^2 = variance \quad variance\ coefficient\ VK = \frac{s}{|x|}\cdot 100\ \%$$

Formula 1: Definition of standard deviation and variance

Authors: Dr. Christina Schönberger, Dr.-Ing. Martin Krottenthaler, Univ.-Prof. Dr. Werner Back, Lehrstuhl Technologie der Brauerei 1, Weihenstephaner Steig 20, D-85354 Freising, Germany
Tables and Figures see Appendix

An F-test is used to test if the standard deviations of two arrays are equal. The one-tailed version only tests in one direction, that is the standard deviation from the first array is either greater than or less than (but not both) the second array standard deviation. The choice is determined by the problem. If the test is ≤ than the values in the tables, the null-hypothesis is accepted, equal variances are given.

$$\text{null hypothesis}: \sigma_1^2 \le \sigma_2^2 \quad \text{Alternative}: \sigma_1^2 > \sigma_2^2 \quad \text{test}: F = \frac{s_1^2}{s_2^2}$$

$$\text{value in F-distribution for } F_{v1;\, v2;\, 1-\alpha} \; (v_1 = n_1 - 1;\; v_2 = n_2 - 1)$$
$$s_{1,2}^{\;2} = \text{variances of the mean values}$$

Formula 2: Null-hypothesis for equal variances

The T-Test is typically used to compare the means of two arrays. It is used to determine whether or not the means in two arrays are significantly different. Whether or not they can be pooled is determined by the F-Test. The null-hypothesis might be that the two array means are equal. If so, this must be converted to the form that the difference between the two array means is not equal to some constant $c_{2n-2;1-\alpha}$, where n is the sample sizes, and x_1 and x_2 are the sample means, and s_1 and s_2 are the sample variances. If equal variances are given, then the formula reduces to formula 3. If the null-hypothesis is accepted, and $t_0 \le c$ then the two means are equal.

$$t_0 = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad \text{for } \sigma_1^2 = \sigma_2^2 \text{ and } n_1 = n_2 = n$$

$$\sigma_1^{\,2} = \sigma_2^{\,2} \text{ equal variances}$$
$$n_1 = n_2 \text{ identical number of the sample sizes}$$

Formula 3: Test of equal means

2.4 Materials

Different German beer types were used for the sensory evaluation. Those beers were heat treated or stored at 8°C for 10 months. The beer types were: Dark beer, lager, pilsner and wheat beer.

3 Results

3.1 Modification of the "trueness-of-type" scheme

The scheme was upgraded as follows: As positive flavours, a combination of the taste qualities and characteristic beer flavours was selected from the flavourwheel (7). Part B shows the additional flavours. In addition to the existing range from 0 – 3 – 0 the range from 4 – 6 was added to the range of 2 – 0 (see Fig. 2). Within the range from 0 – 6 the intensities of each flavour can be determined. With the range from 0 – 3 – 0 the vertical overall score, equivalent to the Flavourplot-score, is given in relation to the best score. The best score, the best balanced beer resp., is given if each flavour which can be percepted is judged with 3 and no additional flavour exists.

3.2 The Flavourplot-profile for different beer-types

Table 4 shows a listing of possible flavours for different beer types to be taken into the Flavourplot-profile. Depending on how the scheme is supposed to be used, as an internal scheme to control only one product, or to get comparisons with other types and products it is possible to opt for the flavours that are regarded as essential and characteristic.

3.3 Comparison of heat treated and naturally aged beer samples

The assessors had to evaluate different beer samples untreated, after the exposure to 24 h of agitation and 4 days at 40°C to simulate storage condition and after a normal storage at 8°C for 10 months. Figure 3 shows the changes in the flavour profiles for the untreated and the heat-treated beer samples. Figure 4 shows the changes in the flavour profiles for the untreated and naturally aged beer samples. Obviously the heat treatment affects the pilsner-type and lager-type beer more than the dark beer or the wheat beer. The heat treated dark beer is rated to taste burnt, the wheat beer is rated to taste more bitter if heat-treated. Especially the pilsner and lager-type beers appear to be susceptible to an oxidative flavour. The heat-treated lager was also rated to taste astringent, solvent-like and rancid.

The Flavourplot-profiles of the naturally aged beer samples (Fig. 4) show that single flavours tend to become more or less intense by aging. The overall profile is not changed as drastically as by the heat treatment. The old and oxidised flavour is not as distinct as in the heat-treated samples. A comparison of the distinct flavours of heat-treated samples and naturally aged samples gives different flavour profiles. Other than the characteristic oxidised flavour, natural aging induces a wider spectrum of changes in flavours. Flavours which occur in naturally aged samples but not in the heat-treated samples are sweet, gassy, diacetyl, grassy and alcoholic (Table 5). A statistical involution drops out because of the small sampling.

3.4 Comparison between DLG-weighted scores and Flavourplot-scores

Table 6 shows that the Flavourplot-scores cannot be related to the DLG-weighted scores. Since the Flavourplot score depends on the number of flavours to be rated and the number of flavours to be considered as essential and characteristic, also relatively low values of the Flavourplot-scores can describe a perfectly balanced product. The Flavourplot-scores should be seen in regard to the Flavourplot-profiles, and a comparison of Flavourplot-scores is only reliable if the same product is evaluated. The Flavourplot-score of the additional flavours gives an insight about possible off-flavours but is given as a separate value to the evaluation of the positive flavours.

3.5 Statistical considerations of the assessors – Pilsner A

To judge the consistency of the assessors beer sample A of the same batch was evaluated on different days at different points of time. There was no information given to the assessors about the beer sample. Table 7 shows the results. The variances coefficient of the DLG-sessions was rather low with a value of 3.62. Given the range of the Flavourplot-scores of the positive flavours, also this variance coefficient turned out to be relatively low with 7.04. In contrast to that the variance coefficient of the negative flavours is high (27.55). The beer sample was assessed as very good, although the scores are merely between 27 and 33.

3.6 Statistical considerations of the assessors – Pilsner B

To judge the consistency of the assessors beer sample B of the same batch was evaluated on different days and at different times. No information was given to the assessors about the beer samples. Table 8 shows the results. The variances coefficient of the DLG-sessions was rather low with a value of 2.58. Given the range of the scores of the positive sums this variance coefficient turned out to be higher compared with Pilsner A with a value of 12.16. The variance coefficient of the negative flavours is also high (33.64). The beer sample was assessed as very good. The scores are as with Pilsner A, merely between 27 and 36, which again supports the decision not to relate the flavour scores with the DLG-weighted scores.

3.7 Statistic evaluation of the flavour intensities (visual comparison)

Table 9 shows the standard deviations for the single flavours for the multiple sessions to pilsner A and pilsner B. It is apparent that standard deviations for bitter, astringent, sour and old/oxidised are comparatively higher, which means that the concentration ranges of the substances that cause these flavours are wider and the perception of the assessors is inconsistent concerning these flavours. The mean values of the multiple sessions were tested using the F- and T-Test. All mean values of the sessions to pilsner A and of the sessions to pilsner B were proved to be equal, albeit some could not be compared since their variances were not equal.

4 Conclusion/Summary

The presented tasting scheme originates from the trueness-of-type scheme. The itemised flavours were chosen from the flavour-wheel and separated in positive and additional flavours as well as the common off-flavours. This Flavourplot-profile is interpreted in two ways. One way is to determine the means for each flavour for a visual comparison in the form of a spider-plot. The other way comprises the sum of the positive flavours and the additional flavours in relation to the best score. To relate the Flavourplot-scores with the weighted scores of the DLG-tasting is not advisable since the Flavourplot-scores vary with the number of flavours considered as essential for the sample.

A comparison of heat-treated beer samples and naturally aged beer samples with the Flavourplot-profile clearly revealed that the heat treating induces flavour changes different from those induced by natural aging, which is to be considered in terms of taste stability. In terms of statistical considerations multiple sessions to the same beer samples showed that the Flavourplot-scores are reproducible. High margins of deviation occur for the additional flavours since the sensitivity of the assessors for certain off-flavours is hedonically distinct. The mean values of the Flavourplot-scores are constant. The mean values of the intensities of the Flavourplot-profile proved to be equal for multiple sessions. Those flavours that were hard to judge for the assessors were bitter, astringent, sour and old/oxidised.

The Flavour-profile is manageable for the assessors and not excessively time consuming. With this Flavourplot-profile used additionally to the DLG-scheme, a vast quantity of sensory data for each beer sample can be gained. According to the requirements of the brewery, the general form of the Flavourplot-profile or an individual form with different flavours can be chosen as an advantageous means for sensory quality management.

Acknowledgments

This research was supported by the Wissenschaftsförderung der Deutschen Brauwirtschaft e. V. B71-II.

5 Literature

1. Bamforth, C.: Beer flavour: Esters, Brewers' Guardian, 2001, 130 (9) 32 – 34.
2. Pfenninger, H.: Brautechnische Analysenmethoden, Band II, 3. Auflage, Freising, Selbstverlag der MEBAK, 1993, 81 – 83.
3. Kaltner, D.: Untersuchungen zur Ausbildung des Hopfenaromas und technologische Maßnahmen zur Erzeugung hopfenaromatischer Biere, Dissertation, TU Munich, 2000, p52.
4. Sensory Analysis Manual, In: Methods of Analysis, London, The Institute of Brewing, 1997.
5. Busch-Stockfisch, M.: Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung, 2002, p2.1, p3.
6. Precht, M.: Bio-Stastistik, 2, 1993.
7. Meilgaard, M. C.: Beer Flavor Terminology – a Case History, In: Applied Sensory Analysis of Foods, Vol I, 1988, 73 – 87.

Paper received February 17, 2004

Appendix

Name: *Sample:* *Date:*

A. Positive Flavours

Please assess the beer and offer a score for each of the following by circling the appropriate number

	Less intense			correct		more intense	
Hoppy	0	1	2	3	2	1	0
Malty	0	1	2	3	2	1	0
Sulphidic	0	1	2	3	2	1	0
DMS	0	1	2	3	2	1	0
Sweet	0	1	2	3	2	1	0
Bitterness	0	1	2	3	2	1	0
Smooth	0	1	2	3	2	1	0
Astringent	0	1	2	3	2	1	0
Full	0	1	2	3	2	1	0

Total A=

B. Negative Flavours

Score the following characters on a 1-3 intensity scale only if you consider them to be present.

Grainy	1	2	3
Diacetyl	1	2	3
Stale	1	2	3
other	1	2	3

Total B=

Trueness of type= $\frac{\text{Total A} - \text{Total B} \times 100\%}{27} = \%$

Fig. 1 "Trueness-of-type" tasting scheme (4)

Name: *Sample:* **A**

Flavourplot-Profile

please mark the accordant intensities

positive flavours	less intense			correct	more intense		
sweet	0	1	2	3	2 (4)	1(5)	0(6)
bitter	0	1	2	3	2 (4)	1(5)	0(6)
hoppy	0	1	2	3	2 (4)	1(5)	0(6)
malty	0	1	2	3	2 (4)	1(5)	0(6)
like roast malt	0	1	2	3	2 (4)	1(5)	0(6)
phenolic	0	1	2	3	2 (4)	1(5)	0(6)
fruity	0	1	2	3	2 (4)	1(5)	0(6)
like flowers	0	1	2	3	2 (4)	1(5)	0(6)
adstringent	0	1	2	3	2 (4)	1(5)	0(6)
gassy	0	1	2	3	2 (4)	1(5)	0(6)
alcoholic	0	1	2	3	2 (4)	1(5)	0(6)

other flavours

	not perceivable	less intense	intense	more intense
Diacetyl	0	1(4)	2(5)	3(6)
DMS	0	1(4)	2(5)	3(6)
rancid	0	1(4)	2(5)	3(6)
grassy	0	1(4)	2(5)	3(6)
solvent like	0	1(4)	2(5)	3(6)
burnt	0	1(4)	2(5)	3(6)
sour	0	1(4)	2(5)	3(6)
umami	0	1(4)	2(5)	3(6)
old/oxidized	0	1(4)	2(5)	3(6)

Fig. 2 Tasting scheme to be referred to as Flavourplot-profile

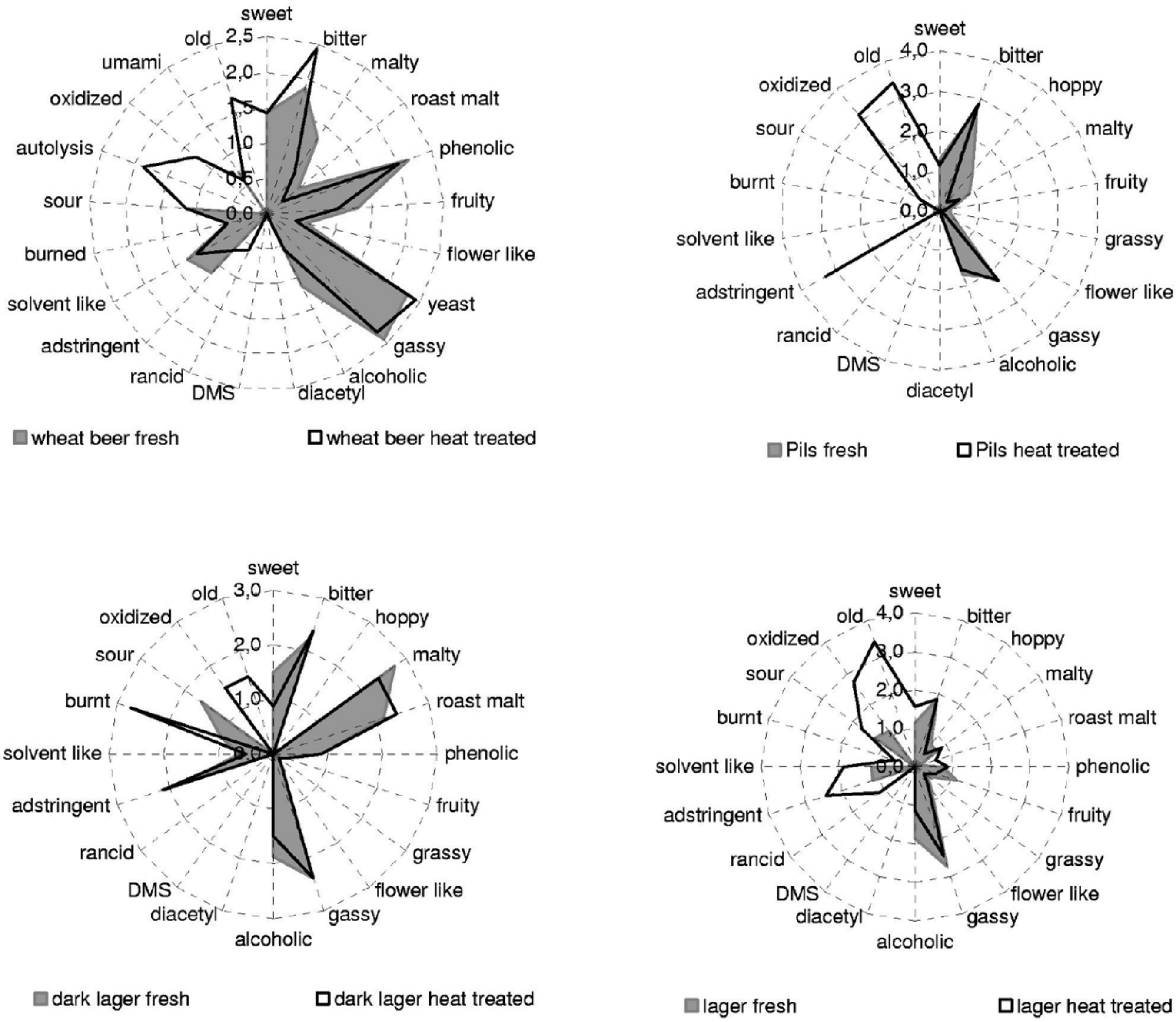

Fig. 3 Flavourplot-profiles of heat-treated and fresh beer samples

Table 1 Reference substances for taste qualities and recommended concentrations

Taste quality	Reference substances	Concentration g/l
sweet	saccharose	6
sour	citric acid	0,6
salty	sodium chloride	1,3
bitter	caffeine	0,3
umami	sodium glutamate	0,5
adstringent	tannic acid	0,05 % (wt/wt)

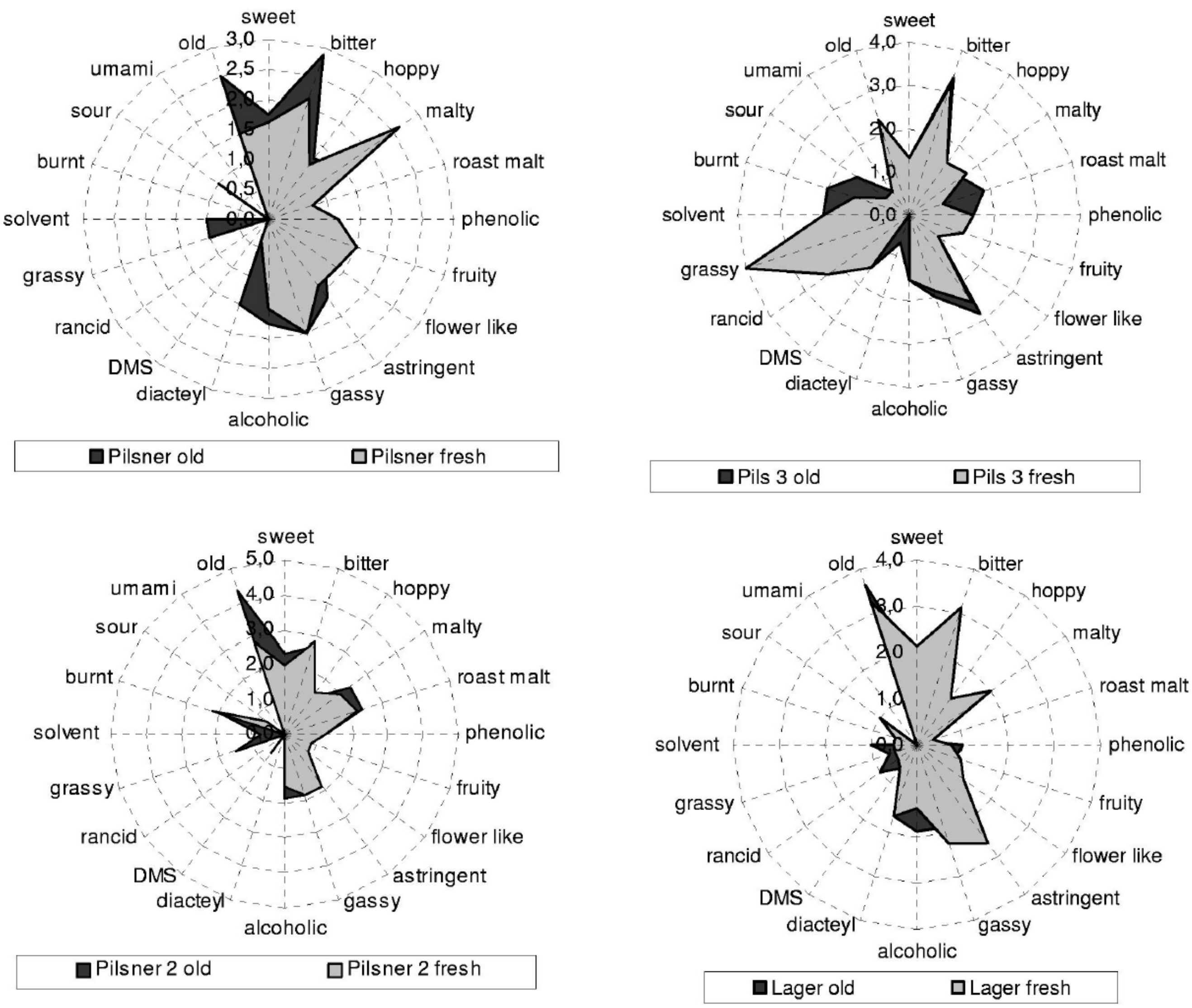

Fig. 4 Flavourplot-profiles of heat-treated and fresh beer samples

Table 2 Reference substances for positive flavours

	Description	Reference substance	Concentration	Threshold
0111	phenolic	eugenol	120 µg/l	40 µg/l
0131	banana, estery	isoamylacetate	3.0 mg/l	1.7 mg/l
0132	apple, fruity	ethylhexanoate	0.75 mg/l	0.25 mg/l
0161	like flowers, like roses	phenylethanol	175 mg/l	125 mg/l
0170	hoppy (citrus)[a]	(R)-linalool[a]	0.30 µg/l	9 µg/l

[a] *defined by Kaltner (3), no reference substance from the flavour wheel*

Table 3 Reference substances for unpleasant flavours

	Description	Reference substance	Concentration	Threshold
0613	rancid, cheesy (old hop)	isovaleric acid	2.5 mg/l	1.5 mg/l
0614	rancid butter, cheesy	butyric acid	5.0 mg/l	1.5 mg/l
0732	DMS, cooked vegetables	dimethylsulfid	150 µg /l	50 µg /l
0620	buttermilk	diacetyl	0.45 mg/l	0.15 mg/l
0724	light-struck “skunky”[a]	3-methyl-2-buten-1-thiol[b]	–	10 ng/l
0230	grassy	cis-3-hexenol	45 mg/l	15 mg/l

[a]*defined by Kaltner (3),*
[b]*not mentioned as a reference substance in the flavour wheel*

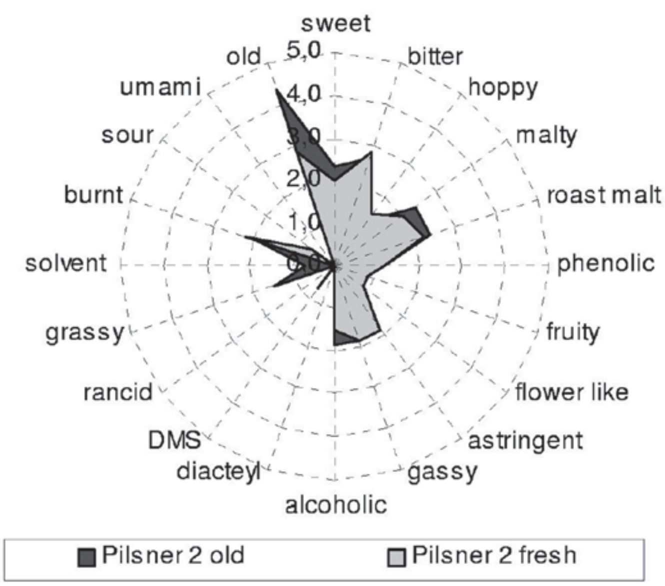

Fig. 5 Flavourprofile-plots for various sessions for the same sample

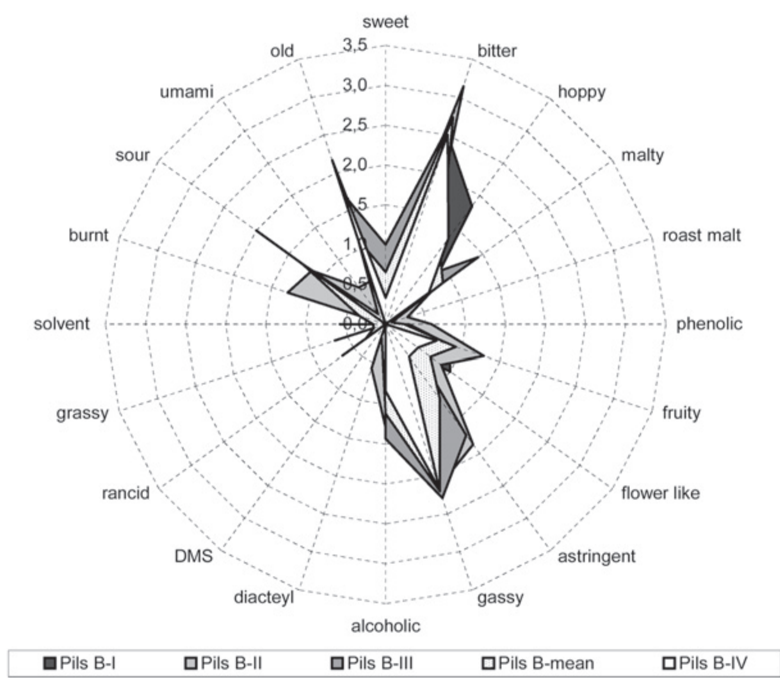

Fig. 6 Flavourprofile-plots for various sessions for the same sample

Table 4 Optional flavours for different beer types for the Flavourplot-Profile

Pilsner-type beers	Light and dark lager	Light / dark ales	Light / dark wheat beers
Positive flavours			
sweet	sweet	sweet	sweet
bitter	bitter	bitter	bitter
hoppy	hoppy	hoppy	hoppy
malty	malty	malty	malty
fruity	roast malt	roast malt	roast malt
grassy	phenolic	phenolic	phenolic
flower-like	fruity	fruity	flower like
gassy	flower-like	caramel	yeast
alcoholic	gassy	flower-like	gassy
	alcoholic	gassy	alcoholic
		alcoholic	
Additional flavours			
diacetyl	diacetyl	diacetyl	diacetyl
DMS	DMS	DMS	DMS
rancid	rancid	rancid	rancid
adstringent	adstringent	adstringent	adstringent
grassy	solvent-like	solvent-like	solvent-like
solvent-like	grassy	grassy	burnt
skunky/light-struck	burnt	burnt	sour
sour	sour	sour	oxidized
oxidised	oxidised	oxidised	autolysis
old	old	old	umami

Table 5 Flavours mentioned for heat-treated and naturally aged beer samples

	Mentioned for heat-treated beer samples %	Mentioned for naturally aged beer samples %
Older/more oxidised	xxxx	xx
More astringent	xxx	xx
More bitter	xxx	xx
More sour	xx	–
More solvent-like	x	xx
More rancid	xx	–
Mor burnt	xx	x
More like roast malt	x	x
Sweeter	–	xx
More hoppy	–	x
More flower-like	–	x
More gassy	–	xx
More malty	–	x
More diacetyl	–	xx
More alcoholic	–	xx
More grassy	–	xx

xxxx mentioned for every sample
xxx mentioned for most samples
xx mentioned for ca. half the samples
x rarely mentioned
– not mentioned

Table 6 Comparison of the DLG-results and the Flavourplot-scores for some beer samples

	Fresh			Naturally aged		
	Positive	Additional	DLG-weighted score	Positive	Additional	DLG-weighted score
Lager	41.29	8.71	3.70	38.26	9.10	3.71
Pilsner 1	47.73	2.78	3.95	41.67	5.30	3.81
Pilsner 2	47.98	9.26	4.00	42.42	9.60	3.87
Pilsner 3	40.40	17.90	3.61	41.41	15.15	3.80

Table 7 Multiple session results for the identical beer sample and statistical figures for the tasting schemes

	Pilsner A-I	Pilsner A-II	Pilsner A-III	Pilsner A-IV	VC	s	M
DLG-weighted score	4.4	4.43	4.07	4.46	3.62	0.16	4.34
Flavourplot-score, positive	33.83	27.78	31.17	32.03	7.04	2.20	31.20
Flavourplot-score, additional	2.52	3.09	4.76	2.64	27.55	0.90	3.25

VC = variance coefficient, s = standard deviation, M = mean value

Table 8 Multiple session results for the identical beer sample and statistic figures for the tasting schemes

Probe	pilsner B	pilsner B	pilsner B	pilsner B	VK	s	M
DLG-weighted score	4.49	4.4	4.35	4.22	2.58	0.113	4.37
Flavourplot-score positive	31.82	26.77	35.93	30.3	12.16	3.794	31.21
Flavourplot-score additional	2.02	4.32	3.896	2.597	33.64	1.0769	3.21

vc = variance coefficient, s = standard deviation, M = mean value

Table 9 Mean values (MV) and standard deviations (SD) of the sample intensities

	MV pilsnerA	SD pilsnerA	MV pilsnerB	SD pilsner B
sweet	0.73	0.92	0.65	0.75
bitter	2.69	**1.42**	2.73	**1.54**
hoppy	0.96	1.24	1.15	1.05
malty	1.00	1.00	0.88	0.91
roast malt	0.04	0.23	0.08	0.27
phenolic	0.38	0.68	0.23	0.59
fruity	0.58	0.99	0.92	0.89
flower like	0.42	0.69	0.69	0.74
astringent	1.31	**1.56**	1.15	**1.59**
gassy	2.00	1.17	2.19	1.17
alcoholic	1.12	1.27	1.12	1.21
diacteyl	0.00	0.00	0.15	0.78
DMS	0.15	0.00	0.00	0.00
rancid	0.31	1.26	0.31	1.09
grassy	0.15	0.92	0.15	0.78
solvent	0.54	1.61	0.15	0.78
burnt	0.15	0.92	0.35	1.23
sour	0.81	**1.79**	1.27	**1.95**
umami	0.15	0.00	0.15	0.78
old/oxidised	1.08	**1.91**	1.12	**1.88**