



# Mapping of best practices and development of testing methods and procedures for identification of characterising flavours in tobacco products

**Final Report**

***RfS Chafea/2014/health/19  
Contract n° 20146202***

**EUROPEAN COMMISSION**

Directorate-General for Health and Food Safety

Directorate B— Health systems, medical products and innovation

Unit B2 — Health in all policies, global health, tobacco control

*E-mail: [SANTE-B2-TOBACCO-CONTROL@ec.europa.eu](mailto:SANTE-B2-TOBACCO-CONTROL@ec.europa.eu)*

*European Commission  
B-1049 Brussels*

**Mapping of best practices  
and development of testing methods  
and procedures  
for identification of characterising  
flavours in tobacco products**

RfS Chafea/2014/health/19

Contract nº 20146202

***Europe Direct is a service to help you find answers  
to your questions about the European Union.***

**Freephone number (\*):**

**00 800 6 7 8 9 10 11**

(\*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

## **LEGAL NOTICE**

This report was produced under the EU Health Programme (2008-2013) in the frame of a service contract with the Consumers, Health, Agriculture and Food Executive Agency (Chafea) acting on behalf of the European Commission.

The content of this report represents the views of HETOC and is its sole responsibility; it can in no way be taken to reflect the views of the European Commission and/or Chafea or any other body of the European Union.

The European Commission and/or Chafea do not guarantee the accuracy of the data included in this report, nor do they accept responsibility for any use made by third parties thereof.

More information on the European Union is available on the Internet (<http://europa.eu>).

Luxembourg: Publications Office of the European Union, 2016

ISBN 978-92-9200-681-5

Doi:10.2818/08983

© European Union, 2016

Reproduction is authorised provided the source is acknowledged.



HETOC (Health Effects Tobacco Composition) Consortium, July 2016

- Consortium leader: National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands
- Consortium member: Wageningen University (WUR), Wageningen, the Netherlands
- Subcontractor: OP&P Product Research (OP&P), Utrecht, the Netherlands

## Abstract

The new EU Tobacco Product Directive 2014/40/EU (TPD) prohibits cigarettes and roll-your-own tobacco having a characterising flavour other than one of tobacco. Although some other jurisdictions have legislation on tobacco flavour, no methods are currently available to assess characterising flavours.

The overall aim of this project was to deliver a test approach for the assessment of characterising odours in tobacco products, by a combination of sensory profiling, chemical-analytical measurements, other methods, or a combination of methods. The tobacco industry and the food industry in general conduct expert panel and consumer research to test the flavour and other sensory characteristics of a product for purposes of successful product development.

In summary, the project consisted of carrying out four work packages (WP), in the time period August 2014-June 2015:

- WP1: Review the literature, with the aim to draft a review of the current methods and approaches that may be suitable for the objective determination of characterising flavours in tobacco products.
- WP2: Identify concept profiles, with the aim to draft three testing approaches, based on the output of WP1, for sensory analysis of tobacco products (including chemical-analytical measurements) applicable to tobacco product regulation.
- WP3: Optimise and peer-review the procedure, with the aim to test, refine and finalise the methodology that had been proposed based on the outcomes of WP1 and WP2.
- WP4: Assess the feasibility and impact, with the aim to check whether the overall aim, to deliver a method to decide whether a tobacco product imparts a characterising flavour other than tobacco, has been reached.

Based on a review of the relevant literature, and discussions with experts from several fields, it was decided that the best approach for assessing characterising flavours would be a combination of sensory profiling and chemical analysis. More specifically, a combination was proposed of a trained expert panel that assesses odours by smelling tobacco samples, with headspace GC-MS.

To test this approach, pilot experiments were performed with a semi-trained expert panel and headspace GC-MS to test the procedure. Based on the findings, a robust and feasible procedure was proposed for assessment by a sensory expert panel, complemented by chemical analysis, laid down in the main document and two SOPs.

It was concluded that an expert panel is a good approach to assess characterising flavours. The pilot in general yielded valid, robust, reliable and reproducible results. Smelling is the preferred starting point, as it captures most of the products with characterising flavours. Products currently marketed on their packaging to contain characterising flavours (cherry, menthol, vanilla) can be distinguished from a reference space of 'normal' commercial cigarettes brands and natural tobacco leaves. Chemical analysis with Headspace GC-MS is a suitable method to identify flavour components and additives. The results can be used to build a flavour library. Comparing results of flavoured cigarettes with those of natural tobacco leaves gives an indication whether a component is added.

In the report, detailed conclusions and recommendations are given to set up a reliable method.

## Contents

<b>1</b>	<b>Executive summary</b> .....	<b>13</b>
<b>2</b>	<b>Introduction</b> .....	<b>16</b>
<b>3</b>	<b>Background</b> .....	<b>17</b>
3.1	<b>References</b> .....	<b>18</b>
<b>4</b>	<b>Methodology</b> .....	<b>19</b>
4.1	<b>WP1: Literature review</b> .....	<b>19</b>
4.1.1	Task 1.1 Development of methodology to retrieve relevant information.....	19
4.1.2	Task 1.2 Review of methods recorded to be used by the tobacco industry for sensory analysis	19
4.1.3	Task 1.3 Review of methods used for assessment of tobacco products by EU Member States and third countries .....	20
4.1.4	Task 1.4 Review of methods for sensory analysis of any type of consumer or food products	20
4.1.5	Task 1.5 Review of alternative methods, not based on sensory analysis .....	20
4.1.6	Task 1.6 draft and review report.....	20
4.2	<b>WP2: Identification of the applicable methods</b> .....	<b>21</b>
4.2.1	Task 2.1 Assessment of the usability/adaptability of sensory analysis.....	21
4.2.2	Task 2.2 Identification of applicable standards and/or best practices for determination of characterising flavours .....	22
4.2.3	Task 2.3 Identification of applicable alternative methods for sensory analysis .....	22
4.2.4	Task 2.4 Identification of three concept profiles for the testing approach .....	23
4.2.5	Task 2.5 Seminar to discuss the three concept profiles.....	24
4.3	<b>WP3: Optimisation of the testing approach on the basis of pilot studies</b> .....	<b>24</b>
4.3.1	Task 3.1 Selection and possibly adaptation of the profile for pilot testing.....	24
4.3.2	Task 3.2 Selection of samples for pilot testing.....	27
4.3.3	Task 3.3 Pilot testing.....	27
4.3.4	Task 3.4 Optimising the study protocol.....	27
4.4	<b>WP4: Feasibility and impact assessment</b> .....	<b>27</b>
4.4.1	Task 4.1 Technical feasibility and quality of outcomes .....	27
4.4.2	Task 4.2 quantification of costs and funding model .....	28
4.4.3	Task 4.3 describe possible modifications and alternative options.....	28
<b>5</b>	<b>Findings</b> .....	<b>29</b>
5.1	<b>WP1: Literature review</b> .....	<b>29</b>
5.2	<b>WP2: Identification of the applicable methods</b> .....	<b>29</b>
5.3	<b>WP3: Optimisation of testing – sensory research</b> .....	<b>31</b>
5.3.1	Results of the pilot study.....	31
5.4	<b>WP3: Optimisation of testing – Chemical analysis</b> .....	<b>61</b>
5.4.1	Products tested .....	61
5.4.2	Final results.....	61
5.4.3	Findings and recommendations .....	63
5.5	<b>WP4: Cost benefit assessment</b> .....	<b>64</b>
5.5.1	Technical feasibility and quality of outcomes .....	64
5.5.2	Quantification of costs and funding model .....	65
5.5.3	Possible modifications and alternative options to be considered for next steps .....	67
<b>6</b>	<b>Overall conclusions and recommendations</b> .....	<b>69</b>

<b>Annex I. WP1: Literature review - Current methods and approaches for the determination of characterising flavours in tobacco products .....</b>	<b>1</b>
<b>1 Introduction.....</b>	<b>1</b>
<b>1.1 The EU Tobacco Products Directive on regulating tobacco flavours.....</b>	<b>1</b>
<b>1.2 Systematic literature selection and review .....</b>	<b>2</b>
<b>1.3 Collection of information .....</b>	<b>2</b>
<b>1.4 Selection and review of the retrieved literature .....</b>	<b>3</b>
<b>1.5 Review set up.....</b>	<b>3</b>
<b>2 Approaches relevant for tobacco sensory analysis .....</b>	<b>3</b>
<b>2.1 Smelling versus smoking tobacco .....</b>	<b>3</b>
<b>2.2 Consumer versus expert testing .....</b>	<b>4</b>
2.2.1 Consumer testing .....	4
2.2.2 Expert testing .....	5
2.2.3 Cross-cultural differences between panel members .....	6
<b>3 Methodologies for measurement and data analysis (relevant for tobacco sensory analysis).....</b>	<b>7</b>
<b>3.1 Discriminative methods to test for sensory product differences in consumers or experts .</b>	<b>7</b>
3.1.1 Undirectional difference tests.....	8
3.1.2 Directional difference tests .....	9
<b>3.2 Descriptive methods using consumer or expert testing to assess sensory product characteristics .....</b>	<b>9</b>
3.2.1 Quantitative Descriptive Analysis.....	10
3.2.2 Free choice profiling .....	11
3.2.3 Flash and ultra-flash profiling.....	11
3.2.4 Ideal profile method.....	12
3.2.5 Check all that apply (CATA) .....	12
<b>3.3 Descriptive methods using expert testing to assess sensory product characteristics .....</b>	<b>13</b>
3.3.1 Flavour Profile .....	13
3.3.2 Spectrum method.....	13
3.3.3 Hybrid Method .....	14
<b>4 Alternative methods .....</b>	<b>14</b>
<b>4.1 Analysing sensory response in the human brain .....</b>	<b>14</b>
4.1.1 Neuro-imaging techniques .....	14
4.1.2 Other methods .....	15
<b>4.2 Sensory testing in laboratory animals.....</b>	<b>15</b>
<b>4.3 Chemical analysis of flavour components .....</b>	<b>16</b>
4.3.1 Flavour component analysis.....	16
4.3.2 Sampling flavour components for GC-MS analysis .....	17
<b>4.4 Responses on tobacco flavour analysis questionnaires from EU- and third countries.....</b>	<b>19</b>
<b>5 Discussion and conclusions .....</b>	<b>25</b>
<b>6 References.....</b>	<b>27</b>
<b>7 Appendix I: Keywords used for literature search.....</b>	<b>33</b>
<b>8 Appendix II: Flowcharts of methods described in this review.....</b>	<b>34</b>
<b>9 Appendix III: Scaling methods .....</b>	<b>35</b>
<b>9.1 Line scaling.....</b>	<b>35</b>
<b>9.2 Category scales .....</b>	<b>35</b>
<b>9.3 Scale for overall quality .....</b>	<b>36</b>



9.4	Category–Ratio (Labelled Magnitude) Scales .....	36
9.5	JAR scaling .....	37
<b>10</b>	<b>Appendix IV: Panel selection and training and test- and facility requirements .....</b>	<b>37</b>
<b>10.1</b>	<b>Panel selection and training .....</b>	<b>37</b>
10.1.1	Panel characteristics.....	37
10.1.2	The panel leader.....	38
10.1.3	Self-generated terminology versus standardized terminology.....	39
10.1.4	Duration and kind of training .....	39
<b>10.2</b>	<b>Test- and facility- requirements.....</b>	<b>40</b>
10.2.1	The test product .....	40
10.2.2	Facility requirements.....	40
10.2.3	Data analysis and reporting.....	40
10.2.4	Discriminative tests .....	40
10.2.5	Descriptive test.....	40
10.2.6	Principal Component Analysis .....	41
10.2.7	Multivariate Analysis of Variance (MANOVA) .....	41
10.2.8	Discriminant analysis or Canonical Variate Analysis .....	41
10.2.9	Generalized Procrustes Analysis (GPA) .....	41
<b>11</b>	<b>Appendix V: Questionnaire on sensory analysis methods for tobacco products .....</b>	<b>41</b>
	<b>Annex II. WP2: Concept profiles.....</b>	<b>1</b>
<b>1</b>	<b>Concept profiles.....</b>	<b>1</b>
<b>2</b>	<b>Assessment of methods and practices .....</b>	<b>1</b>
<b>2.1</b>	<b>Assessment of sensory analysis .....</b>	<b>1</b>
2.1.1	Assessment of sensory testing methods .....	2
2.1.2	Assessment of approaches/best practices in sensory analysis .....	3
2.1.3	Assessment of alternative methods to sensory analysis.....	4
2.1.4	Conclusions of the assessment of methods and practices.....	6
<b>2.2</b>	<b>Alternative methods to sensory analysis .....</b>	<b>8</b>
2.2.1	Chemical analysis methods .....	8
2.2.2	Alternative methods.....	8
<b>3</b>	<b>Identification of three basic approaches and methods.....</b>	<b>9</b>
<b>3.1</b>	<b>Reasoning choosing the three approaches and methods .....</b>	<b>9</b>
<b>3.2</b>	<b>Overview basic approaches .....</b>	<b>12</b>
3.2.1	Approach 1 .....	13
3.2.2	Approach 2 .....	16
3.2.3	Approach 3 .....	19
<b>4</b>	<b>Procedure; concept profiles and possible orders.....</b>	<b>21</b>
<b>4.1</b>	<b>Explanation with the procedure flow charts .....</b>	<b>22</b>
4.1.1	Concept profile 1 .....	22
4.1.2	Concept profile 2 .....	22
4.1.3	Concept profile 3 .....	22
<b>5</b>	<b>Preliminary detailed basic approaches.....</b>	<b>23</b>
<b>5.1</b>	<b>Approach 1.....</b>	<b>23</b>
<b>5.2</b>	<b>Approach 2.....</b>	<b>25</b>
<b>5.3</b>	<b>Approach 3.....</b>	<b>27</b>
	<b>Annex III: WP3 - MAIN document .....</b>	<b>1</b>

<b>1</b>	<b>Purpose and scope</b> .....	<b>1</b>
<b>2</b>	<b>General</b> .....	<b>1</b>
<b>3</b>	<b>Definitions</b> .....	<b>1</b>
<b>4</b>	<b>Control and classification of the tobacco product</b> .....	<b>4</b>
<b>5</b>	<b>Advisory expert panel</b> .....	<b>5</b>
5.1	Panel leaders.....	6
5.2	Panel selection .....	6
5.3	Panel training.....	7
5.4	Panel training frequency .....	7
5.5	Test session frequency and panel maintenance .....	7
5.6	General test set-up.....	8
5.7	Product purchase .....	8
5.8	Product profiling .....	9
5.9	Processing of data by the panel leader .....	9
5.10	Product space, reference space and product testing.....	9
<b>6</b>	<b>Chemical analysis</b> .....	<b>11</b>
6.1	Chemical analysis prerequisites .....	12
6.2	General test set-up.....	12
6.3	Product profiling .....	12
6.4	Processing of data by the researcher .....	12
<b>7</b>	<b>Risk assessment</b> .....	<b>12</b>
<b>8</b>	<b>References</b> .....	<b>14</b>
	<b>Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1)</b> .....	<b>1</b>
<b>1</b>	<b>Purpose and scope</b> .....	<b>1</b>
<b>2</b>	<b>General</b> .....	<b>1</b>
<b>3</b>	<b>Definitions</b> .....	<b>1</b>
<b>4</b>	<b>Standard operating procedure</b> .....	<b>3</b>
4.1	<b>Panel set up</b> .....	<b>3</b>
4.1.1	Recruitment and hiring of panel leaders.....	3
4.1.2	Role of the panel leaders.....	4
4.1.3	Supporting personnel .....	4
4.1.4	Recruitment of participants .....	4
4.1.5	Information meeting for potential candidates/panellists .....	5
4.1.6	Selection of participants based on in and exclusion criteria.....	5
4.1.7	Screening of participants on olfactory function, verbal and social skills. ....	5
4.1.8	Panellist selection.....	7
4.2	<b>Panel training</b> .....	<b>7</b>
4.2.1	Determine training schedule .....	7
4.2.2	Instructions before the training .....	7
4.2.3	Attribute (generation) training sessions .....	8
4.2.4	Consensus training sessions .....	8
4.2.5	Test training sessions .....	8
4.2.6	Monitoring panel performance .....	9
4.2.7	Panel feedback .....	10
4.2.8	Panel maintenance, robustness and sustainability control .....	10

<b>4.3</b>	<b>Test phase</b> .....	<b>11</b>
4.3.1	Setting up the reference space .....	11
4.3.2	Test sessions.....	11
4.3.3	Materials.....	12
4.3.4	Test room .....	12
4.3.5	Number of samples .....	12
4.3.6	Sample selection.....	12
4.3.7	Sample preparation.....	12
4.3.8	Randomization.....	13
4.3.9	Test set-up- product assessment .....	13
4.3.10	Test procedure .....	13
4.3.11	Product profiling.....	14
<b>5</b>	<b>Statistics</b> .....	<b>15</b>
5.1	Data processing procedure.....	15
5.2	Data analysis .....	15
<b>6</b>	<b>References</b> .....	<b>20</b>
<b>7</b>	<b>Appendix I. Declaration of honour</b> .....	<b>21</b>
<b>8</b>	<b>Appendix II. Health questionnaire</b> .....	<b>22</b>
<b>9</b>	<b>Appendix III. Screening form tobacco odour recognition</b> .....	<b>27</b>
<b>10</b>	<b>Appendix IV. Screening forms sniffin stick test</b> .....	<b>28</b>
<b>11</b>	<b>Appendix V. Normative scores sniffin stick test</b> .....	<b>33</b>
<b>12</b>	<b>Appendix VI. EyeQuestion example</b> .....	<b>34</b>
<b>Annex V: WP3 – Standard Operating Procedure 2 (SOP 2)</b> .....		<b>1</b>
1	Purpose and scope.....	1
2	General.....	1
3	Terms and definitions .....	1
4	Standard Operation Procedure .....	2
4.1	Preferred qualities chemical analyst.....	2
4.2	Requirements of the lab facility.....	2
<b>Annex VI. WP4: Questionnaire sent to industry</b> .....		<b>1</b>
1	Questionnaire on costs aspects involved in the analysis of tobacco products .....	1
<b>Annex VII. Minutes of seminar</b> .....		<b>1</b>
1	Minutes of characterising flavours in tobacco products seminar.....	1
<b>Annex VIII. Medical Ethical Committee application</b> .....		<b>1</b>
<b>Annex IX. Cost estimate</b> .....		<b>1</b>
1	Estimation of costs for implementing the methodology as described in WP3 .....	1
1.1	Cost-estimation for the identification of tobacco products with a characterising flavour by smelling the product using a sensory expert panel.....	1
1.1.1	Cost estimation setting-up the panel (estimated timeframe: 6 months).....	3
1.1.2	Cost estimation for establishing the reference space for cigarettes and roll-your-own tobacco products in the first year (timeframe: 1 month) .....	4

1.1.3	Cost estimation for maintaining the panel.....	5
1.1.4	Cost estimation for assessment of products in subsequent 4 years (2016-2020) by monthly meetings of the expert panel.....	6
1.1.5	Total cost estimate for setting up and running the panel.....	7
<b>1.2</b>	<b>Cost estimation for chemical analysis of tobacco products to identify products with a characterising flavour .....</b>	<b>8</b>
1.2.1	Cost estimation for assessment of reference products in the first year (timeframe: ± 1 month) .....	8
1.2.2	Cost estimation for assessment of products over a period of 5 years (2015-2020) .....	8
<b>Annex X. Pilot proposal.....</b>		<b>1</b>
<b>1</b>	<b>Sensory tests.....</b>	<b>1</b>
1.1	Study aim .....	1
1.2	Study design.....	2
1.3	Time line .....	2
1.4	Study population.....	3
1.5	Statistical analysis .....	9
<b>2</b>	<b>Chemical analysis.....</b>	<b>11</b>
2.1	Aim .....	11
2.2	Experimental design.....	11
2.2.1	Time line .....	12
2.2.2	Products.....	12
2.2.3	Method.....	12
<b>3</b>	<b>Budget (pilot).....</b>	<b>12</b>
<b>4</b>	<b>References.....</b>	<b>13</b>
<b>Annex XI. Course of the panel training.....</b>		<b>1</b>
<b>Annex XII. Extended results of pilot study.....</b>		<b>1</b>
<b>1</b>	<b>CATA results screening I.....</b>	<b>1</b>
<b>2</b>	<b>Preliminary list of attributes generated by the panel .....</b>	<b>6</b>
<b>3</b>	<b>Results first test training session week 3 panel performance .....</b>	<b>7</b>
<b>4</b>	<b>Results test-training session week 5 panel performance.....</b>	<b>8</b>
<b>5</b>	<b>R-Scripts.....</b>	<b>9</b>

## 1 Executive summary

The new EU Tobacco Product Directive 2014/40/EU (TPD) prohibits cigarettes and roll-your-own (RYO) tobacco having a characterising flavour other than one of tobacco. Other jurisdictions already implemented legislation with respect to flavour regulation. Brazil (RDC ANVISA Nº 14) and Canada (BILL C-32) banned most flavours altogether. The US Food and Drug Administration (FDA) banned cigarettes containing certain characterising flavours in their Tobacco Control Act (Sec. 907, tobacco product standards): *“...a cigarette or any of its component parts (including the tobacco, filter, or paper) shall not contain, as a constituent (including a smoke constituent) or additive, an artificial or natural flavour (other than tobacco or menthol) or an herb or spice, including strawberry, grape, orange, clove, cinnamon, pineapple, vanilla, coconut, liquorice, cocoa, chocolate, cherry, or coffee, that is a characterising flavour of the tobacco product or tobacco smoke.”*

The new TPD takes a similar approach by prohibiting cigarettes and roll-your-own tobacco having a characterising flavour other than one of tobacco. Characterising flavour means a *“clearly noticeable smell or taste other than one of tobacco, resulting from an additive or a combination of additives, including, but not limited to, fruit, spice, herb, alcohol, candy, menthol or vanilla, which is noticeable before or during the consumption of the tobacco product.”* In this definition, ‘tobacco’ means: *“leaves and other natural processed or unprocessed parts of tobacco plants, including expanded and reconstituted tobacco.”* Tobacco products other than cigarettes and RYO are currently exempted from this prohibition but this exemption can be removed if there is a substantial change of circumstances in terms of sales volume and use among young people (Art 7(6) and 2(28) of the TPD).

The TPD does not ban additives or combination of additives that have the potential to result in a characterising taste altogether, but the level of the additives or their combination needs to be decreased below the level resulting in a characterising flavour. Additives deemed necessary for the manufacture of tobacco products are exempted, *“for example sugar to replace sugar that is lost during the curing process, should be allowed, as long as they do not result in a characterising flavour or increase the addictiveness, toxicity or CMR properties of the product. An independent European advisory panel should assist in such decision making. The application of this Directive should not lead to discrimination between different tobacco varieties, nor should it prevent product differentiation.”*

Although some countries have legislation on tobacco flavour, as described above, no methods are currently available to quantify characterising flavours. The current project aims to develop such a method. The tobacco industry and the food industry in general conduct expert panel and consumer research to test the flavour and other sensory characteristics of a product for purposes of successful product development.

The overall aim of this project is to deliver a test approach for the assessment of characterising flavours in tobacco products, by a combination of sensory profiling, chemical-analytical measurements, other methods, or a combination of methods.

In summary, the project consisted of carrying out four work packages, in the time period August 2014-June 2015:

- WP1: Review the literature, with the aim to draft a review of the current methods and approaches that may be suitable for the objective determination of characterising flavours in tobacco products.
- WP2: Identify concept profiles, with the aim to draft three testing approaches, based on the output of WP1, for sensory analysis of tobacco products (including chemical-analytical measurements) applicable to tobacco product regulation.

- WP3: Optimise and peer-review the procedure, with the aim to test, refine and finalise the methodology that had been proposed based on the outcomes of WP1 and WP2.
- WP4: Assess the feasibility and impact, with the aim to check whether the overall aim, to deliver a method to decide whether a tobacco product imparts a characterising flavour other than tobacco, has been reached.

Based on a review of the relevant literature, and discussions with experts from several fields, it was decided that the best approach for assessing characterising flavours would be a combination of sensory profiling and chemical analysis. More specifically, a combination was proposed of a trained expert panel that assesses odours by smelling tobacco samples, with headspace GC-MS.

To test this approach, performed pilot were experiments with a semi-trained expert panel and headspace GC-MS to test the procedure. Based on the findings, a robust and feasible procedure was proposed for assessment by a sensory expert panel, complemented by chemical analysis, laid down in the main documents and the two SOPs.

It was concluded that an expert panel is a good approach to assess characterising flavours. The pilot in general yielded valid, robust, reliable and reproducible results. Smelling is the preferred starting point, as it captures most of the products with characterising flavours. Products marketed to contain characterising flavours (cherry, menthol, vanilla) can be distinguished from a reference space of 'normal' commercial cigarettes brands and natural tobacco leaves. Chemical analysis with Headspace GC-MS is a suitable method to identify flavour components and additives. The results can be used to build a flavour library. Comparing results of flavoured cigarettes with those of natural tobacco leaves gives an indication whether a component is added.

In the report, detailed conclusions and recommendation are given to set up a reliable method. Overall, the method as performed is a good method to determine characterising odours in tobacco products. The panel was sufficiently trained to differentiate between products. When the panel would have been trained more extensively, panellists should also be able to distinguish between the products that are more similar. The products assumed to have a characterising odour were also defined as such. However, the borderline products fell into the reference space, which may be partly due to the beforementioned fact that the panel was not yet able to distinguish products that were more similar. For the final panel, the borderline products may not cluster together with a group of 'reference' products. In case it is decided that these products should be in the reference space, the clustering settings could be adjusted. In that way, only products that are most different fall outside the reference space. Independent of how well trained the panel is, there may always be borderline products.

The notion of 'clearly noticeable' as described in the definition of a charactering flavour needs yet to be defined. Based on the chosen methodology and as result of the pilot study, the following definition has been used: The panel will generate attributes, in consensus, for a wide range of tobacco products (marketed as flavoured and unflavoured). These attributes are noticeable smells. Then, a product space and reference space is created, thereby determining 'tobacco smell' and its boundaries. It can then be tested whether a product is overall significantly different from the reference space/products. If so, this means the product is perceived as 'different from tobacco', and this can be traced back to significant differences in specific attributes (noticeable smells). If the attribute is rated significantly higher in a specific test product than in the reference products, thus higher than a 'noticeable smell', it can be concluded that this is a 'clearly noticeable smell'.

This method is suitable to assess characterising odours in other tobacco products as well due to the elimination of visual cues. However, an updated or renewed attribute lists need to be established

when using the method for other tobacco products (i.e. other than cigarettes, RYO, and tobacco leaves).

Regarding chemical analysis, the accuracy of headspace GC-MS appeared to be much better than expected, with an error margin of around 5%. Therefore, headspace GC-MS is a suitable method to identify flavour components and additives. It is recommended to use pure tobacco leaves or research cigarettes containing no additives (such as CM7) as baseline for chemical analysis, since commercially available brand cigarettes, such as Cig-RP1, contain additives as well. To determine which additives are present in commercially available products with or without a characterising flavour, it is required to compare these products to leaves. In this way, components present in the leaves are considered reference, while all other components are most likely additives, and can be used to build up a chemical flavour library. When a large flavour library is obtained, patterns can be found. When combined with sensory data, detection limits can be determined and insight can be gained into the perception and intensity of a certain flavour.

## 2 Introduction

The new EU Tobacco Product Directive 2014/40/EU (TPD) foresees that Member States shall prohibit the placing on the market of tobacco products with a characterising flavour<sup>1</sup>. These provisions initially apply to cigarettes and roll-your-own tobacco. For determining whether a tobacco product has a characterising flavour, the TPD requires the Commission to adopt implementing acts laying down uniform rules for the procedures for this determination. In addition, an independent advisory panel shall be established at Union level that can be consulted by Member States and the Commission.

Although some countries have legislation on tobacco flavour, and tobacco industry performs sensory research, no methods are currently available to assess characterising flavours. Therefore, the overall aim of this project is to deliver a test approach for the assessment of characterising odours in tobacco products, by a combination of sensory profiling, chemical-analytical measurements, other methods, or a combination of methods. As such, the project contributes to the development of a validated method to identify characterising flavours in burnt and unburnt tobacco products applicable for regulatory purposes in the EU.

In summary, the project consisted of carrying out four work packages, in the time period august 2014-june 2015:

- WP1: Review the literature, with the aim to draft a review of the current methods and approaches that may be suitable for the objective determination of characterising flavours in tobacco products.
- WP2: Identify concept profiles, with the aim to draft three testing approaches, based on the output of WP1, for sensory analysis of tobacco products (including chemical-analytical measurements) applicable to tobacco product regulation.
- WP3: Optimise and peer-review the procedure, with the aim to test, refine and finalise the methodology that had been proposed based on the outcomes of WP1 and WP2.
- WP4: Assess the feasibility and impact, with the aim to check whether the overall aim, to deliver a method to decide whether a tobacco product imparts a characterising flavour other than tobacco, has been reached.

---

<sup>1</sup> The following definition applies as set in the TPD: 'characterising flavour' means a clearly noticeable smell or taste other than one of tobacco, resulting from an additive or a combination of additives, including, but not limited to, fruit, spice, herbs, alcohol, candy, menthol or vanilla, which is noticeable before or during the consumption of the tobacco product.



### 3 Background

Sensory product characteristics such as taste, aroma, and respiratory tract sensations (mouth feel, impact) are important for smoking satisfaction, product acceptance, and the desire to smoke (1,2). Cigarettes without nicotine reduce tobacco craving, desire to smoke, and tobacco withdrawal symptoms (3-5). Smokers identify taste and aroma as important factors in the pleasure derived from smoking and for their choice of cigarette brand (6-9). For instance, the sweetness of cigarette smoke is closely related to satisfaction and pleasantness (10).

According to SCENIHR, additives considered attractive may lead to brand preference or a higher consumption of tobacco products (11).

Some countries already implemented legislation with respect to flavour regulation while taking different approaches. Brazil (RDC ANVISA Nº 14) and Canada (BILL C-32) banned most flavours altogether. Canadian legislation prohibits the use of many different classes of additives in the manufacture of cigarettes, little cigars, and blunt wraps that contribute to the flavour of tobacco products. First, additives that have flavouring properties or that enhance flavour, e.g. those identified by the Flavour and Extract Manufacturers Association (FEMA) Expert Panel in its lists of GRAS (Generally Recognized as Safe) flavouring substances. Some additives are excluded from this rule, such as the humectants glycerol and propylene glycol, menthol, guar gum, citric acid. Secondly, spices, seasonings and herbs, as well as sugars and sweeteners (excluding starch), are not allowed.

As youth continued to use flavoured products, including cigars, an amendment of June 4 2015 (entering into force 180 days later) also places flavour restrictions on cigars weighing more than 1.4 g, but not more than 6 g and on cigars having physical characteristics similar to those of cigarettes or little cigars. Some Canadian provinces have passed additional legislation on other flavoured tobacco products such as smokeless tobacco, bidis, and pipe tobacco.

Another approach is to set limits on individual flavours. France (Article L3511-2 of the code de la santé publique) had in the past set upper limits to the use of additives in aromatized cigarettes that give smoke a sweetish or acidulated taste. In light of this Article, Decree no. 2009-1764, Art. D. 3511-16 had fixed the use of vanillin and ethyl vanillin at 0.05% of tobacco weight, and sweetener applied at the filter of the cigarette at the analytic detection threshold.

As it is difficult to assess the effect of each additive separately, another solution might be to focus on specific characteristics of the entire product, for instance flavour in tobacco or tobacco smoke. The US Food and Drug Administration (FDA) banned cigarettes containing certain characterising flavours in their Tobacco Control Act (Sec. 907, tobacco product standards): *“...a cigarette or any of its component parts (including the tobacco, filter, or paper) shall not contain, as a constituent (including a smoke constituent) or additive, an artificial or natural flavour (other than tobacco or menthol) or an herb or spice, including strawberry, grape, orange, clove, cinnamon, pineapple, vanilla, coconut, liquorice, cocoa, chocolate, cherry, or coffee, that is a characterising flavour of the tobacco product or tobacco smoke.”*

The new TPD takes a similar approach by foreseeing that Member States shall prohibit the placing on the market of tobacco products with a characterising flavour. These provisions initially apply to cigarettes and roll-your-own tobacco.

Characterising flavour means a *“clearly noticeable smell or taste other than one of tobacco, resulting from an additive or a combination of additives, including, but not limited to, fruit, spice, herb, alcohol, candy, menthol or vanilla, which is noticeable before or during the consumption of the tobacco product.”*

In this definition, 'tobacco' means: *"leaves and other natural processed or unprocessed parts of tobacco plants, including expanded and reconstituted tobacco."* Tobacco products other than cigarettes and RYO are currently exempted from this prohibition relating to ingredients as long as there is no substantial change of circumstances in terms of sales volumes or consumption patterns of young people.

The TPD does not ban additives or combinations of additives resulting in a characterising taste altogether, but the level of the (combination of) additives or the combination needs to be decreased below the level resulting in a characterising flavour.

Although some countries have legislation on tobacco flavour, as described above, no methods are currently available to quantify characterising flavours. The current project aims to develop such a method.

### 3.1 References

1. Levin ED, Behm F, Carnahan E, LeClair R, Shipley R, Rose JE. Clinical trials using ascorbic acid aerosol to aid smoking cessation. *Drug Alcohol Depend.* 1993;33(3):211-23.
2. Rose JE. Nicotine and nonnicotine factors in cigarette addiction. *Psychopharmacology (Berl).* 2006;184(3-4):274-85.
3. Pickworth WB, Fant RV, Nelson RA, Rohrer MS, Henningfield JE. Pharmacodynamic effects of new de-nicotinized cigarettes. *Nicotine Tob Res.* 1999;1(4):357-64.
4. Robinson ML, Houtsmuller EJ, Moolchan ET, Pickworth WB. Placebo cigarettes in smoking research. *Exp Clin Psychopharmacol.* 2000;8(3):326-32.
5. Rose JE, Behm FM, Westman EC, Johnson M. Dissociating nicotine and nonnicotine components of cigarette smoking. *Pharmacol Biochem Behav.* 2000;67(1):71-81.
6. Levin ED, Behm F, Rose JE. The use of flavor in cigarette substitutes. *Drug Alcohol Depend.* 1990;26(2):155-60.
7. Leatherdale ST, Ahmed R, Barisic A, Murnaghan D, Manske S. Cigarette brand preference as a function of price among smoking youth in Canada: are they smoking premium, discount or native brands? *Tob Control.* 2009;18(9):466-73.
8. DiFranza JR, Eddy JJ, Brown LF, Ryan JL, Bogojavlensky A. Tobacco acquisition and cigarette brand selection among youth. *Tob Control.* 1994;3(4):334-8.
9. Talhout R, Sleijffers A, Van Amsterdam JGC, Opperhuizen A. *Wat rookt de Nederlandse jeugd en waarom?* Bilthoven: RIVM, 2009.
10. Jaffe AJ, Glaros AG. Taste dimensions in cigarette discrimination: a multidimensional scaling approach. *Addict Behav.* 1986;11(4):407-13.
11. SCENIHR. *Addictiveness and Attractiveness of Tobacco Additives.* Brussels: European Union, 2010.

## 4 Methodology

In summary, the methodology consists of carrying out four work packages, in the time period August 2014-June 2015:

- WP1: Literature review.
- WP2: Identify concept profiles.
- WP3: Optimised and peer-reviewed procedure.
- WP4: Feasibility and impact assessment.

Below, the methodologies for the four work packages are described.

### 4.1 WP1: Literature review

The aim of WP1 was to draft a **review** of the current methods and approaches for the determination of characterising flavours in tobacco products, which are potentially applicable for standardized control. WP1 was subdivided in six tasks:

- Task 1.1: Development of methodology to retrieve relevant information.
- Task 1.2: Review of methods recorded to be used by the tobacco industry for sensory analysis.
- Task 1.3: Review of methods used for assessment of tobacco products by EU Member States and third countries.
- Task 1.4: Review of methods for sensory analysis of any type of consumer or food products.
- Task 1.5: Review of alternative methods, not based on sensory analysis.
- Task 1.6: Draft and review report.

All consortium members participated in this WP, led by RIVM, and its timeline was M1-M4.

#### 4.1.1 Task 1.1 Development of methodology to retrieve relevant information

To retrieve relevant information scientific literature has been searched, as well as grey literature such as tobacco documents, industry websites, and ISO documents. In addition, regulators and scientists were consulted via a questionnaire. A search strategy was developed to retrieve relevant information dealing with the topic of the report from sources such as scientific literature and tobacco industry documents by using appropriate initial search terms followed by review and selection of relevant papers.

A systematic search was performed based on several keywords or keyword combinations (Appendix I). In addition, with a few key documents a snowball approach was performed to obtain detailed information about a subject. For example, using this approach an article (e.g. Reinbach et al. 2014) on sensory profiling methods contains references to other sources about the same subject (e.g. Lawless et al. 2010, or Le et al. 2008) leading to additional information.

#### 4.1.2 Task 1.2 Review of methods recorded to be used by the tobacco industry for sensory analysis

Although no ISO or CORESTA methods are available for tobacco sensory analysis, tobacco industry has internal standards for sensory research. Apart from scientific literature, tobacco industry sources have been searched.

As appropriate data sources for the literature review, scientific literature databases were used (PubMed and Scopus) as well as tobacco industry sources. Sources such as digital libraries containing tobacco industry documents (tobaccodocuments.org, legacy tobacco documents) were included. These documents not only contain information on advertising, manufacturing, marketing and sales, but also on scientific research activities. Other literature sources on tobacco research and product development that were consulted are websites of tobacco manufacturers, the tobacco research institute and the Fawky Abdallah Company.

#### **4.1.3 Task 1.3 Review of methods used for assessment of tobacco products by EU Member States and third countries**

As additional source to retrieve information on sensory analysis methods for tobacco products, and other methods useful for assessing characterising flavours, a questionnaire (Appendix V) covering these topics has been sent to EU regulators via the email lists of the EU Expert Group on Tobacco policy and the Expert Subgroup on Ingredients. Regarding third countries, a questionnaire was sent to countries that regulate ingredients or have legislation on flavours/flavourings, i.e. the USA, Canada and Brazil, as well as to selected sensory scientists.

Relevant resulting information, for instance papers, personal communications, legislation documents, have been included in the review.

#### **4.1.4 Task 1.4 Review of methods for sensory analysis of any type of consumer or food products**

For literature on methods for sensory analysis of consumer or food products not only scientific literature databases were consulted (§4.1.2), but also EU regulations and ISO standards regarding this topic. ISO methods have been defined for sensory analysis of consumer or food products and can also be applied for tobacco sensory analysis.

#### **4.1.5 Task 1.5 Review of alternative methods, not based on sensory analysis**

As alternative methods, instrumental methods, animal experiments, neuro-imaging studies, and electrophysiological methods, such as EEG (Hummel, Phillip Morris), have been identified. Scientific literature has been consulted, as well as internal tobacco industry sources.

#### **4.1.6 Task 1.6 draft and review report**

The review provides a comprehensive overview of methods used for sensory analysis of tobacco, other consumer or food products, and alternative methods for the determination of characterising flavours. Methods applicable for sensory assessment by both smelling and smoking the tobacco products have been included.

Articles were selected based on their relevance; if both title and abstract seemed relevant, articles were read. Selection criteria for final inclusion of articles in the review are relevance for smoked tobacco products in burnt and/or unburnt form (e.g. smoked or unsmoked cigarettes), quality of data, useful for the purpose of the product, possibility to translate approach into a method suitable for legislation.

For the applicability in standardized protocols and suitability for legislation it would be advantageous if the analysis method of choice was already successfully used for products or even referred to in EU legislation. Therefore, the focus was to select articles on analysis methods that are either currently used to determine the composition of tobacco products or on established methods to determine characterising flavours in other consumer products or food.

For all methods, several aspects such as pros and cons, accuracy, costs, time-investment and ethical issues have been described. Aspects that are applicable to the sensory evaluation of tobacco such as type of assessors, selection of assessors, test facilities, and type of method have been discussed. In the discussion, the main characteristics of each method described are summarized and evaluated.

#### 4.1.6.1 Expert subgroup on ingredients meeting

A first draft of the literature was delivered by 31 October, in order for DG SANTE and Chafea to prepare for the meeting of the expert subgroup on ingredients on 7 November.

The findings were presented during the expert meeting by a HETOC delegation, together with discussion points and our preliminary views on the three concept profiles. First selection of brands to be tested was presented, for comments.

## 4.2 WP2: Identification of the applicable methods

The aim of WP2 was to draft, based on the output of WP1, best practices for sensory analysis of tobacco products (including chemical-analytical measurements) applicable to tobacco product regulation, to assess them and select three concept profiles for the testing approach.

Results of WP2 are the **three concept profiles** selected, as well as the minutes of the dedicated seminar and the input of the participants on these methods. WP2 was subdivided in five tasks:

- Task 2.1: Assessment of the usability/adaptability of sensory analysis.
- Task 2.2: Identification of applicable standards and/or best practices for determination of characterising flavours.
- Task 2.3: Identification of applicable alternative methods for sensory analysis.
- Task 2.4: Identification of three concept profiles for the testing approach.
- Task 2.5: Seminar to discuss the three concept profiles.

All consortium members participated in this WP, led by WUR, and its timeline was M3-M5.

Based on the information gathered in WP1, it was assessed which methods are useful to measure characterising flavours, including their specific strengths and weaknesses. Based on that, the three most promising concept profiles for methods to assess characterising flavours were proposed and these methodologies have been discussed during the seminar.

### 4.2.1 Task 2.1 Assessment of the usability/adaptability of sensory analysis

The applicability of methods for sensory analysis indicated in WP1 was assessed by the criteria in Table 1. Scores were assigned to these criteria. Criteria importance was considered before concluding which methods were suitable to include in the concept profiles.

**Table 1. Weight of criteria used to assess sensory methods**

<b>Most important</b>	<b>Important</b>	<b>Less important</b>	<b>Not very important</b>
Fits Regulatory needs	Specificity	Ease for panellist to perform the method	Required expertise on data analysis
Applicability burnt and unburned tobacco products	Sensitivity	Time and costs	Required hard and software
	Reproducibility		

## 4.2.2 Task 2.2 Identification of applicable standards and/or best practices for determination of characterising flavours

The applicable standards/best practices signalled in WP1 were assessed by the criteria in Table 2 and Table 3. Product utilization (smoking or smelling) and panel characteristics (consumer or expert panel) were assessed. Criteria importance was considered together with the method to be used to detect characteristic flavour in tobacco products, before a final conclusion was made on which product utilization and panel characteristic to choose.

**Table 2. Weight of criteria used to assess ways of product utilization (smoking vs smelling)**

Most important	Important	Less important	Not very important
Fits Regulatory needs	Specificity	Ease for panellist to perform the method	Requirements test facilities
	Sensitivity	Ethical burden	Applicable younger age groups
	Reproducibility		

**Table 3. Weight of criteria used to assess consumer vs expert panels**

Most important	Important	Less important	Not very important
Fits Regulatory needs	Specificity	Time & costs	Suitability weekly & periodical testing
	Sensitivity	Ethical burden	Applicable younger age groups
	Reproducibility		Selection procedure
			Suitability discriminative test
			Suitability CATA, QDA, Spectrum
			Suitability at home testing

## 4.2.3 Task 2.3 Identification of applicable alternative methods for sensory analysis

### 4.2.3.1 Chemical analysis

The applicability of chemical analysis methods for sensory analysis signalled in WP1 was assessed by the criteria in Table 4. Methods that received the lowest score on one of the criteria were not considered as a possible method for the assessment of characteristic flavour in tobacco products. Criteria importance was considered before a final conclusion was made on which method to use to detect characteristic flavour in tobacco products.

**Table 4. Weight of criteria used to assess chemical analysis methods**

Most important	Important	Less important	Not very important
Fits Regulatory needs	Specificity	Labour intensive	Required expertise on data analysis
Applicability burnt and unburned tobacco products	Sensitivity	Sustainability	Required hard and software
	Reproducibility		
	Accuracy		

### 4.2.3.2 Other alternative methods

The applicable alternative methods for sensory analysis signalled in WP1 were assessed by the criteria in Table 5. Methods that received the lowest score on one of the criteria were not considered as a possible method for the assessment of characteristic flavour in tobacco products. All of the

alternative methods received the lowest score on at least one criterion, and therefore no alternative methods were considered as possible method to detect characteristic flavours.

**Table 5. Weight of criteria used to assess alternative methods**

Most important	Important	Less important	Not very important
Fits Regulatory needs	Specificity	Ease for panellist to perform the method	Required expertise on data analysis, and required software
Applicability burnt and unburned tobacco products	Sensitivity		Required hard- software
	Reproducibility		

#### 4.2.4 Task 2.4 Identification of three concept profiles for the testing approach

Three concept profiles have been developed for the testing approach. Several issues have been addressed in the development of these concept profiles, specifics depending on the choice of methods:

- Reference/standard
- What is a criterion for successful detection/identification of the characterising flavour?
- Consumers or Experts
- Age group
- Smoking and/or smelling
- Sensory panels / alternative methods

Setting maximum content levels for those additives or combination of additives that result in characterising flavours is foreseen in the Directive. However, within the current experimental set-up of testing around 5 brands with different characterising flavours, it would only be possible to conclude that a brand has a characterising flavour at the amount present in the brand. To set limits, also brands need to be tested with subtler flavours in order to determine the lowest level at which participants can still detect a flavour.

A clear definition of baseline is also crucial here. One good option is to define a product space in which plain tobacco can be differentiated from tobacco with a different flavour, but this is not dichotomous. There will be a large area with overlap where it can probably not be decided with sufficient power whether there is a relevant difference. Discussions about refinement are only relevant when it is known if the method is able to differentiate and when the resolution is known.

Therefore, a follow-up testing strategy has been proposed for further validation/refinement of the testing approach in WP4. This will include the calibration of reliable standard sensory profiles of various types of products based on a robust sample of relevant tobacco and related products marketed throughout the EU ensuring maximal robustness of the proposed procedure in 'real life' setting and possibly, if needed, to set qualitative or quantitative thresholds when a detected flavour should be considered as characterising.

In the presentation to the meeting of the EU expert subgroup on ingredients on 7 November, the consortium presented its preliminary views regarding the three concept profiles, and performed task 2.1-2.3 (§4.2.1, §4.2.2, §4.2.3) subsequently.

First, conclusions were drawn based on the assessment of the methods performed in these tasks. Based on those conclusions, the best methods were identified and the reasoning behind choosing the three concept profiles were explained.

Next, flow charts were made that depicted all open options within the concept profiles that need to be discussed during the seminar. All steps within the flowchart were explained in the paragraph after the flow chart to clarify the options within each concept profile.

As a last part, the three preliminary detailed profiles (sensory analysis and the chemical analysis methods) were described, explaining:

- Methods
- Procedure
- Validity, reliability and robustness of the results
- Considerations
- Advantages and disadvantages

#### **4.2.5 Task 2.5 Seminar to discuss the three concept profiles**

A seminar was held on 12-13 January, at RIVM, involving experts in sensory analysis and regulators. This seminar was held back to back with the interim meeting. The main purpose of the seminar was to discuss the concept profiles, and their respective strengths and weaknesses. Invitations have been sent, together with the letter of comfort provided by SANTE. Participants received the concept profiles and the literature review before the meeting, on 19 December.

The minutes of the seminar can be found in Annex VII.

### **4.3 WP3: Optimisation of the testing approach on the basis of pilot studies**

The aim of WP3 was to test, refine and finalise the methodology that had been proposed based on the outcomes of WP1 and WP2, based on a pilot study.

Results of WP3 are an optimised and peer-reviewed **study design**, and advice on further research. WP3 was subdivided in four tasks:

- Task 3.1: Selection and possibly adaptation of the profile for pilot testing.
- Task 3.2: Selection of samples for pilot testing.
- Task 3.3: Pilot testing.
- Task 3.4: Optimising the study protocol.

All consortium members participated in this WP, led by WUR, and its timeline was M5-M10.

The Medical Ethical Committee (Dutch: Medisch Ethische Toetsingscommissie, METC) approved of the experiments (Annex VIII).

#### **4.3.1 Task 3.1 Selection and possibly adaptation of the profile for pilot testing**

##### **4.3.1.1 Introduction**

The pilot study was performed according to the pilot proposal (Annex X). The purpose of the pilot was to obtain relevant information about the procedure to assess products with characterising flavours, including test set up, facilities and logistics. This report includes a detailed description of participant selection and screening, experience gained during the pilot including the training sessions, information on the assessment of the panel performance and recommendations for future operation of such a panel include additional considerations (e.g. tobacco product assessment via smoking).



This pilot test was limited in terms of panel training time (7 weeks) and the number of tobacco products that are assessed by the panel (20 products). Because of this, a preliminary list of potential attributes was obtained as well as a preliminary product space as final study results.

#### **4.3.1.2 Study design**

The study consisted of 3 phases:

- phase I: screening of potential candidates
- phase II: training of the selected panel members
- phase III: testing of 20 different tobacco products

A sensory panel assessed odours of unburned tobacco products by rating odour intensity according to an attribute list (generated by the panel). The method as described is a Quantitative Descriptive Analysis sensory method (QDA).

#### **4.3.1.3 Screening**

##### *Screening 1*

Odour recognition skills were assessed with use of flavoured tobacco products. Ten tobacco products were offered to the participants in non-transparent cups to eliminate visual cues. Participants had to smell the sample and indicate the main odour they perceived from a multiple-choice list (in line with the check-all-that-apply (CATA) method). Participants with the highest scores (indicating the odour as described on the package of the tobacco product) were selected (1).

This test was performed at two time points with the same stimuli and given to the participant in the same order but samples were coded differently. For this test almond aroma was added in 4 concentrations to (0.10 mg) RYO-RP2 roll-your-own tobacco. The participant had to rate the 4 samples according to the odour intensity on a line scale with anchors going from 'weak' to 'strong'. The participants that were best able to rate the products in intensity corresponding with the rank order (i.e. the sample with the lowest concentration of the flavour/odour has been rated lowest on the line scale) were selected for the second screening.

##### *Screening 2*

Olfactory function was assessed using an odour stimuli test, the so-called 'Sniffin Sticks' developed by Hummel and colleagues (2). This test consists of pen-like odour dispensing devices with odours considered to be familiar. The smell test contains three parts: an odour discrimination (DIS) and identification test (ID) and odour threshold test (THR). Participants were selected if they passed the smell test, i.e. if they met the normative standards, and were among the 18 best performing panellists (3).

#### 4.3.1.4 Training

The content of each week and training session can be found below.

**Table 6. Training content**

Week	Training session number	Content
1	1,2	Attribute generation through reversed grouping
2	3,4	Practicing with attribute list
3	5,6	Real test exercise, assessment of 10 products in duplicate
4	7,8	Discussion results week 3. Panellist practice rating the difficult products and attributes.
5	9,10	Real test exercise, assessment of 10 products in duplicate (different from products used in week 3)
6	11,12	Discussion results week 5. Panellist practice rating difficult products and attributes.
7	13,14	Panellist practice rating difficult products and attributes

#### 4.3.1.5 Testing

During the test-phase of the study panellists assessed 20 products in triplicate. The products that were assessed can be found in Table 7. The sensory test was performed in sensory booths under standardized conditions (odour free, good lightning) and conducted by the computer program EyeQuestion (Logic8, version 3.7.6). Panel members received the samples in a monadic manner and randomized order. For each sample panel members rated the intensity of 13 odour attributes, as generated by the panel (§5.3.1.4, week 7). Panel members indicated the strength of the odour on a 100 mm visual analogue scale (line scale) with at 10mm and 90 mm of the scale opposite terms (weak and strong). Participants were allowed to smell the sample as often as they liked and rate the attributes at their own pace. Between smelling the different tobaccos, the participants had to smell their arm to neutralize their smell (20 seconds).

**Table 7. Products assessed by the panel**

	Tobacco product	Number	Brands
<b>Products hypothesised to determine the reference space (reference products)</b>	Original filter cigarettes	6	Cig-RP1 (American blend), Cig-RP2, Cig-RP3 , Cig-RP4, Cig-RP5, Cig-RP6
	Roll your own (original)	2	RYO-RP1 (medium strength), RYO-RP2 (full strength)
	Tobacco leave	3	Virginia, burley, oriental
	Filter cigarette without additive	1	Coresta Monitor 6 (CM6)
<b>Products hypothesised to be outside reference space (test products)</b>	Menthol cigarettes	2	Cig-TP1, Cig-TP2
	Vanilla flavoured cigarettes	2	Cig-TP3 Hp 19 (expected low, expected borderline significant), Cig-TP4 (expected high)
	(Other) potentially flavoured cigarettes	3	Cig-TP5, Cig-TP6 (expected borderline significant), Cig-TP7
	Roll your own (flavoured)	1	RYO-TP1 hand rolling vanilla
		Total = 20	

### 4.3.2 Task 3.2 Selection of samples for pilot testing

The sample selection for the pilot testing is described in §4.3.1.

### 4.3.3 Task 3.3 Pilot testing

The pilot study was performed according the pilot proposal described in §4.3.1. The purpose of the pilot was to obtain relevant information about the procedure to assess products with characterising flavours, including test set up, facilities and logistics.

### 4.3.4 Task 3.4 Optimising the study protocol

Based on the results of our pilot studies, the methodology for assessing characterising flavours was optimised, leading to the test approach and methods as described in Annex III, Annex IV, and Annex V.

## 4.4 WP4: Feasibility and impact assessment

The aim of WP4 was to check whether the overall aim, to deliver a method to decide whether a tobacco product imparts a characterising flavour other than tobacco, has been reached. In this work package, an impact analysis is performed on the feasibility, sustainability, costs and benefits of implementing the methodology described in WP3. The method needs to work in practice, has to be cost-efficient and defensible for its instant application for regulatory purposes.

Results of WP4 are the **cost estimation**, and a discussion on the technical feasibility and quality of outcomes. WP4 was subdivided in three tasks:

- Task 4.1: Describing the technical quality and usability of the developed methodology.
- Task 4.2: Quantification of the costs involved applying the developed methodology in practice and proposals for funding models.
- Task 4.3: Possible modifications of the developed methodology and follow-up experiments for further validation will be proposed.

All consortium members participated in this WP, led by RIVM, and its timeline was M7-M10.

### 4.4.1 Task 4.1 Technical feasibility and quality of outcomes

Based on the outcomes of the project, it is described what is needed to gain sufficient quality and robustness of the method to be defensible in the context of regulatory decision-making and subsequent challenges in court. Furthermore, the requirements for technical sustainability are described, such as:

- Maintaining a sufficiently large pool of available panel members
- Maintaining the quality of the panel by training and providing test sessions
- Maintaining other facilities

Here, the quality and robustness of the data is discussed, and further justification is provided regarding the selected methodology.

It is important that the methodology developed fits to the legal requirements of Directive 2014/40/EU and allows for uniform decision-making on products at EU-level.

#### **4.4.2 Task 4.2 quantification of costs and funding model**

Compliance implications such as total costs of the proposed testing method for the regulators and other stakeholders such as industry will depend on regulatory choices. The costs for several options for assessing characterising flavours by sensory analysis combined with chemical analysis were estimated. Regarding procedure, it was assumed that tests would be carried out at one independent testing facility in Europe where all products are tested.

A 'bottom-up' approach was used for cost estimation: cost components and their elements are identified, cost elements are estimated, and estimates are summed to obtain total direct cost. Costs were estimated based on our expert knowledge of the components below, together with additional literature research and seeking advice of other experts, such as the financial department at the RIVM.

- Start-up costs, e.g. buying equipment, software, setting up a test facility/laboratory, training personnel, recruiting, selecting and training a panel. If sensory tests include actual smoking, the test facility cannot be used for other sensory tests such as for food due to third-hand smoke.
- Fixed maintenance costs: e.g. depreciation of equipment, additional training of personnel, additional training and maintaining skills of panel, maintaining testing facility (e.g. hiring, maintenance of building, heat, light, etc.), quality control, overhead.
- Costs per test, e.g. buying samples, panellist reward, other materials.
- Different scenarios for the amount of tests that need to be performed.

Cost estimations have been made for the different methodologies, which allow identification of tobacco products with a characterising flavour: smelling the product using a sensory expert panel, chemical analysis of the product.

The cost calculations have been performed with different sample numbers of products to be tested (for regulatory purposes). To illustrate a possible range it was decided to choose a low number of products to be assessed (10 on average on a yearly basis) and a high number (100 on average on a yearly basis).

Furthermore, costs for several different frequencies of maintenance meetings/testing sessions were considered, such as biweekly and every 3 months, while the detailed cost calculation was performed for monthly meetings.

A questionnaire on cost aspects involved in the sensory analysis of tobacco products was sent to the tobacco industry to gather information on various types of costs involved in sensory analysis (Annex VI). The questionnaire involved several aspects regarding sensory analysis of tobacco products such as costs involved in setting up, training, operating and maintaining a sensory panel.

#### **4.4.3 Task 4.3 describe possible modifications and alternative options**

In this task, it was assessed how the test approaches could be adapted and improved.

## 5 Findings

### 5.1 WP1: Literature review

A systematic literature search has been carried out, in order to describe current methods used for the analysis of flavours in food or consumer products that can be applied for tobacco sensory analysis. In addition, methods used for the analysis of the composition of tobacco products are described. In this literature review not only scientific data sources but also digital libraries providing information on methods used by the tobacco industry and data sources providing information on current European Union regulatory methods were consulted. A draft version of this document was sent to the participants of the seminar, as background document for this meeting. In addition, an adapted version was sent to tobacco industry by DG SANTE to check for completeness of information presented. For comments received from industry, please refer to the minutes of the meeting with stakeholders on procedures to determine tobacco products with characterising flavours, organised by DG Sante: [http://ec.europa.eu/health/tobacco/docs/ev\\_20150306\\_mi\\_en.pdf](http://ec.europa.eu/health/tobacco/docs/ev_20150306_mi_en.pdf)

Overall, standardized methods for the sensory evaluation of flavours or characterising flavours in tobacco products as well as guidance on how to address the evaluation of whether a product has a characterising flavour are still lacking.

In the review, the different methods that are able to provide information on this are described. The formation of test panels is required to provide information on human perception of tobacco product flavour. Aspects relevant to tobacco sensory panels such as type of panel (expert vs. consumer), selection of panel members, test facilities and type of testing method are discussed. Quantitative or qualitative test methods involve detection by difference testing or descriptive testing. Data analysis methods, data reporting and different types of scaling techniques to encompass a range of intensities used to assess and categorize product attributes are described.

In addition to sensory analysis using human subjects, alternative methods are described as well. The methods involve analysis of sensory responses in the brain by neuro-imaging, sensory testing in laboratory animals and chemical analytical analysis of flavour components. In the discussion the main characteristics of each method described have been summarized and evaluated.

For the assessment of characterising flavours several methods are suitable either separately or in combination. As characterising tobacco product flavour requires sensory information from consumers and/or trained assessors, a first step must always be to perform sensory evaluation of tobacco products. In addition, flavour components can be identified and quantified using chemical analysis.

### 5.2 WP2: Identification of the applicable methods

From the assessment of the methods we concluded that the best sensory methods for our research aim are QDA and/or CATA. For QDA, a trained / expert panel is recommended, as consumers typically have difficulties to reliably rate intensity of attributes ratings. For CATA, a consumer panel is preferred, as this is a relatively easy method where no direct information about intensities of the attributes is obtained, and thus no training is necessary.

When setting up and training a sensory expert panel (decided after the seminar), the QDA method will provide information about intensities of (characterising) flavours. Furthermore, an expert panel is useful to generate relevant vocabulary/attributes, and to obtain a product space. A trained panel scores higher on the criterion of suitability for regulatory needs. A trained panel can be seen as a calibrated machine/measure, as we can control panel performance by checking discriminability,

repeatability and group consensus for the panel as a whole and for panellists and attributes (i.e. flavours) specifically.

When setting up a sensory consumer panel, the CATA method will provide information about which (characterising) flavours are present in a tobacco product. Furthermore, a consumer panel gives information about how 'regular consumers' (either smokers or non-smokers) perceive the flavour or odour of tobacco products. When panel members from different nationalities are included, cultural differences can be taken into account.

Consumer information may be used to confirm/complement findings of an expert panel ("do untrained consumers also perceive these flavours?"). When a consumer panel is used as a first approach, vocabulary/ attributes could be derived from flavourists (i.e. by flavour chemist, someone who uses chemistry to engineer flavours) instead of generated through the expert panel.

In addition, it was concluded that chemical analysis is only useful once data from a sensory panel is combined with analytical data and upper limits of specific flavourants are determined.

When upper limits are set, chemical analysis can be performed as a first step or screening which has the advantage that fewer tests need to be performed with expensive expert and consumer panels. When the flavour library is established, sensory panels still need to be used in cases whereby the upper limit of flavour compounds is not exceeded. This is necessary because a product may impart new characterising flavours for which no upper limits are established yet.

Whether the tobacco products should be assessed by smelling, or smoking, or both, depends on regulatory needs, and the composition of the tobacco products (e.g. the presence of products with characterising flavours that would escape detection if assessment were carried out by smelling only). For ethical reasons, adolescents can only smell the tobacco products. However, assessment through smelling only does not fully reflect the way tobacco products are used; therefore, an adult consumer panel smoking the cigarettes could be an option. In case the consumers are adults, tests can be performed at home.

Experts can assess the cigarettes through smelling (the way the consumer first comes in contact with the tobacco product) and through smoking (how the product is used).

We also concluded that chemical analysis is preferred to set upper levels for additives or combination thereof causing the characterising flavour. Headspace GC-MS would be the best method to determine flavour compounds in tobacco products. Headspace GC-MS is not the most accurate method (error margin 10-15%), but is less labour intensive and it is more easy to reproduce results, which is therefore most suitable for regulatory needs, in contrast to the DS GC-MS method. To obtain the most accurate amount of flavour compounds, DS GC-MS (dissolvent liquid injections) can be performed as second option in case of results that fall within the error margin of the headspace GC-MS method. Upper limits can be derived when chemical analysis data is combined with sensory (expert panel) data.

We presented the aforementioned options during the seminar, with various experts present, and came to similar conclusions about usefulness, (dis)advantages and feasibility of the proposed methods and approaches.

During the seminar it was mentioned that using adolescents as a test population is not necessary as the perception of the flavours does not differ between adults and adolescents. Furthermore, using adolescents the tobacco products cannot be assessed by smoking due to ethical constraints. The assessment of the tobacco products should always be done under standardized conditions, therefore

at home assessments by consumers is not advisable. A consumer panel and expert panel are both valid and reliable panels to assess the tobacco products on characterising flavours. Using an expert panel has as an advantage that the assessments can easily be repeated (without recruiting a large consumer panel), while a disadvantage of an expert panel are the maintenance sessions that need to be held in order for the panel to perform well and have reliable test results. Using a consumer panel has as advantage that it does not require a start-up period in which panellist are screened and trained furthermore it does not require maintenance sessions. The disadvantage of a consumer panel is that intensity of the flavours is not assessed in a direct manner.

As we anticipated that some type of panel study would have to be performed in WP3, we already filed a Medical Ethical Committee application at the start of the project (Annex VIII).

### **5.3 WP3: Optimisation of testing – sensory research**

#### **5.3.1 Results of the pilot study**

##### **5.3.1.1 Recruitment**

Participants were recruited through the participant data base, Facebook and twitter account of the human nutrition department of Wageningen University (Annex X). 163 respondents replied by email, and one by phone to subscribe for the information meeting. People who had shown interest were send an email with an invitation to subscribe for the information meeting and were asked to check the in- and exclusion criteria from the study to determine their eligibility. In addition, they were asked to check the training schedule to see whether they were available on those days and times. From the initial 164 respondents, 72 came to the information meeting and signed informed consent by which they declared to be informed about the study's purpose and content and agreed on participation. From the 72 people who came to the information meeting 71 were eligible according to the in- and exclusion criteria, and were invited for the first screening round. For the first screening round 64 people were included (7 participants did not subscribe for the screening or did not show up). For the second screening we selected the best 28 participants based on the scores of the first round (§4.3.1.3). Based on the total score of the second screening we selected the 18 best participants to take part in the panel. For an overview of the recruitment steps see Figure 1.

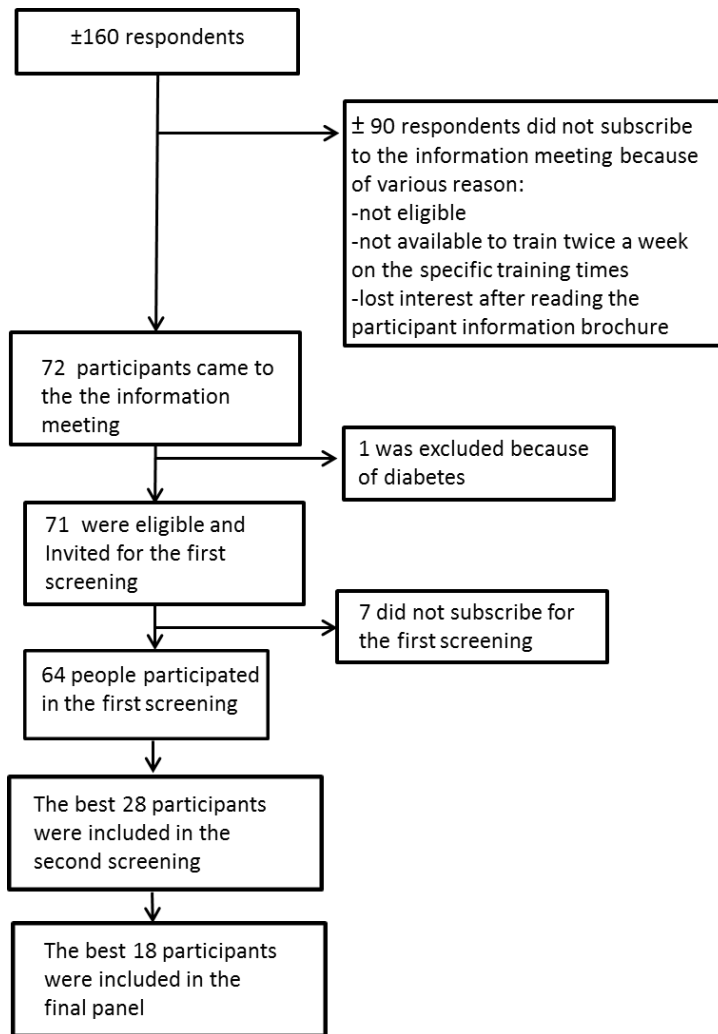


Figure 1. Recruitment flowchart



### 5.3.1.2 Screening

#### 5.3.1.2.1 First screening round

##### Participants orientated results

During the first screening participants had to indicate intensities of 4 samples on a line scale in duplicate. The rank ordered intensity scores had to correspond with the rank order strengths of the four samples. For all intensities correctly indicated, participants obtained a point. In addition, participants had to indicate odours from a multiple-choice list they perceived when smelling 10 samples with and without flavours. Participant did this twice, once when smelling the cigarette stick and once smelling a glass bottle in which the tobacco from the cigarette (without filter and paper) was placed. This was done to see whether odours were perceived differently or stronger when smelling in a blinded way or when smelling the cigarette stick directly.

Results can be found in Annex XII. For every correctly indicated odour participants received a point. Scores of both tests were combined. The maximum score for both test combined was eighteen; however, the highest score obtained by the participants was 16 (correctly identified 8 out of 10 products and correctly identified the strength for each concentration twice 8). Minimum score was zero however lowest score obtained by the participants was a score of 5. When selecting the best 28 out of 64, the lowest score of the 28 best participants was (the cut-off) was 11 points (see Table 8). However, 5 participants who initially where among the 18 highest scoring participants were excluded based on being an outlier (see Figure 2) for the intensity scores. Outliers where the participants who deviated once or twice the standard deviation from average, in Figure 2 these participants are indicated by a 'star' or 'circle' (TabakD8, D2, D20, D14, D7, D9). Because intensity ratings were done twice, participants can be an outlier twice.

**Table 8. Number of scoring points obtained in the first screening round**

	Min. (points)	Max. (points)	Mean	SD
<b>Intensity scores</b>	0	8	5	2
<b>CATA scores</b>	4	9	6	1.3
<b>Total score (CATA and Intensity combined)</b>	5	16	11	2.6

##### Product orientated results

For the intensity-scoring part of the first screening round almond aroma was added in 4 concentrations (0 ml, 0.5 ml, 2 ml, and 4 ml) to (0.10 mg) RYO-RP2 roll-your-own tobacco. The concentrations added aroma were based on a pilot experience (n=6). The participants were asked to rate the 4 samples according to the odour intensity on a 100 mm line scale with anchors going from 'weak' to 'strong'. ANOVA multiple-comparisons post-hoc test; least significant difference (LSD) was performed to see if participants received a differences between the four strengths. Participants were able to significantly ( $p < 0.001$ ) distinguish between the first 3 strengths but were not able to make a distinction between the third and the fourth strength ( $p = 0.234$ ). Number of times the panellist indicated the right intensity (corresponding with the rank order) and indicated the correct odour from the multiple choice list can be seen in Figure 3 and Figure 4.

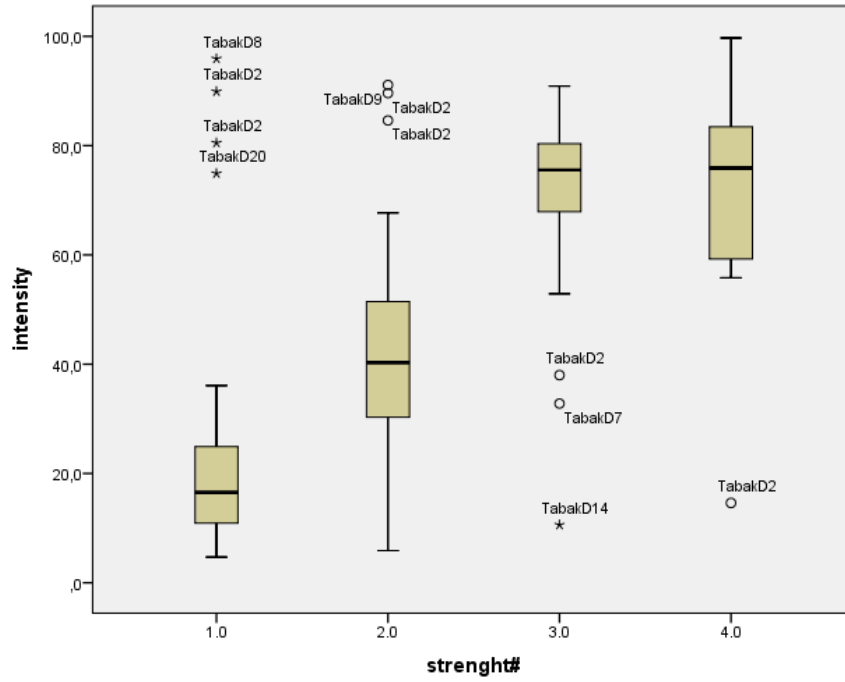


Figure 2. Box plots of intensity scores. For each strength (one is no added aroma, four contained the most added aroma), outliers are indicated with a circle (deviates one time the standard deviation) or star (deviates twice the standard deviation).

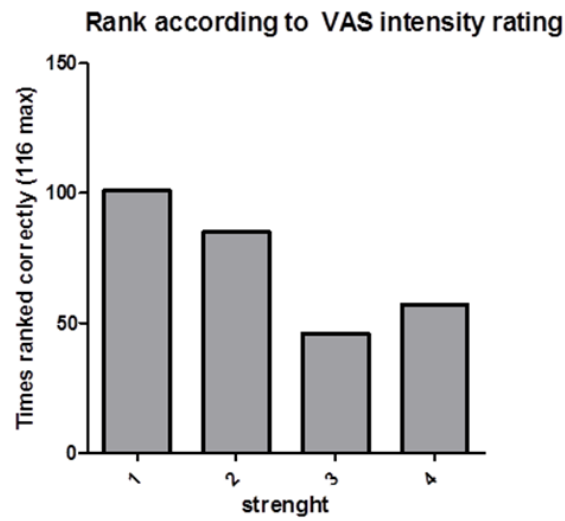
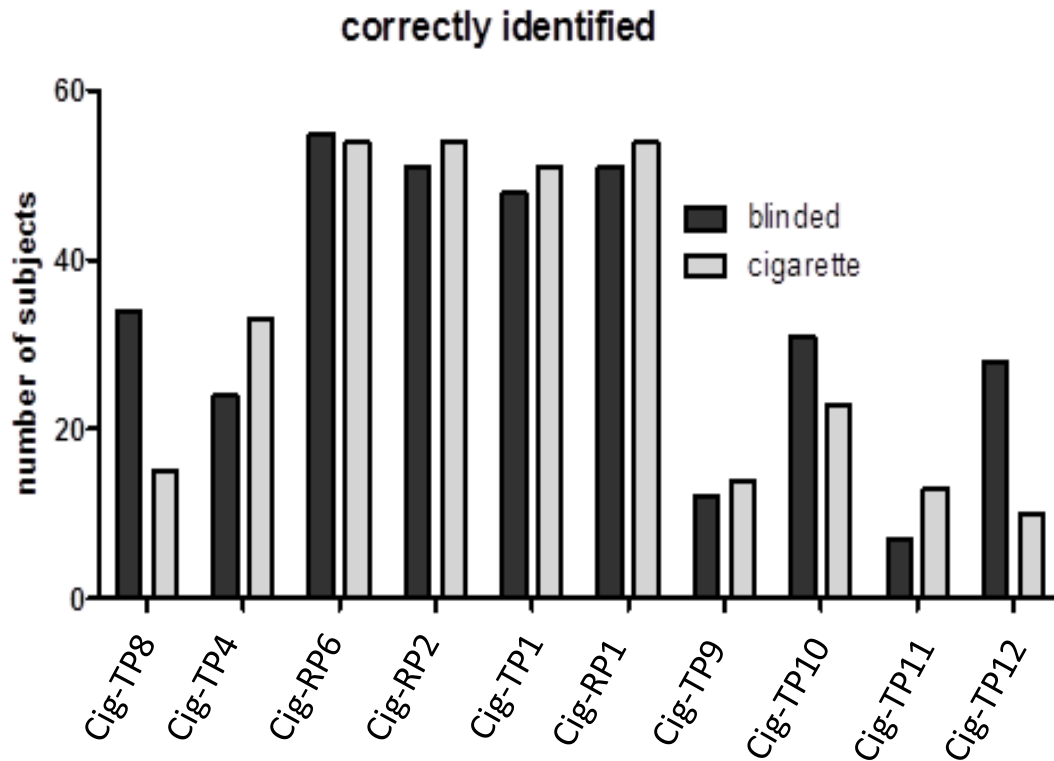


Figure 3. Number of times participants correctly indicated the rank order per strength (strength 1 is no added flavour; strength 4 contains the most added flavour). Max is 116, as 64 participants rated all intensities twice.



**Figure 4.** Number of participants that identified the odour mentioned on the cigarette package. For the products that did not have a flavour indicated on the package the assessment was considered correct if participants did not identify a flavour (such as mango, strawberry, vanilla, apple, menthol). Results are shown for smelling the cigarette stick directly (grey bars) or smelling a glass bottle in which the tobacco was put without cigarette filter or paper (black bars). The limitations of this method must be underlined, in particular the fact that the labelling as such cannot be considered a decisive factor for determining whether a product has characterising flavour or not.

#### **5.3.1.2.2 Second screening round**

From the 28 participants who participated in the second round we selected the best 18 based on the combined score for the Sniffin' Stick threshold, discrimination and identification test provided that participants had normative scores, which are above the 10<sup>th</sup> percentile for age and gender (see Annex IV for explanation of the Sniffin' Stick test and the normative scores).

The maximum score of the threshold test that can be obtained is 16, the highest score among the 28 participants was 14.75 and the lowest threshold was 7.75. However, the low threshold score belonged to the panel member who originally was excluded, but re-included (as the 19<sup>th</sup> best) because a participant that dropped out due to scheduling issues (see Table 9). Not taken into account this panel member, the lowest threshold was 9.25, which is equal to the 50<sup>th</sup> percentile of the best performing age and gender group (female 16-35).

For the discrimination and identification score the highest score that could be obtained was for both tests a score of 16. Among the 28<sup>th</sup> participants the highest scores were 16 and 15 for the discrimination and identification test respectively. The lowest discrimination score of the panellists that were included in the panel was 13, this score was equal to the lowest identification score.

Combining the scores of the three tests the highest score was 41.8 and the lowest score was 35.7, maximum score that could be obtained for this test was 47 (see Table 9). All panellists were among or above the 50<sup>th</sup> percentile for normative scores for age and gender.

**Table 9. Sniffin' Stick test scores of the 28 participants**

	Min. (points)	Max. (points)	Mean	SD
<b>Threshold test</b>	7.25	14.75	9.5	1.5
<b>Discrimination test</b>	10	16	14	1.5
<b>Identification test</b>	10	15	13	1.0
<b>Total score (threshold, discrimination, identification combined)</b>	32.5	41.8	36.8	2.2

### 5.3.1.3 Panel characteristics

The youngest panellist was 18 years old and the oldest 52. Most panel members were between 20-25 years old, student and non- smoker (see Table 10). The fact that few smokers were included may be due to the exclusion of smokers because of the extensive screening on smell ability and this is known to be decreased in smokers, however only 4 smokers were screened in total.

**Table 10. Panel characteristics**

<b>Mean Age (years) ± SD</b>	24 ± 9
<b>Male (n)</b>	3
<b>Female (n)</b>	15
<b>Number of smokers (n)</b>	2
<b>Number of people with hay fever (n)</b>	1

### 5.3.1.4 Training specifics and outcomes

Training sessions were scheduled twice a week. The eighteen panellists were divided over two groups due to scheduling issues, as there was no time and day during which all panellists were available and because of the availability of one panel leader, which made following group discussions too difficult if the group trained as a whole. For a detailed description of the course of the training, see Annex XI. During the first week, both groups received a general introduction to the activities and focussed on the attribute generation. During the second week, both groups scored intensities of all attributes for 5 products, using + and – signs, and the list of generated attributes was shortened.

The preliminary attribute list that was used for the training test sessions of week 3 can be found in Annex XII.

#### Week 3: Training 5 and 6, performance test session.

During these training sessions panellists assessed the products individually in the sensory booths similar to a real-test situation. Panellists assessed 10 products per training session, and testing was performed in duplicate. The results of the panel performance can be found in Annex XII. In general, it was found that:

- The panel discriminated and was in consensus for most attributes but not for the odours: rooibos, liquorice, coconut, honey, flowers, leather, salty, dried fruit, old wet hay, almond and coffee. Possible reasons are that 1) they did not understand the odour attribute; 2) the odour is not present in any of the tested tobacco products.
- Panellists limit their ratings to the extremes of the scale rather than using the whole scale, as instructed. Therefore, in the next weeks, scales were used with anchors on the outer sides (0 mm, 100mm) of the line-scale.

#### Week 4: Training 7 and 8.

During week 4, both groups ranked products and attributes difficult to assess as determined during week 3. In addition, odour attributes were further clarified using examples of odours of real (food) products, called references.

Week 5: Training 9 and 10, performance test session.

During these training sessions panellists assessed the products individually in the sensory booths similar to a real-test situation. Panellists assessed the other 10 products (not assessed during week 3). Testing was done in duplicates. Results of panel performance can be found in Annex XII.

It was found that some of the attributes were not used to describe the products (product column), i.e. are not significant. However, it should be kept in mind that this is not only dependent of the panel performance but also on the products used. For example, red fruit odour has not been used, which makes sense because Cig-TP5 is most typical for this odour and that product was not assessed during these training test sessions but was assessed during week 3.

Week 6: Training 11 and 12.

Based on the panel performance of week 5 this trainings' focus was on the attributes: tobacco, hay, old wet hay, tea, honey and wood as least consensus was wound for these attributes. References or examples of odours (real products) were used for each attribute besides tobacco. Strength was indicated using + and – signs of each odour for 5 products.

Week 7: Training 13 and training 14

Both groups assessed and compared in week 7 all twenty products. Furthermore, the attribute list was finalised.

Final attribute list.

<i>Smokey/burned</i>	<i>Honey</i>	<i>Clove</i>
<i>Vanilla/caramel</i>	<i>Liquorice</i>	
<i>Coconut</i>	<i>Hay</i>	
<i>Chocolate/cocoa</i>	<i>Red fruit</i>	
<i>Nutty</i>	<i>Menthol/mint</i>	
<i>Raisins</i>	<i>Tea</i>	

**5.3.1.5 Final test results**

Statistics were performed with the statistical software R, with use of the packages FactoMineR and SensoMineR. R-scripts can be found in Annex XII.

**5.3.1.5.1 Panel Performance**

A mixed model ANOVA was used to determine panel performance (for more explanation see Annex IV).

In general, the global F-test of ANOVA model: *attribute intensity ~ Product + panellist + Session + Product:panellist + Product:Session + panellist:Session*

Results following the ANOVA test can be found in Table 11.

If the **product effect** is significant ( $p < 0.05$ ), it means that the intensity values of a specific attribute, evaluated by the panellists, are significantly different between products, therefore the panel has been able to discriminate the products with respect to that specific sensory attribute.

The **Panellist effect** is almost always significant in sensory analysis. This means that, although panellists receive training to rate each descriptor, they do not use the scale identically. These individual differences are taken into account by the ANOVA model.

In conclusion, while individual differences in scale use exist, they are statistically controlled.

Concerning the **session effect**, the situation desired would be not having any significant effect, which means that from one session to another the panel would assess the products in the same way.

In conclusion, the session effect is only significant for hay and raisin, meaning that all other attributes are assessed in a consistent and repeatable manner, so those would be the attributes that you focus on if you would continue training the panel

If the **interaction product:panellist** is significant, there is no consensus among the panel to evaluate each product. This can be due to two different situations:

- a) Panellists disagree on the relative intensity the products classification: a trend cannot be pointed out.
- b) The gap of ratings between two judges is different from one product to another: some give always higher marks than others but they all rank the products in the same way. If the interaction **product:panellist** is not significant ( $p > 0.05$ ), there is a panel consensus to the assessment parameters, and thus the **panel is reproducible**.

In conclusion, this means that in our case panellist do not agree on the intensity of the odours, so odour intensity A is not perceived equally intense by all the panellist. However, this is not problematic and common for the QDA method. In our case, we are interested in the relative, i.e. attribute-specific, difference between the products (within subject assessment). Therefore, as long as panellist A rates product X more intense than product Y, and this is done similarly by the whole panel the real or true intensity is of less importance.

To study the **panel repeatability**, the interaction **panellist:session** can be considered. If it is significant, it indicates that some panel members use the scale different from one session to the other.

In conclusion, panellists do not have the same mean for clove, coconut and tea but is repeatable for all the other attributes (this is repeatability in judging the odour descriptors).

If the **product:session** interaction is significant, it means that the intensity values of the attribute evaluated are significantly different depending on the type of product, and on the sessions when it was evaluated, therefore the panel is not repeatable from one session to another.

In conclusion, the performance of the tobacco panel is repeatable for all attributes except for vanilla odour.

Overall, the panel performance in terms of repeatability and discriminability is satisfactory.

#### **5.3.1.5.2 Results test-products**

##### Adjusted means

In Table 12, the adjusted means for all odour intensities per product can be found. All odours attributes with an intensity of 50 or higher are coloured yellow. These are most typical odours for this product relatively to all the other products. For example, Cig-TP5 has an intense red fruit and clove odour compared to all the odour products. Similarly, RYO-RP2 has an intense smokey/burned odour. Indicated in red are the products that were hypothesised to be outside the reference space, i.e. hypothesised to be products with a characterising flavour (see also Table 7).

**Table 11. Results of the ANOVA Model showing the significances of the F-tests for attribute (Y), product, panellist session and their interaction. In dark grey the significant ( $p < 0.05$ ) outputs are indicated**

	product	panellist	session	product:panellist	product:session	panellist:session
Glove	1.599e-161	1.799e-30	0.3318	8.004e-82	0.8549	0.02375
Mint	5.304e-152	8.543e-13	0.7492	1.651e-12	0.6971	0.1675
Redfruit	4.19e-146	9.307e-70	0.3	4.661e-34	0.348	0.9915
smokeBurn	2.152e-121	2.441e-58	0.1911	1.633e-20	0.5162	0.05131
Vanilla	2.356e-90	8.456e-29	0.2943	8.728e-13	0.0003199	0.3414
Coco	2.178e-57	1.94e-40	0.4481	9.833e-26	0.3507	0.02816
Hay	1.767e-54	6.589e-55	0.02899	4.527e-26	0.3271	0.5393
Cacao	5.991e-41	8.702e-12	0.2021	9.589e-15	0.3205	0.1559
Tea	2e-40	4.4e-31	0.07091	4.519e-20	0.5387	0.04877
Raisin	2.858e-30	1.427e-11	0.008456	5.796e-08	0.6345	0.2873
liquorice	5.083e-25	1.183e-49	0.6091	5.642e-18	0.9572	0.8652
Nuty	2.835e-24	2.393e-22	0.5283	0.3365	0.8885	0.3877
Honey	3.397e-13	8.344e-24	0.09599	0.001641	0.7773	0.07005

EUROPEAN COMMISSION

Table 12. Adjusted means for the intensity scores as indicated by the panel for every product and every odour attribute on a line scale from 0-100 mm

Product	Attribute												
	Smokey burned	Vanilla	Coconut	Cocoa	Nutty	Raisin	Honey	Liquorice	Hay	Red fruit	Mint	Tea	Clove
<b>Cig-TP2</b>	8.15	6.12	1.78	5.70	3.14	9.87	8.05	37.72	18.90	3.10	71.71	10.17	1.84
<b>Cig-TP7</b>	8.68	23.29	8.4	19.85	10.87	19.81	13.74	7.99	29.50	2.31	4.53	28.5	1.0
<b>Cig-TP4</b>	6.35	89.34	63.8	17.19	5.41	2.87	9.26	2.66	7.35	2.58	0.97	6.20	2.38
<b>Burley</b>	29.40	4.78	2.82	20.89	9.306	6.82	3.70	5.55	69.28	1.16	9.19	8.80	1.50
<b>Cig-RP6</b>	8.84	30.16	10.72	46.64	12.15	19.14	15.72	13.72	27.94	3.19	3.02	19.50	1.29
<b>Coresta monitor</b>	8.98	6.206	3.53	6.6	4.50	22.61	16.50	4.33	36.99	3.87	2.29	63.92	2.48
<b>Cig-TP5</b>	1.14	5.12	2.42	1.79	4.0	6.66	3.20	1.95	2.44	92.66	1.86	8.84	90.1
<b>RYO-RP1</b>	91.26	6.09	1.29	3.09	4.18	2.09	2.09	4.30	18.17	0.86	2.19	2.99	1.61
<b>Cig-RP2</b>	11.76	7.97	4.95	23.60	10.82	19.08	13.93	15.42	46.16	4.77	8.22	21.781	1.69
<b>Cig-TP6</b>	8.42	23.23	9.73	37.63	13.82	17.57	21.50	10.88	25.95	1.97	4.46	22.10	1.87
<b>Cig-RP3</b>	35.68	7.93	4.02	16.24	10.03	10.77	7.77	10.78	60.05	2.84	2.10	12.55	2.02
<b>RYO-TP1</b>	37.60	70.60	15.9	10.60	5.76	7.60	7.32	3.01	12.50	9.3	3.6	5.8	4.6
<b>Cig-RP4</b>	5.9	37.17	12.11	51.08	14.14	10.98	16.30	12.02	21.70	3.39	2.65	14.01	1.45
<b>Cig-TP3</b>	9.73	13.19	6.27	23.85	8.33	19.587	14.26	12.58	41.50	2.44	3.84	23.82	1.46
<b>Cig-TP1</b>	3.15	4.02	2.64	4.89	2.26	3.67	5.32	32.27	8.63	2.86	90.57	5.16	1.26
<b>Cig-RP1</b>	10.53	18.94	8.20	51.37	28.72	17.42	13.17	13.02	36.2	1.7	1.97	12.29	1.76
<b>Oriental</b>	9.87	5.08	3.40	14.25	7.57	25.11	9.73	6.45	70.27	3.34	2.67	15.47	2.06
<b>Cig-RP5</b>	7.27	28.40	8.50	34.17	10.6	18.4	16.52	20.40	25.35	2.41	9.08	17.41	2.00
<b>Virginia</b>	12.97	6.03	1.75	8.43	4.426	47.86	18.85	4.75	49.6	5.16	2.01	25.84	1.4
<b>RYO-RP2</b>	91.56	6.54	1.13	3.3	3.43	2.5	1.78	3.62	16.25	1.09	1.69	3.32	1.52



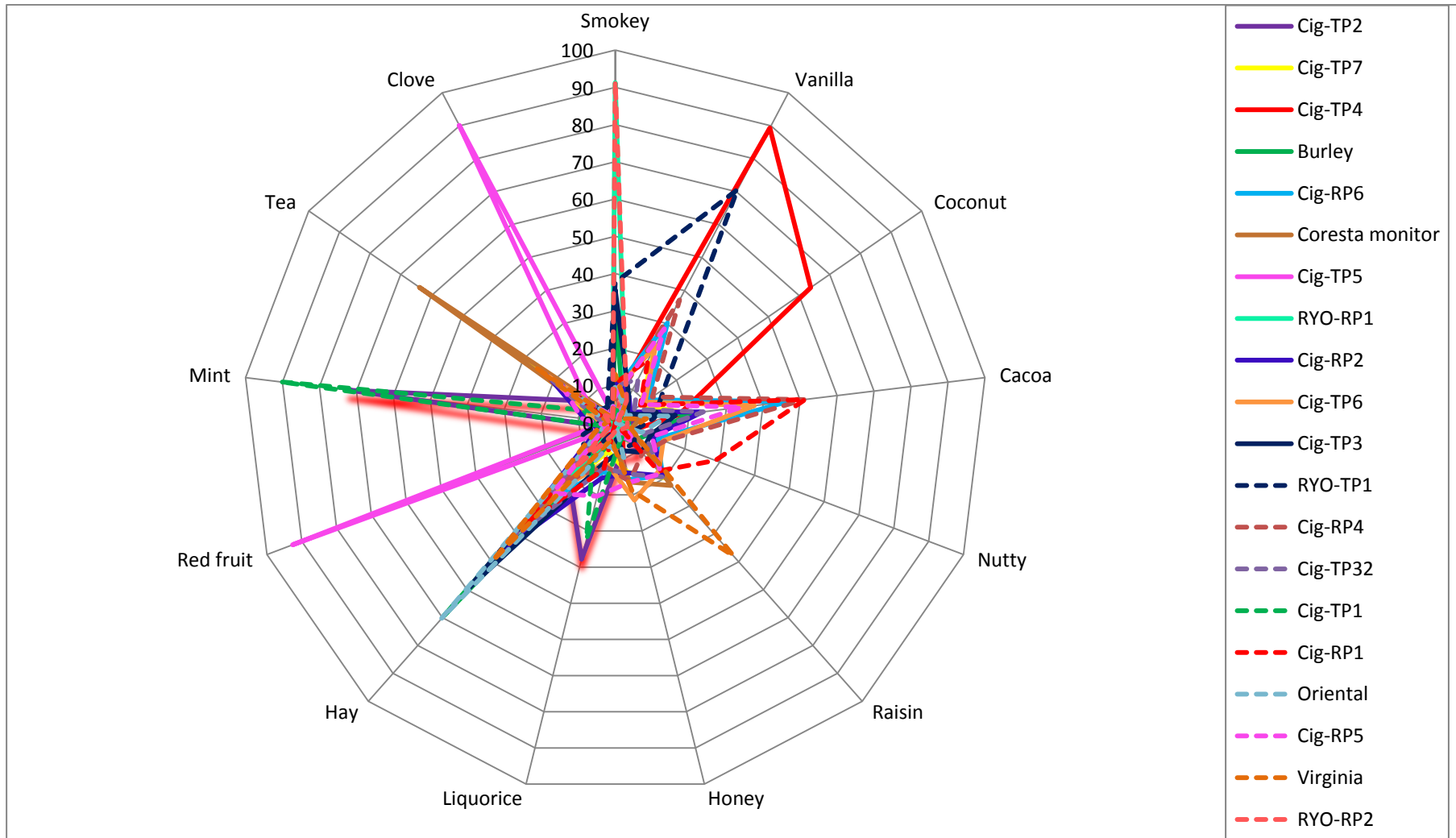


Figure 5. Spider plot of data shown in Table 12

Principal component analysis (PCA)

With the use of MANOVA analysis (see Appendix III. Screening form tobacco odour recognition), we checked whether the attributes were used to describe the products. Only attributes that distinguish between the products are taken into account in the PCA analysis, when a panel is fully trained and calibrated, generally, all attributes are significant as can be seen in Table 13. This table shows the Vtest output and the belonging P-values for each odour attribute. All odour attributes are significantly used by the panel to indicate difference between the products.

**Table 13. MANOVA output, attribute discrimination ability**

Attribute	Vtest	P-value
<b>Red fruit</b>	Infinitive	0.00
<b>Mint</b>	Infinitive	0.00
<b>Clove</b>	Infinitive	0.00
<b>Smokey/Burned</b>	36.60	<0.001
<b>Vanilla</b>	29.47	<0.001
<b>Coconut</b>	24.19	<0.001
<b>Hay</b>	23.62	<0.001
<b>Tea</b>	19.43	<0.001
<b>Cocoa</b>	18.99	<0.001
<b>Raisin</b>	15.10	<0.001
<b>Liquorice</b>	15.05	<0.001
<b>Nutty</b>	8.71	<0.001

We obtained the following product maps by performing PCA with the first and second principal component (or dimensions), and the second and third principal components (see Figure 6, Figure 8). The corresponding factor maps are shown in Figure 7 and Figure 9.

Principal component 1 and 2

Principal component 1 is mainly determined by the attributes red fruit, clove and smoke/burned. These attributes are most specific for the products Cig-TP5 (red fruit, clove) and RYO-RP2 and RYO-RP1 (smoke/burned). The second principal component is driven by the attributes coconut and vanilla, which are attributes that are most specific for Cig-TP4 and RYO-TP1. RYO-TP1 is also rated intense for smokey odour, which is why the product lays in-between the first and second principal component (Figure 6 and Figure 7). Most other products cluster in the centre of the map, meaning that none of the attributes is really specific for this group of products or that the differences between these products are explained by other principal components.

Principal component 2 and 3

The second principal component is of course still driven by the attributes coconut and vanilla; however now only Cig-TP4 is specific for these attributes as RYO-TP1 is within the group cluster as it is more explained by the smoked attribute which is not explained by neither the 2<sup>nd</sup> or 3<sup>rd</sup> component which are depicted on this product map. The third component is driven by red fruit, clove on one extreme, menthol, and liquorice on the other extreme. Cig-TP5 is again most specific for red fruit and clove. Cig-TP2 and Cig-TP1 are most specific for menthol and liquorice (see Figure 8 and Figure 9). Most other products cluster in the centre of the map, meaning that none of the attributes is really specific for this group of products or that the differences between these products are explained by other principal components.

### Eigenvalues of the PCA

In Table 14 the eigenvalues of each dimension of the principal component product space are shown. This is the amount of explained variance per dimension or principal component. In this case, up to 5 principal components explain most of the variance in the data, meaning, that there are 5 products or odours that are so different or distinguishable that they cannot be merged together with other principal components. In the product spaces below (Figure 6, Figure 7, Figure 8, and Figure 9) dimensions 1, 2, and dimensions 2, 3 are depicted in the product space.

**Table 14. Eigenvalues of each dimension of the principal component product space**

	Dimension								
	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9
<b>Variance %</b>	28.28	19.31	17.23	14.80	10.25	4.61	2.87	1.12	0.76

### Hierarchical cluster analysis

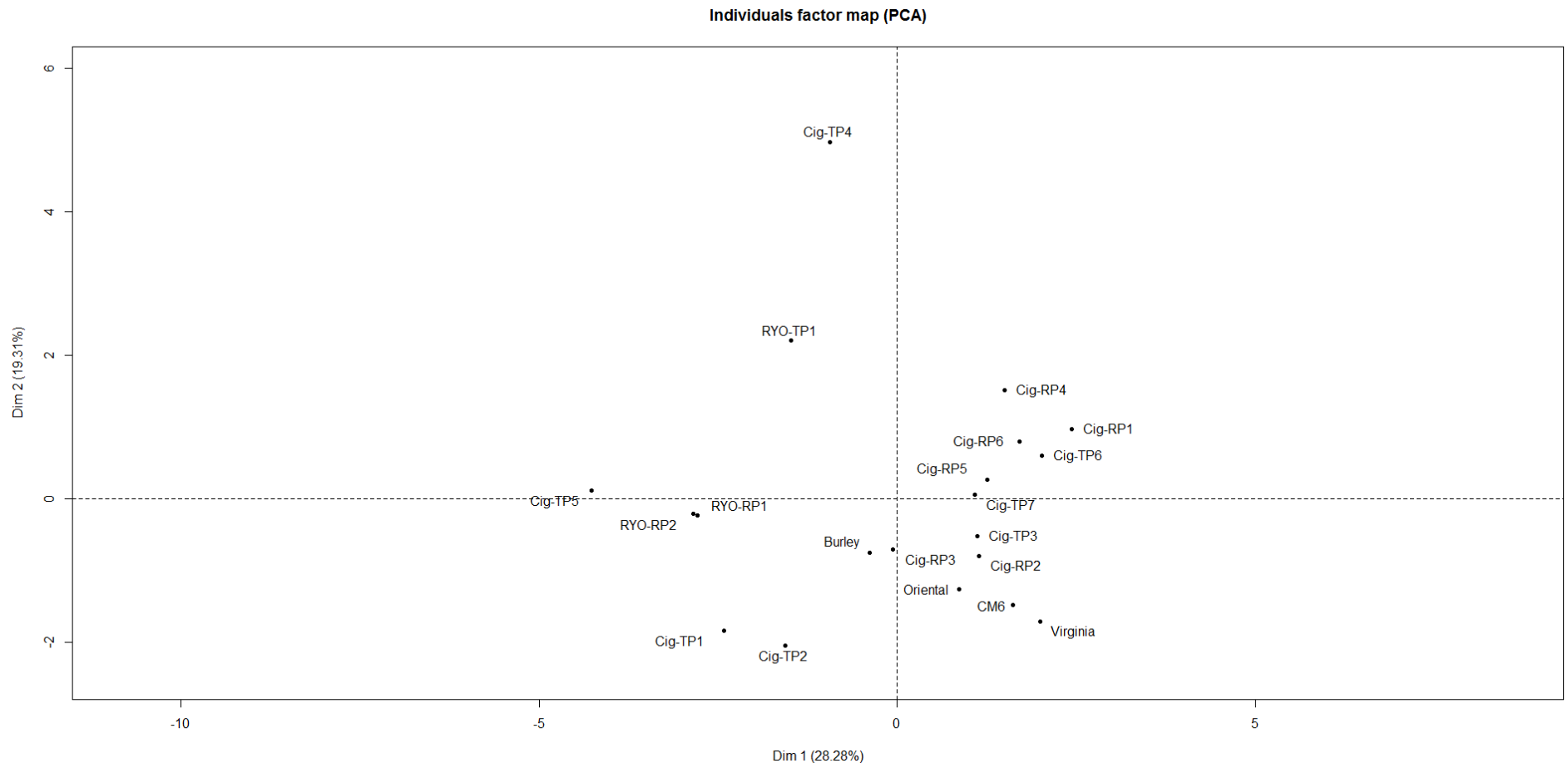
Figure 10 and Figure 11 show the results of the hierarchical cluster analysis. Products with the same colour belong to the same cluster. In the table next to the product space the overall means and standard deviations are shown together with the mean and standard deviation specific for that cluster (i.e. based on which attributes these clusters are formed).

Figure 10 shows four clusters. The first cluster consists of only the Cig-TP5 cigarette brand. This cluster is based on the clove, and red fruit attributes that are significantly different compared with all other clusters. Cluster number 2 consists of Cig-TP4 and RYO-TP1 and the two roll your own (RYO-RP1, RYO-RP2). This cluster is based on the attributes vanilla, coconut, and smokey/burned odour. Vanilla and smokey/burned attributes are together in one cluster due to RYO-TP1Cig-RP4, which scores, intense for both attributes. Cluster number 3, is based on the attributes menthol and Liquorice and Cig-TP1 and Cig-TP2 are specific for these attributes and therefore fall into the same cluster. The fourth cluster constitutes of all the other products.

Figure 11 shows six clusters. In this case, again the first cluster is formed by only the Cig-TP5 and is due to the attributes clove and red fruit. The second cluster is formed by the menthol products, and the third cluster is formed by products that are intense for the smokey/burned attribute. The fourth cluster consists of products that are intense for the vanilla/coconut attributes. The fifth cluster constitutes of products that are Virginia like, and products that are typical American blend products form the sixth cluster.

### Confidence ellipses

Figure 12 and Figure 13 show confidence ellipses around the products; the better the panel is trained the smaller the confidence ellipses will be and the better the panel is able to distinguish between products. In both plots the following products are significantly ( $p < 0.05$  and  $p < 0.01$ ) different from all other products: Cig-TP5, Cig-TP4, RYO-TP1, Cig-TP1 and Cig-TP2. In addition, the RYO products; RYO-RP1 and RYO-RP2 form their own cluster meaning that RYO products are too different from cigarettes (at least for the brands we have assessed during this pilot) to be plotted in the same product space.



**Figure 6. Principal Component analysis, PC1 and PC2 product space**

Variables factor map (PCA)

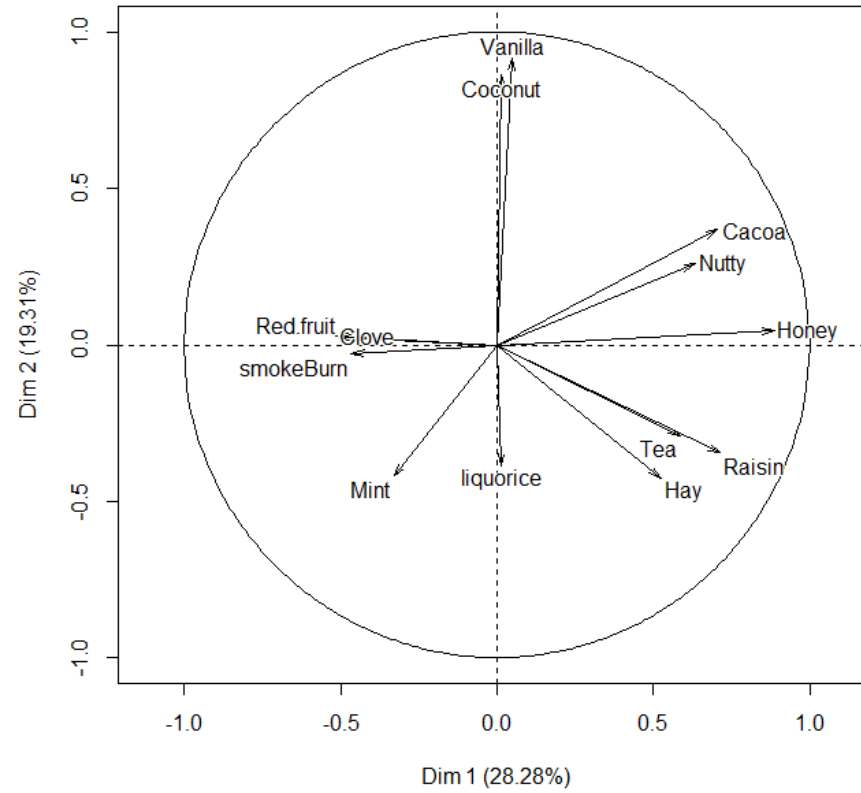
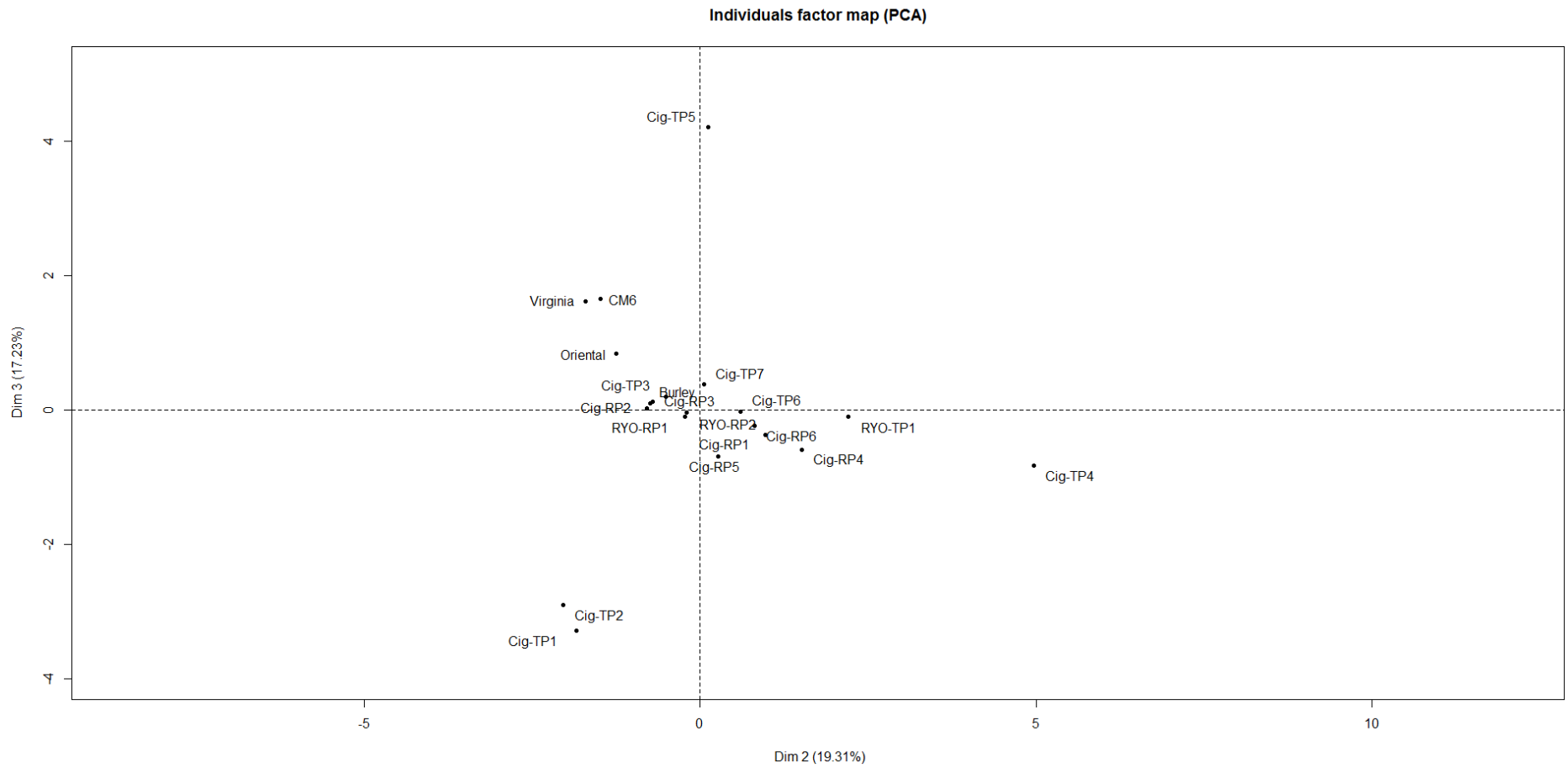


Figure 7. Principal Component analysis, PC1 and PC2 factor map



**Figure 8. Principal component analysis, PC2 and PC3 product space**

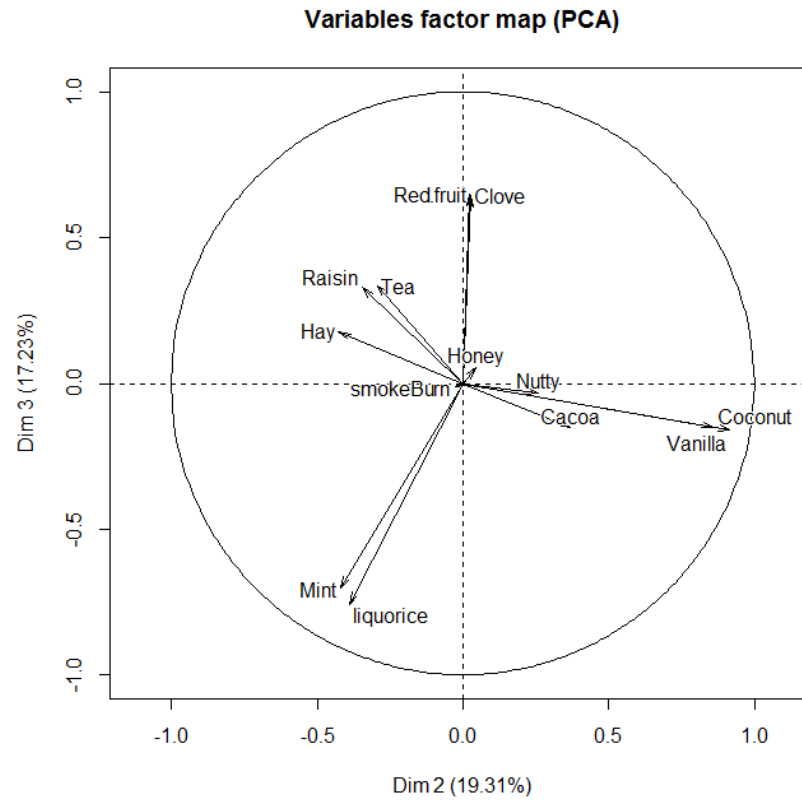


Figure 9. Principal component analysis, PC2 and PC3 factor map

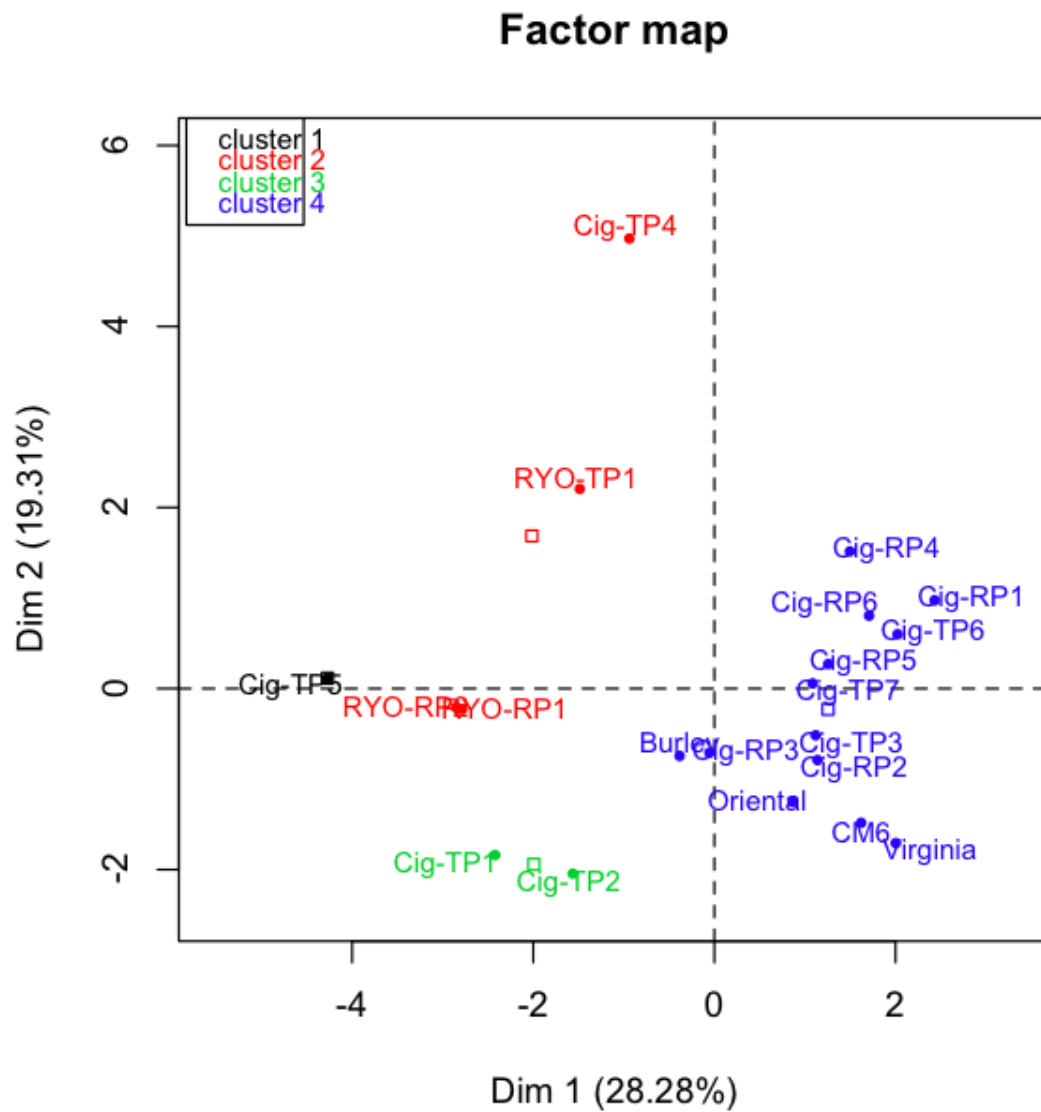


Figure 10. Hierarchical cluster analysis with four clusters

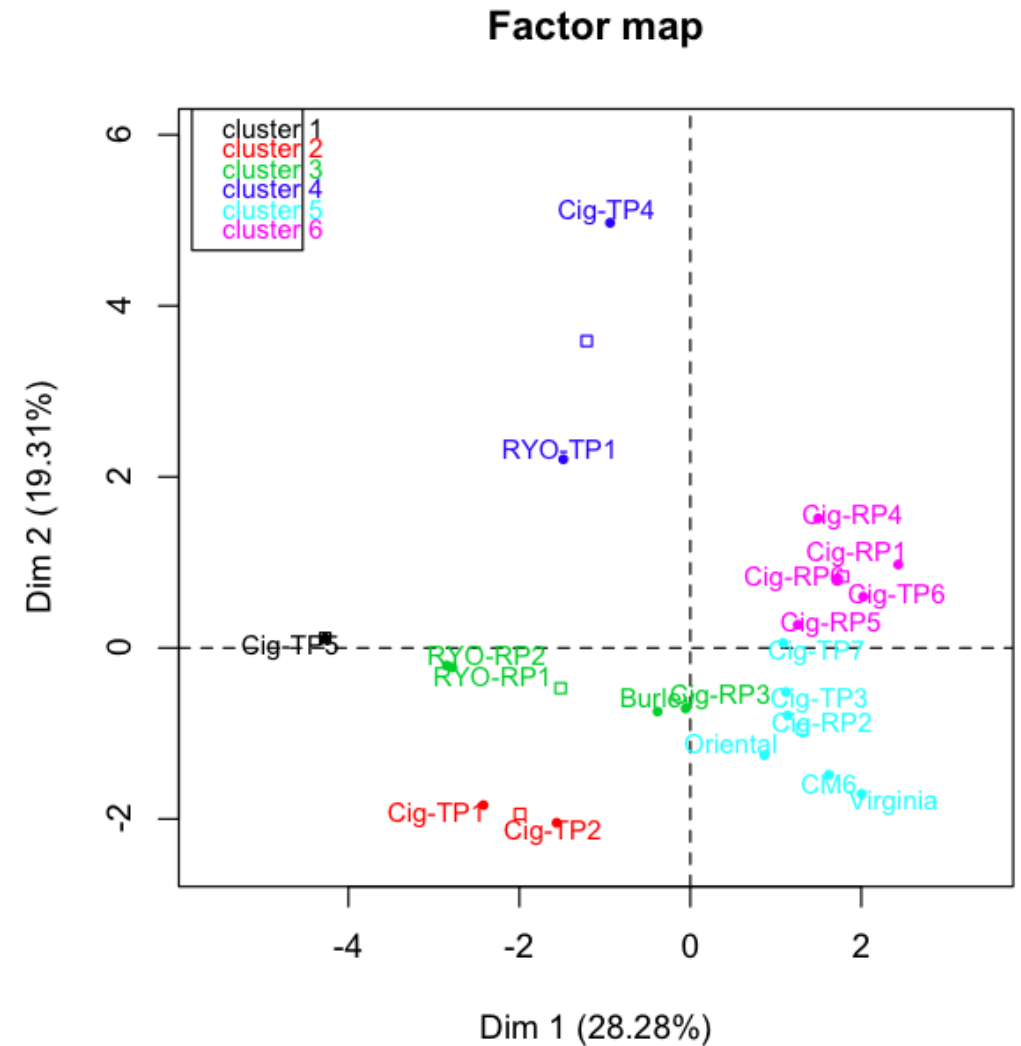
Table 15. Significant attributes for each cluster

Cluster	Attributes	Mean in category	Overall mean	Sd in category	Overall Sd	P.value
#1	Clove	90.09	6.27	0	19.23	<0.001
	Red Fruit	92.66	7.55	0	19.6	<0.001
#2	Smoke/burn	56.68	20.36	36.44	25.58	0.002
	Vanilla	43.14	20.0	37.41	22.45	0.025
	Hay	1359	31.24	4.15	19.36	0.05
	Honey	5.11	10.93	3.25	5.75	0.03
	Raisin	3.76	14.52	2.22	10.46	0.03
#3	Mint	81.14	11.43	9.43	23.54	<0.001
	Liquorice	34.99	11.20	2.7	9.35	<0.001
#4	Hay	41.58	31.24	15.94	19.36	<0.001
	Honey	13.98	10.93	4.51	5.75	<0.001
	Raisin	19.63	14.52	9.46	10.46	0.003
	Cocoa	27.27	20.05	14.86	15.79	0.006
	Nutty	11.18	8.67	5.83	5.85	0.01
	Tea	21.96	16.39	13.31	13.23	0.01



**Table 16. Significant attributes for each cluster**

Cluster	Attributes	Mean in category	Overall mean	SD in category	Overall SD	p value
#1	Clove	90.09	6.27	0	19.23	<0.001
	Red Fruit	92.66	7.55	0	19.6	<0.001
#2	Mint	81.14	11.43	9.43	23.54	<0.001
	Liquorice	34.99	11.20	2.73	9.35	<0.001
#3	Smokey	61.96	20.36	29.53	25.58	<0.001
	Honey	3.83	10.93	2.38	5.75	0.007
#4	Vanilla	79.95	20.0	9.38	22.45	<0.001
	Coconut	39.88	8.67	23.93	13.27	<0.001
#5	Raisin	25.68	14.52	10.14	10.46	0.002
	Tea	29.81	16.40	15.75	13.22	0.004
	hay	45.67	31.24	12.75	19.36	0.03
#6	Cocoa	44.17	20.05	7.05	15.79	<0.001
	Nutty	15.89	8.67	6.53	5.8	0.001
	Honey	16.63	10.93	2.7	5.75	0.013



**Figure 11. Hierarchical cluster analysis with six clusters**

Confidence ellipses for the mean points

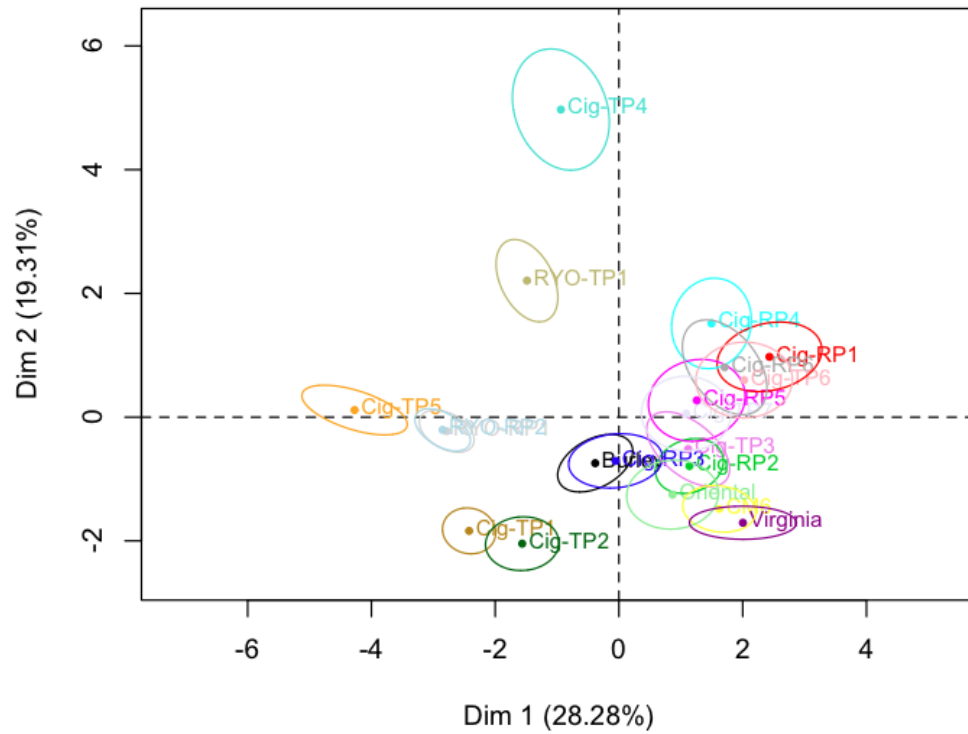


Figure 12. 95% confidence ellipses (alpha=0.05)

Confidence ellipses for the mean points

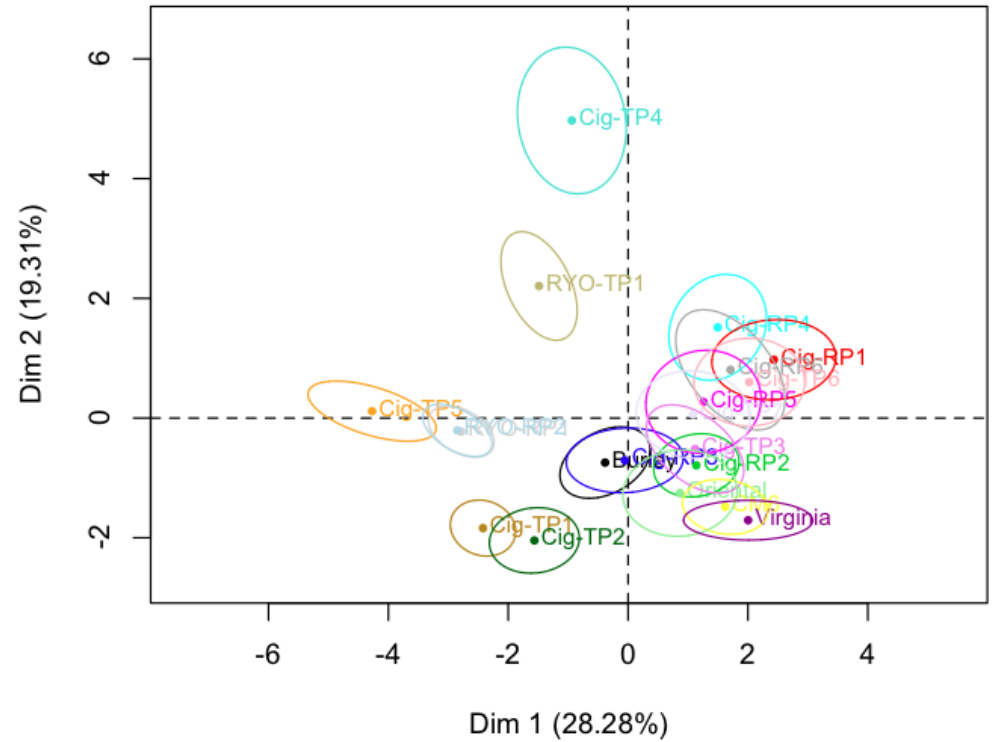
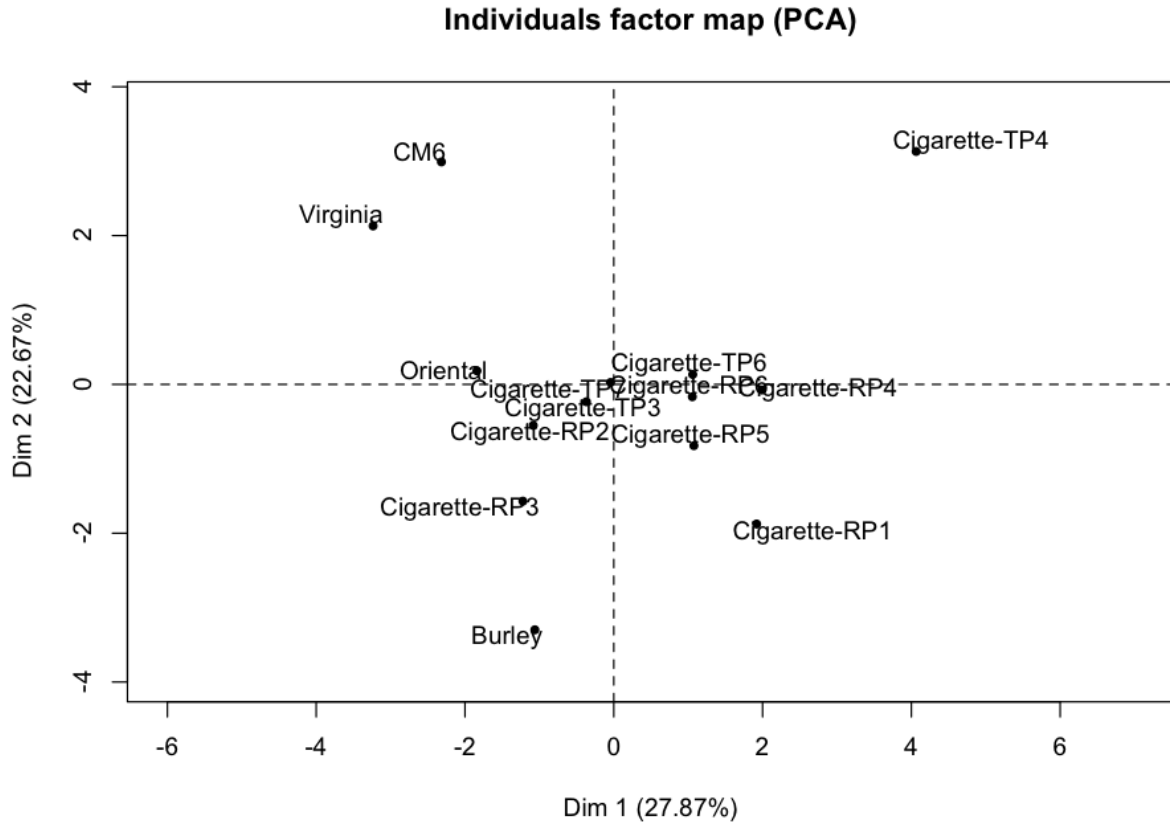


Figure 13. 99% confidence ellipses (alpha=0.01)

**5.3.1.6 Case Study**

In this example, Cig-TP4 was chosen as a test product and was plotted together with the reference products to determine whether or not it had a characterising odour. Based on the cluster analysis (Figure 10) the products from cluster 4 were chosen as reference products (considered not to have a characterising odour). Products included in the reference space are CM6, Virginia, Oriental, Cig-RP2, Cig-RP3 , Cig-RP5, Cig-RP1, Cig-RP4, Cig-TP6, Cig-RP6 , Cig-TP3, Burley and Cig-TP7.



**Figure 14. Total product space: the reference space plus the test product Cig-TP4**

### Confidence ellipses for the mean points

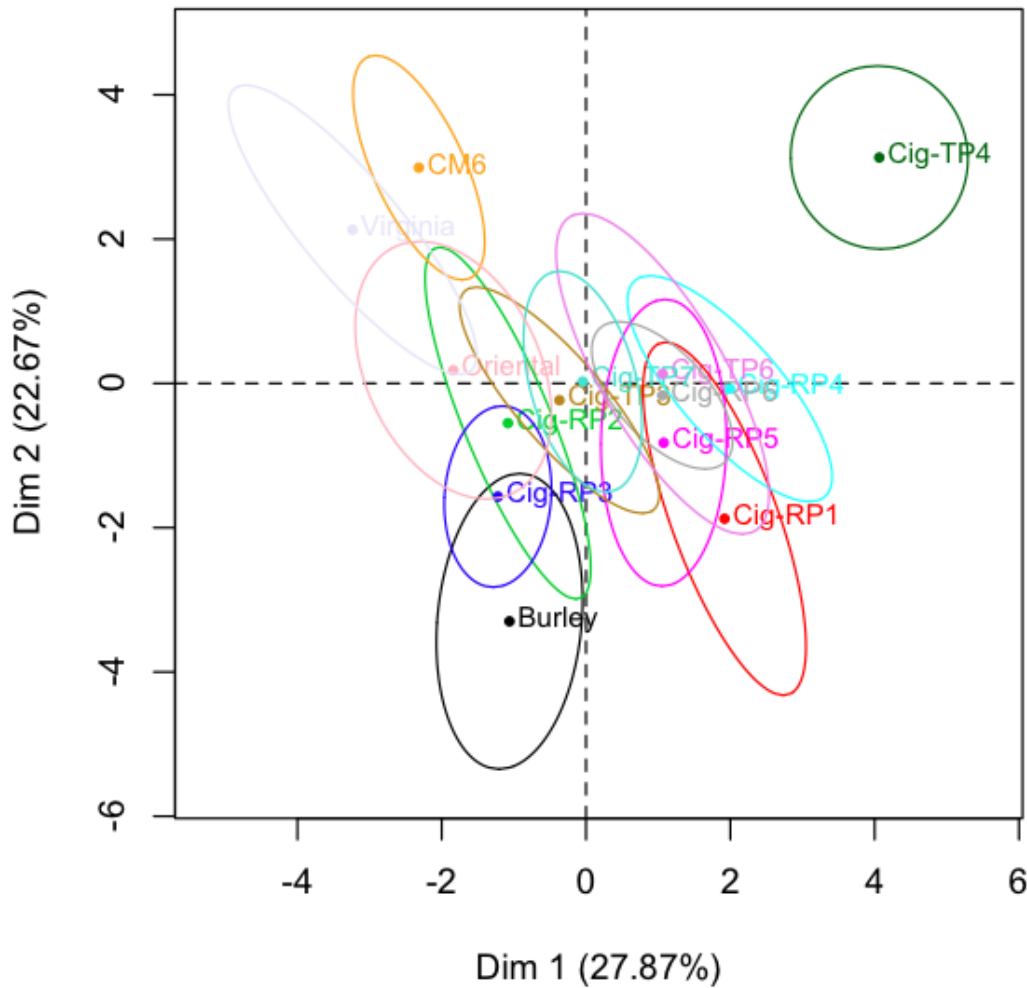


Figure 15. 99% Confidence ellipses case study

To determine whether Cig-TP4 is similar to products in the reference space, confidence ellipses (99%) were drawn. From Figure 15 it can be concluded that Cig-TP4 is significantly different from all reference products (ellipse does not overlap with any of the other product ellipses) based on one or more attributes. To determine on which attributes Cig-TP4 is significantly different from the reference group, pairwise comparison Bonferroni LSD of the attribute means is performed (Figure 16, Table 17). The Bonferroni LSD test corrects for multiple comparison by dividing the alpha (familywise error rate) by the number of comparisons. This is a prerequisite as the attributes of multiple products are compared, decreasing the change of obtaining a false positive outcome.

The Bonferroni formula to calculate whether the pairwise comparison is significant:

$$\frac{|\bar{y}_i - \bar{y}_j|}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} > t_{v, \alpha/2k}$$

T is the value from the t distribution for  $y$  degrees of freedom (in this case, 14 products are being compared, so 13 degrees of freedom)  $\alpha/2k$  is confidence level,  $S_p$  is the pooled standard deviation (mean square error of 2-way ANOVA),  $\bar{Y}$  is the mean value for a certain attribute of the products that are being compared, and  $n$  is the sample size of the two different products (ratings of all panellists in triplicate).

Figure 16 shows the mean values (based on the ratings of all panellists) of each attribute for each product. In this case, all the reference products and the test product: Cig-TP4. On the y-axis, the visual analogue scale (VAS) score in mm and on the x-axis each attribute in a separate column. The red bar indicates whether the attribute is significantly used to describe at least one of the products (i.e. whether the attribute is used to differentiate between Cig-TP4 and the reference space).

From this figure it can be concluded that Cig-TP4 (number 7 indicated by the red circle) scores much higher, compared with the reference products, for vanilla and coconut odour attributes.

Table 17 shows the mean values of each attribute for each product. When the mean values in the same column are not indicated by the same letter, the means are significantly different. This is based on the Bonferroni value indicating that when two means are at least that value different that the means are significantly different.

#### Conclusion

From Table 17 it can be concluded that Cig-TP4 is a tobacco product significantly different ( $\alpha 0.05/k$ ) from all the other products in the reference space based on the characterising vanilla/coconut odour, as the difference with all the other products from the reference space is much larger than the Bonferroni LSD value.

Based on the chosen methodology and as result of the pilot study, the following has been defined: The panel will generate attributes, in consensus, for a wide range of tobacco products (marketed as flavoured and unflavoured). These attributes are noticeable smells. Then, a product space and reference space is created, thereby determining 'tobacco smell' and its boundaries. It can then be tested whether a product is overall significantly different from the reference space. If so, this means the product is perceived as 'different from tobacco', and this can be traced back to significant differences in specific attributes (noticeable smells). If the attribute is rated significantly higher in a specific test product than in the reference products, thus higher than a 'noticeable smell', it can be concluded that this is a 'clearly noticeable smell'.

EUROPEAN COMMISSION

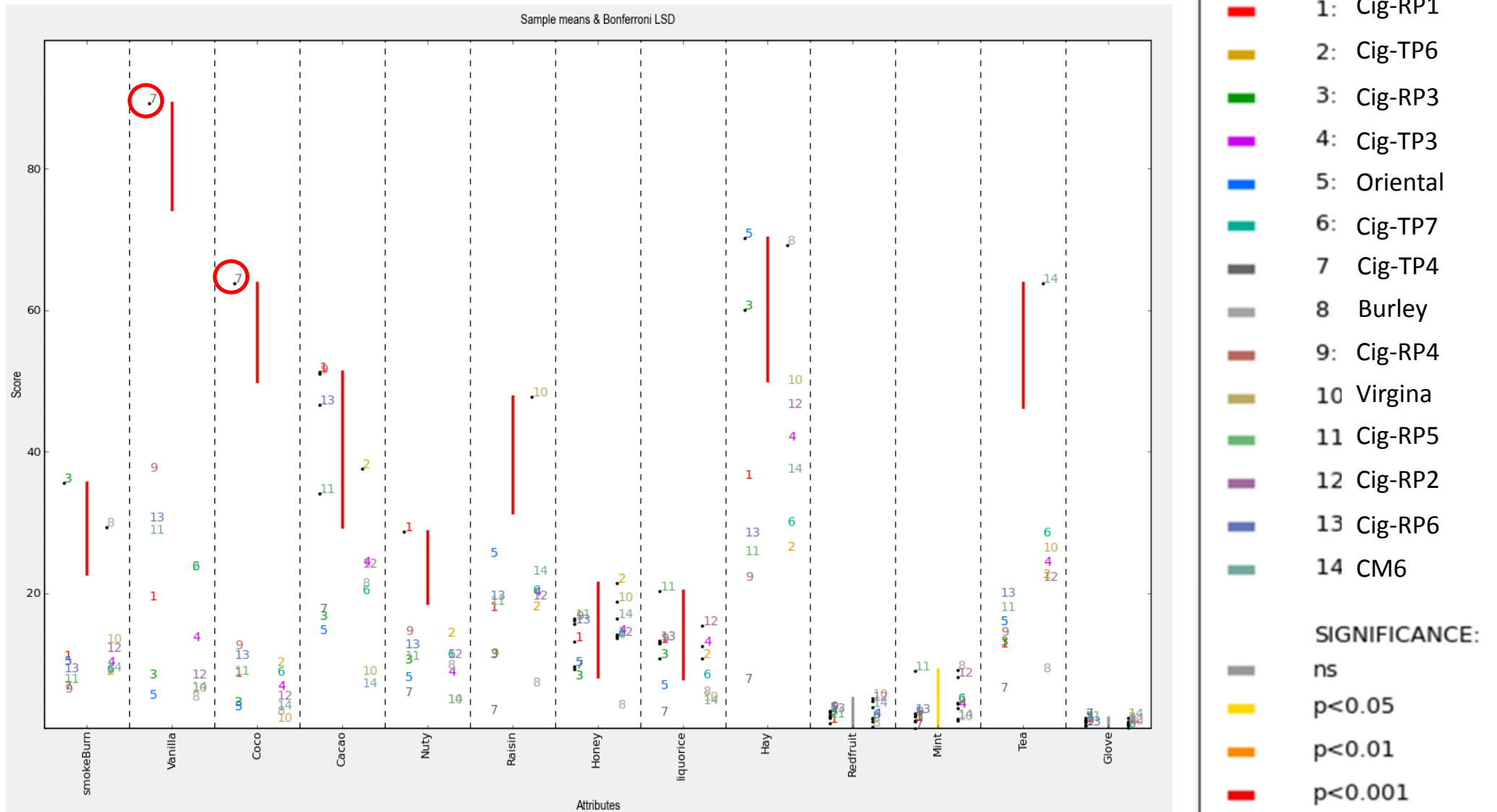


Figure 16. Visualisation of the Bonferroni LSD scores for all attributes listed in Table 17. Comparing all reference products and including test product Cig-TP4.

EUROPEAN COMMISSION

**Table 17. Mean scores for all reference products and test product Cig-TP4 for all attributes and Bonferroni LSD score difference for significance**

Product	Smokey/ Burned	Vanilla	Coconut	Cacao	Nutty	Raisin	Honey	liquorice	Hay	Red fruit	Mint	Tea	Clove
Cig-RP1	10.53 <sup>a</sup>	18.94 <sup>a</sup>	8.20 <sup>a</sup>	51.37 <sup>b</sup>	28.72 <sup>b</sup>	17.42 <sup>ab</sup>	13.17 <sup>a</sup>	13.02 <sup>ab</sup>	36.20 <sup>ac</sup>	1.70 <sup>a</sup>	1.97 <sup>a</sup>	12.29 <sup>bc</sup>	1.76 <sup>a</sup>
Cig-TP6	8.42 <sup>a</sup>	23.23 <sup>a</sup>	9.73 <sup>a</sup>	37.63 <sup>bc</sup>	13.82 <sup>a</sup>	17.57 <sup>ab</sup>	21.49 <sup>a</sup>	10.88 <sup>ab</sup>	25.95 <sup>a</sup>	1.97 <sup>a</sup>	4.46 <sup>a</sup>	22.09 <sup>bc</sup>	1.87 <sup>a</sup>
Cig-RP3	35.68 <sup>b</sup>	7.93 <sup>a</sup>	4.02 <sup>a</sup>	16.24 <sup>ac</sup>	10.03 <sup>a</sup>	10.77 <sup>ab</sup>	7.77 <sup>a</sup>	10.78 <sup>ab</sup>	60.05 <sup>bd</sup>	2.84 <sup>a</sup>	2.10 <sup>a</sup>	12.55 <sup>bc</sup>	2.02 <sup>a</sup>
Cig-TP3	9.73 <sup>a</sup>	13.19 <sup>a</sup>	6.27 <sup>a</sup>	23.85 <sup>ac</sup>	8.33 <sup>a</sup>	19.59 <sup>ab</sup>	14.26 <sup>a</sup>	12.58 <sup>1ab</sup>	41.49 <sup>a</sup>	2.44 <sup>a</sup>	3.84 <sup>a</sup>	23.86 <sup>bc</sup>	1.46 <sup>a</sup>
Oriental	9.87 <sup>a</sup>	5.08 <sup>a</sup>	3.40 <sup>a</sup>	14.25 <sup>a</sup>	7.57 <sup>a</sup>	25.11 <sup>b</sup>	9.73 <sup>a</sup>	6.45 <sup>ac</sup>	70.27 <sup>b</sup>	3.34 <sup>a</sup>	2.66 <sup>a</sup>	15.47 <sup>bc</sup>	2.06 <sup>a</sup>
Cig-TP7	8.68 <sup>a</sup>	23.29 <sup>a</sup>	8.38 <sup>a</sup>	19.85 <sup>a</sup>	10.87 <sup>a</sup>	19.81 <sup>ab</sup>	13.74 <sup>a</sup>	7.99 <sup>ac</sup>	29.50 <sup>a</sup>	2.31 <sup>a</sup>	4.53 <sup>a</sup>	28.05 <sup>bc</sup>	1.01 <sup>a</sup>
Cig-TP4	6.35 <sup>a</sup>	89.34 <sup>c</sup>	63.82 <sup>b</sup>	17.19 <sup>a</sup>	5.41 <sup>a</sup>	2.87 <sup>a</sup>	9.26 <sup>a</sup>	2.66 <sup>a</sup>	7.35 <sup>e</sup>	2.58 <sup>a</sup>	0.97 <sup>a</sup>	6.02 <sup>a</sup>	2.38 <sup>a</sup>
Burley	29.36 <sup>b</sup>	4.78 <sup>a</sup>	2.815 <sup>a</sup>	20.88 <sup>ac</sup>	9.31 <sup>a</sup>	6.82 <sup>ab</sup>	3.71 <sup>b</sup>	5.55 <sup>ac</sup>	69.28 <sup>b</sup>	1.16 <sup>a</sup>	9.19 <sup>a</sup>	8.80 <sup>ab</sup>	1.49 <sup>a</sup>
Cig-RP4	5.90 <sup>a</sup>	37.17 <sup>ab</sup>	12.11 <sup>a</sup>	51.08 <sup>b</sup>	14.14 <sup>a</sup>	10.98 <sup>ab</sup>	16.26 <sup>a</sup>	13.02 <sup>ab</sup>	21.70 <sup>ae</sup>	3.39 <sup>a</sup>	2.65 <sup>a</sup>	14.01 <sup>abc</sup>	1.45 <sup>a</sup>
Virginia	12.97 <sup>a</sup>	6.03 <sup>a</sup>	1.752 <sup>a</sup>	8.43 <sup>a</sup>	4.43 <sup>a</sup>	47.86 <sup>c</sup>	18.85 <sup>a</sup>	4.75 <sup>ac</sup>	49.60 <sup>bcd</sup>	5.16 <sup>a</sup>	2.01 <sup>a</sup>	25.84 <sup>bc</sup>	1.41 <sup>a</sup>
Cig-RP5	7.27 <sup>a</sup>	28.40 <sup>ab</sup>	8.48 <sup>a</sup>	34.17 <sup>bc</sup>	10.63 <sup>a</sup>	18.40 <sup>ab</sup>	16.52 <sup>a</sup>	20.40 <sup>b</sup>	25.35 <sup>ae</sup>	2.41 <sup>a</sup>	9.08 <sup>a</sup>	17.41 <sup>abc</sup>	2.00 <sup>a</sup>
Cig-RP2	11.76 <sup>a</sup>	7.97 <sup>a</sup>	4.94 <sup>a</sup>	23.58 <sup>ac</sup>	10.82 <sup>a</sup>	19.09 <sup>ab</sup>	13.93 <sup>a</sup>	15.42 <sup>bc</sup>	46.16 <sup>c</sup>	4.77 <sup>a</sup>	8.22 <sup>a</sup>	21.78 <sup>abc</sup>	1.69 <sup>a</sup>
Cig-RP6	8.84 <sup>a</sup>	30.16 <sup>ab</sup>	10.72 <sup>a</sup>	46.64 <sup>b</sup>	12.15 <sup>a</sup>	19.14 <sup>ab</sup>	15.72 <sup>a</sup>	13.27 <sup>ab</sup>	27.94 <sup>a</sup>	3.19 <sup>a</sup>	3.02 <sup>a</sup>	19.48 <sup>abc</sup>	1.29 <sup>a</sup>
CM6	8.98 <sup>a</sup>	6.21 <sup>a</sup>	3.53 <sup>a</sup>	6.63 <sup>a</sup>	4.48 <sup>a</sup>	22.61 <sup>a</sup>	16.50 <sup>a</sup>	4.33 <sup>ac</sup>	37.00 <sup>c</sup>	3.87 <sup>a</sup>	2.29 <sup>a</sup>	63.92 <sup>d</sup>	2.48 <sup>a</sup>
Bonferroni LSD**	12.93	15.05	13.91	21.94	10.14 <sup>a</sup>	16.49	13.28	12.41	20.17	5.61	9.29	17.66	2.22 <sup>a</sup>

\* Attribute mean values in the same column not indicated with the same letters are significantly different from each other.

\*\* The difference between two mean values is significant when they differ more than the Bonferroni LSD value.

### 5.3.1.7 Overall conclusion and recommendations for future development steps

#### 5.3.1.7.1 Overall conclusion

The method as performed in the pilot study is a good method to determine characterising odours in tobacco products.

The panel was sufficiently trained to differentiate between products. When the panel would have been trained more extensively, panellists should also be able to distinguish between the products that are more similar. Furthermore, the attribute list was reduced from 18 to 13 attributes at the end of the panel training as panellists agreed that they were not able to distinguish products for the subtler odour attributes and that they were not in consensus about them. The products, which we assumed had a characterising odour, were also defined as such. The products that we assumed did not have a characterising odour were clustered together and selected as reference products/space (except for the RYO, RYO-RP2 and RYO-RP1, who fell outside this cluster). However, the products that were a-priori considered borderline, i.e. Cig-TP3 and Cig-TP6 (§4.3.1.5) fell into the reference space. This may be partly due to the before mentioned argument that the panel was not yet able to distinguish products that were more similar. For the final panel, the borderline products may not cluster together with a group of 'reference' products. In case it is decided that these products should be in the reference space the number of clusters could be adjusted. In that way, only products that are most different fall outside the reference space (i.e. Cig-TP5, Cig-TP4, menthol products). Independent of how well trained the panel is, there may always be borderline products.

Based on the chosen methodology and as result of the pilot study, the following has been defined: The panel will generate attributes, in consensus, for a wide range of tobacco products (marketed as flavoured and unflavoured). Therefore, these are noticeable smells. Then, a product space and reference space is created, thereby determining 'tobacco smell' and its boundaries. It can then be tested whether a product is overall significantly different from the reference space. If so, this means the product is perceived as 'different from tobacco', and this can be traced back to significant differences in specific attributes (noticeable smells). If the attribute is rated significantly higher in a specific test product than in the reference products, thus higher than a 'noticeable smell', it can be concluded that this is a 'clearly noticeable smell'.

This method is suitable to assess characterising odours in other tobacco products as well due to the elimination of visual cues. However, an updated or renewed attribute lists need to be established in when using the method for other tobacco products (i.e. other than cigarettes, RYO, and tobacco leaves).

#### 5.3.1.7.2 Recommendations

Below, recommendations for the future are given. We distinguish between key recommendations, which are necessary alterations for the final method (§5.3.1.7.2.1); and possible considerations, which may be nice to take into account for the future (§5.3.1.7.2.2).

##### 5.3.1.7.2.1 Key recommended alterations for the final method based on pilot study experiences

- Number of products during one test session  
The pilot study, in which ten tobacco products were tested in 1.5 hour, demonstrated that this was not too tiring, given that panellists have at least 30 seconds of resting/neutralizing between samples and have a 5 minutes break after evaluating five samples. Most panellists finished the task within one hour, and therefore 12 products are estimated to be the maximum number of products tested within 1.5 hour.
- Smokers vs. non smokers



People who smoke have a very different perception of the odours of cigarettes and roll your own compared to people who do not. During the pilot study, it was noticed that smokers generally have a broader idea of tobacco odours (as they already have experience with the products) compared to non-smokers.

Furthermore, some of the panellists that were smokers were able to recognize the odour of the sample and relate it to a product. For example, panellists who were (ex) smokers were able to identify the 'Cig-TP4' odour and Cig-TP5.

This does not indicate that for the expert panel smokers should be used because the idea is that panel members gradually learn this after being exposed to many different tobacco types during the training. Using smokers for an expert panel will therefore only have the advantage that the panel may be trained faster compared to when using non-smokers because they are already familiar with the odours. However, finding smokers who pass the screening test (i.e. odour threshold test) may be difficult, and given the purpose of the TPD (to make smoking less attractive for young people), it might be more useful to only recruit non-smokers.

- Personality screening

For the pilot test, participants were not screened for personality characteristics. During the training sessions, it was noticed that some of the panel members were not ideal to have in the panel due to either being too dominant in discussions or too passive. Furthermore, motivation of the panel members was an important factor. Some of the panellists were much better in remembering what was agreed upon during previous training sessions compared to others and were much more interested in the process of becoming a panel member, which was noticed by questions that they asked relating to panel performance. In addition, motivation was of importance as occasionally panellists did not show up and some of the panellists were always late. Therefore, it is recommended to screen for personality and motivation and this should therefore be part of the standard operating procedure.

- Number of panellists

Eighteen panellists were included in the pilot panel. Normally, a panel consists of 12 panellists however due to the relatively short training period (7 weeks) this number was increased by 6 to increase statistical power. During the panel-training phase, some disadvantages of using 6 more panellists were noticed. For example, the more people are in the panel the more difficult it is to come to consensus as all of the panellist have their own idea about 1) how the odour should be named, 2) how intense the odour should be rated. Furthermore, due to the large number of panellist and the availability of only one panel leader it was impossible to have all 18 panellists train at the same time because of scheduling issues and because it would have been too difficult to follow group discussions. However, having 9 panellists at two different time points is not a preferable way of training as it is difficult to get the two groups on the same line. For the pilot study the groups were therefore not fixed (i.e. panellists often switched from group) to prevent possible large differences between two groups. However, during the training it was noticed that panellists were confused about decisions made previously that were not agreed upon by members of the other group.

Because of this, it is recommended to have a panel of 18 panellists who train together as a group and two panel leaders. The best 12 panellists can then be used for test-sessions; the remaining 6 are needed to have 12 available panellists in all cases, and to cover dropouts.

**Sample selection and preparation**

- Blinding the products

During the pilot, the tobacco was now offered in a bottle where the cigarette filter and paper of the cigarette was removed. It is generally thought that odours when placed in a bottle come out more strongly. However, during the screening, participants (n=64) indicated from a list which odours they perceived when smelling the tobacco in a bottle and when smelling the tobacco stick (§4.3.1.3). Although intensity was not measured in a direct manner during the screening, the intensity can indirectly be indicated from the number of participants that perceived the odour (i.e. if an odour is very strong then all subjects will perceive it).

From these results, we concluded that there is little to no difference in odour percept when smelling tobacco from a bottle and smelling a tobacco stick. Therefore, for the pilot study we concluded that placing the tobacco in a glass bottle is the best method to eliminate visual cues and to perceive the odours present in the tobacco. Furthermore, the intensity score will be counterbalanced as the reference products will be tested in the same manner. However, we recommend further investigating this by performing a study where the product is assessed in various ways (smelling the tobacco stick, smelling the tobacco stick when placed in a small box, or smelling tobacco when placed in a glass bottle). In addition, intensity is to be measured in a direct manner (i.e. use intensity scales such as a visual analogue scale or nine point scale from not perceived at all to perceived as very intense).

- Standardized manner of assessing products

Initially panellist all have a different way of assessing the products and this should be standardized as this may induce differences in rating the odours. For example, some of the panellist shake the bottle before opening, resulting in 'dust' coming of the tobacco making ratings in hay higher compared to the ratings of panellist that did not shake the bottle. Another example is that some of the panellist did not close the bottle after smelling the sample. Resulting in crossover odours from sample to sample and less intense ratings compared to panellist who did close the bottles. Furthermore, some of the panellists only smell the sample once and then rate most of the attributes except for the ones they find difficult. Only for the difficult attributes, they smell the sample again. The panellist should be instructed to open the bottle and smell, then rate the odours that come out as most intense. After that, they should smell the bottle before rating each single odour attribute.

- Number of products for the training

At the start of the panel set up, the training should be conducted with only a limited number of different tobacco products or cigarettes. Panel members need to be able to have all products together and compare them to one another. This is because they have to learn the attributes by having in mind a certain product that is most characterising for a particular odour. For example, for the attribute smoky the panel always referred to the samples with RYO-RP2 roll your own tobacco as reference. For the pilot test 20 different tobaccos or roll-your own were included, this were too many different products for the panel to remember. Therefore, it is recommended to start out with 10 products as a maximum. Once the panel is able to repeatedly rate these products in consensus, new products can be introduced. This process should be repeated until the panel has been trained with a wide variety of different tobaccos covering all typical tobacco odours.

**Panel performance**

- Number of attributes and panel training duration/ reliability of product characteristics identification

Many of the odour attributes were merged, as the panel was not trained sufficiently to distinguish between, for example, chocolate and cocoa, and vanilla and caramel. Discriminatory performance of the panel will further increase with additional training.

Therefore, training the panel thoroughly (which will take at least several months, estimation approximately half a year) before starting to test products on characterising odours is recommended.

#### **Statistics and reference space**

- Including pure tobacco leaves as a reference  
Tobacco leaves were included in the pilot test to verify how commercially available products deviate from pure tobacco. During the panel training, it was noticed that the samples containing the pure tobacco (Oriental, Virginia and Burley) had different smell depending on the leaves used (i.e. variability from leaf to leaf). Therefore, variations between samples of the same leaf type should be taken into account when including the tobacco leaves as a reference. This can be achieved by combining different samples of the same leaf type into a single sample.
- Separate reference space for roll-your-own tobacco and cigarettes  
From the results of the pilot study, it seems that roll-your-own tobacco forms its own cluster based on 'smokey or burned' attributes. This is most likely due to the fact that roll your own tobacco leafs are mostly fire cured (<http://www.ryomagazine.com/tobacco.htm>). Because of this, it seems advisable to have a separate reference space for roll-your-own products.
- Randomization of the order of the products  
As described in the method, products should be counter-balanced. The experimental design should be designed in such a way that it takes care of both order and carry-over effects. This way, the presentation order is not fully random as the presentation is according to an algorithm, in which each sample is tested the same number of times at each position, and is following each of the other samples equally. Only when the panellists are trained really well, the experimental design might not be of importance.

#### **5.3.1.7.2.2 Possible considerations for the future**

- Assessment of products by a consumer panel  
A consumer panel may be used in addition, or as alternative to, an expert panel, in order to determine how relevant findings from the expert panel are for regular consumers, i.e. smokers. E.g., will consumers perceive similar odours from the products, and distinguish between products in a similar way as an expert panel does. A consumer panel could use CATA. However, it is unclear how useful this addition is for regulatory purposes, and it needs to be determined how such a method could be used in detail.
- Assessments of tobacco products through smoking  
Another expert panel may be trained in smoking tobacco products. With the assessment through smelling, most of the products containing a characterising flavour will be indicated as such. However, products may be developed that do only release a flavour upon combustion. A smoking panel may be established to be able to assess this type of products.
- The spectrum method  
In case the real intensity is of importance rather than the relative importance of differences between products regarding the odour, the spectrum method may be a more suitable method than QDA. However, from general experience it is known that people (trained or

untrained) are not able to indicate precise intensities of odours. People are only able to indicate the relative difference (QDA) or rank order. If, in the future a method is developed to train people in indicating intensities of odours (not very realistic) the spectrum method could be the best alternative. However, training a spectrum expert panel for assessing odours may take more than two years.

- Confirm variance in RYO products is smaller than the variance in tobacco products  
When setting up the reference space, and RYO tobaccos are assessed, it can be determined whether the variation between RYO tobaccos is indeed smaller than in cigarette tobacco. This outcome should then determine the amount of products that need to be tested to set up the reference space for the RYO products. When the variation of the group of samples no longer increases it means that the right amount of samples (i.e. representing products on the market) are included to determine the reference cluster.
- Update reference space  
The panel may become better in assessing the products due to the test or maintenance sessions after the initial training of the panel. This may also change the way the panel assesses the reference products (i.e. the panel becomes more sensitive). As described in §7 of Annex III, this is not likely, given the amount of training that the panel already receives before product testing. To take this variation in assessing products over time into account the reference space may need to be updated. Moreover, it is also possible that new products enter the market with possibly new flavours that are not yet taken into account in the product and reference space.

### 5.3.1.8 References

1. OPP Research. Workshop setting up a sensory panel. 2001.
2. Hummel T, Sekinger B, Wolf SR, Pauli E, Kobal G. 'Sniffin' Sticks': Olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chem Senses*. 1997;22(1):39-52.
3. Hummel T, Kobal G, Gudziol H, Mackay-Sim A. Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. *Eur Arch Oto-Rhino-L*. 2007;264(3):237-43.

## 5.4 WP3: Optimisation of testing – Chemical analysis

### 5.4.1 Products tested

In the pilot experiment, three products were chemically tested as an example to determine whether this method is suitable for measurement of flavour additives in tobacco products. The products that were tested are Cig-TP1, Cig-RP1, and American Virginia flue cured 2013. These products were selected to compare a flavoured product with a commercially available product assumed to impart no characterising flavour and one of the pure tobacco leaves. Virginia is the most common tobacco leave in cigarettes, hence this type of leave was chosen.

### 5.4.2 Final results

Figure 17 shows the chromatogram of Cig-TP1, Figure 18 the chromatogram of Cig-RP1, and Figure 19 the chromatogram of American Virginia flue cured 2013 tobacco.

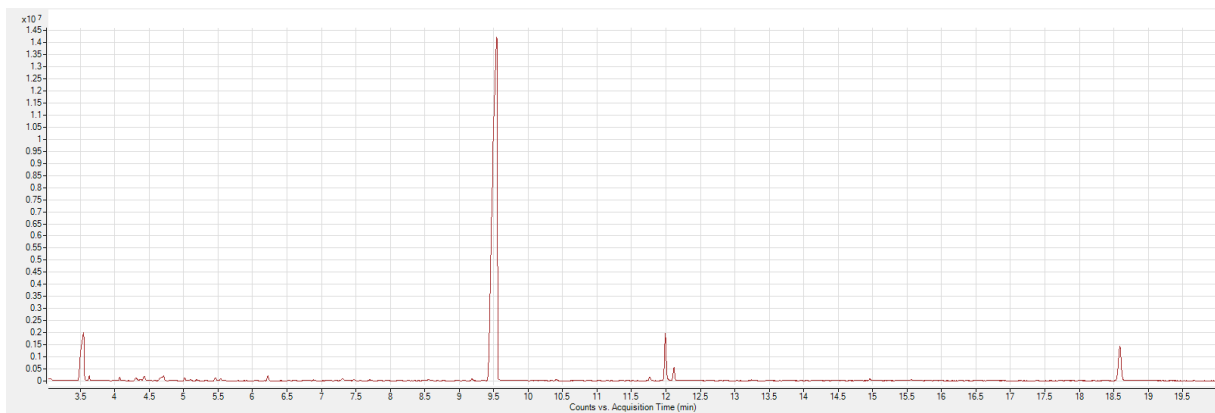


Figure 17. Chromatogram of Cig-TP1

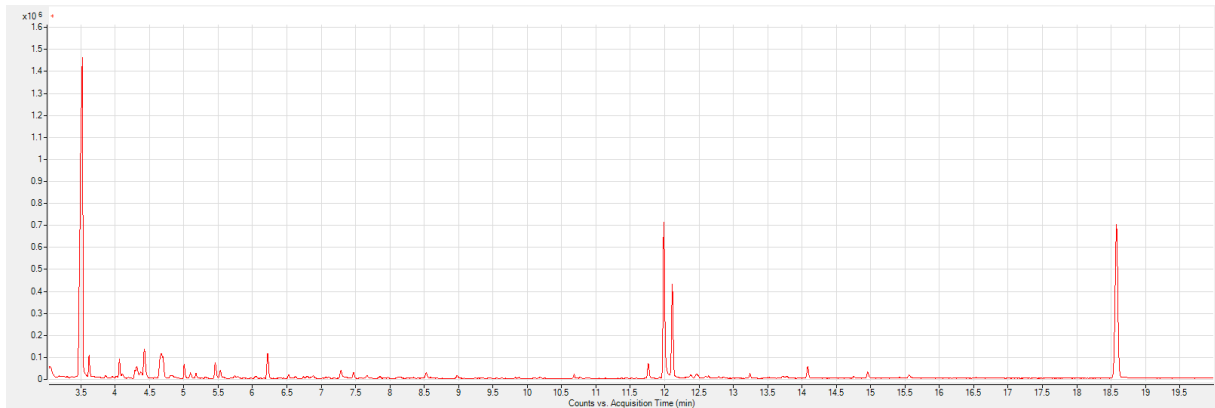


Figure 18. Chromatogram of Cig-RP1

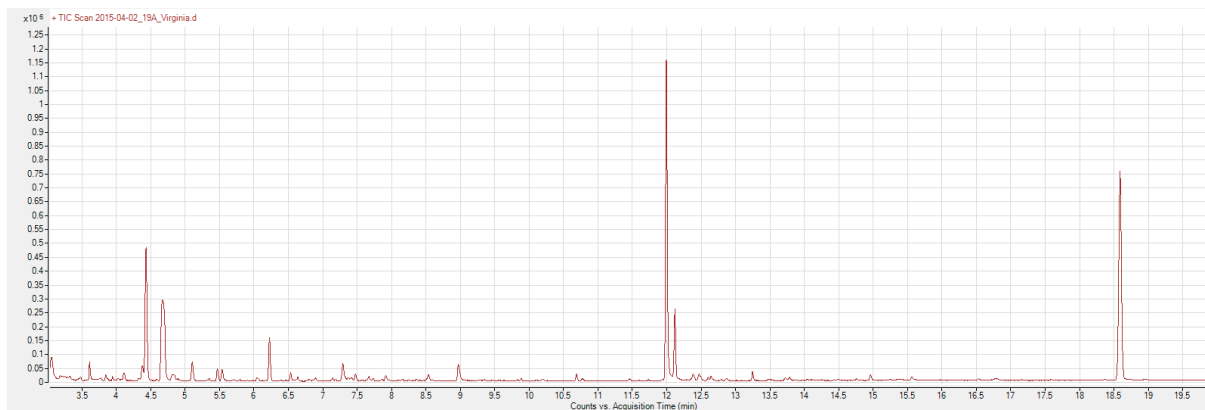


Figure 19. Chromatogram of American Virginia flue cured 2013 tobacco

Figure 20 shows the identified components and their CAS number measured in the three tobacco products. For each component, the box contains an ‘x’ when the component was present in all triplicate measurements of that tobacco product. The final column represents the components’ odour description. The odour description of 20 components was found in the Leffingwell Flavour Base 9 – Tobacco version; odour descriptions of 8 other components was found on the internet; the odour description of the remaining 8 components is unknown (either it has never been investigated or the component has no flavour). Components marked in blue represent additives (only present in the commercially available products); the components marked in orange are the components present in Virginia leaves; the component marked in red is excluded, since it is known that this component does not contribute to the flavour of the product.

Chemical component	CAS number	Marlboro Menthol	Marlboro red	American Virginia flue-cured 213	Description
1 1,2-propanediol, 1-acetate	627-69-0	x	x		unknown
2 1,2-propanediol diformate	53818-14-7	x	x		unknown
3 1,2-propanediol, 2-acetate	6214-01-3	x	x		unknown
4 menthol	1490-04-6	x			Cooling; less cooling than (-)-menthol, with musty-minty notes
5 lactic acid	050-21-5		x		Weak, sour, buttermilk odor and taste
6 menthone	14073-97-3	x			Minty-herbaceous (not green); dry woody notes
7 neomenthyl acetate	2230-87-7	x			Fresh, minty, slightly fruity for (+)-neomenthyl acetate
8 1h-pyrrole-2-carboxaldehyde, 1-methyl-	1192-58-1	x			Somewhat burnt, roasted, cracker, nut
9 butylated hydroxytoluene	128-37-0		x		Antioxidant for flavor oils - See comments:
10 3-isoxazolamine, 5-methyl-	1072-67-9	x			unknown
11 pyrazine, 2-ethyl-6-methyl-	13925-03-6	x			Roasted, hazelnut, cocoa, cooked potato like notes
12 benzaldehyde	100-52-7	x	x		Odor of bitter almond oil; characteristic sweet cherry taste
13 pyrazine, methyl-	109-08-0	x	x		Green, nutty, cocoa, musty, potato, fishy-ammoniacal notes
14 megastigmatrienone	38818-55-2	x			Tobacco/incense aroma
15 6,8-nonadien-2-one, 8-methyl-5-(1-methylethyl)-, (e)-	54868-48-3	x	x	x	unknown
16 neophytadiene	504-96-1	x	x	x	Almost odorless
17 pyridine	110-86-1	x	x	x	Fishy type odor
18 benzeneacetaldehyde	122-78-1	x	x	x	Flowery odour
19 benzyl alcohol	100-51-6	x	x	x	Faint, sweet, almond fruity aroma; sweet, but somewhat chemical taste
20 phenylethyl alcohol	060-12-8	x	x	x	Has a floral type odor and an floral type flavor
21 megastigmatrienone	38818-55-2	x	x	x	Tobacco/incense aroma
22 hept-5-en-2-one <6-methyl->	110-93-0	x	x	x	Oily-green, herbaceous odor; green fruity taste
23 furfural	098-01-1	x	x	x	Has a bready type odor and an brown type flavor
24 2-furanmethanol	098-00-0	x	x	x	Has a bready type odor and an burnt type flavor
25 furfural <5-methyl->	620-02-0	x	x	x	Has a caramellic type odor and an sweet type flavor
26 2,3-pentanedione	600-14-6	x	x	x	Oily-buttery, fatty odor; butter, cream, milk taste
27 furan-3-one <2-methyl-, tetrahydro->	3188-00-9	x	x	x	Sweet and solvent-like odour, nutty and astringent flavour with a creamy almond nuance
28 naphthalene, 1,2,3,4-tetrahydro-1,1,6-trimethyl-	475-03-6	x	x	x	unknown
29 butyrolactone	096-48-0		x	x	Weak, slightly caramellic, nutty, buttery odor
30 5,9-undecadien-2-one, 6,10-dimethyl-, (e)-	3796-70-1	x		x	A fresh, rose, magnolia-type odor
31 pyrrole <2-acetyl->	1072-83-9	x		x	Has a musty type odor and an nutty type flavor
32 4h-pyran-4-one, 2,3-dihydro-3,5-dihydroxy-6-methyl-	28564-83-2			x	unknown
33 ethanone, 1-(2-furanyl)-	1192-62-7			x	Has a balsamic type odor and an nutty type flavor
34 1,4-benzenediol, 2-methyl-	095-71-6			x	unknown
35 2-buten-1-one, 1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)-, (e)-	23726-93-4			x	Has a floral type odor and an woody type flavor
36 nicotine	054-11-5	x	x	x	Odorless

Figure 20. Results of chemical analysis on Cig-TP1, Cig-RP1, and American Virginia flue-cured 2013 tobacco

### 5.4.3 Findings and recommendations

- Accuracy of headspace GC-MS appeared to be much better than expected, around 5%. Therefore, headspace GC-MS is a suitable method to identify flavour components and additives. This method can be used in other laboratories if the equipment is present.
- The components marked blue, i.e. the additives, can be used to build up a chemical flavour library. Therefore, the method tested in this pilot experiment can be considered a first step in building a flavour library. When a large flavour library is obtained, patterns can be found. The components that are present in all brands of a certain flavour are considered characteristic for that particular flavour. These components can be further examined by determination of a detection threshold using sensory experiments.
- Since chemical data on flavours do not provide information concerning human perception, multivariate analysis techniques such as PCA and PLS can be used to verify systematic associations between specific chemical and sensory data. It is doubtful whether chemical analysis can ultimately completely replace sensory analysis, but increased insights into chemical-sensory associations may result into a more important role for chemical analysis, for example in screening large number of (known) samples.
- It is recommended to use pure tobacco leaves or research cigarettes containing no additives (such as CM7) as baseline for chemical analysis, since commercially available brand cigarettes, such as Cig-RP1, contain additives as well. To determine which additives are present in commercially available products with or without a characterising flavour, it is required to compare these products to a baseline consisting of leaves and/or research cigarettes.
- When the chemical compound measured (e.g. menthol) in a tobacco sample is not present in natural tobacco (e.g. burley, Virginia, oriental) this means that this component is most likely a tobacco additive. Sensory analysis is needed to determine if for example menthol is a characteristic flavour in a sample. Using chemical analysis, the amount of menthol can then be quantified in this sample and compared to the amount of menthol in the natural tobacco and reference space samples (Annex V).

The SOP for chemical analysis (Annex V) represents a method to identify certain flavour components in tobacco samples. In this way, the flavour that the sensory panel smells can be chemically verified.

## **5.5 WP4: Cost benefit assessment**

### **5.5.1 Technical feasibility and quality of outcomes**

The main document and the standard operating procedures (Annex III, Annex IV, and Annex V) describe methods for sensory and chemical analysis. Both methods need to be performed under standardization quality norms (ISO guidelines) to ensure high technical quality and reproducibility by trained and skilled personal using up to date equipment.

It can be concluded from the work described in WP3 that it is technically feasible to use a combination of sensory- and chemical analysis to identify tobacco products with a characterising flavour. Based on sensory analysis it is technically feasible to decide whether a tobacco product imparts a characterising flavour other than tobacco noticeable before or during the consumption of the tobacco products. Based on chemical analysis it is technically feasible to determine how a characterising flavour as detected by sensory analysis is generated and whether it derives from a particular additive or combination thereof.

By carefully analysing the complete procedure described in the SOPs, we defined technical steps that are of high importance for the quality of the results. For sensory analysis, the quality of outcomes depends on the panel performance. The panellists should provide objective measurements and maintain high standards in delivering good and objective data sets. Panellists should be carefully selected and need to be trained for a longer period, e.g. at least half a year, to ensure reliable and reproducible results. Panel performance needs to be monitored before every test session (e.g. once a month) to ensure good panel performance.

For chemical analysis, the reproducibility of the data will be high when working under ISO norms, and performed by a research scientist trained to work with up to date equipment fit for the purpose. Quality checks need to be regularly performed using internal standards of chemical components and by generating calibration curves.

To sustain quality of the method over several years, quality assessments need to be performed regularly. Further development and improvement of the methodology is ensured by keeping the sensory panel- and chemical analytical facilities operational. This could be done by performing regular testing of products, follow-up experiments and further validations.

A period of five years is taken for depreciation of chemical analytical equipment, equipment and software needed for sensory analysis and experimental facilities. After this time point, quality of the methodology can no longer be sustained. Equipment and facilities need to be renewed and updated to the newest standards.

For technical sustainability, maintenance of a sufficiently trained panel over time is important. To increase panel loyalty, group activities can be organized once or twice a year.

From a legal perspective, both methods (expert or consumers) are well defensible. Industry may argue that sensory perception of consumers may be different to highly trained experts. However, differences in sensitivity may be adjusted for. Another point of consideration is the fact that a characterising flavour is most likely caused by the addition of chemical components (additives) to the tobacco, but possibly, it could also result from genetic modification of the tobacco plant. Developments regarding genetic modification of the tobacco used in specific products could be monitored.



### 5.5.2 Quantification of costs and funding model

The estimations in Annex IX. Cost estimate are presented separately for costs necessary to set up the panel, costs of determining the reference space and costs of longterm operation and maintenance of the panel. The cost calculation is based on assumption that the panel is set up to assess tobacco products for the period of 4 years following the initial set-up period of about half a year.

In the first half year, costs will involve setting-up the facilities, screening of candidate panellists and training of the panel. Once the panel is trained, it will be operational to test products and to determine the reference space. Assessment of products and/or panel maintenance will take place in the form of regular meetings (e.g. once a month). During these meetings, the panel performance is monitored first (maintenance test session). Based on these results the panel leader decides whether the panel performance meets the standards and whether additional maintenance meetings may be needed.

Costs are calculated for individual phases of the process as well as for different types of methodologies. All cost aspects involved in the proposed method are taken along including setting up, training and maintainance of an operational sensory panel.

Costs are categorized as follows:

1. Start-up costs (e.g. setting up facilities, purchase of equipment)
2. Fixed costs (e.g. recruitment, selection and training of panellists)
3. Variable costs (e.g. linked to sample number to be tested and costs to maintain the panel operational)

After several orienting cost estimations had been made for various methodological and organisational approaches it was decided to further specify costs of the developed methodology (i.e. smelling by an expert panel and chemical analysis, see Annex IX. Cost estimate). While various frequencies of maintenance meetings were considered in internal cost-estimations the presented data consider monthly frequency of the panel meetings unless additional meetings must be scheduled for testing of samples..

For the cost assessment, a time frame of 5 years (2016-2020) was taken for calculations. The costs have been split in one-off initial costs (such as setting up and training the panel) and running costs in the course of subsequent 4 years (2016-2020). Establishment of the initial reference space will be performed in the first year during the first testing phase. Reference space needs to be defined separately for both main product categories (e.g. cigarettes and roll-your-own tobacco). Costs are assumed for reference spaces consisting of 50 cigarette brands, respectively 25 roll-your-own products. Later on, further refinement of the reference space may be considered as part of maintenance sessions.

To gather information on various types of costs involved in sensory analysis a questionnaire on cost aspects involved in the sensory analysis of tobacco products was send out to the tobacco industry (Annex VI. WP4: Questionnaire sent to industry). The questionnaire involved several aspects regarding sensory analysis of tobacco products. Information provided by the tobacco industry confirms our cost estimate covers all aspects involved in setting up, training, operating and maintaining a sensory panel.

Costs for starting up a facility for sensory analysis will lie around €130.000 (for a five years' period). Costs for setting up a panel (mainly screening of the candidates and training the panel) will be around €160.000. Initial determination of the reference space was estimated at €47.000. Costs for

panel maintenance are estimated to be around €150.000 yearly. When including product testing, these costs will be slightly higher, around €200.000 per year. The costs per product when analysing around 10 products per year will be approximately €20.000. The unit price is then decreasing with growing number of products analysed per year. Personnel costs make up more than half of the calculated costs and their variation may impact significantly the values presented above. In our calculation, we used standard personnel costs in our establishments and no long-term contracts were foreseen for the panellists.

In addition to the costs discussed above, some tasks need to be carried out by all stakeholders involved in implementation of the testing procedure including the European Commission, national regulators and the tobacco manufacturers/importers. These costs have not been estimated yet, as the procedure is not yet finalised.

Compliance costs for regulators (MS) and/or the European Commission (COM) include (depending on the final procedure):

- Signalling and selection of products on the market that may have a characterising flavour. Products may be selected based on e.g. assessing their ingredient composition, package and marketing of the product, and discussions on internet fora;
- Sampling tobacco products from the market (selling points and distributors) for testing purposes and transporting them;
- Sending products to the central test facility, together with a request for testing
- Evaluating the test report and decide on next steps;
- If so decided, send opinion of panel to manufacturer and collect reactions;
- Administration of all these steps.

Compliance costs for the tobacco industry include:

- Replying to information requests from COM/MS;
- Evaluating the test report and decide on next steps;
- Withdraw product from the market when this is the outcome of previous steps

It should be noted that there are also other elements in the procedure, before a decision is taken to test a product. For instance, the fact that a product that is suspected of having a characterising flavour is on the market, will be notified to the manufacturer. The manufacturer may decide to withdraw the product from the market, after discussions with the MS.

In addition, COM and other MS need to be informed on actions taken.

It is foreseen in the Tobacco Products Directive that Member States and the Commission may charge proportionate fees for the assessment whether a tobacco product has a characterising flavour to the manufacturer/importer. These fees could be based on the costs for the establishment and maintenance of the panel as well as the testing process (the actual costs per product).

#### **5.5.2.1 Analysis responses tobacco industry to the questionnaire on cost aspects involved in the sensory analysis of tobacco products**

Seven tobacco manufacturers and the confederation of European cigarette manufacturers (CECCM) responded to the questionnaire; see the Table 18 for a short overview of the manufacturers that responded to the questionnaire.

**Table 18. Manufacturers that responded to the questionnaire**

CECCM	Karelia Tobacco company
Imperial tobacco	Landewyck Tobacco
Japan Tobacco International	Philip Morris International
British American Tobacco	Bulgartabac

All manufacturers completed the questionnaire with either an individual company response or a more general response as formulated by the CECCM/ESTA.

Only few responses indicated actual costs: mostly total costs were indicated or costs were not specified at all. When costs were indicated, general numbers were provided without details on the specific types of costs. Reasons given by the companies for not providing these details are diverse. Manufacturers indicated that this information is commercially sensitive, that costs vary largely depending on the analysis and are therefore difficult to define and/or that a third-party agency performs sensory research. In the latter case, the data might not be as easily available.

In the first part of the questionnaire, general questions were asked regarding the sensory analysis of tobacco products carried out by the company. The tobacco companies indicated that they make use of either consumer-, expert- or semi-trained panels. Expert panels were generally trained in house and composed of employees. This results in relatively low personnel costs. Both smelling (including unburnt tobacco) and smoking experiments are performed. A reference product is not used since assessments are often comparisons between products. The type of screening and training methods, as well as procedures to define panel performance, give information on the costs involved and matches our estimate.

In the second part of the questionnaire, the procedure and costs for setting-up a sensory panel were addressed. Information provided on panel size and number of panel leaders gives information on the costs involved. In addition, information on the working conditions needed for sensory assessment gives an indication of costs involved. The need to work under quality norms as well as the use of ventilated facilities and associated costs confirm the information we have on this. One company indicated that compensation for panel members during training and product assessment was paid, confirming our information.

In the third and last part of the questionnaire, costs on maintaining and/or running the panel are addressed. Costs for maintaining and running the sensory panel vary depending on the amount of samples assessed and type of analysis performed.

Information provided by the tobacco industry confirms that our cost estimate covers all aspects involved in setting up, training, operating and maintaining a sensory panel.

### **5.5.3 Possible modifications and alternative options to be considered for next steps**

Based on the findings of WP1, and WP2, it was decided that smelling experiments have a higher priority than smoking. Therefore, in the current method, the human perception of tobacco flavours is assessed by smelling tobacco taken from cigarettes and RYO. An additional advantage of this method is that it can easily be used for assessing other tobacco types as well, should a regulatory need arise due to changes in prevalence of use.

Smoking could be added as a second step, given that some flavours may arise from smoking, but it must be considered that it would more than double the costs of the assessment. It is anticipated that the costs of a smoking panel will be much higher than that of a smelling panel, for instance because less samples can be assessed in a certain time period, and because a dedicated room is necessary for smoking experiments due to the smell of tobacco smoke that remains present for a long time.

Regarding chemical analysis, in order to be able to define thresholds for a specific flavour, the chemical components that form the basis of the flavour should not only be identified but also quantified. In this way, a detection threshold or threshold above which a flavour is considered characterising (based on results of prior sensory analysis) can be established. In addition, the detection of components in burnt form (after smoking) could be considered.

## 6 Overall conclusions and recommendations

The overall aim of this project is to deliver a test approach for the assessment of characterising odours in tobacco products, by a combination of sensory profiling, chemical-analytical measurements, other methods, or a combination of methods. Based on review of the relevant literature, and discussions with experts from several fields, we decided that the best approach for assessing characterising flavours would be a combination of sensory profiling and chemical analysis. More specifically, we proposed a combination of a trained expert panel that assesses odours by smelling tobacco samples, with headspace GC-MS.

To test this approach, we performed pilot experiments with a semi-trained expert panel and headspace GC-MS to test the procedure. Based on our findings, we proposed a robust and feasible procedure for assessment by a sensory expert panel, complemented by chemical analysis, laid down in the main documents and the two SOPs.

Based on our findings, we conclude that:

- An expert panel is a good approach to assess characterising flavours. The pilot demonstrated that limited training already produced fairly good results in terms of validity, robustness and reliability. However, the fact that the panel is still learning adds to the panel variance. To further improve the performance, and to further reduce the variance, and to ensure that the results are sufficiently valid, robust, reliable and reproducible, the panel training needs to be more extensive, e.g. 6 months.
- Smelling is the preferred starting point, as it captures most of the products with characterising flavours. At a later stage, it could be reconsidered if a smoking panel needs to be added.
- Products marketed to contain characterising flavours (cherry, menthol, vanilla) can be distinguished from a reference space of 'normal' commercial cigarettes brands and natural tobacco leaves. 'Clearly noticeable' is here defined as significantly different from reference space (§5.3.1.7.1), which contains only noticeable smells (as generated by panel consensus) Chemical analysis with Headspace GC-MS is a suitable method to identify flavour components and additives. The results can be used to build a flavour library. Comparing results of flavoured cigarettes with those of natural tobacco leaves gives an indication whether a component is added. If the component is not present in natural tobacco leaves, it is most likely added. To set thresholds, components could also be quantified using standards.

In Annex V, we give some more detailed conclusions and recommendation to set up a reliable method.

## **Annex I. WP1: Literature review - Current methods and approaches for the determination of characterising flavours in tobacco products**

In order to improve the functioning of the internal market and to protect the health of European citizens, in particular, by reducing the smoking prevalence among young people, the new EU Tobacco Products Directive (TPD) 2014/40/EU prohibits cigarettes and roll-your-own tobacco having a characterising flavour other than that of tobacco.

A systematic literature search has been carried out in order to describe current methods used for the analysis of flavours in food or consumer products that can be applied for tobacco sensory analysis. In addition, methods used for the analysis of the composition of tobacco products are described. In this literature review, not only scientific data sources but also digital libraries providing information on methods used by the tobacco industry and data sources providing information on current European Union regulatory methods were consulted.

Sensory analysis to determine a characterising flavour of a tobacco product before and upon its intended use involves smoking and/or smelling the product in burnt and/or unburnt form. The different methods that are able to provide information on this are described. The formation of test panels is required to provide information on human perception of tobacco product flavour. Aspects relevant to tobacco sensory panels such as type of panel (expert vs. consumer), selection of panel members, test facilities and type of testing method are discussed. Quantitative or qualitative test methods involve detection by difference testing or descriptive testing. Data analysis methods, data reporting and different types of scaling techniques to encompass a range of intensities used to assess and categorize product attributes are described.

In addition to sensory analysis using human subjects, alternative methods are described as well. The methods involve analysis of sensory responses in the brain by neuro-imaging, sensory testing in laboratory animals and chemical analytical analysis of flavour components. In the discussion, the main characteristics of each method described have been summarized and evaluated.

For the assessment of characterising flavours, several methods are suitable either separately or in combination. As characterising tobacco product flavour requires sensory information from consumers and/or trained assessors, a first step must always be to perform sensory evaluation of tobacco products. In addition, flavour components can be identified, and quantified, using chemical analysis.

### **1 Introduction**

#### ***1.1 The EU Tobacco Products Directive on regulating tobacco flavours***

In order to improve the functioning of the internal market and to protect the health of European citizens, in particular, by reducing the smoking prevalence among young people, the new EU Tobacco Products Directive (TPD) 2014/40/EU prohibits cigarettes and roll-your-own tobacco having a characterising flavour other than that of tobacco. Tobacco products are commonly made attractive in order to encourage their use (18). The taste of tobacco can be made more attractive by adding flavours for example to mask and reduce the harshness of tobacco smoke. Cigarette smokers identify flavour as an important factor in the pleasure derived from smoking and for their choice of cigarette brand (19,20). For instance, the sweetness of cigarette smoke appeared closely related to

satisfaction and pleasantness (21). There is growing evidence that young people and female smokers are more attracted and vulnerable to the use of flavoured cigarettes (22-25). Teenagers who had ever tried flavoured tobacco products were more likely to be current smokers than those who had never tried flavoured tobacco products (26). It has been suggested that certain flavoured tobacco products may not only influence smoking initiation, but also result in greater dependence on nicotine or worse smoking cessation outcomes (27).

This literature review is a part of the project, which is intended to support the Commission in implementing the TPD with regard to laying down uniform rules for the procedures for determining whether a tobacco product imparts a characterising flavours and establishment of an independent advisory panel at EU-level. The overall aim is to deliver a method to decide whether tobacco product on the market imparts, either in unburnt or burnt form, a characterising flavour other than tobacco.

Flavour can be described as the perception of chemical compounds that affect both the sense of smell and taste and is detected by the olfactory and gustatory system. A characterising flavour is defined in Art. 2(25) of the TPD *“a clearly noticeable smell or taste other than one of tobacco, resulting from an additive or a combination of additives, including, but not limited to, fruit, spice, herbs, alcohol, candy, menthol or vanilla, which is noticeable before or during the consumption of the tobacco product.”* In this definition, tobacco means *“leaves and other natural processed or unprocessed parts of tobacco plants, including expanded and reconstituted tobacco.”*

This literature review will describe current methods and approaches, which could be applied to determine characterising flavours in tobacco products. While the focus will be on available scientific and technical literature on all relevant aspects of sensory analysis, alternative methods were also considered.

## **1.2 Systematic literature selection and review**

We developed a search strategy to retrieve relevant information dealing with the topic of the report from sources such as scientific literature and tobacco industry documents by using appropriate initial search terms followed by review and selection of relevant papers.

## **1.3 Collection of information**

As appropriate data sources for the literature review, we used scientific literature databases (PubMed and Scopus) and tobacco industry sources. We included sources as digital libraries containing tobacco industry documents (tobaccodocuments.org, legacy tobacco documents). These documents not only contain information on advertising, manufacturing, marketing and sales, but also on scientific research activities. Other literature sources on tobacco research and product development that we consulted are websites of tobacco manufacturers, the tobacco research institute and the Fawky Abdallah Company. Throughout this review, tobacco industry documents are highlighted with an asterisk behind the reference ([ref]\*).

As additional source to retrieve information on sensory analysis methods for tobacco products, and other methods useful for assessing characterising flavours, a questionnaire (Appendix V of this annex) covering these topics has been sent to EU regulators via the email lists of the EU Expert Group on Tobacco policy and the Expert Subgroup on Ingredients. Regarding third countries, we sent our questionnaire to countries that regulate ingredients or have legislation on flavours/flavourings, such as the USA, Canada and Brazil. We also sent the questionnaire to selected sensory scientists.

For literature on methods for sensory analysis of consumer or food products, we consulted not only scientific literature databases, but also EU regulations and ISO standards regarding this topic. ISO

methods have been defined for sensory analysis of consumer or food products and can also be applied for tobacco sensory analysis.

A systematic search was performed, based on several keywords or keyword combinations (Appendix I of this annex).

#### **1.4 Selection and review of the retrieved literature**

Articles were selected based on their relevance. First, titles were screened, and abstracts were read in case of relevant titles. Papers were read if both abstract and title seemed relevant.

Selection criteria for final inclusion of articles in the review are relevance for smoked tobacco products in burnt and/or unburnt form (e.g. smoked or unsmoked cigarettes), quality of data, useful for the purpose of the product, possibility to translate approach into a method suitable for legislation.

For the applicability in standardized protocols and suitability for legislation it would be advantageous if the analysis method of choice was already successfully used for products or even referred to in EU legislation. Therefore, we focussed our selection of articles on analysis methods that are either currently used to determine the composition of tobacco products or on established methods to determine characterising flavours in other consumer products or food.

#### **1.5 Review set up**

This review will provide a comprehensive overview of methods used for sensory analysis of tobacco-, other consumer or food products and alternative methods for the determination of characterising flavours.

Sensory analysis to determine a characterising flavour of a tobacco product before and upon its intended use involves a combination of smoking and smelling the product.

In this review methods recorded to be used by the tobacco industry, methods for assessment of tobacco products by EU member states and third countries, methods for sensory analysis of any type of consumer or food products and alternative methods not based on sensory analysis will be described. For all methods aspects as pros and cons, accuracy, costs, time-investment and ethical issues will be described. Methods are applicable for sensory assessment by both smelling and smoking the tobacco product.

Aspects that are applicable to the sensory evaluation of tobacco such as type of assessors, selection of assessors, test facilities, type of method will be discussed. Success of such human testing approaches will depend on validation of such methods using state-of-the-art sensory practices. In the discussion, the main characteristics of each described method will be summarized and evaluated.

## **2 Approaches relevant for tobacco sensory analysis**

In this chapter, we describe general principals of sensory methods relevant for tobacco flavour analysis. Tobacco product flavour can be assessed for instance by smelling and smoking. This assessment can be performed by consumer or expert testing using a wide variety of methods. The details of such methodologies will be described in more detail in §3 of this annex.

### **2.1 Smelling versus smoking tobacco**

To determine the characterising flavours of cigarettes, the cigarettes can be smoked and smelled. Smelling cigarettes could be a suitable alternative for smoking as it does not include health hazards



associated with smoking. Smelling experiments could include smelling of the cigarette package, tobacco, the side stream smoke of burning cigarettes, or the Cambridge filters. Cambridge filters contain the particle fraction of tobacco smoke, substances that are formed during tobacco smoking. The substances are captured on the Cambridge filters when smoking a cigarette with a smoking machine.

By smelling the package and isolated tobacco, assessors can provide information concerning the flavour of unburnt cigarettes. The smell of side stream smoke represents the odour of burnt cigarettes, but in this approach, the flavour dilutes immediately in the surrounding air. Therefore, it is difficult to recognise flavours in side stream smoke. In addition, smelling side stream smoke (second hand smoking) presents a large health risk.

Another possibility to examine the flavour of burnt cigarettes is using Cambridge filters (28). The filters collect the total particulate matter in the mainstream smoke (29). Therefore, smoke components present in tar, and to a lesser extent vapour components are trapped on the filter. Examples of components in tobacco and smoke that contribute to tobacco flavour are carbohydrates, acids, (poly)phenols, carbonyls, pyrazines, pyridines, and various aromatic components such as vanillin, benzaldehyde, and phenyl ethyl alcohol (30, 31).

Consumers and trained assessors can both perform smelling and smoking tests, following several methods as described in the following paragraphs. The advantage of smelling tests with unburnt cigarettes is that they can be performed by specific samples of consumers for example teenagers and young adolescents, (target population) and by non-smokers. However, for ethical reasons this group of people cannot be asked to smoke. A disadvantage of smelling tests is the lack of being able to judge characteristics that do (partly) determine the concept of a flavour such as taste-, nasal- and oral feelings. One cannot be sure without pilot testing whether flavours perceived through smelling are the same as flavours perceived when smoking a cigarette. The tobacco industry combines smelling the package (closed and open), smelling the cigarette, and ‘puffing’ (without inhalation) and smoking cigarettes when assessing quality of tobacco products (32, 33)\*.

**Table 19. Comparison of smoking and smelling**

	<b>Advantage</b>	<b>Disadvantage</b>
<b>Smoking</b>	Way of utilization	Longer training process
<b>Smelling</b>	First step in using the product	Shorter training process

## **2.2 Consumer versus expert testing**

### **2.2.1 Consumer testing**

In a review on consumer response to tobacco products, both field and laboratory testing methods for assessing responses to tobacco products have been described (34)\*. Field-testing is usually performed among consumers. It involves one-stick (cigarette) testing, and in-home product testing. Tobacco companies may ask targeted consumers their brand perception after smoking a single cigarette or perform large-scale consumers testing on product acceptability. A recent tobacco industry peer-reviewed study on potential reduced tobacco exposure products describes such a consumer acceptability tests (35)\*: *“Groups of 50 smokers of specific popular brands of commercial cigarettes were recruited. Smokers were aged between 21 and 60 years, and the recruitment target was 50% male and 50% female subjects. A number of studies were performed: direct, single blind comparisons of experimental cigarettes or a tobacco blend control against smokers’ usual brand at a central laboratory location; single blind analysis of experimental cigarettes compared to smokers’ usual brands or tobacco blend controls at a central laboratory location; and open label, home use studies with limited numbers of experimental cigarettes. Detailed study plans were explained to all*

*participants and informed consent documents were required to be signed as a condition of enrolment.*” Although these tests have some value for quick screening for further development work, e.g. on liking, taste, ease of draw, they are not suitable for an in-depth assessment as attributes like harshness, blend characteristics, and other sensory aspects (36).

For in-home product testing, tobacco industry provides pre-recruited respondents with one or more packs of the test product to use for a limited time and provide ratings on overall satisfaction, purchase interest, and several sensory attributes. Products can be send/given to participants in form of a monadic presentation (participants receive the product one by one, 1 panellist only evaluates 1 product) or by carrying out sequential monadic tests (products are given one by one but 1 panellist assesses multiple products): The tobacco industry most often uses the pure monadic presentation (37)\*. Philip Morris uses a large European consumer panel to perform test on quality control, and new product and concept development (37)\*. Combining consumer panels from 9 different countries in Europe they received 55,600 responses from panellist in 1 year (1997)(37)\*. Advantage of in-home testing is a more naturalistic smoking environment compared to laboratory testing, and the possibility to test larger number of subjects. Disadvantage is that there is no control over smoking behaviour and compliance to the testing paradigm.

Conference proceedings on chemical and sensory aspects of tobacco flavour report that consumers can be tested via mail, in-home test coupled with an interview, or at a central location (31). Consumer perceptions and reactions are typically tested in consumer panels or large scale consumer surveys for attributes like strength, harshness, smoothness, tobacco taste, aftertaste, mouth drying, impact, irritation, sweet taste, bitter taste, cooling sensation, menthol taste, mouthful of smoke, off taste, ease of draw, mouth coating etc. (36).

### **2.2.2 Expert testing**

Expert panellists are able to use technical or highly specific attributes to describe products (38). It requires training of panel members to judge products in a standardized manner using the same terminology (39) (Appendix IV). An expert panel usually includes 10-12 panellists and is suitable for frequent testing, for example for quality control. Many laboratory-based assessments performed by tobacco industry include sensory evaluation expert panels (34)\*. Expert panels can be used to evaluate the effect of specific product characteristics on sensory response (including, taste and aftertaste, odour, mouth feel, kick). For example, a review on tobacco industry documents by Abdallah describes how an expert panel is trained and used as a research tool for evaluating the smoke flavours of experimental tobacco blends (30)\*. He proposes using highly trained smoke panellists as research tools for evaluating the smoke flavours of experimental tobacco blends (40)\*. Expert panels are also used in quality trademark assignment of olive oil and butter. When the panel was sufficiently trained the sensory test was adopted as a trade standard, to detect defects and to qualify the intensity for the classification of different grades of olive oils and to assess butter based of defect quality assessment (41, 42).

The two main advantages of using an expert panel is that obtained results can easily be reproduced and when the panel is proper trained there is consensus between panel members and small variability, therefore less panel members are needed compared to a consumer panel (38). Generally it is thought that using experts for product testing gives more reliable and valid results than when judged by an untrained panel or consumer panel (43). However, a study of T. Worch et al. found that expert and consumer panels are similar in discrimination ability and reproducibility but more consensus was found between panel members from the expert panel, which is likely due to their experience with the product, and their training. The higher variability when performing analysis with consumers can be compensated by using a sufficiently large consumer panel. Therefore T. Worch and colleagues conclude that the performance of consumers does not differ from the performance of a

trained expert panel and that consumers could also be used for classical profiling (38). An important factor to consider when choosing to use either an expert or a consumer panel is the origin of the data and the studied population. Examples of studies that found comparable results are a study that investigated the categorization of beers, where the authors did not find an effect of sensory training on the organization of the system of categorization of different beers. They also concluded that because of the chemosensory knowledge of professional tasters they might perceive novel products differently than the average consumer (44). A study, that described flavours of wine, found that trained experts used more words to describe an odour than non-experts did. In addition, experts were better in matching their own descriptors and repeatability was also better for them (45). Another study that looked at differences between an expert panel and a consumer panel in describing attributes of sauce/gravy found that the two types of panels obtained similar results when instructed to evaluate the same set of products on similar attributes. However, findings of the expert panel were less variable. This study concluded that consumer panels could be used to describe attributes of products (46).

### 2.2.3 Cross-cultural differences between panel members

Cultural and demographic differences (e.g. language and familiarity with the product, or age and gender) may influence olfactory function (47) and thus may play a role in how consumers perceive and evaluate tobacco flavours. However, such cultural differences are mainly found for recognition, identification, associations or pleasantness ratings, and not specifically in the analytical scoring of attributes. Moreover, the aim of training a panel to identify and correctly label specific attributes is meant to overcome such possible cultural or demographic differences, until the panel as a whole reaches consensus and rates similarly (22). For example, a panel with highly trained panellists may optimally reflect differences in specific attributes, but may offer less information about whether these differences actually contribute to “characterising flavours” to a naïve consumer (which may even vary per culture). In contrast, a consumer panel requires less training, can be set-up relatively easy in different cultures, and offer more information on flavours that would be considered “characterising” or attractive in that culture. On the other hand, this panel will be less suited to detect differences in specific attributes. Any method using sensory panels will have this ‘risk’. Only methods such as chemical-analytical methods, will not show cultural differences if present, but on the other hand also cannot capture human experience.

Moreover, possible cultural differences between consumer panels may not lead to differences in discriminating between samples, or choosing and rating attributes (e.g. for profiling of dark chocolate in two panels, from Austria and Germany “the panels distinguished between the samples on a comparable level and used a comparable terminology with respect to type and frequency of the descriptors” (48). In the quality evaluation of milk and milk products results obtained by assessors from different regions of a member state and different member states are compared where applicable. If significant differences are observed an investigation should aim to arrive at comparable results (41). Additionally, if differences arise, they can be taken into account in the statistical analysis, as suggested by Sebastien Le (49) “...allows evaluating the sensory description of all the panels in the same framework even if the number of panellists is not the same and even if the descriptors used in each panels are not the same”.

At present, researchers at WU are conducting a cross-cultural study to sensory characterise a range of food products in the Netherlands as well as Malaysia, to determine if and how products differ between countries, as well as determine whether cultural differences exist between trained panels (SVT study). This study indicates that – even though products are branded the same - differences in product content and thus sensory profile exist between cultures. For instance, a Yakult drink from Malaysia differs in content and sensory attributes from a Yakult drink in the Netherlands. This may also apply to tobacco products, in that a commercial brand cigarette in the Netherlands is perhaps

not exactly the same as in France or Poland. This should be taken into account when determining the 'reference flavours'.

As known from tobacco literature Philip Morris has been using different test methods (monadic or sequential monadic) for different countries, depending on the amount of testing planned in that country (37)\*. In Europe, PM has established a large consumer panel, which does the entire blind consumer product, testing (established in 1993, in 9 countries, interviews are conducted in all European languages). Consumer test are used for monitoring, maintenance and new product and concept development, paying considerable attention to cultural differences between markets in how attributes are scored, with the focus on hedonic differences between cultures. Moreover, only Marlboro FF was tested in all markets, different subsets of (competition) brands were tested in a subset of countries.

A document of Philip Morris describes "Differences in culture and language can have a profound effect on the types of descriptors and scales that are appropriate. These differences need to be understood before testing is initiated. Liking is an example of a scale where cultural/language differences make a profound difference in the type of scale that is used. The end points that are used for English speaking countries are "dislike extremely" at the low end and "like extremely" at the high end. In many countries where the language is derived from Latin or Germanic roots (for example, French, Spanish, German), the endpoints are typically "don't like at all" to "like extremely". While the English version has been found by some researchers to work in non-English speaking countries, it does not seem to be the natural wording for those countries. The French/Spanish version will not work in English speaking countries, since all products that are disliked are rated at the lowest scale point, regardless of the degree of disliking. This causes unequal variances between products that are liked and those that are disliked" (37)\*.

### **3 Methodologies for measurement and data analysis (relevant for tobacco sensory analysis)**

This chapter outlines the process of how products are assessed, how data is analysed and results are obtained. It should be noted that all sensory methods described in this chapter are applicable for sensory assessment by both smelling and smoking the tobacco product. We will describe the relevant analytical discriminative and descriptive sensory methods that are currently used for tobacco or food- and consumer products.

Sensory methods can be classified into two groups, analytic methods such as discrimination (difference), descriptive (describing) tests, and affective methods such as acceptance- and preference tests. Affective tests involve the acceptability of a product, or preference of one product over another. These tests are not useful for our purpose since they are mainly used for product improvement and new product development and not for comparison of differences between samples or providing descriptions of the sensory qualities.

#### **3.1 Discriminative methods to test for sensory product differences in consumers or experts**

Discriminative methods are used to explore perceptual differences between almost similar samples, and can be performed with consumer or expert panels.

Basically, there are two kinds of difference tests (50), 'un-directional difference tests' to examine if a sensory difference exists between samples, without specification of the nature of the difference and (51) 'directional difference tests' to explore whether the two samples differ on a specific attribute. In case of directional difference tests, panel members are instructed to make a decision based on a

specific attribute. In general, all tests described below can be un-directional or directional based on the instructions given to the participants, and knowledge of the experimenter of the aspects that differ. Directional difference tests are statistically much more powerful than un-directional tests.

A large number of difference tests exist, of which the Triangle test is most often used. However, not all of these methods are applicable for tobacco research. In the following paragraph, the most relevant methods that can be applied in tobacco research will be discussed: The Duo-trio test, Triangle test, A-not-A test, tetrad test, 2-AFC, and difference from control test. The sensitivity and reliability of a test depends on the sample size and the set-up of the test. For each method, we discuss whether or not the method would be suitable and relevant to determine whether a tobacco product imparts a characteristic flavour other than tobacco. For further details about these test methods we refer to the following books (39, 52).

### 3.1.1 Unidirectional difference tests

For example, Triangle and Duo-Trio tests are frequently applied in tobacco industry research but mainly for comparative reasons (is variant A the same as natural tobacco) (34)\*. British American Tobacco used the duo-trio test, where a smoker has to choose which of two cigarettes is equal to a control cigarette (53)\*. Although it does not seem very relevant to measure if two brands differ from each other it could be useful to estimate the perceptual difference between two samples by means of Thurstonian scaling (calculate the number of correct responses, or possibly ordered categories of response) when ranking methods are applied. This will result in a measure of the perceptual difference ranging from 0 to 3. However, as for any difference test, a well-defined reference is needed.

The triangle test method is used when the test objective is to determine whether a sensory difference exists between two products. The triangle test is particularly useful in situations where treatment effects may have produced product changes that cannot be characterised simply by one or two attributes. Subjects receive three coded samples, and are instructed to taste and examine the products from left to right, and they are told that two of the samples are identical and that one of the samples is different. After tasting all samples, subjects have to choose the odd sample.

For this method, no training is needed but subjects need to be familiar with the procedure of the triangle test and with the product being tested. In general, 20-40 subjects are needed for this test. The power of the test is low when you want to conclude that two products do not differ from each other (i.e. that they are similar). In that situation, 200 or more judgments are needed.

A tobacco industry document describes the use of triangle tests odour evaluations of Cambridge filter pads moistened with mineral oils to capture side stream smoke odours (28)\*. Differences between odours of Flue-cured, Oriental and Burley tobacco were detected by expert panels of 5-12 members (both smokers and non-smokers). Differences in blends of two of these tobacco types could also be detected, but not those of mixtures of three (28).

The tetrad method is an alternative method for the triangle test, subjects receive four samples for evaluation—two samples from product A and two samples from product B. Since there are four samples, the power of this test is much higher. Like any difference test, the Tetrad can only measure if two samples are different or similar (54).

Another test to measure if two samples are different or similar is the A-not-A test. It is used when products have for instance a strong and/or lingering flavour, and can therefore not be measured with the triangle test. For this test, subjects have to be familiarized with samples 'A' and 'not A'. These samples are only available until the start of the test. Next, subjects evaluate one, two or a set of

samples and indicate whether it is sample 'A' or 'not A'. Subjects have to be trained to recognize sample 'A' and sample 'not A', but this is not a very extensive training. According to the literature, about 10-50 subjects have to be 'trained'. As subjects need some training to understand the test procedure and to recognize sample 'A' and sample 'not A', this test method will take some extra preparation time. Still, this is a relatively cost- and time-efficient method (39).

### **3.1.2 Directional difference tests**

In the paired comparison- or two-alternative forced choice (2-AFC) test, two samples are compared and subjects are asked in which way two samples differ on a particular sensory characteristic (which one is the sweetest?).

This test can be conducted with naïve consumers; no training is needed. They have to be able to identify the specified sensory attribute. This training is not very extensive but might differ per product and also depends on the sensory attribute. The exact number of required subjects is dependent on whether the test is one- or two-sided and also depends on the values chosen for the test-sensitivity parameters. The power of this test is much higher than the power of the unidirectional difference tests.

This test might be interesting for this project as subjects could for example be asked for which of the two tobacco samples sensory attribute X is higher (e.g. the chocolate or strawberry odour). It is a feasible and cost-efficient method (39).

The difference from control test is another type of directional difference test. It is used to explore whether one or more samples differ from a control and to estimate how large that difference is. The difference from control test is actually a basic directional difference test with an additional question about the size of the difference. The test is used most often in quality control situations. Subjects need to be trained to be able to identify the sensory attribute(s) of interest (39).

The difference and acceptance tests measure either liking, acceptability or the likelihood that two products are perceptibly different or identical. With regard to this project, these tests would not identify characterising flavours other than tobacco without a well-defined reference. The reference that determines when a tobacco flavour is different can however not be precisely defined since natural unflavoured tobacco mixtures are variable. Therefore, it is necessary to perform more analytical test for example by describing attribute intensities. Methods that are able to do this will be described in more detail in the paragraph below.

## **3.2 *Descriptive methods using consumer or expert testing to assess sensory product characteristics***

Descriptive analysis aims to provide a quantitative measure of the sensory properties for a set of products by means of a sensory panel (55). A descriptive test provides word descriptors of product characteristics. These word descriptors can be used for comparing product similarities and differences. Furthermore, results can be used to relate specific product ingredient to the sensory attributes of a product (46). For example to profile flavour analytically, the technique of descriptive analysis with the usage of an expert panel has been applied to a variety of products including wines, cider, beer, and whisky, fruits and vegetables and cheese (56). Similar to our research aim with cigarettes, the wine industry uses odour descriptive methods to determine classification of natural and additive odours. Wine odour intensities are measured with an expert panel at different concentrations levels as determined through chromatography. The odour intensity level determines classification of odours, whether they are naturally present in wine or when odours are considered to be an additive (56).

In the following paragraphs, we will review the major methods of descriptive analysis methods. For descriptive testing with consumers or experts, fast descriptive or verbal based methods are used such as quantitative descriptive analysis, the free choice profiling, flash and ultra-flash profiling, ideal profile and CATA (39). Descriptive testing with experts can additionally be performed using flavour profile, spectrum method and hybrid method.

The different descriptive methods can be performed with a consumer or expert panel. The main difference between consumers and experts in descriptive tests are the attributes used to describe product characteristics. Experts have been trained and use different (often more technical) attributes than consumers and have less variability in their ratings. To compensate for this, consumer panels are much larger than expert panels (a factor 10 is normal). We will discuss the feasibility, sensitivity and accuracy of the methods, and their suitability for our research aim: to determine whether a tobacco product imparts a characteristic flavour.

In Figure 21 an overview of the major methods used for descriptive profiling in order of the training intensity needed to perform the method (55).

### **3.2.1 Quantitative Descriptive Analysis**

Quantitative descriptive analysis (QDA)<sup>®</sup> is a method developed in 1970 by Stone and Sidel (57). With this method individual panellist identify and quantify sensory properties of a product (31, 34)\*. Repeated judgments have to be made and the resulting scores can be statistically analysed to determine differences among products (31).

Quantitative descriptive analysis is performed with 10-12 semi-trained panellists. The panellist should be product users or likers and have discriminating abilities. Panellists are trained to understand the methodology used. During the training sessions, panellists are exposed to a wide range of products to facilitate accurate concept formation. To describe the product attributes consumer terminology is used developed by the panellist through consensus with specific definitions, anchors and reference standards agreed upon by all panellist. For most foods approximately 30 attributes should be developed (55). Rating of the intensity of the attributes is done on a 15 cm VAS-scale, panellist rate according to their own scale, references standards are not often used (10% of the time) usually the reference is a verbal definition.

The accuracy (bias) and precision (variability) of the QDA<sup>®</sup> method are dependent of panel performance and should be measured over time. Bias can be controlled by performing a control test whereby panellists need to match the accepted intensity of the attributes of a control or reference product. Variability of panel members (within measurement of one panel member) can be measured through repeated ratings of the same control or reference product. When both bias and variability are low for all panel members individually, the QDA<sup>®</sup> method is considered an accurate and precise measurement instrument (13).

Quantitative descriptive analysis is used to describe dairy products (58, 59) grapes and wines (60, 61) fruits and vegetables (62-64) and is described in more detail in sensory methodology books (13, 39). A slightly different version of this method is used in the olive industry to classify different quality trademarks. However, for this purpose changes were made to the originally QDA method as a different scaling was used which is combining this method with the hybrid method (§3.3.3 of this annex) (65).

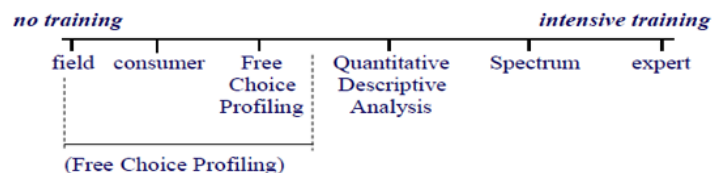


Figure 21. Overview of descriptive methods (66)

### 3.2.2 Free choice profiling

This method is characterised by panel members using their own vocabulary with inconsistent definitions and is mainly used for product development (39). Panellists evaluate the product in their own way and describe all stimuli that interest them. Panellist rate product attributes on an intensity scale according to their personal rating standards. This method has been used by the tobacco industry to understand consumers' perception on the characteristics of cigarettes, “white” cigarettes as well as kretek, to provide the basic for product development (67)\*. For the free choice profiling method a large number of consumers are needed (39). However, no panel training is needed for this method, which makes this method time efficient. This method fits our research aim in a way that it is not important to identify the precise tobacco flavour, but only to determine if the tobacco flavour differs from a reference, for example from natural tobacco. Advantage of this method is that it allows diversity of points of view and therefore panel members of different countries using different languages can be included in studies using this method.

### 3.2.3 Flash and ultra-flash profiling

This method provides a relative sensory positioning of a set of products and is based on the free choice profiling and ranking methods (68). The flash profile consists of three sessions. During the first session, panel members have to observe, smell and taste the products and generate a set of attributes. A prerequisite of this method is that products that need to be assessed differ sufficiently to permit ranking the products. During the second session, a list of all the generated attributes is given to the panel members so they can update their own list by adding or replacing attributes. During the last session, the panel members have to order the products from least to most on their chosen attributes. Flash profiling is popular among food companies and has been applied in studies to fruit products (69) (70, 71), chewing gum (72), bread texture(73), and lemon ice tea (74).

Napping or projective mapping is a sorting technique whereby consumers are asked to place each product on a surface (i.e. large blank sheet or paper) according to their similarities and dissimilarities. The criteria used to place the product is up to him/her, no structure or point of view or attributes are imposed by the panel leader (39).

The ultra-flash profiling combines the Napping method (75) with Flash profiling. For this method panel members are asked to place products on a square piece of paper or computer map according to their similarities and differences. In addition to the ranking panel members have to describe with the usage of their own vocabulary on what attribute the products differ or are similar (in case products are placed close to each other (76).

No panel training is needed for this method however panel members chosen for the flash profile should be sensory evaluation experts or product experts (39). Product maps produced by the flash profiling and Quantitative Descriptive Analysis (QDA®) method (§3.2.1 of this annex) are similar (39, 77). Advantage of this method is that it allows diversity of points of view and therefore panel members of different countries using different languages can be included in studies using this method.



Another advantage of these methods is that results are not based on the assumption that odour or flavour percept can be analysed and reported using independent descriptors. Intensity ratings of separate odours as done with the QDA<sup>®</sup> and Spectrum method<sup>®</sup> might not be adequate for complex attributes such as odour mixtures, and using these methods would suggest that the odour experience is a collection of independent analysable parts while this is generally not true (78). However, a study of Lelievre and colleagues looked at the validity of fast descriptive methods and concluded that these methods are more applicable for studies focused on panel members' behaviour rather than studies that aim to find precise and reliable description of product attributes (79). This is in accordance with several other studies that concluded that fast descriptive methods are less suited than conventional descriptive analyses when detailed information of sensory properties of products is required as these methods only give rough estimates (69, 76) (80).

Another weakness is that the interpretation of sensory terms is sometimes difficult due to the large number of terms and lack of definition (81). Because this product relies on ranking it cannot be used for a large number of products due to fatigue (77).

### **3.2.4 Ideal profile method**

The Ideal Profile Method is similar to QDA<sup>®</sup> except that the products are not rated by trained assessors but by naive consumers; it resembles Just About Right scaling but instead of an implicit ideal the consumers are asked to rate their ideal intensities explicitly. This method is however not suitable for our research question as we are not trying to find the ideal or optimum flavour of tobacco.

### **3.2.5 Check all that apply (CATA)**

The various methodologies to capture consumer perceptions (CATA, Napping) are generally easier to perform and less time-consuming than traditional descriptive analysis. Consumers or trained assessors must indicate the presence or absence of descriptors in products with a predefined list with the instruction to check all descriptors that apply to the product in question (in that respect it is identical to conventional profiling except that they will not rate the perceived intensities but only the presence or absence). One of the suggested drawbacks of CATA is that this method produces relatively impoverished dichotomized data (1/0), which allegedly would mask relative differences between specific attributes. Another suggested drawback is the fact that the respondents can skip part of the attributes by not paying attention to all of them (82).

Different investigations have compared simple YES/NO CATA with CATA combined with intensity ratings (which is actually classical profiling with the option to skip attributes), and found hardly any differences. Skipping attributes can be prevented if presented one by one, with the simple question whether they applied or not, and this is advised for tobacco flavour research. When we would perform the CATA method for tobacco flavour research, we advise this more accurate procedure of presenting attributes one by one instead of a list with attributes.

Comparison of these different methods (CATA, mapping, QDA, ideal profile) shows a high similarity between methods as confirmed by the RV coefficients ranging from 0.90 to 0.97. These results suggest that the precision and reproducibility of sensory information obtained by consumers with CATA is comparable to that of Napping and traditional profiling. The forced choice option is preferred to the scaling option because the primary aim is to identify the different flavours in the samples. Although rating could be applied with naïve or trained assessors, it is less practical with naïve consumers and rating does not improve the results substantially. Regardless of question format, consumers, on average, perceived the tests as easy and not tedious (83).

The results of CATA can be analysed uni- or multivariate. Each product can be characterised by the frequency with which the attributes have been selected. Differences between products can be analysed by non-parametric tests. A Combination of the Stuart-Maxwell frequency test and the McNemar test has been recommended to test difference between products (84). The data can also be analysed by Correspondence analysis or Principal component analysis (PCA), which results in a multivariate product space. By computing confidence ellipses around the products, clusters of identical products can be defined. When enough samples have been tested, the resulting product space can be used as a reference and profiles of new products can be projected in the space as illustrative (the reference space does not change) to see in which category they belong (85).

### ***3.3 Descriptive methods using expert testing to assess sensory product characteristics***

#### **3.3.1 Flavour Profile**

This method focuses on the flavour characteristics of the product, their intensities and order of perception. In addition, Panellists have to give an overall score of the flavour impression of a product (55). Samples should be served to the panellist in the same form as consumers would receive it (39). The flavour profile method uses 4-6 panellist, highly trained in language and methodology (86). To describe product attributes technical terminology is used with specific definitions and reference standards, as agreed upon by all panel members. Panellists use the same rating scale and then discuss to come to a consensus.

To rate the flavour attributes, panellist assign a number or symbol to the intensity of the flavour from 0=undetectable to 3=strong (Appendix III of this annex). Due to the fact that results are achieved via panel consensus, no statistical analysis is possible. A solution for this would be the Profile Attribute Analysis (PAA) by which panellist individually score the intensities on a 7 point scale. These results could be analysed using conventional statistical procedures (39). Results from the flavour profiling of a product are reproducible when the panellist are trained appropriately, the more training panel members receive the more accurate are the results (39). A study comparing the flavour method analysis with expert panellist to mass spectrometry found good comparison between the two methods (87). This method is described in multiple books that describe sensory methodology and is also used independent from industry (13, 39). To train the panellist to define the flavours of a product a 2- to 3 week program is needed (39).

#### **3.3.2 Spectrum method**

This method is developed by Gail Civille in 1970 (88). This method distinguishes itself from other methods by the extensive use of reference lists, specialized panel training and scaling procedure (13, 39, 89, 90) (Appendix IV of this annex). The spectrum method® is based on the texture profile method but does not only focus on the texture attributes of products but on all product attributes. For this method 12-15 panellist are needed, they should be highly trained in language and methodology and are expected to have a basic knowledge of physiology and product development (90). Terminology to describe attributes is derived from the panellists; however, no panel specific vocabulary is developed to describe sensory attributes. Instead, technical terminology from a standardized lexicon of terms is used. The terms used to describe product characteristics are determined before testing and remain the same over the course of testing products within the same product category. Ratings are done on a 15-point numerical scale and panellist use the same universal scaling. Scales are standardized and anchored with multiple reference points. The use of reference points makes this method suitable for comparison with instrumental data (39, 90). Panellist are extensively trained, standard times for each stage are 15-20 h for terminology

development, 10-20 h for introduction to scaling, 15-40 h for initial practice, 10-15 h for small product differences and 15-40 h for final calibration (50, 90).

Because of the extensive training panellist use the scale identically; therefore with this method resultant data are absolute values instead of relative scores as with the QDA<sup>®</sup> method (39).

The Spectrum Method<sup>®</sup> has been applied in several studies (91-95) and many other studies use aspects of this method in sensory research (90).

### 3.3.3 Hybrid Method

This method mainly distinguishes itself through combining attribute ratings with a quality scale (Appendix III of this Annex) (96). The quality scale combines overall line ratings on for example a 10-point scale, with attribute categories that can be used to reject a certain product on the basis of a particular attribute. An important part of this method is the fact that the boundaries of the “accept” and “reject” range are determined before training, preferably in a consumer study but at least with authority input (in our study the EU commission could fulfil this role) (39).

This method is for example applied to determine the quality of butter. To assess the quality, panellist have to rate butter samples bases on appearance, consistency and flavour/aroma and have to give points (range 1-5) from very good to very poor whereby different attributes fall under a certain class. For example, attributes specific for class 1 are the following: very poor flavour/aroma, acidic, mould flavour, over salted, oily/fishy. For class 5 these are ideal type, highest quality (absolutely pure finest aroma). If the butter sample has a score lower than 3 (out of 5) the butter gets the description ‘poor product’ (41).

The defined reference standards or tolerance range should be shown to subjects with the usage of reference products to establish the concept boundaries (97). The hybrid method uses standard technical terminology and specific definitions and reference standards. For this method, 6 panellists are needed who should be highly trained in language and methodology. Panellists rate the attributes individually and no consensus discussion is held (39).

## 4 Alternative methods

In this chapter, alternative methods to sensory analysis using assessors are described. The methods involve analysis of sensory responses in the brain by neuro imaging, sensory testing in laboratory animals and chemical analytical analysis of flavour components.

### 4.1 *Analysing sensory response in the human brain*

#### 4.1.1 Neuro-imaging techniques

Neuro-imaging techniques such as functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) can be used to provide insight in brain processes related to flavour perception. Physiological, emotional, cognitive and sensory responses caused by flavours and odours can be tested, also in relation to ‘reward dose’ in the brain. Structural MRI measures anatomical structures in the brain, which may change as a consequence of learning (repeated exposure) or training, but that are not expected to change in response to (tobacco) flavour stimuli in the current setting/research question.

The brain integrates sensory inputs such as taste, touch and smell, and the resulting neural activation can be studied by e.g. fMRI and PET (98). Many brain areas are involved, such as primary sensory cortices (insula/operculum for taste; piriform cortex for smell), amygdala, and the orbifrontal cortex.

Odours, tastants, and trigeminal stimuli (chemical irritants) are processed within the olfactory network, gustatory network, and trigeminal network, which are interacting networks (99). The widespread network involved in the processing of odorants, tastings, and chemical irritants recruits several key cerebral areas, including those involved in coding for emotions, memories, and reward. Reward consists of the psychological components learning, affect, and motivation (100).

For characterising flavours, neuro-imaging techniques may be used to study the effects of specific additives, aromas or whole smoke on the primary sensory or reward networks.

Though it is possible to present odours, tastes, trigeminal stimuli or flavours to participants lying in an MRI scanner, it is not a very realistic setting. Moreover, though it is possible to determine brain areas involved in sensory or reward processes, it is hitherto extremely difficult to compare and discriminate between similar stimuli (such as different tobacco flavours). For example, Howard et al have shown that, within the piriform cortex, odorants with more (or less) similar fMRI patterns were perceived as more (or less) alike (101). Moreover, learning processes may change these coding patterns in the brain (102). Similar results have been found for the coding of taste qualities in the brain (103). Interestingly though, a study by van den Bosch et al has shown that overlapping but different patterns of brain activation can be related to liking or disliking, and this pattern of brain activity may even predict the participants' (dis)liking of grapefruit juice with an 88% success rate (104).

In conclusion, structural MRI is not a suitable method to determine characterising flavours, but may possibly be used in a future setup that is focussed on changes in the brain related to learning processes.

Likewise, functional MRI is not a feasible method for the current research question (to determine characterising flavours), but is of interest when focussing on changes in the brain related to learning processes (repeated exposure).

#### **4.1.2 Other methods**

In past research projects conducted by Philip Morris, electroencephalography (EEG), pattern reversal evoked potential (PREP), and chemo-sensory event-related potential (CSERP) were used to measure physiological, sensory, and cognitive responses related to nicotine and cigarette additives (105)\*, and to flavourants commonly found in tobacco, such as menthol. As far as we know, EEGs, PREP and CSERP were never used to monitor responses to inhaled smoke directly. Instead, the measures were typically used to verify changes in specific components of EEG/ERP after smoking that may indicate specific effects of tobacco ingredients such as nicotine on the smoker's cognitive state, e.g. his ability to concentrate. Alternatively, the methods were used to verify perceptual differences between different types of the same flavourant found in tobacco, such as natural and synthetic versions of menthol. The results showed that CSERPs may show superior sensory discriminability compared to standard sensory tests (106)\*. Chemosensory ERP studies (olfactory, trigeminal and gustatory) have also been performed by the group of Thomas Hummel. Such studies enable identification of different components that relate to sensory characteristics of a stimulus (e.g. concentration, duration of presentation) or evaluation processes. Even though EEG/CSERPs have produced interesting results, the fact that they are typically performed on flavourants presented in isolation and not as part of the complex smoke mixture, makes it unlikely that EEG/CSERPs will be a successful tool for the investigation of characterising flavours in cigarette smoke.

#### **4.2 Sensory testing in laboratory animals**

The Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR) concludes that currently there are no animal models to assess tobacco product attractiveness (107). It may however

be possible to assess the preference of animals for the taste or smell of specific additives by e.g. conditioned place preference or self-administration paradigms. This is based on the hypothesis that a well-tasting additive will activate the reward system of the animal and therefore the animal will self-administer more of this product by pressing a lever or nose poking.

From several rodent studies there is evidence that adding sugar or sweeteners to nicotine solutions increases nicotine intake, but the evidence is not conclusive (100). Palmatier et al. describe an animal model to test a discriminable taste, grape or cherry, paired with sucrose that is meant to model a familiar appetitive taste stimulus associated with reward (108). Nicotine pre-treatment reinforced the rewarding effect, probably via learning mechanisms. Although this model needs to be further developed, it has some value to show that characterising tastes may work synergistically with nicotine (36).

A flavour preference can be conditioned by association with a positive consequence. Reinforcement of such preference is done by combining the flavour with post-ingestive consequences of nutrients such as glucose and fructose (109). Studies also report a neutral flavour to become strongly preferred in mammalian species using such methods (110). In addition, taste aversion can be conditioned in this way.

It is most likely not possible to assess the taste and smell of whole smoke, but only that of (simple mixtures of) additives and smoke components. An animal cannot be taught to self-administer tobacco smoke, although smoke condensate may be an alternative since it has been shown to be more reinforcing (111). In addition, although it may be possible to test the appeal of such additives to animals, it will be difficult to distinguish effects due to taste from effects due to other properties of the additive that may activate the reward system. However, one could also argue that mere self-administration is most important, not the mechanisms behind it. Although it is known that rodents and humans share the same olfaction receptors (112), one would need to ascertain that animals have the same taste preference as humans (36).

### **4.3 Chemical analysis of flavour components**

Although tobacco manufacturers sent ingredient lists to governments, individual components and additives of tobacco can also be validated by chemical analysis. To validate an objective method for the purpose of detecting products with characterising flavours, it has to be correlated with subjective data. Sensory testing is required to determine the limit of detection of humans. In addition, although objective chemical analysis identifies the individual compounds of tobacco, perhaps components at low concentrations combine to form a flavour that could only be detected and identified subjectively. Therefore, chemical analysis often is validated with sensory testing to determine flavours (30)\*.

Different techniques that can be used for sampling and subsequent analysis of chemical components will be described in the next paragraphs.

#### **4.3.1 Flavour component analysis**

Several analytical methods, such as gas chromatography (GC), infrared spectroscopy (IR), nuclear magnetic resonance (NMR), and mass spectrometry (MS), can be used to analyse tobacco ingredients. Since tobacco additives such as flavours are often present at trace levels, analytical methods need to be sensitive, such as gas chromatography-mass spectrometer (GC-MS) methods. For instance, a method has been reported where 23 flavour additives have been determined using GC-MS (113). The detection limit of flavours was in the range of 0.46 mg/kg tobacco. Gas chromatography coupled with Ion-Mobility-Spectrometry (GC-IMS), combines the high selectivity of the chromatographic separation with a sensitive IMS detector.

Besides chemical analysis with GC-MS, flavour components can also be analysed using electronic noses (e-noses) (114). E-noses are based on classical sensor-arrays with different sensor types, or based on instrumental analysis with mass spectrometers or gas chromatography (115). Instrumental analysis combines common analytical GC-MS measurements with a chemical identification system. An electronic nose works similar to the human olfactory system to generate an odour specific pattern. Patterns can be recognised with principal component analysis. These measurements can be translated into 'odour maps', representing a chemical or olfactive fingerprint. Research has been performed where an e-nose method is combined with artificial neural networks, which causes results to be improved and four different types of cigarettes to be recognised (116). However, this type of sensory analysis provides less information than assessors do, and cannot be used to chemically identify or quantify single flavour components.

Philip Morris International has also performed combined sensory and analytical research to assess taste and aroma based on electronic nose research. The method they used to determine the odour impression of the separated component is called gas chromatography-mass olfactometry (GC-O) (117)\*. This technique consists of a combination of GC analysis with human input, and is also called GC-sniffing. The flavour sample passes through a column for GC analysis and then is split into two detectors: a human detector where one could smell the aroma that is released from the GC column at that moment and a GC-MS detector which provides the chromatographic peak belonging to that aroma compound. In this way, the peaks in the chromatogram form the fingerprint of 'olfactory relevant compounds' could be correlated with the sensory assessment.

The technique was also further developed to a multidimensional system by the tobacco company Souza Cruz (118)\*. During application of the technique the sample is firstly separated on a capillary column as in conventional GC, then on the basis of the aromas smelt at the sniffing port/olfactometer, a retention time interval is selected where a flavour of interest elutes together with other compounds. The same chromatographic process is repeated, but the compounds found at the retention time interval of interest are collected and re-injected into another chromatographic column of different polarity where its separation is optimised. This enables separation of mixture aroma components in a more efficient way.

This technique is considered not to be useful in our research, because with this human detector one would smell all separate flavour compounds one by one. However, we are interested in the total flavour instead of all separate components, to assess whether they result in a product with a characterising flavour.

#### **4.3.2 Sampling flavour components for GC-MS analysis**

Chemical analysis of volatiles such as flavours can be performed with several methods: headspace (HS), solid-phase micro-extraction (SPME), solvent assisted flavour evaporation (SAFE), direct solvent extraction (DSE) and steam distillation (SDE) analysis. All techniques are based on different principles for sample collection, and have particular strengths and weaknesses. Headspace is based on (heat assisted) diffusion of volatiles to the gas phase, SPME is based on extraction of volatile organic components on a solid matrix, and extraction using organic solvents is the basis of SAFE and DSE.

In a recent article, different candy flavourings in several tobacco products were described. In this research flavour chemicals and their levels were identified and determined using DSE followed by GC-MS analysis (119). The evaluation by the EU of milk products such as butter involves a combination of sensory analysis and the determination of product content by analysing specific compounds using GC or HPLC (41).

SAFE is a technique used to isolate aromas in complex matrices under vacuum condition. The distillation products are collected in a flask cooled with liquid nitrogen and then analysed with GC-MS. Due to the low temperature and pressure condition, SAFE is considered a “gold standard”, since potential alteration of the sample is avoided. In contrast to headspace where only volatile components are extracted, SAFE allows extraction of both volatile and non-volatile sample components. In this way, SAFE provides a complete profile of the sample components. It must be taken into account that probably only a small number of these contribute significantly to the flavour of the sample. Therefore, a profile provided by the SAFE technique will be complex to analyse due to the huge number of compounds. Another fast and sensitive technique to extract volatile compounds such as flavours is SPME. SAFE and SPME were used to characterise aroma compounds in apple cider (120). Comparing these techniques, the SPME technique showed a higher recovery for highly volatile compounds than SAFE. In flavour research, SAFE, SPME, and headspace extraction are generally recognised as effective approaches.

An older paper on methods and techniques in tobacco flavour research describes chemical-analytical measurements using headspace extraction, combined with sensory evaluation (121) Headspace analysis is also useful to analyse flavours. In this technique, the flavour tobacco sample is heated causing flavour components to become volatile. The formed vapour will be analysed using GC-MS. With GC-MS, all components of tobacco can be identified and quantified. Therefore, intensity levels of the components can be determined. According to the author of this paper, sensory evaluation is still superior to instrumental methods, however, there is a relation between specific compounds measured in tobacco and tasting properties such as taste, strength, aroma, and tobacco flavour.

Since the headspace technique only captures volatile components, such as flavours, no non-volatile components are able to cause noise in the chromatogram. This makes the headspace technique simpler and clearer than SAFE, where also peaks from non-volatile components appear in the chromatogram. In addition, headspace analysis is fast, easy to use, and clean since no chemicals are involved.

SPME is comparable to headspace. However, because SPME is more sensitive than headspace, SPME will show more components in the resulting chromatogram, which makes it more complex to analyse. Therefore, headspace is probably also simpler and clearer than SPME. When the headspace technique appears to be insufficiently sensitive to tobacco flavours, it might be advisable to use a more sensitive measurement, i.e. SPME.

Hasebe and Suhara have performed headspace GC-MS vapour analysis of three different tobacco types. The selected peaks of the resulting chromatogram as well as the results from expert panel studies were identified with principal component analysis. They correlated both data sets to find the main determinants of flavour and smoking quality and show this method can be used to analyse vapours of different tobacco types with good reproducibility (117)\*.

The headspace technique is commonly used in tobacco research for a long time. In addition, this technique is most commonly used in the food industry to measure flavours in food. One example out of many articles in which headspace analysis is used in the food industry, is the research of Isabel Montesinos and Mercedes Gallego. They performed research on halogenated volatile organic compounds (VOCs) in beverages using headspace gas chromatography. They determined four VOCs and ten volatile THMs (iodinated trihalomethanes) in 100 types of samples (122).

Another example is the research in which headspace SPME is used to identify volatile compounds of eleven typical Italian monocultivar extra virgin olive oils. As a result, forty-eight compounds of olive oil were determined by headspace GC-MS. The results from sensory panel tests and the presence of specific volatile compounds measured with GC-MS were evaluated. In many cases a quantitative

correspondence between the score of its aroma according to sensory analysis and the amount of a volatile compound was found (123).

Several compounds and the flavours they liberate are not detected by usual methods of volatile flavour analysis whose initial stage is the production of a DSE or Headspace extract. The compounds are called reaction flavours and are often complex mixtures that liberate flavours during cigarette burning at temperatures in the order of 300-400°C. Tobacco companies are known to use these compounds for flavouring cigarettes (118). Analysis on these compounds can be performed by pyrolysis.

Concluding, SPME and SAFE are less specific and reproducible, and more labour-intensive than headspace and DSE analysis. Headspace is more specific, sustainable since no chemicals are involved, and less labour-intensive than DSE. However, quantification with headspace is slightly more difficult than with DSE.

#### **4.4 Responses on tobacco flavour analysis questionnaires from EU- and third countries**

Responses to questionnaires (Appendix V of this annex) were received from 22 EU member states (MS), from three third countries, and from four scientists. We included any relevant reference to literature in our review above, as well as any suggestions towards methods that we did not yet include.

Overall, the replies received confirm that standardized methods for the sensory evaluation of flavours or characterising flavours in tobacco products as well as guidance on how to address the evaluation of whether a product has a characterising flavour are still lacking.

Regarding methods for tobacco flavour assessment in general, respondents mention that tobacco industry use in-house and research methods to assess tobacco flavour, some of them reported in published literature. There are in-house established sensory analysis methods that have been used for the assessment of flavours in tobacco products for years. These include the use of expert sensory panels, e-nose technology, discrimination analysis (such as sensory triangle testing), sensory quality grading/index and intelligent systems. There are a number of research methodologies, which have been employed for the sensory evaluation of flavours in tobacco products in peer-reviewed publications. These include individual methods, such as the use of sensory panels, e-nose systems, multivariate/chemometric analysis, different types of sensor methodologies, consumer panels, and intelligent systems, as well as a combination of methods in which two or more of such methods are combined. Regarding specific sensory methods, method BSS 8389-85 from the International Tobacco Tasting Association (ITTA) in Bulgaria was mentioned by a MS, which is a method to quantify aroma, taste and strength attributes of inhaled tobacco smoke on a 1-10 scale. In Vietnam, tobacco industry uses some standards about cigarette/filter cigarette sensory testing such as TCN 26-01:2003, and TCVN 4286-86. These standards are more similar to a scorecard using in beer or wine tasting rather than the descriptive method recognized by ISO.

Regarding methods for other products such as food that could be useful in the development of methods to assess characterising flavours, member states refer to e.g. ISO methods for conducting sensory analysis of food products. General conditions would also apply to the assessment of tobacco products. Specifically, methods by ISO Technical Committee on Food Products - ISO/TC 34 could be employed for the evaluation of flavours in tobacco products and act as a guide to setting up a sensory panel. The Afnor developed a commission "Analyse sensorielle : AFNOR/V09A" which is developing norms on methodology of sensorial analysis (124).



Regarding non-sensory methods, e.g. chemical analysis, it is mentioned that they can only be used in combination with sensory methods, as they do not provide a description of the human experience of flavour. There are established GC, IR and MS methods for the chemical evaluation of flavours, e.g. headspace GC-MS. In addition, electric nose and/or electric tongue are used as a sensory method for assessment of flavour in food.

Advantages and disadvantages of potential methods to be used for the determination whether a tobacco product has a characterising flavours

In the table below, we summarise the methods described in this literature review. Advantages and disadvantages of the different methods are summarized and the question if the method is suitable for our research question is answered. Paragraph numbers referring to the sections in the review are indicated.

Table 20. Advantages and disadvantages of the methods summarised in the literature review

Paragraph Number	Method	Advantage	Disadvantage	Suitable for our research question? Yes/No
2	Methods relevant for tobacco sensory analysis – smelling versus smoking			
2.1	Smelling: package, burned and unburned cigarettes and Cambridge pads	<ul style="list-style-type: none"> <li>- can be done with adolescents, which is the relevant population when considering measures to prevent smoking initiation</li> <li>- from our pilot experiment characterising flavours are better detectable by smelling compared to when the tobacco product is smoked</li> <li>- no health concerns</li> <li>- can be done in standard sensory booths.</li> <li>- this is how adolescents or people in general first come in contact with a cigarette, where they decide to smoke/not smoke.</li> </ul>	<ul style="list-style-type: none"> <li>- It might be possible that certain flavours are only detectable when the cigarette is smoked by the participant</li> <li>- might be a weakness when legislation is based on smelling while the use of the product is smoking.</li> </ul>	Yes
	Smoking: cigarettes and roll-your-own	<ul style="list-style-type: none"> <li>- the natural way of using the product</li> </ul>	<ul style="list-style-type: none"> <li>- health concerns</li> <li>- can only be done with adults,</li> </ul>	Yes
2.2	Consumer versus expert testing			
2.2.1	Consumer testing	<ul style="list-style-type: none"> <li>- do not need to be trained</li> <li>- Fast results</li> <li>- representative for the population</li> </ul>	<ul style="list-style-type: none"> <li>- less suitable for “daily/weekly” use compared to an expert panel</li> <li>- difficulties rating intensity of attributes</li> </ul>	Yes
2.2.2	Expert testing	<ul style="list-style-type: none"> <li>- measurement can be easily repeated</li> <li>- less expensive than a consumer panel</li> </ul>	<ul style="list-style-type: none"> <li>- time intensive</li> </ul>	Yes
3.1	Discriminative methods			
	Triangle method, A-not-A method, Tetrad method, 2 Alternative forced choice (AFC), duo-trio method, Difference from control	<ul style="list-style-type: none"> <li>- Easy, gives yes/no answer</li> </ul>	<ul style="list-style-type: none"> <li>- do not provide information about the magnitude or direction of the difference.</li> <li>- requires a single suitable reference product or gold standard</li> <li>- ‘standard’ (unflavoured) tobacco products already differ too much</li> <li>- do not provide information on whether the difference is based on a <i>characterising</i> flavour</li> </ul>	No

EUROPEAN COMMISSION

3.2	Descriptive methods using consumer or expert testing			
3.2.1	Quantitative descriptive analysis (QDA)	<ul style="list-style-type: none"> <li>- detailed information about product characteristics</li> <li>- consumer vocabulary</li> <li>- gives information about flavour intensity</li> <li>-panel performance can be monitored</li> </ul>	<ul style="list-style-type: none"> <li>- time intensive and costly</li> <li>- does not automatically lead to a yes/no answer</li> </ul>	Yes
3.2.2	Free choice profiling	<ul style="list-style-type: none"> <li>- Participants use their own vocabulary to describe products</li> <li>- unbiased as participants are not redirected in finding specific flavours</li> </ul>	<ul style="list-style-type: none"> <li>- Consumers might use different words that mean the same</li> <li>-Consumers might use the same words that have different meanings for them</li> <li>- Panel leader has to group words, is not objective</li> </ul>	Yes
3.2.3	Flash and ultra-flash profiling	<ul style="list-style-type: none"> <li>- Participants use their own words to describe products.</li> <li>- Comparison of products by placing them on a map.</li> </ul>	<ul style="list-style-type: none"> <li>- don't provide information about based on which criteria participants placed the products</li> <li>- might be less sensitive compared to other methods</li> </ul>	No
3.2.4	Ideal profile method	<ul style="list-style-type: none"> <li>- Participants also give information on whether or not this is their ideal intensity of a flavour.</li> </ul>	<ul style="list-style-type: none"> <li>- method does not give answer to our research aim</li> <li>-relates to preference testing</li> </ul>	No
3.2.5	Check all that apply (CATA)	<ul style="list-style-type: none"> <li>- can be done by consumers and panellist</li> <li>-easy to standardize as you have a predefined attribute list.</li> <li>-easy to perform by consumers</li> </ul>	<ul style="list-style-type: none"> <li>- no information about flavour intensity</li> <li>-new flavour=new attribute added to the list.</li> <li>-Long list of attributes, need to group flavours</li> </ul>	Yes
3.3	Descriptive methods using expert testing			
3.3.1	Flavour profiling	<ul style="list-style-type: none"> <li>- Score flavour attributes</li> </ul>	<ul style="list-style-type: none"> <li>- Intensity of flavours is based on group consensus.</li> <li>-more subjective to bias compared with other methods</li> </ul>	No
3.3.2	Spectrum	<ul style="list-style-type: none"> <li>- detailed information about product characteristics</li> <li>- Panellists are trained to rate intensity in the</li> </ul>	<ul style="list-style-type: none"> <li>- technological terminology</li> <li>-very time intensive</li> <li>-does not automatically lead to a yes/no</li> </ul>	No

EUROPEAN COMMISSION

		same way.	answer	
3.3.3	Hybrid	- combining quantitative measures with a quality scale. - work with clear cut-offs	- We do not have clear cut of ranges (yet). -extensive training	Yes
2.2.3	Considerations regarding the population to involve for tobacco sensory analysis in EU countries			
	National expert panel, using a common language	- more easy to generate attributes - training might be faster as people will be more familiar with terms in their own language - Results of an expert panel can be applied across different cultures, as they are highly trained to detect specific attributes, regardless of their previous familiarity with it.	- terms have to be translated to English, words might not have the <i>exact</i> same meaning	Yes
	Mixed culture expert panel, using the English language	- terms do not have to be translated	-might be more difficult to train, although panellist need to learn the precise meaning of each word also in their national language	Yes
	National consumer panel	-	- only represents one population, we do not know whether results can be translated to other countries	Yes
	Mixed culture consumer panel	- can test products with a consumer panel from the country the cigarette will be sold on the market, or use a mixed consumer panel	-	Yes
4	Alternative methods			
4.1	Neuro-imaging analysis	- Interesting from a scientific viewpoint	- Too costly, not discriminative	No
4.2	Laboratory animals	-	- Cannot be used to give answer to our research question.	No
4.3	Chemical analysis			
	Infrared spectroscopy (IR)	- Sustainable	- Not specific enough - Not sensitive enough - Less accurate - Labour-intensive	No
	Nuclear magnetic resonance (NMR)	-	- Not specific enough - Less accurate - Less sensitive - Labour-intensive	No

EUROPEAN COMMISSION

		- Less sustainable	
E-nose	- Not labour-intensive - Sustainable	- Not specific - Less accurate - Less sensitive	No
Olfactometry – GC-MS	- Very sensitive - Sustainable - Chemical and sensory analysis combined	- Less accurate - Labour intensive - Provides no sensory information concerning the flavour of the product as a whole - Less specific - Flavours consisting of more than one chemical component cannot be recognized	No
Solvent assisted flavour evaporation (SAFE) - GC-MS	- Sensitive - Less labour-intensive - No potential alteration of the sample	- Less specific (both volatile and non-volatile sample components) - Less accurate - Less sustainable (use of liquid nitrogen)	No
solid-phase micro-extraction (SPME) - GC-MS	- Very sensitive - Sustainable (no chemicals involved) - Less labour-intensive - Only volatile sample components such as flavours	- Less specific - Less accurate (more difficult to quantify)	Yes
Direct solvent extraction (DSE) - GC-MS	- Very accurate (easy to quantify) - Very sensitive - Specific - Less labour-intensive - Measurement of all tobacco components	- Less sustainable (use of chemicals) - Measurement of all tobacco components: complex results	No
Headspace - GC-MS	- Very specific - Very sensitive - Accurate - Not labour-intensive - Sustainable - Easy to use - Only volatile sample components such as flavours	- Less accurate than DSE GC-MS	Yes

## 5 Discussion and conclusions

The aim of this literature review is to provide scientific information on test methods that can be used to determine characterising flavours of tobacco products. A flowchart of the methods described in this review along with paragraph numbers can be found in Appendix II of this annex. We discuss all methods currently used for the analysis of characterising flavours in food or consumer products that can be applied for tobacco as well. In addition, methods used for the analysis of the composition of tobacco products are described. The information as presented in this review is comprehensive and complete since multiple data sources were consulted. Information was not only gathered from scientific literature but also from digital libraries providing information on methods used by the tobacco industry and containing information on European Union regulations. In addition, a questionnaire was sent out to regulators and scientist in several EU Member States and third countries to provide insight on methods used for assessment of tobacco products in these countries. The results from the questionnaire overlap with the topics we discuss in this literature review. Current methods used under EU regulation to determine characteristic flavours in consumer products are defined for virgin olive oil and butter. These methods were established for the classification of the oil based on organoleptic characteristics, and to determine butter quality based on defects (41) (125).

Here we discuss the characteristics of the methods described in this review and their applicability to determine characterising flavours in tobacco products in a standardized manner. Not only sensitivity, reliability and feasibility for standardization of the method are requirements for the method, but also ethical concerns, time investment, and costs are considered.

It can be concluded that currently the analysis of sensory response in the human brain by neuroimaging methods and sensory testing in laboratory animals using behavioural studies are not suitable to answer the research question. Chemical analysis of flavour components and methods for tobacco sensory analysis are and will be mentioned in more detail in the remaining part of this discussion.

The definition of a clearly noticeable flavour other than tobacco is not clearly defined. Natural tobacco is present in a range of different classes (burley, oriental etc.) that differ in amounts of tobacco components including those responsible for odour and taste. Most cigarettes are comprised of a mixture of tobacco types resulting in a quite complex and variable 'tobacco specific' background flavour. Flavourings are also often added in order to balance the natural tobacco taste.

Perception of flavour is not only determined by smell and taste but is also influenced by extrinsic factors such as branding, colour and packaging. When cigarettes of their preferred brand are given to consumers, they are likely to associate certain expectations with it. In order to represent a realistic smoking situation it is recommended to use branded cigarettes with or without packaging.

Characterising flavour of a tobacco product requires sensory information from consumers and/or trained judges. To determine relative differences between samples and controls it is important to have the same tobacco type as reference, preferably by comparing several brands with so called characterising tastes to brands with no additives or no predicted characterising taste.

Quantitative or qualitative test methods used in sensory analysis involve detection by difference testing or description testing and scaling techniques to encompass a range of intensities or categorize product attributes. Depending on the test method, large or subtle product differences can be assessed and classified. One can determine the exact identity of the type of flavour, the overall intensity of a flavour or simply the presence or absence of a flavour.

The overall intensity of an aroma or flavour is determined by overall feelings and sensations that the product will have on the consumer. This can be dependent on the testing situation as well. Using specific measuring methods, more subtle differences can be revealed by describing the product details in more depth. Flavour groups can be identified by assessing product attributes such as olfactory sensations (vanilla, fruity, floral, skunky), taste sensations (salty, sweet, sour, bitter, umami) or nasal feelings (cool, pungent) or oral feeling factors (heat, cool, burn, metallic, astringent). Quantitative methods express the degree to which a characteristic is present and is expressed by assigning a value on a scale (intensity score). The validity and reliability of the analysis is dependent on the selection of scaling technique, which should encompass full range of intensities but be sufficiently sensitive to pick up small differences.

Characterising flavours will be assessed differently by experts than by consumers. Expert panels are formed and trained to fully understand the terms used to describe product characteristics and to use them consistently thereby they are efficient in providing reproducible results with a small variability. Test with expert panels are mostly carried out in a testing facility, but can also be carried out in special testing environments in order to combine the sensory information with images of activity in certain brain regions (MRI). For smoking test, a special facility with a proper ventilation system needs to be chosen. Consumer panels can be asked to assess the product in a testing facility or assess the product at home.

Since several studies show that young people are more attracted and vulnerable to the use of flavoured cigarettes, this subgroup could be selected as part of expert or consumer panels. Tests with consumer- and expert panels can be repeated to check validity and precision of the method for product evaluation. In case of limited time and a limited number of measurements, descriptive consumer panels are more efficient compared to expert panels; the higher number of panellists then outweighs the loss of accuracy. Furthermore, descriptive tests with expert panels require the selection and training of a panel. This will take at least several months and is only efficient if the panel is assessing a large number of samples over a long period of time. The number of training sessions required is dependent on panel performance, the level of accuracy one wants to achieve and difficulty of rating the test product. Expert panels are trained to generate reproducible results. They are highly trained to detect specific attributes in an independent manner. Therefore, cultural differences within the panel should not influence test results.

Testing procedures to define tobacco products with a characterising taste will be harmonized across the European Union. Sensory analysis to determine a characterising flavour of a tobacco product before and upon its intended use involves a combination of smoking and smelling the product. Smelling the burnt product in the form of side stream smoke is not preferable regarding the health consequences involved. For ethical reasons this can only be done by smokers and not by using non-smokers. Non-smokers can however smell the tobacco of cigarettes without health hazard. The advantage of using non-smokers is that they taste and smell better than smokers do, but the disadvantage is that they cannot be asked to smoke the product and therefore only assess the unburnt product. When the unburnt smell is very appealing it is very likely that the burnt product also displays a characteristic flavour. Beside the unburnt product, the burnt product can also be smelled in a way that does not pose a health hazard to the assessor. This can be done by smelling Cambridge filter pads containing the substances of tobacco smoke, which are formed during tobacco burning.

The results retrieved using subjective sensory profiling (smoking and smelling tobacco products by either consumers or experts) can be related to the chemical analysis of the flavour components. In the laboratory, flavour component(s) can be captured and analysed for their chemical composition. A flavour can be caused by one chemical component, but also by several components. In addition, flavour enhancers and attenuators can modify the intensity and characteristic of a flavour. Using this

type of analysis, the flavour component(s) can be identified and quantified in a standardized way. Quantification of flavour components in specific products allows for determining the concentration for which a flavour is defined by sensory analysis as 'characterising'. Such a limit is very useful for regulatory purposes, and in this set-up, ideally, chemical analytical measurements can replace sensory analysis.

Sensory analysis methods combined with chemical-analytical measurements seem applicable and well suited for tobacco product regulation regarding characterising flavours.

## 6 References

1. Lewis, M.J. and O. Wackowski, *Dealing with an innovative industry: a look at flavoured cigarettes promoted by mainstream brands*. Am J Public Health, 2006. **96**(2): p. 244-51.
2. Levin, E.D., F. Behm, and J.E. Rose, *The use of flavour in cigarette substitutes*. Drug Alcohol Depend, 1990. **26**(2): p. 155-60.
3. Leatherdale, S.T., et al., Cigarette brand preference as a function of price among smoking youths in Canada: are they smoking premium, discount or native brands? Tob Control, 2009. **18**(6): p. 466-73.
4. Jaffe, A.J. and A.G. Glaros, Taste dimensions in cigarette discrimination: a multidimensional scaling approach. Addict Behav, 1986. **11**(4): p. 407-13.
5. Kaleta, D., et al., Use of flavoured cigarettes in Poland: data from the global adult tobacco survey (2009-2010). BMC Public Health, 2014. **14**: p. 127.
6. Rosenbloom, J., et al., A cross-sectional study on tobacco use and dependence among women: Does menthol matter? Tob Induc Dis, 2012. **10**(1): p. 19.
7. Rising, J. and K. Wasson-Blader, *Menthol and initiation of cigarette smoking*. Tob Induc Dis, 2011. **9 Suppl 1**: p. S4.
8. Minaker, L.M., et al., Flavoured tobacco use among Canadian students in grades 9 through 12: prevalence and patterns from the 2010-2011 youth smoking survey. Prev Chronic Dis, 2014. **11**: p. E102.
9. Farley, S.M., et al., *Teen use of flavoured tobacco products in new york city*. Nicotine Tob Res, 2014. **16**(11): p. 1518-21.
10. Besaratinia, A. and S. Tommasi, *The lingering question of menthol in cigarettes*. Cancer Causes Control, 2014.
11. Gains, L.H.M., L.S., *Sensory Evaluation of Sidestream Odor Using Transfer Testing Methodology*. Beitrage zur tabakforschung international 1987. **14**(1): p. 53-59.
12. Vaughan, C., et al., Automated determination of seven phenolic compounds in mainstream tobacco smoke. Nicotine Tob Res, 2008. **10**(7): p. 1261-8.
13. Abdallah, F.M., Cigarette product development. Blending and processing know-how, sensory testing of cigarette smoke. 2004, Raleigh: Tobacco reporter. 135.
14. Leffingwell, J.C., Symposium of the 42nd Tobacco Chemists' Research Conference: Chemical and Sensory aspects of Tobacco Flavour. Recent advances in tobacco science, 1988. **14**.
15. Jewell, J., Central Operations - Best Practice: Procedure for Checking for Discernible Taint in Finished Products. Legacy Tobacco Documents Library, 2000. **British American Tobacco**.
16. Tobacco, B.A., *Sensory analysis of cigarettes*. Legacy Tobacco Documents Library, 1989.
17. Rees, V.W., et al., Assessing consumer responses to potential reduced-exposure tobacco products: a review of tobacco industry and independent research methods. Cancer Epidemiol Biomarkers Prev, 2009. **18**(12): p. 3225-40.
18. McAdam, K.G., The use of a novel tobacco-substitute sheet and smoke dilution to reduce toxicant yields in cigarette smoke. Food Chem Toxicol, 2011. **49**(8): p. 1684-96.
19. Talhout R. van de Nobelen S., K.A.S., Scientific assessment of tobacco products attractiveness: characterising flavours. Submitted, 2014.
20. Morris, P., Blind product testing guidelines. 1998.



21. Worch, T., S. Le, and P. Punter, *How reliable are the consumers? Comparison of sensory profiles from consumers and experts*. Food Quality and Preference, 2010. **21**(3): p. 309-318.
22. Lawless, H. and H. Hildegarde, *Sensory Evaluation of Food principles and practices*. second edition ed. 2010, New York, Dordrecht, Heidelberg, London: Springer.
23. Abdallah, F.M., Sensory Testing of Cigarette Smoke. Panel Selection, Training, and Use, in Department of Crop Science. 1975, North Carolina State University: Raleigh. p. 203.
24. (EC), C.R., Methods for the analysis and quality evaluation of milk and milk products. Official journal of the european union, 2008. **273**.
25. Angerosa, F. and C. Campestre, Chapter 14 Sensory Quality: Methodologies and Applications, in Handbook of Olive Oil: Analysis and Properties. 2013, Springer Science Business Media New York p. 524-556.
26. Stone, H. and J.L. Sidel, *Sensory evaluation practices*. . 1993, California: Academic Press.
27. Lelievre, M., et al., Beer-Trained and Untrained Assessors Rely More on Vision than on Taste When They Categorize Beers. Chemosensory Perception, 2009. **2**(3): p. 143-153.
28. Lawless, H.T., Flavour Description of White Wine by Expert and Nonexpert Wine Consumers. Journal of Food Science, 1984. **49**(1): p. 120-123.
29. Moskowitz, H.R., et al., Descriptive Panels/Experts versus Consumers, in Viewpoints and Controversies in Sensory Science and Consumer Product Testing. 2008, Food & Nutrition Press, Inc. p. 109-123.
30. Wysocki C.J., P.J.D., Gilbert A.N., Geographic, cross-cultural, and individual variation in human olfaction" in Smell and Taste in Health and Disease. T.V. Getchell, ed., New York: Raven Press, 1991: p. 287-314.
31. Thamke I, D.K., Rohm H, *Sensory description of dark chocolates by consumers*. LTW Food and Science technology, 2009. **42**: p. 534-539.
32. Le S, P.J., Husson F, Methodology for the comparison of sensory profiles provided by several panels: Application to a cross-cultural study. Food Quality and Preference, 2008. **19**(2): p. 179-184.
33. Lawless, H.T. and H. Heymann, *Sensory evaluation of food principals and practices*. 1998, New York: Chapman and Hall.
34. Harr, M.E., *Probabilistic Estimates for Multivariate Analyses*. Applied Mathematical Modelling, 1989. **13**(5): p. 313-318.
35. Meilgaard, M.C., Civille, G.V., Carr, B.T, Sensory Evaluation Techniques (4th ed.) 2007.
36. Jenkins, C.R., *British American Tobacco Company Limited Product Seminar*. Tobacco documents 760076123-760076408, 1997.
37. Stone, H., et al., Chapter 5 - Discrimination Testing, in Sensory Evaluation Practices (Fourth Edition). 2012, Academic Press: San Diego. p. 167-231.
38. Herbert Stone, R.N. Bleibaum, and H.A. Thomas, Chapter 6 - Descriptive Analysis, in Sensory Evaluation Practices.
39. Reynolds, A.G., *Viticulture and wine quality, in managing wine quality*. 2010, woodhead publishing limited: Cambridge.
40. Stone, H., et al., *Sensory Evaluation by Quantitative Descriptive Analysis*. Food Technology, 1974. **28**(11): p. 24-&.
41. Meilgaard, M., C. G.V, and C. T.B, *Sensory evaluation techniques*. 4th ed. 2007, New York: CRC Press.
42. Chapman, K.W., H.T. Lawless, and K.J. Boor, Quantitative descriptive analysis and principal component analysis for sensory characterization of ultrapasteurized milk. Journal of Dairy Science, 2001. **84**(1): p. 12-20.
43. Mcewan, J.A., J.D. Moore, and J.S. Colwill, *The Sensory Characteristics of Cheddar Cheese and Their Relationship with Acceptability*. Journal of the Society of Dairy Technology, 1989. **42**(4): p. 112-117.
44. Campo, E., et al., Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: An application to Burgundy Pinot noir wines. Food Quality and Preference, 2010. **21**(1): p. 44-55.

45. Vilanova, M., A. Masa, and J. Tardaguila, Evaluation of the aroma descriptors variability in Spanish grape cultivars by a quantitative descriptive analysis. *Euphytica*, 2009. **165**(2): p. 383-389.
46. Ng, M., et al., Using quantitative descriptive analysis and temporal dominance of sensations analysis as complementary methods for profiling commercial blackcurrant squashes. *Food Quality and Preference*, 2012. **25**(2): p. 121-134.
47. Leighton, C.S., H.C. Schonfeldt, and R. Kruger, Quantitative Descriptive Sensory Analysis of Five Different Cultivars of Sweet Potato to Determine Sensory and Textural Profiles. *Journal of Sensory Studies*, 2010. **25**(1): p. 2-18.
48. Krumbein, A., P. Peters, and B. Bruckner, Flavour compounds and a quantitative descriptive analysis of tomatoes (*Lycopersicon esculentum* Mill.) of different cultivars in short-term storage. *Postharvest Biology and Technology*, 2004. **32**(1): p. 15-28.
49. Commission, E.U., Regulation EC 640/2008, in No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. 2008, Official Journal of the European Communities. p. 178, 11-16.
50. Dijksterhuis, G.B., *Multivariate data analysis in sensory and consumer science*. food and nutrition press inc., 1997.
51. Santosa, A., *Research Report on the Results of Project Komodo-2*. Legacy Tobacco Documents Library, 2000. **British American Tobacco**.
52. Sieffermann, J.M., Le profil Flash: Un outil rapide et innovant d'évaluation sensorielle descriptive. *L'innovation: de l'idée au succès* Lavoisier, Tec & Doc, 2000: p. 335-340.
53. Delarue, J. and J.-M. Sieffermann, Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 2004. **15**(4): p. 383-392.
54. Blancher, G., et al., French and Vietnamese: How do they describe texture characteristics of the same food? A case study with jellies. *Food Quality and Preference*, 2007. **18**(3): p. 560-575.
55. Tarea, S., G. Cuvelier, and J.M. Sieffermann, *SENSORY EVALUATION OF THE TEXTURE OF 49 COMMERCIAL APPLE AND PEAR PUREES*. *Journal of Food Quality*, 2007. **30**(6): p. 1121-1131.
56. Delarue, J. and E. Loescher, *Dynamics of food preferences: a case study with chewing gums*. *Food Quality and Preference*, 2004. **15**(7-8): p. 771-779.
57. Lassoued, N., et al., Baked product texture: Correlations between instrumental and sensory characterization using Flash Profile. *Journal of Cereal Science*, 2008. **48**(1): p. 133-143.
58. Veinand, B., et al., Highlight of important product characteristics for consumers. Comparison of three sensory descriptive methods performed by consumers. *Food Quality and Preference*, 2011. **22**(5): p. 474-485.
59. Louw, L., et al., Validation of two Napping® techniques as rapid sensory screening tools for high alcohol products. *Food Quality and Preference*, 2013. **30**(2): p. 192-201.
60. Perrin, L., et al., Comparison of three sensory Napping (R) procedure: Case of methods for use with the ten wines from Loire valley. *Food Quality and Preference*, 2008. **19**(1): p. 1-11.
61. Valentin, D., et al., *Quick and dirty but still pretty good: a review of new descriptive methods in food science*. *International Journal of Food Science & Technology*, 2012. **47**(8): p. 1563-1578.
62. Lawless, H.T., *Descriptive analysis of complex odors: reality, model or illusion?* *Food Quality and Preference*, 1999. **10**(4-5): p. 325-332.
63. Lelievre, M., et al., What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 2008. **19**(8): p. 697-703.
64. Saint-Eve, A., E.P. Kora, and N. Martin, Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality and Preference*, 2004. **15**(7-8): p. 655-668.
65. Albert, A., et al., Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality and Preference*, 2011. **22**(5): p. 463-473.

66. Jaeger, S.R., et al., Check-all-that-apply (CATA) responses elicited by consumers: Within-assessor reproducibility and stability of sensory product characterizations. *Food Quality and Preference*, 2013. **30**(1): p. 56-67.
67. Reinbach, H.C., et al., Comparison of three sensory profiling methods based on consumer perception: CATA, CATA with intensity and Napping (R). *Food Quality and Preference*, 2014. **32**: p. 160-166.
68. Meullenet JF, X. R, and F. C., *Multivariate and probabilistic analyses of sensory science problems*. Blackwell Publishing, Ames, IA, 2007: p. 256.
69. Ares, G., et al., Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers. *Food Quality and Preference*, 2014. **36**: p. 111-121.
70. Sjöström, L.B., et al., *The descriptive analysis of flavour; Food Acceptance Testing Methodology*. 1954, Quartermaster Food and Container Institute: Chicago. p. 25-61.
71. Gasperi, F., et al., The mozzarella cheese flavour profile: a comparison between judge panel analysis and proton transfer reaction mass spectrometry. *Journal of the Science of Food and Agriculture*, 2001. **81**(3): p. 357-363.
72. Civille, G.V. and B. Lyon, *ASTM Lexicon Vocabulary for Descriptive Analysis*. American Society for Testing and Materials, Philadelphia., 1996.
73. Meilgaard, M.C., G.V. Civille, and B.T. Carr, *Sensory evaluation techniques (2nd ed.)*. 1991, Boca Raton, FL: CRC Press.
74. Murray, J.M., C.M. Delahunty, and I.A. Baxter, *Descriptive sensory analysis: past, present and future*. Food Research International, 2001. **34**(6): p. 461-471.
75. Johnsen, P.B. and C.V. Civille, *A standardised lexicon of meat WOF descriptors*. *Journal of Sensory Studies*, 1986. **1**: p. 99-104.
76. Johnsen, P.B., C.V. Civille, and J.R. Vercellotti, *A lexicon of pond raised catfish flavour descriptors*. *Journal of Sensory Studies*, 1987. **2**: p. 85-91.
77. Johnsen, P.B., et al., *Development of a lexicon for the description of peanut flavour*. *Journal of Sensory Studies*, 1998. **3**: p. 9-17.
78. Civille, G.V. and K.N. Oftedal, *Sensory evaluation techniques - Make "good for you" taste "good"*. *Physiology & Behavior*, 2012. **107**(4): p. 598-605.
79. Krinsky, B.F., et al., *The development of a lexicon for frozen vegetable soybeans (edamame)*. *Journal of Sensory Studies*, 2006. **21**(6): p. 644-653.
80. Beckley, J.P. and D.R. Kroll, *Searching for sensory research excellence*. *Food Technology*, 1996. **50**(2): p. 61-63.
81. Nakayama, M. and C. Wessman, *Application of Sensory Evaluation to the Routine Maintenance of Product Quality*. *Food Technology*, 1979. **33**(9): p. 38-&.
82. Lundstrom, J.N., S. Boesveldt, and J. Albrecht, *Central Processing of the Chemical Senses: an Overview*. *ACS Chem Neurosci*, 2011. **2**(1): p. 5-16.
83. Berridge, K.C. and T.E. Robinson, *Parsing reward*. *Trends Neurosci*, 2003. **26**(9): p. 507-13.
84. Collins, A.C., et al., *Oral Nicotine Self-Administration in Rodents*. *J Addict Res Ther*, 2012. **S2**.
85. Howard, J.D., et al., Odor quality coding and categorization in human posterior piriform cortex. *Nat Neurosci*, 2009. **12**(7): p. 932-8.
86. Li, W., et al., Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. *Science*, 2008. **319**(5871): p. 1842-5.
87. Schoenfeld, M.A., et al., Functional magnetic resonance tomography correlates of taste perception in the human primary taste cortex. *Neuroscience*, 2004. **127**(2): p. 347-53.
88. van den Bosch, I., et al., To like or not to like: neural substrates of subjective flavour preferences. *Behav Brain Res*, 2014. **269**: p. 128-37.
89. Rabinoff, M.D., et al., *Pharmacological and Chemical Effects of Cigarette Additives*. American Journal of Public Health, 2007: p. AJPH.
90. Gullotta, F.P., *Summary of research conducted at Philip Morris Legacy Tobacco Documents Library*, 1994.

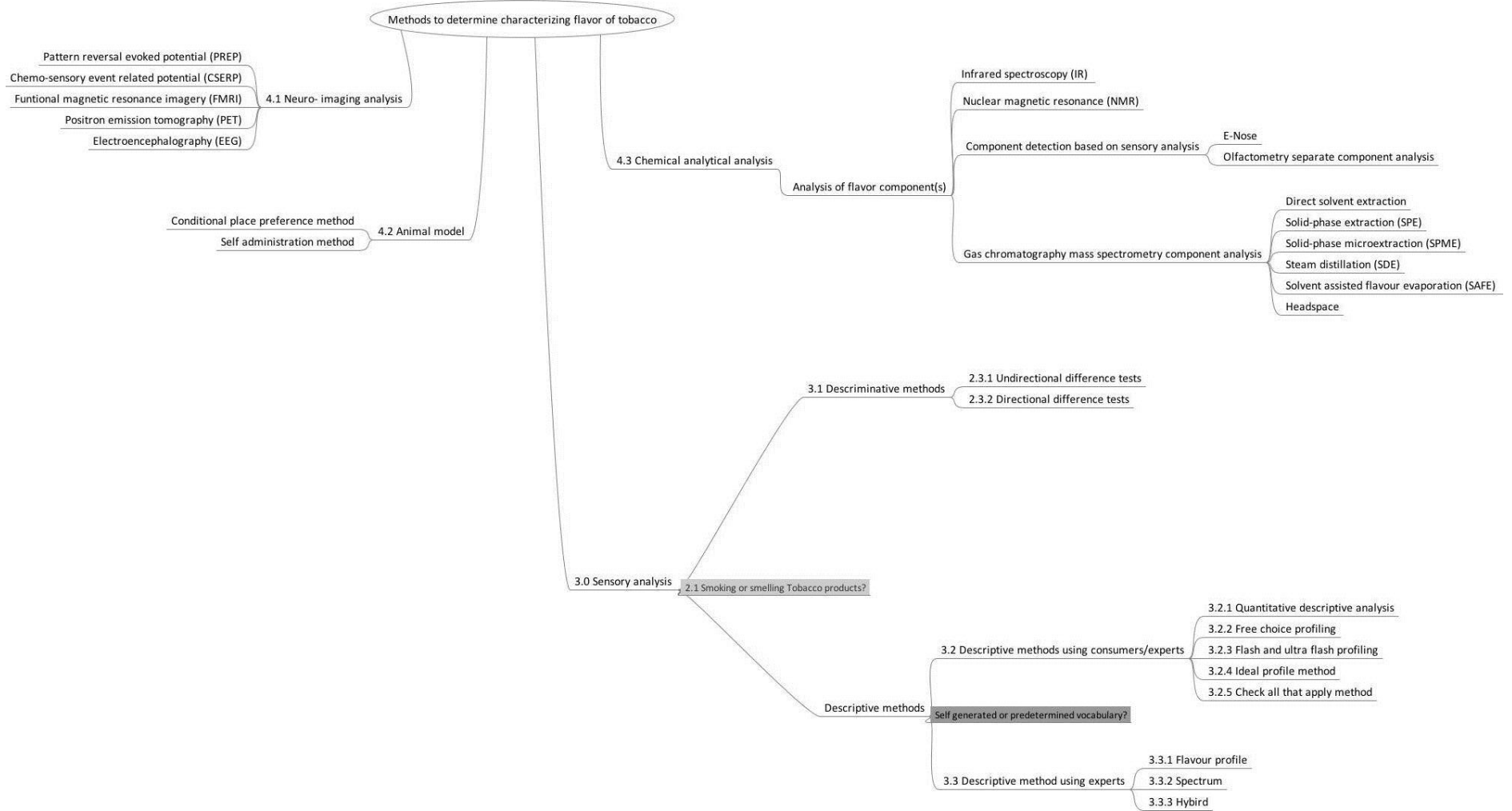
91. SCENIHR. Addictiveness and Attractiveness of Tobacco Additives. 2010; 1-119]. Available from: [http://ec.europa.eu/health/scientific\\_committees/emerging/docs/scenihr\\_o\\_031.pdf](http://ec.europa.eu/health/scientific_committees/emerging/docs/scenihr_o_031.pdf).
92. Palmatier, M.I., et al., Effects of nicotine on olfactogustatory incentives: preference, palatability, and operant choice tests. *Nicotine Tob Res*, 2013. **15**(9): p. 1545-54.
93. Sclafani, A. and K. Ackroff, Nutrient-conditioned flavour preference and incentive value measured by progressive ratio licking in rats. *Physiol Behav*, 2006. **88**(1-2): p. 88-94.
94. Figueroa, J., et al., Flavour preferences conditioned by protein solutions in post-weaning pigs. *Physiol Behav*, 2012. **107**(3): p. 309-16.
95. Costello, M.R., et al., Comparison of the reinforcing properties of nicotine and cigarette smoke extract in rats. *Neuropsychopharmacology*, 2014. **39**(8): p. 1843-51.
96. Yarmolinsky, D.A., C.S. Zuker, and N.J. Ryba, *Common sense about taste: from mammals to insects*. *Cell*, 2009. **139**(2): p. 234-44.
97. Zhang, Y., et al., Simultaneous determination of 23 flavour additives in tobacco products using gas chromatography-triple quadrupole mass spectrometry. *J Chromatogr A*, 2013. **1306**: p. 72-9.
98. Luo, D., H.G. Hosseini, and J.R. Stewart, *Application of ANN with extracted parameters from an electronic nose in cigarette brand identification*. *Sensors and Actuators B: Chemical*, 2004. **99**(2-3): p. 253-257.
99. Rock, F., N. Barsan, and U. Weimar, *Electronic nose: current status and future trends*. *Chem Rev*, 2008. **108**(2): p. 705-25.
100. Frauendorfer F.; Christlbauer M.; Hufnagel J.C.; Schaller J.-P.; Glabasnia A.; Gadani, F., *Flavour evolution investigations in a tobacco gene-to-smoke project*. Proceedings of the 12th Weurman Symposium, 2008: p. 4.
101. Hasebe, H.S., S. , The Quality Estimation of Different Tobacco Types Examined by Headspace Vapor Analysis Legacy Tobacco Documents Library, 1999. **21**(2): p. 213-222.
102. Cruz, P.-R.S., *Nonconventional Analytical Techniques Applied to Tobacco and Cigarette Smoke Flavour Analysis*. Collection Brown & Williamson, 1993.
103. Brown, J.E., et al., *Candy Flavourings in Tobacco*. *New England Journal of Medicine*, 2014. **370**(23): p. 2250-2252.
104. Xu, Y., W. Fan, and M.C. Qian, Characterization of aroma compounds in apple cider using solvent-assisted flavour evaporation and headspace solid-phase microextraction. *J Agric Food Chem*, 2007. **55**(8): p. 3051-7.
105. Podlejski J, O.W., *Methods and techniques in the research of tobacco flavour*. . *Die Nahrung*., 1983. **27**(5): p. 429-36.
106. Montesinos, I. and M. Gallego, How the Inclusion of Treated Water in Beverages Influences the Appearance of Halogenated Volatile Organic Compounds. *J Agric Food Chem*, 2014.
107. Cecchi, T. and B. Alfei, Volatile profiles of Italian monovarietal extra virgin olive oils via HS-SPME-GC-MS: newly identified compounds, flavours molecular markers, and terpenic profile. *Food Chem*, 2013. **141**(3): p. 2025-35.
108. AFNOR, *Analyse sensorielle : AFNOR/V09A*. 2014.
109. EU, The international olive council's method for the organoleptic assessment of virgin olive oil Official Journal of the European Union, 2013.
110. (ASTM), A.S.f.T.a.M., Standard test method for unipolar magnitude estimation of sensory attributes. , in Designation E 1697-05. In: Annual Book of ASTM Standards, Vol. 15.08. 2008, American Society for Testing and Materials: Conshohocken, PA. p. 122-131.
111. Lawless, H.T. and J.G. Malone, The discriminative efficiency of common scaling methods. *Journal of Sensory Studies*, 1986. **1**: p. 85-96.
112. Lawless, H.T. and G.J. Malone, *A comparison of scaling methods: Sensitivity, replicates and relative measurement*. *Journal of Sensory Studies*, 1986b. **1**: p. 155-174.

113. Hayes, J.E., A.L. Allen, and S.M. Bennett, Direct comparison of the generalized visual analog scale (gVAS) and general labeled magnitude scale (gLMS). *Food Quality and Preference*, 2013. **28**(1): p. 36-44.
114. Giovanni, M.E. and R.M. Pangborn, Measurement of taste intensity and degree of liking of beverages by graphic scaling and magnitude estimation. . *Journal of Food Science*, 1983(48): p. 1175-1182.
115. SOSP. Society of sensory professionals: Just about right scales. 2014.
116. Johnson, L. and Z. Vickers, Avoiding the centering bias or range effect when determining an optimum level of sweetness in lemonade. *Journal of Sensory Studies*, 1987. **2**(4): p. 283-292.
117. Vickers, Z., Sensory specific satiety in lemonade using a just right scale for sweetness. . *Journal of Sensory Studies*, 1988. **3**(1): p. 1-8.
118. Sarah, K., H. Tracey, and H. Joanne, *Sensory evaluation: A practical Handbook*. 1 ed. 2009: Wiley-Blackwell.
119. Carpenter, R., D. Lyon, and T. Hasdell, Who Are the Right People for Sensory Analysis?, in *Guidelines for Sensory Analysis in Food Product Development and Quality Control*. 2000, Springer US. p. 71-93.
120. Sulmont, C., et al., Comparative training procedures to learn odor descriptors: Effects on profiling performance. *Journal of Sensory Studies*, 1999. **14**(4): p. 467-490.
121. Delgado, C. and J.X. Guinard, How do consumer hedonic ratings for extra virgin olive oil relate to quality ratings by experts and descriptive analysis ratings? *Food Quality and Preference*, 2011. **22**(2): p. 213-225.
122. Jolliffe, I.T., *Principal Component Analysis* Springer Series in Statistics. 2002.
123. Research, O.P.P. Generalised Procrustes Analysis applications in sensory evaluation and instrumental analysis. GPA course 2004.

## 7 Appendix I: Keywords used for literature search

Keyword 1	Keyword 2
Sensory research	tobacco
Sensory analysis	Consumer panel
flavour	olive oil
Taste receptor	butter
Sensory evaluation	Expert panel
Identification key aroma (compounds)	Consumer product
Characteristic flavour	smoking
Characterising taste	Analytical chemical
Flavour release	Tobacco additives
Sensory assessment	smoking
Flavour perception	Laboratory animals
Chemo-sensory event	rodents
Characterising aroma	EU legislation
Characterising odour	olfactometry
trigeminal	Neuro-imaging
somatosensory	Tobacco product attractiveness
Sensory perception	Smoking panel
Discriminative taste	cigarettes
Sensory evaluation system	smoke
Sensory evaluation method	Transient receptor potential channels
Sensory evaluation	TRP channel
Distinguishable aroma	Smoking test
	Tobacco additive(s)

# 8 Appendix II: Flowcharts of methods described in this review



## 9 Appendix III: Scaling methods

### 9.1 Line scaling

Line scales are used to rate intensities of product attributes. Panellists are asked to mark the line to indicate the intensity of a particular product attribute. The line scale is often called Visual analogue scale (VAS-scale) and is used for example in the QDA method. Comparing the line mark with a categorical scale, panellists make a more continuous choice and are less limited. Most of the line scales have labels only at both ends of the line, although there are also line scales that use reference points that represent the value or a standard. In that way, the test products are scaled relative to the references (39). The line scale as suggested according to the ASTM procedure includes the following labels across the line scale; threshold, slight, moderate and strong (126).

Line scaling is not more sensitive compared to other scaling techniques (127, 128). However, line scale data can be statistically analysed using parametric test; this is not possible for categorical scales. Lines scales are often used when working with an expert panel as after training panel members are able to distinguish more intermediates than for example in a 3 or 5 categorical scale. The disadvantage of a line scale but also for a 9- or 7- point categorical scale is that people have the tendency to avoid the end of the scale. When setting up an expert panel, panel members should be trained in using the utmost ends of the scale (39). A study assessing solutions that differ in sweetness found no differences in results found when using a labelled magnitude scale (as recommended by the ASTM standard) compared to the standard visual analogue scale. Both scales were able to measure the differences in sweetness. Participants did indicate that the usage of the labelled scale was easier. In addition, the labelled scale gave data that are more categorical, and that were not normally distributed, while this was not the case using the visual analogue scale without labels (129).

### 9.2 Category scales

The category scale consists of discrete response alternatives to indicate increasing sensation intensity. The alternatives can be placed in a horizontal or vertical line and can be indicated with either numbers or verbal descriptors or check boxes (39). Category scales limited the number of alternative responses, when compared with for example the line scale. Seven to fifteen categories are often used to describe intensities of attributes; the number of categories is dependent on application and number of graduations the panellist should be able to distinguish. In the spectrum method a 15-point scale category is used but allows intermediate points, resulting in a 150-point scale (13). Disadvantage of using categorical scales is that people have the tendency to use their favourite number more than others (130). A solution for this problem would be using a horizontal row of check boxes without numbers, anchored on outer sides with for example not sweet at all to very sweet. Another way of overcoming this problem is by training panel members to use the extremes of the categorical scale by using reference products (39). Category scales are as sensitive as line scales (127) however category scales are most suited for consumer studies than for studies with expert panels (39). An example of a categorical scale is the flavour profile scale. Panellists assign a symbol or number to the intensity or other attribute of a certain product see figure 22.

Rating	Explanation
0	Not present
) (	Just noticeable
1	slight
2	moderate
3	Strong

Figure 22. Flavour profile scale



### 9.3 Scale for overall quality

The quality scale is a categorical scale with categories from 1 till 10. Within these categories, classifications are made: reject unacceptable, acceptable, match see Figure 23. This scale is usually used for key components or attributes of products on which the product can be rejected if these attributes deviate from the standard (39). For example if a product clearly departs from the standard on a specific attribute the product gets a score 1 or 2. If the samples are different from the standard but in an acceptable range they receive scores of 6-8 and when the samples are nearly identical to the reference they receive a 9 or 10 (39).

1	2	3	4	5	6	7	8	9	10
Reject		Unacceptable			Acceptable			Match	

Figure 23. Scale for overall quality

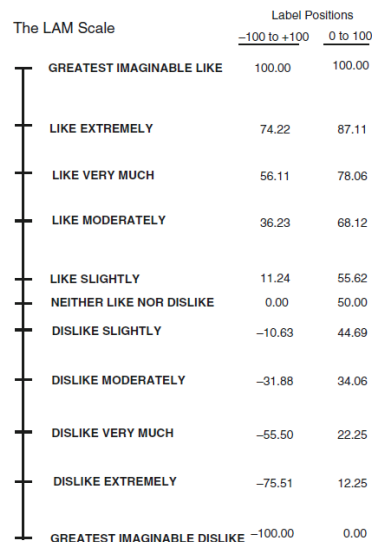


Figure 24. Labelled magnitude scale

### 9.4 Category–Ratio (Labelled Magnitude) Scales

This scale is a vertical line with marked verbal anchors that are spaced according to calibration with reference standards using ratio-scaling instructions (see Figure 24). This scale is part of the hybrid method and is a category-ratio scale. This scale can be used as an alternative for a 9 point category scale although both scales perform similarly in descriptive and differentiating functions (22). Panellists are asked to make a mark on the line to indicate their perception or strength of a specific attribute. Because of the labelled anchors panel members might have a similar idea of the intensity and therefore would be working on the same “scaling” (39).

To assess olive oil quality a standard method is developed including a scale that is a combination between the scale for overall quality and the labelled magnitude scale (see Figure 25). Panellists have to rate intensity of attributes that are important to determine the quality of olive oil on a line scale. The median of all panellist ratings determines in which classification the olive oil falls.

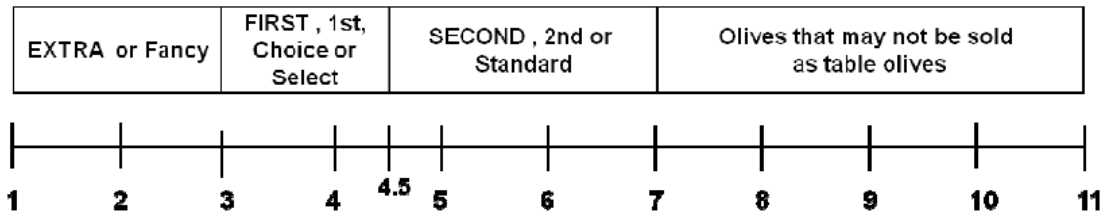


Figure 25. Scale for olive oil quality assessment

## 9.5 JAR scaling

Just about right (JAR) scales measure the appropriateness of the level of a specific attribute, and are used to determine the optimum levels of attributes in a product (50). In consumer testing of packaged goods, consumers are often asked whether a sensory characteristic of a product (e.g., saltiness) is too high, too low, or just about right. Such “attribute diagnostics” are included to assist researchers in understanding why consumers like or dislike a product and to guide product development efforts aimed at increasing consumer acceptability.

While there are many variations of JAR scales, such scales typically consist of five or seven points, ranging from too little to too much for a given characteristic (84).

One end point is labelled as “much too little”, the other end point as “much too much” and the middle point as “just right” or “just about right” (131). There are many variations of JAR scales. The most prevalent for is a 5 point category scale where categories are labelled (e.g. not nearly salty enough, not salty enough, just about right, too salty, much too salty). However, an unstructured line scale anchored with “not nearly sweet enough” at the left, “just right” at the centre, and “much too sweet” at the right was used to optimise sweetness in lemonade (132, 133).

Although JAR scales have been criticized on several grounds, including that these diagnostic questions place too great a demand on consumers to know what they ideally would like and that consumers have to have a consensus understanding of the attributes in question (84), they continue to be extremely popular, both among sensory and market research professionals. The JAR and hedonic scales information can be combined to provide directional information for product reformulation or optimisation, but analysis of the JAR data can pose some problems.

## 10 Appendix IV: Panel selection and training and test- and facility requirements

### 10.1 Panel selection and training

#### 10.1.1 Panel characteristics

All descriptive tests use 20 or fewer panel members with an optimum of 12 and a minimum of 10 panel members (55). Including more panel members does not improve product differentiation (55). Depending on the size of the difference between the products (slightly noticeable differences vs. clear difference) more or less panel members are needed (55).

General inclusion criteria for a descriptive smoking panel are:

- 18-55 years at the day of screening
- Healthy as defined by the F1 health questionnaire
- Smoker for the past five years
- Current smokers, Smoking  $\geq$  10 cigarettes a day

- The smokers' own brands are non-mentholated tobacco
- Participant should be used and willing to inhale during smoking.
- Available for a minimum period of 6 months.

Subjects are excluded from participation in a smoking panel if the subject is:

- Using medications that are known to affect taste and smell perception
- Pregnant or having plans to get pregnant
- Is lactating
- Has problems with tasting, smelling or swallowing

General inclusion criteria for a descriptive smelling panel:

- Participants can be of any age; it would make sense to use the target age (12-18).
- Panel members should be healthy
- Available for a minimum period of 6 months

General exclusion criteria for a descriptive smelling panel:

- Using medications that are known to affect smell perception
- Problems with tasting, smelling or swallowing

Selection of panel members is based on multiple characteristic factors, a panel member should be able to work in a team, have cosmopolitan preference, be positive and not over-bearing, a good listener and communicator, committed and flexible (39). Availability is another important factor, for example for the flavour profile method and QDA method panellist have to be available for years as it takes time and money to train the panel sufficiently.

Before selecting panel members, additional screening test should be performed to measure taste and odour sensitivity, the ability to discriminate between products and naming/creativity of sensory properties of a product. Examples of test are (39):

- Taste-strip test, detection and recognition
- Taste intensity ranking JND taste ca. 30% conc. Step, sensitivity
- Sniffin' sticks, odour recognition and threshold testing
- Odour memory test
- Vocabulary test, Naming/creativity e.g. sensory properties of a product

Only 10-30% of the participant passes the screening tests and can be included in the expert panel (134).

### **10.1.2 The panel leader**

The role of the panel leader is to make sure that tests are carried out appropriately. Successful panel leaders has the following common characteristics (39, 135):

- Active interest in people and the ability to earn their respect
- The ability to lead in a natural manner
- The ability to organize work, time and resources

Depending on the method used, the panel leader has different roles. For example, for the flavour profile method the panel leader is an active participant in developing the language and evaluation of the phases of the study by moderate interactions between panellist and leading the group towards an conclusion or general opinion.

It is the task of the panel leader to keep introducing the same products until reproducible results are obtained (39). However, other methods for example Spectrum method use a passive panellist that facilitates but is not active in terminology development.

### 10.1.3 Self-generated terminology versus standardized terminology

Methods differ in how the terminology used to describe attributes is developed or determined. Some methods generate the vocabulary by panel discussion and consensus. For the quantitative descriptive method panellist should be exposed to a wide range of the product assortment that is of interest, during the training panellist first individually decide on attributes that could be used to describe the difference between 2 products of the same product group. After that, through panel consensus a standard vocabulary is decided on including verbal definitions and sometimes references as decided by all panellist. During the training period also the sequence for evaluating each attribute is decided upon (39). The advantage of generating terms to describe attributes through this method is that all panel members clearly understand and know the meaning of the terms. A study of Sulmont et al. found that panel performance is better when using the consensus training to develop vocabulary (136). In practice, often a combination is used, some descriptors are provided and others are developed by panel members and added through suggestions by the panel leader from a word list.

Allowing panels to generate their own consensus may lead to misleading results when findings are generalized to the population (39), although vocabulary derived from panel members is likely to provide a better representation of product characteristics as perceived by the consumer (55). Developing for example a QDA vocabulary where terminology of panel members is used takes approximately 8-10 hours before agreement is reached on the attributes (55). However, using technological terminology panel members has to be highly trained in order to fully understand the meaning of the descriptor used to describe a product attributes (39). The most important factor of the terminology used is consensus among panel members, the existing language can be changed and the original word or words are placed in the definition section. For inexperienced subjects, the time required to learn an existing technological language can require as many as four hour sessions (55).

### 10.1.4 Duration and kind of training

The purpose of training sessions is to familiarize subjects with descriptive language and use this to evaluate products in a standardized manner by either smoking or smelling tobacco products. Depending on the test method, used panel training can differ extensively. Before performing the QDA method, panellist will meet daily for 5 consecutive days for approximately 90 minutes each session varying from 15-50 sessions (39).

Summed up the standardized method to train a panel for descriptive analysis (39):

1. Train the judges by providing the panellist with a wide range of products in the specific category and ask participants to describe differences in attributes between the products.
2. In most training situations, the most helpful references are usually a products raw materials. It is the panel leaders responsibility to determine the correct materials (55).
3. Assess attributes by generating descriptors and reference material to use as standards to describe differences among the products, usually by coming to some consensus. Reference material can play a major role in developing appropriate terminology and to evaluate intensities (39). Another way is to provide the panellist with a range of products within the category and a list of standardized descriptors and references to describe the products(90). Scorecards for many foods will have 30 or more attributes. This should not be surprising if one considers that there will be attributes for each modality—that is, for appearance, aroma, taste, texture, and aftertaste (55).
4. Determine the order of rating the attributes
5. After the terminology is developed the training sessions should focus on getting familiar with the test protocol, discrimination tests and intensity scaling. The training to evaluate cigarettes should include training on how to smoke in a standardized way. Extensive training leads to greater consistency and accuracy.
6. Determining panel members' reproducibility during training

The thesis of Fawky Abdallah, and a series of articles by Abdallah, evaluates the sensory response to smoke flavours, defines the criteria used for selecting smoke panels, and describes how panellists were trained for taste tests of cigarettes (30, 40).

## **10.2 Test- and facility- requirements**

### **10.2.1 The test product**

There are several options for testing cigarettes: having panel members or consumers smoke the cigarettes, or smell side stream smoke, unburned and burned cigarettes, cigarette package and smelling Cambridge filter pads that captured side stream smoke.

For example, Gains and Marmor describe the use of triangle tests odour evaluations of Cambridge filter pads moistened with mineral oils to capture side stream smoke odours (28)\*. Differences between odours of Flue-cured, Oriental and Burley tobacco were detected by expert panels of 5-12 members (both smokers and non-smokers). Differences in blends of two of these tobacco types could also be detected, but not those of mixtures of three.

When we decide that panel members have to smoke the cigarettes to evaluate the characterising flavour, cigarettes should be marked to use as physical method to control the amount of stimulus (smoke) as too much stimulus will lead to fatigue and too little to inadequate ratings (39). A maximum of three cigarettes per session is recommended (30, 40).

For descriptive methods it is necessary for judges to replicate their judgments up to six times, for consistency checks of the individual panellist and the complete panel (39). In a study where they assessed the quality of olive oil, samples were evaluated in triplicate (137)

### **10.2.2 Facility requirements**

For expert panel descriptive analysis all smoking and/or smelling tests should be conducted under controlled conditions in sensory booths. When participants have to smoke cigarettes in order to evaluate the characterising flavour there should be special attention given to the ventilation requirements to evacuate cigarette smoke as efficiently as possible. Surface materials of for example furniture should be easily cleaned to avoid unwanted build-up of smoking residues.

For both smelling and smoking experiments, the lighting, temperature and humidity of the test room should be controlled. The temperature of the room should range between 20 °C and 25 °C and relative humidity of the room should range between 30%-60% to assure comfort of the assessor and their complete attention to the sensory test (42).

### **10.2.3 Data analysis and reporting**

Depending on the method and scale, different statistical analysis methods can be applied. Some of the sensory methods produce qualitative descriptive results that cannot be statistical analysed for significance. For example, from the flavour profile method qualitative data are obtained due to panel consensus. Considering the aim of the present study methods that use scales that only produces qualitative descriptive data are not sufficient as we cannot test whether findings are found by change or not, and our method needs to be robust. For this study aim we need a method that gives us quantitative data, for example with the QDA method intensity ratings are obtained of a visual analogue scale and these kind of data can be analysed using variance and multivariate statistical techniques.

### **10.2.4 Discriminative tests**

Data from discrimination test can be analysed using binominal test, chi-square test or a Z-test on proportions. With these tests, it can be determined whether the results of the study were found by change or whether panellists actually *noticed* a difference between samples.

### **10.2.5 Descriptive test**

Data from descriptive test can be analysed in multiple ways.

### 10.2.6 Principal Component Analysis

This multivariate technique is used to simplify data components and describes the relationships among multiple dependent variables (attributes) and the objects (product of interest). Purpose of this method is to find primary sensory variables (138). The PCA needs to be performed on the mean data for product averaged across panellist and replications. With this method, the original variables are structured into new dimensions to simplify the data and to help interpreted the results. PCA gives you a graphical representation of the relationships among variables and objects. This technique is mainly useful for dependent or correlated variables as common in sensory research. From ANOVA analysis, we can obtain the descriptors that significantly differ among samples. However, they may describe the same underlying characteristic, which you can identify by performing a PCA analysis.

### 10.2.7 Multivariate Analysis of Variance (MANOVA)

With this method, it can be determined whether there is a significant difference among treatments when comparing all dependent variables (descriptors). The difference with univariate analysis of variance (ANOVA) is that MANOVA tests for differences between all dependent variables simultaneously whereas with ANOVA one dependent variable is tested at the time limiting a type I error. When MANOVA f-test does not indicate significance, then samples do not differ across their descriptors (dependent variables) and no individual comparison should be made.

### 10.2.8 Discriminant analysis or Canonical Variate Analysis

The discriminant analysis can be used to classify products; the canonical variate analysis (CVA) can be used to separate groups of products. Discriminant analysis is not often used in sensory studies but used for classification based on chemical and instrumental analysis. Canonical is more often used and this analysis just like PCA gives you a two or three-dimensional graph of the relationships within and between products. CVA is most useful to obtain information about between-product and within product variation. CVA is basically the same as ANOVA however it is a multi-dimensional mean separation technique. If you know that product attributes significantly differ from each other (MANVO), you can use CVA to obtain a graphical map of the sample mean separation. A study of Heymann and noble found that CVA gave superior results above PCA analysis, because CVA accounts uncertainties and error correlations in the raw data.

### 10.2.9 Generalized Procrustes Analysis (GPA)

The GPA analysis is a statistical technique whereby consensus configuration from two or more data sets is compared. These data sets can be data from individual panel members. Like the PCA, the GPA provides a simplified configuration based on patterns of correlated variables. The consensus map is interpreted similar to the PCA map and shows relationships between products and attributes. Data obtained from the free choice profiling can be analysed with the generalized Procrustes analysis (139). From this analysis we obtain a consensus picture of the data of all individual panellist in two or three-dimensional space. These results show large differences between product but this method is not sensitive for small differences and results cannot be tested for significance (39). In contrast with Principal component analysis (PCA), the amount of variance explained does not give an indication for the significance or fit of the final solution, with a permutation test the odds are estimated that a random dataset would give a similar percentage consensus variance.

## 11 Appendix V: Questionnaire on sensory analysis methods for tobacco products

Dear colleagues,

The new EU Tobacco Product Directive (TPD) prohibits the placing on the market of cigarettes and roll-your-own tobacco with a characterising flavour other than tobacco. Such products may increase the attractiveness of tobacco products and thus may facilitate the initiation of tobacco consumption or affect consumption patterns. Although some countries have legislation regarding the use of flavours, no methods are operational to measure characterising flavours in order to take regulatory decisions regarding products containing such flavours.

Therefore, the Consumers, Health and Food Executive Agency (CHAFAEA) has contracted the HETOC Consortium to develop a method (based on subjective sensory profiling, chemical-analytical measurements, other methods, or a combination of methods) to decide whether a tobacco product imparts a characterising flavour other than tobacco.

The first step of this method development is to provide an overview of methods that are relevant for sensory analysis of tobacco products. To retrieve information on specific methods used for assessment of tobacco product flavour in your country, we kindly ask you to complete the following questionnaire. We will be grateful for a partially completed questionnaire if you are unable to answer some of the questions.

Your feedback will be included in our overview. Any information you provide will be helpful for our inventory, and enable evidence-based future regulation.

We are looking forward to receiving your input at your earliest convenience, but not later than **October 28<sup>th</sup>**. Please return your questionnaire by replying to this mail. Should you have any further questions, please do not hesitate to contact Dr. Reinskje Talhout (HETOC@rivm.nl).

Thank you for your time. Kind regards, the HETOC Consortium.

National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands  
 Wageningen University (WUR), Wageningen, the Netherlands  
 OP&P Product Research (OP&P), Utrecht, the Netherlands

*This questionnaire was produced under the Health Programme (2008-13) in the frame of service contract n° 2014 62 02 Concerning the Request for Specific Services N° Chafea/2014/health/19 for the implementation of Framework Contract N° EAHC/2013/Health/23 – Concerning mapping of best practices and development of testing methods and procedures for identification of characterising flavours in tobacco products with the Consumers, Health and Food Executive Agency (Chafea) acting under the mandate from the European Commission. The content of this report represents the views of HETOC CONSORTIUM and is its sole responsibility; it can in no way be taken to reflect the views of the European Commission and/or Chafea or any other body of the European Union. The European Commission and/or Chafea do not guarantee the accuracy of the data included in this report, nor do they accept responsibility for any use made by third parties thereof.*

**Contact details of the person who completed this questionnaire**

Country  
 Name:  
 Function:  
 Name of Institution:  
 Address:  
 Telephone:  
 Email:

**Question 1: are you aware of any established/officially recognized sensory methods (e.g. subjective sensory profiling) currently used for the assessment of flavours in tobacco products, either in your country, or in other countries? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 2: are you aware of any methods as referred to in question 1, currently being developed? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 3: do you have any suggestions regarding the composition of a sensory panel to be used for the assessment of flavours in tobacco products (regarding e.g. smoking status, expert or consumer, age)?**

Answer:

**Question 4: are you aware of any established/officially recognized other methods (e.g. chemical-analytical measurements, other methods, or a combination of methods) that could be used for the assessment of flavours in tobacco products, either in your country, or in other countries? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 5: are you aware of any methods as referred to in question 4, currently being developed? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 6: in your country, are there any established/officially recognized methods (e.g. subjective sensory profiling, chemical-analytical measurements, other methods, or a combination of methods) that can be used for the assessment of flavours in other products? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 7: are you aware of any methods as referred to in question 6, currently being developed? If so, please specify and include references to relevant documents and/or send them as attachments.**

Answer:

**Question 8: do you know of any experts or groups in your country/other countries working on sensory research or other relevant methods to assess characterising flavours?**

Answer:

**Question 9: do you have any other remarks or information relevant to the assessment of characterising flavours or the assessment of products containing those flavours?**

Answer:





**Thank you for your time.**

## Annex II. WP2: Concept profiles

### 1 Concept profiles

This section describes how the conclusions of the literature review are transferred into the development of the basic approaches. Based on the systematic literature search of work package 1 the methods are assessed and basic approaches identified. The information as presented provides an overview of the determination of the best method to detect characterising flavours in tobacco products.

§2 of this annex covers tasks 2.1-2.3 “identification of the applicable methods” and 2.4 “identification of three basic approaches” of work package 2. In this section, methods, product utilization and panel characteristics signalled in WP1 are assessed according to the predetermined criteria as described in the technical offer. The methods, product and panel characteristics are scored from – – to ++ on each criterion based on the information gathered in the literature review. Methods are scored individually on all criteria, the assessments of the scores are done by at least two researchers for each category and final scores were discussed in a team meeting.

- *Methods, Product* and *panel* characteristics are taken into consideration if the method, product or panel characteristic does not score – – on a criterion.
- We chose the most suitable method(s) for the three basic approaches, to detect characterising flavour, by weighing the importance of each criterion.

In §3 of this annex, three basic approaches are identified based on the ratings described in chapter one and conclusions of the literature review. Each approach is depicted in a flow chart, showing all possible options within the profile, followed by an explanation of the options. The approaches are independent but combinations of the profiles in different orders are possible.

A proposed work plan for work package 3 is established in line with the three basic approaches to test tobacco products on characterising flavours (see §5 of this annex for preliminary detailed basic approaches).

### 2 Assessment of methods and practices

#### 2.1 Assessment of sensory analysis

In this paragraph, sensory methods are assessed according to the pre- determined criteria as considered relevant by the researchers based on the literature review. The methods are scored individually from – – to ++ on each criterion. Methods will be considered as a possible method to detect characterising flavours in tobacco products, when they do not score – – on one of the criterion. Weight of the criteria, as indicated at the second row of the table, will be taken into account when the final method is chosen for the basic approaches. This will be further explained in paragraph 2.1.4.

2.1.1 Assessment of sensory testing methods

Table 21. Assessment of sensory methods

Compulsory			Essential			Important		Relevant		
Method	Fits Regulatory needs	Results of the method can be applied across all EU member states*	Applicability tobacco products in both burnt and unburned form	Specificity	Sensitivity	Reproducibility	Ease for panellist to perform the method low score is difficult	Time and costs -- is high costs	Required expertise-data analysis low score is high expertise	Required hard-software low score is much needed
<b>Discriminative methods</b>	-	Consumers:+/- expert:++	--	++	+/-	++	++	+	+	+
<b>Free choice profiling</b>	-	consumer: +/- expert: +	++	+/-	+/-	+/-	+	+	-	+
<b>Flash and ultra-flash profiling</b>	-	consumer:+/- expert: +	burned: -- unburned:++	-	-	-	+	burned: - unburned:+	-	+/-
	Compulsory		Essential	Important	Relevant		Compulsory		Essential	Important
Method	Fits Regulatory needs	Results of the method can be applied across all EU member states*	Applicability tobacco products in both burnt and unburned form	Specificity	Sensitivity	Method	Fits Regulatory needs	Results of the method can be applied across all EU member states*	Applicability tobacco products in both burnt and unburned form	Specificity
<b>Ideal /just-about- right profile method</b>	--	consumer: - expert:--	-	-	+	+/-	+	-	+	+
<b>Check All That Apply (CATA)</b>	++	consumer: ++ expert: ++	++	+	+	+	++	+	+	+
<b>Flavour profiling</b>	--	consumers:- expert:-	++	+	+/-	consumer:+/- expert: +	+/-	consumer: +/- expert: -	+	+

<b>Quantitative</b>	++	consumer:--	++	+	consumer:--	consumer: --	+/-	consumer:+/-	+	+
<b>Descriptive</b>		expert:++			expert: +	expert: ++		expert: -		
<b>Analysis (QDA)</b>										
<b>Spectrum</b>	-	consumer:--	++	+	expert: ++	expert: ++	--	expert: --	+	+
<b>Hybrid</b>	++	consumer:--	--	++	expert: ++	expert: ++	--	expert: -	+	+
		expert:++								

\* Provided that when the method is performed with an expert panel, performance is sufficient, and when done with a consumer panel, the panel is representative for the population

### 2.1.2 Assessment of approaches/best practices in sensory analysis

In this paragraph the product utilization and panel characteristics are scored individually from -- to ++ on each criteria. Criteria are chosen based on the literature review. Product utilization and panel characteristic will not be considered as a possible method to detect characterising flavours when scored - - on one of the criteria. Product and panel characteristics are chosen based on the weight of each criterion as indicated at the second row of the table. This will be further explained in paragraph 2.1.4.

**Table 22. Assessment of product utilization**

	<b>Compulsory</b>	<b>Essential</b>		<b>Important</b>			<b>Relevant</b>	
Utilization	<b>Fits regulatory needs</b>	<b>Specificity</b>	<b>Sensitivity</b>	<b>Reproducibility</b>	<b>Ease for panellist to perform the method</b>	<b>Ethical burden</b> --is high ethical burden	<b>Requirements for test facilities</b> -- is much needed	<b>Applicability younger age groups</b>
<b>Smelling</b>	+	+	+	+	+	+	++	+
<b>Smoking</b>	+	+	+/-	+/-	-	--	-	Younger age groups are not allowed to smoke
<b>Smelling+ smoking</b>	++	++	++	+/-	-	--	-	

\* Provided that when the method is performed with experts, panel performance is sufficient and when done with a consumer panel the panel is representative for the population

Table 23. Assessment of consumer vs expert panel

Panel characteristics	Compulsory	Essential		Important			Dependent of the method chosen			Relevant					
	Fits Regulatory needs	Results of the method can be applied across	Specificity	Sensitivity	Reproducibility	Time & costs -- is costly & time intensive	Ethical burden if smoking -- high ethical burden	Suitability Weekly testing	Periodical testing 2 times a year	Suitability at home testing	Suitability discriminative test	Suitability CATA method	Suitability QDA, Spectrum	Applicable younger age groups	Selection Procedure Low score is much time
Consumers*		**	+	-	+/-	+/-	+/-	--	++	+	++	++	--	+/-	+
	+														
	/-														
experts	+	++	+/-	+	++	++	-	++	+/-	--	++	++	++	--	-
	+														

\* One can have semi trained consumers (i.e. panel members receive two training sessions). This will increase reproducibility and therefore also regulatory needs to “+”, however for the training sessions panel members need to visit the sensory laboratory at location. Ratings of all other criteria stay the same. \*\* This is dependent on the population that is used as a consumer panel.

### 2.1.3 Assessment of alternative methods to sensory analysis

In this paragraph, alternative methods are assessed according to the pre- determined criteria based on the literature review. The methods are scored individually from – – to ++ on each criterion. Methods will be considered as a possible method to detect characterising flavours in tobacco products, when they do not score – – on one of the criterion. Weight of the criteria, as indicated at the second row of the table, will be taken into account when the final method is chosen for the basic approaches. This will be further explained in paragraph 2.1.4.

Table 24. Assessment of the chemical analysis methods

Compulsory		Essential				Important			Relevant	
Method	Fits regulatory needs	Applicability tobacco products in both burnt and unburned form	Accuracy (error margin) -- score is high uncertainty	Specificity	Sensitivity	Reproducibility	Labour intensive -- score is very labour intensive	Sustainability environmentally safe ++ score is more safe	Required expertise -data analysis --is high expertise	Required hard-software -- is much needed
E-nose	--	+/-	+/-	--	+/-	+/-	++	++	++	+
IR	--	+	+/-	-	-	+/-	+/-	++	--	+/-
NMR	--	+	+/-	-	+/-	+/-	+/-	+/-	--	-
O GC-MS	--	++	+/-	+/-	++	+/-	+/-	++	+/-	-
DS GC-MS	+	++	++	+	++	++	+	+/-	++	+/-
SAFE GC-MS	+/-	++	+/-	+/-	+	+/-	+	++	+/-	+/-
SPME GC-MS	+/-	++	+/-	+/-	++	+/-	+	++	+	+/-
Headspace GC-MS	++	++	+	++	++	++	++	++	++	+/-

Table 25. Assessment of alternative methods

Compulsory		Essential			Important		Relevant	
Method	Fits regulatory needs	Applicability tobacco products	Specificity	Sensitivity	Reproducibility	Ease for panellist to perform the method low score is difficult	Required expertise -data analysis -- is high expertise	Required hard-software low score is much needed
Neuro-imaging analysis	--	--	-	-	+/-	+/-	-	--
Laboratory animals	--	--	-	--	--	Not applicable	-	-

## 2.1.4 Conclusions of the assessment of methods and practices

### 2.1.4.1 Sensory methods

#### Possible methods:

Check All That Apply (CATA) method can be considered as method to detect characterising flavours in tobacco products. This method scores high on all criteria, fits regulatory needs, and is suitable for research with tobacco products. No direct measure of flavour intensity is obtained here, making this method suitable for consumers as well, because no training is needed.

Quantitative Descriptive Analysis (QDA) can be considered as a method to detect characterising flavours. This method can only be performed with a trained panel, and therefore takes more time, and is more expensive compared to the CATA method. But it is also more reliable/consistent/better suited for legislative purposes. However direct information is obtained about the intensity of the flavour (making cut-offs possible) and flavour vocabulary is determined through panel consensus.

Free choice profiling can be considered as a suitable method. However, this method is not very specific, sensitive and reproducible as subjects use their own vocabulary. The check- all- that- apply- method and Quantitative Descriptive Analysis (QDA) method score higher on these criteria.

Hybrid may also be considered a suitable method, but we do not (yet) have a golden standard on which we can base clear cut-off points for making a grouped scale.

#### Not suitable:

Discriminative methods should not be considered because these methods cannot be applied to tobacco products as they already differ too much from brand to brand. Therefore, detecting an “odd one” out of 3 or 5 is too easy and no information about possible characterising flavours can be obtained.

Flash and ultra-flash profiling method should not be considered because these methods cannot be applied to tobacco products as they need to be grouped and ranked according to similarities and differences and this requires frequent re-tasting/smoking. This is difficult to standardize as people need to re- taste to make new groupings when a new product is introduced. The re-tasting requires a lot of smoking of different cigarettes in a short amount of time, this could alter flavour perception therefore the results will not be reliable.

Ideal /just- about- right profile method should not be considered because we are not interested in consumer preferences, which is the main aim of these methods. This project aims to find a method that is capable of detecting characterising flavours in tobacco products, not to determine the ideal flavour.

Flavour profiling method should not be considered because the whole process on deciding whether or not a tobacco product contains a characterising flavour is based upon group consensus. This is not an objective method and prone to bias. Other methods require individual ratings, and with statistical analysis it is decided if a product contains a characterising flavour, this way of decision making is objective and not subjective as with flavour profiling. Therefore, this method is not suitable for legislation purposes.

Spectrum method should not be considered even though it scores high on the most important criteria. The reason for not considering this method is that it requires a lot of training and is difficult for panellist to perform. This method differs from QDA in two ways; it uses a predetermined vocabulary

instead of vocabulary generated by the panel members. The second way the spectrum method differs from QDA is that it aims to have panel consensus on the intensity of the flavours. This is not a necessity as we are interested in the relative difference between products.

#### **2.1.4.2 Approaches/ best practices in sensory analysis**

##### **Product utilization (Smelling vs smoking):**

For the most important criteria such as sensitivity and reproducibility, smelling scores better compared to smoking. Findings of a smelling experiments are more easily reproducible compared to results from flavour assessment through smoking because characterising flavours seem more easy to identify when smelling an unburned cigarette compared to smoking, where an overwhelming tobacco/smoky flavour is present. We concluded this from a small pilot test (Appendix II of Annex I).

Also for the less important criteria, smoking scores worse compared with smelling because of the higher ethical burden and facility requirements. Smoking experiments cannot be done with adolescents and only with adults that are frequent smokers.

Combining smelling with smoking gives the best scores on regulatory needs because we are interested in flavour of tobacco products before and during smoking, this also results in higher scores on specificity and sensitivity as some characterising flavours might not be detected when either smelling or smoking is performed as assessment form.

##### **Expert vs consumer panel:**

A trained panel scores higher on the important criteria: regulatory needs. A trained panel can be seen as a calibrated machine/measure, as we can control panel performance by checking discriminability, repeatability and group consensus for the panel as a whole and for panellists and attributes (i.e. flavours) specifically. If none of the panel members scores significantly different on the before mentioned 3 factors (discriminability, repeatability and consensus) for any of the attributes this means that the results are reliable. With a consumer panel, sensitivity can be measured through power calculation. Specificity of a consumer panel can be increased when semi-trained (i.e.  $\pm 2$  training sessions) in which they learn the flavour belonging to the flavour attributes with the use of references. For example, melon and apple might be difficult to distinguish; to prevent incorrect ratings these training sessions could help learning the names of particular flavours.

In addition, a trained or expert panel is more time- and cost-efficient when experiments need to be repeated in a short time period, while if experiments are done for example, twice a year a consumer panel is more time- and cost-efficient as this does not require panel maintenance.

A consumer panel can perform tests at home; however, no control can be exerted over the exact procedure that consumers will follow at home, making the results somewhat less reliable.

Testing at home location can only be done when consumers are adults because young adolescents cannot perform sensory tests independently, and therefore there is a large risk of parental interference, which will bias the results. Trained panel members can also not perform tests at home because they need to be trained together frequently (up to 3 times a week at the start of the training). In addition, results are easily biased when tests are performed at home, while with expert panels you aim to have an objective method.

For the less important criteria, consumer panels cannot be used for the Quantitative descriptive analysis method because consumers are not capable of rating intensities accurately without training.



Expert panels cannot include adolescents because of ethical reasons (train adolescents to become tobacco experts), furthermore, panellists have to be really committed and this cannot be expected from (young) adolescents.

## **2.2 Alternative methods to sensory analysis**

### **2.2.1 Chemical analysis methods**

#### **Possible methods:**

Headspace GC-MS is the best and DS GC-MS the second best method for flavour analysis, because they are most sensitive, specific, and reproducible. Headspace is more sustainable and less labour intensive for flavour analysis compared to DS GC-MS. However, headspace GC-MS is not the most accurate method (error margin 10-15%), but since it is less labour intensive in contrary to the DS GC-MS method, headspace can be used as screening method, while for quantification DS is more useful. The disadvantage of DS GC-MS is the use of chemicals that lead to noise within data, and therefore the method is less specific.

#### **Not suitable:**

E-nose, IR, and NMR are not sensitive and specific enough, so not usable for this research. Olfactory GC-MS is considered not to be useful in our research, because with this human detector one would smell all separate flavour compounds one by one. However, we are interested in the total tobacco flavour instead of all separate components.

SAFE GC-MS is not as sensitive as headspace and DS GC-MS, and is therefore not the most suitable method for this research.

In addition, SPME GC-MS is not most suitable, since it is more labour intensive and requires a higher expertise than headspace GC-MS.

### **2.2.2 Alternative methods**

#### **Laboratory animals**

Experiments with laboratory animals are not suitable for this research aim as we are interested in human flavour perception before and during smoking of tobacco products.

#### **Neuro-imaging analysis**

Although it is possible to determine brain areas involved in sensory or reward processes, it is hitherto extremely difficult to compare and discriminate between similar stimuli (such as different tobacco flavours) to each other. Therefore, neuro-imaging analysis is not a suitable method to detect characterising flavours in tobacco products.

### 3 Identification of three basic approaches and methods

#### 3.1 Reasoning choosing the three approaches and methods

In this paragraph, we explain the line of reasoning before coming to the three basic approaches. We choose the approaches by combining scores of the methods as done in chapter one of this report together with the main conclusions of the literature review.

##### Methods

- We assessed sensory methods and concluded that the quantitative descriptive analysis method and the Check All That Apply (CATA) method are the most suitable methods to answer our research question (§2 of this annex). These methods score high on all criteria, fit regulatory needs, are suitable for research with tobacco products.
- Chemical-analytical measurements of tobacco products can help identify flavour components in a standardized and objective manner. However, a flavour can be evoked by one chemical component (menthol) but also by several components (strawberry) and flavour enhancers and attenuators are involved in modifying the intensity and characterising of a flavour. This type of analysis need to be combined with sensory data to establish threshold levels of flavour components before it can be used as a screening method to identify characterising flavour.
- We assessed the chemical analytical methods and concluded that headspace GC-MS and DS GC-MS are the most suitable methods to analyse flavour components in tobacco products, because these methods are most sensitive, specific, and reproducible. Headspace GC-MS would then be the best method to determine flavour compounds in tobacco products and these data can be combined with sensory data to establish a flavour library that could be used to determine upper limits for specific flavour compounds. Headspace GC-MS is not the most accurate method (error margin 10-15%), but it is less labour intensive and therefore more suitable for this research, in contrary to the DS GC-MS method.
- A combination of a sensory method with a chemical method would be suitable for testing the characterising flavour as the results of the sensory profiling by smoking and smelling tobacco products can be related to the analysis of the flavour components. Expert panels provide detailed information about different kinds of odours/flavours and their intensities. Data obtained from an expert panel is most suitable to link to chemical analytical data. Quantification of flavour intensity allows for determining a limit whereby a flavour is defined by sensory analysis as 'characterising'. Such a limit is very useful for regulatory purposes. Ideally, in this way chemical analytical measurements can be performed not in addition to sensory analysis but in place of.
- Sensory analysis of characterising flavour of a tobacco product is subjective, and therefore requires sensory information from consumers and/or trained judges.
- We assessed alternative methods do determine characterising flavour in tobacco products and concluded that neuroimaging methods and tests using laboratory animals are not suitable.

### Sensory panel

- **Expert vs consumer:** Expert panels are formed and trained to fully understand the terms used to describe product characteristics and to use them consistently thereby they are efficient in providing reproducible results with a small variability. Expert panels require the selection and training of a panel, which will take at least several months (estimation: approximately half a year). In addition, a long-standing expert panel needs to be maintained. The balance between these time and cost aspects, and the test frequency and number of samples to be tested, is a consideration. Panel members need to be trained continuously; how often training sessions need to be scheduled is depending on panel performance, which should be checked after each test and training session. The extensive training phase is associated with costs, as panel members need to be paid per training session. Furthermore, ethical issues need to be considered when training sessions involve smoking cigarettes, as this is a health hazard for panel members. Using an expert panel is the best method to obtain a good set of attributes and a good product space provided that a representative sample of products that are on the market is used to establish the attributes and product space. Product space depicts all measured products in a 2D space. The distance between the samples indicates differences and similarities between products. These distances are based on all product characteristics or variables that are measured (see statistical analysis of detailed basic approaches §5 of Annex IV) and is most suitable to link with analytical data to ultimately establish thresholds.

Descriptive consumer panels are more efficient compared to expert panels in case of limited time, and when a very limited number of measurements needs to be done periodical. The difference in variation in the final data between expert- and consumer panel can be compensated for by the size of the panel. Consumer panels are useful to obtain information on how regular consumers perceive the (characterising) flavour of the tobacco product and to include cultural differences. However, it seems not very likely that the population of one country might consider a flavour as characterising while in another country the population does not. Establishing expert panels in every European country is not a possibility as the costs will be tremendous and because the EU commission cannot impose all member state to conduct an expert panel for the assessment of tobacco products.

- **Smokers or non-smokers as assessors:** The advantage of using non-smokers is that their taste and smell normally is better than that of smokers, the disadvantage being they cannot be asked to smoke the product and therefore they can only assess the product through smelling. When the unburnt smell is very characterising, it is likely that the burnt product also displays a characterising flavour. Beside the unburnt product, the burnt product can also be smelled in a way that does not pose a health hazard to the assessor. This can be done with smelling Cambridge filter pads containing the substances of tobacco smoke, which are formed during tobacco burning.
- **Population:** Since several studies show adolescents to be more attracted and vulnerable to the use of flavoured cigarettes, this subgroup could be selected as part of consumer panels. However, minors cannot be used for smoking experiments or tobacco expert panels.

### Sample

- **Smoking vs smelling:** When training sessions involve smoking cigarettes the ethical part of setting up an expert panel should be considered, as this is a health hazard for panel members. Furthermore, when performing smoking test, a special facility with a proper ventilation system needs to be chosen/build (smoking test can be performed at Wageningen university until July 2015), unless smoking tests are performed with consumers at home. Smelling the burnt product in the form of side stream smoke is not preferable as health consequences are still the case. When these health consequences are accepted (for example when using adult smokers) the

assessment of tobacco products should rather be done the way the product is used; by smoking. For ethical reasons smoking tests can only be done by adult smokers and not by non-smokers or adolescents. Non-smokers can however smell the tobacco of cigarettes without health hazard. When cigarettes are assessed through smelling, we can say something about the odour of the product. When the cigarette is smoked we can say something about the taste.

- Reference: Suitable references are brands with no additives or no predicted characterising taste; in case of a smelling experiment tobacco leaves can be used as a reference.
- Location: Consumer panels can be asked to assess the product in a testing facility or assess the product at home. This cannot be done with expert panels because they need to be trained together to be in consensus with each other. In addition, an expert panel is trained to have a validated measure; this is in contrast with performing tests at home location that most likely causes results to be biased. When tests are performed in a home situation no control over test conditions is guaranteed and data could be easily biased. At home tests can be done for smoking and smelling tests.

### Conclusion

The best sensory methods for our research aim are QDA and CATA. Experts should perform QDA as consumers have difficulties with reliable intensity ratings. As the panel is trained it is illogical to perform CATA as this is a relatively easy method where you do not obtain direct information about intensities, therefore this method is preferred in case of a consumer panel.

A sensory expert panel performing the QDA method provides information about intensities of (characterising) flavours. Furthermore, an expert panel is useful to generate vocabulary/attributes and to obtain a product space. Data from an expert panel can be combined with analytical data to determine upper limits for characterising flavours.

When upper limits are set, chemical analysis can be performed as a *first step* or *screening* which has the advantage that *fewer* tests need to be performed with expensive expert and consumer panels. When the flavour library is established sensory panels still need to be used in cases whereby the upper limit of flavour compounds is *not* exceeded. This is necessary because product may impart new characterising flavours for which no upper limits are established yet.

Experts can assess the cigarettes through smelling (the way the consumer first encounters the tobacco product), and through smoking (how the product is used).

A sensory consumer panel performing the CATA method provides information about which flavours are present in tobacco flavours and intensities can indirectly be derived from the amount of consumers that indicate that a certain flavour is present in a tobacco product. Furthermore, a consumer panel gives information about how regular consumers perceive the flavour or odour of tobacco products. When panel members from different nationalities are included, cultural differences can be taken into account.

Consumer information may be used to confirm/complement findings of an expert panel (“do untrained consumers also perceive these flavours”), however this severely increases costs and time investment. When a consumer panel is used as a first approach, vocabulary/ attributes could be derived from flavourists (i.e. by flavour chemist, someone who uses chemistry to engineer flavours) instead of generated through the expert panel.

From an ethical perspective, adolescents should only be allowed to smell the tobacco products. They should not be trained to participate in an expert panel, and should only perform tests at the

laboratory under standardized conditions. However, assessment through smelling does not reflect the way tobacco products are used; therefore, an adult consumer panel smoking the cigarettes could be an option. In case the consumers are adults, tests can be performed at home.

Chemical analysis is preferred to set upper levels for additives or combination thereof causing the characterising flavour. Headspace GC-MS would be the best method to determine flavour compounds in tobacco products. Headspace GC-MS is not the most accurate method (error margin 10-15%), but is less labour intensive and it is more easy to reproduce results, which is therefore most suitable for regulatory needs, in contrast to the DS GC-MS method. To obtain the most accurate amount of flavour compounds, DS GC-MS (dissolvent liquid injections) can be performed as second option in case of results that fall within the error margin of the headspace GC-MS method. Upper limits can be derived when chemical analysis data is combined with sensory (expert panel) data.

### **3.2 Overview basic approaches**

The following three concept methods can be applied to determine characterising flavours in tobacco products. We came to this decision by eliminating methods that are not suitable for our research as explained in paragraphs 2.1.4 .

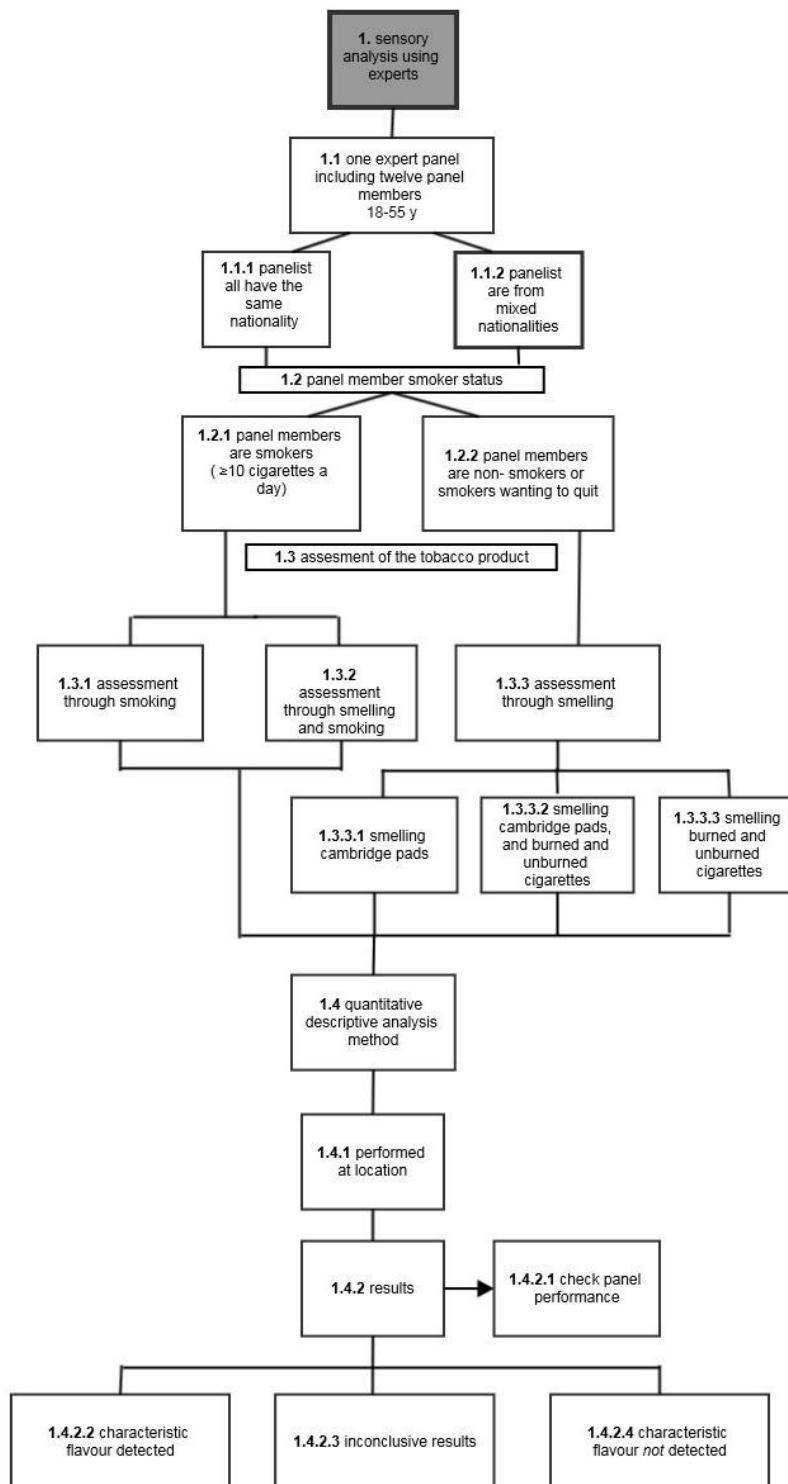
The proposed profiles are only concept approaches; multiple options within the profiles and combinations of profiles are still open for discussion, and will be decided upon at the expert seminar Jan 12-13 2015.

The three concept profiles each have their own specifics that are critical for the determination of characterising flavours in tobacco products.

- The expert panel is needed to obtain attributes/vocabulary and the appropriate product space.
- The consumer panel is needed to obtain information about consumer perception; “are characterising flavours perceived by regular consumers”.
- The chemical analysis is needed to identify flavour compounds that, in combination with findings of the sensory (expert panel), can be used to establish a flavour library specific for tobacco products that will help us determine upper levels for flavour components, which could be used as a first screening tool at a later stage.

In this paragraph each basic approach is first depicted in a flowchart, followed by the explanation of each step in the flow chart. We encourage the reader to keep the flowchart and explanation next to each other when reading this chapter.

3.2.1 Approach 1



### **3.2.1.1 Explanation with the flow chart approach 1**

#### **1. Sensory analysis using experts**

##### **1.1 Population**

One expert panel including twelve panel members aged 18-55 years; both male and female will be included. Only the participants with good smell and taste sensitivity (based on normative data) will be selected as panel members. Members are adults because it would be unethical to include minors to become tobacco odour experts, and parents would most likely not allow this. Furthermore, panel members need to be committed; this cannot be expected from young adolescents. As the ability to taste and smell decreases from the age of 55 years, older participants should not be included in an expert panel.

##### **1.1.1 & 1.1.2 Panellists all have the same nationality or panellists are from mixed nationalities**

An advantage of using panellist with the same nationality is that they might be easier to come to an agreement about flavours and their intensities, because they have the same cultural background and previous experiences with familiar flavours. Because of that, they might need fewer training sessions compared to a panel with mixed nationalities. Once the panel is trained and panel performance is sufficient there will be no difference between panellists from one or multiple nationalities.

Having the same language as a mother tongue is not a prerequisite; it might be easier for discussion due to the otherwise possible language barriers, however when people speak English sufficiently this should not be a problem. Furthermore, attributes should be generated in English, so no translation need to be made for legislation on European level.

Results of an expert panel can be applied across different cultures, since panel members are highly trained to detect specific attributes, regardless of their previous familiarity with it, and may be used for future regulatory purposes of, in this case, tobacco products.

#### **1.2 Panel member smoker status**

##### **1.2.1 Panel members are smokers**

When panel members are smokers (smoking  $\geq 10$  cigarettes a day, the assessment of the tobacco product can be done through smoking. This should be approved by a medical ethical committee, provided that smokers do not want to quit smoking. A disadvantage of using smokers being that their taste and smell ability might be decreased due to smoking tobacco, however we only include smokers that have sufficient taste and smell ability determined through a screening.

##### **1.2.2 Panel members are non- smokers or smokers wanting to quit**

When panel members are non-smokers assessment of cigarettes can only be done through smelling. The advantage of using non-smokers is that their taste ability is not decreased due to smoking; however, we do screen for taste and smell function in both cases they need to have normative smell and taste function.

#### **1.3 Way of assessment of the tobacco product**

##### **1.3.1 Assessment through smoking**

Assessing the flavour of tobacco products through smoking would be a straightforward decision as this is the natural way of using the products. This might be the best strategy to base legislation on.

##### **1.3.2 Assessment through smelling and smoking**

When assessment is done through smelling and smoking the tobacco products, we include flavour perception as perceived when first coming into contact with the product and when the tobacco product is used in the way it is used.

### **1.3.3 Assessment through smelling**

Assessing tobacco products through smelling would have the advantage that no special sensory smoking laboratory facility needs to be built and no ethical concerns are involved when no adolescents are used. From our small pilot study, participants perceived flavours more obvious than when smoking the tobacco product as flavours were not 'masked' due to the smoke.

Panel members could assess the tobacco products based on smelling packages, cigarettes (burned and unburned) and Cambridge pads (Annex I). Smelling tobacco packages and cigarettes is the first step in tobacco use. Furthermore, training the panel to assess flavours through smelling will take less time compared to smoking.

#### **1.3.3.1/2/3 Smelling Cambridge pads burned and unburned cigarettes**

Smelling a Cambridge pad is not a validated method, however when assessment is done through smelling this comes most close to actually smoking the cigarette. Volatile and non-volatile flavour compounds stick to the tar that will be captured on the Cambridge pads. While smelling un-burned tobacco products is the flavour perceived in the first step in tobacco use, burned tobacco products might contain different flavours (intensities) due to lighting the cigarette.

## **1.4 QDA**

From the method assessment as described in §2 we concluded that in case of a trained panel Quantitative Descriptive Analysis (QDA) is the best method to use for the determination of characterising flavours in tobacco products.

### **1.4.1 Performed at location**

Performing the tests at location has the advantage that conditions are standardized. With an expert panel you aim for a calibrated measure, performing tests at home would increase bias, which increases the risk of erroneous results. Furthermore, panels often need to train together in order to come to consensus about product attributes.

### **1.4.2 Results**

#### **1.4.2.1 Panel performance**

Panel and panellist performance should be controlled. Panellists should not score significantly different for repeatability, discriminability and group consensus for any of the attributes.

#### **1.4.2.3 Characteristic flavour detected**

Product falls outside the confidence ellipse of the range of tobacco products without a characterising flavour. Product is deemed to have a characterising flavour and should be removed from the market.

#### **1.4.2.4 Inconclusive results**

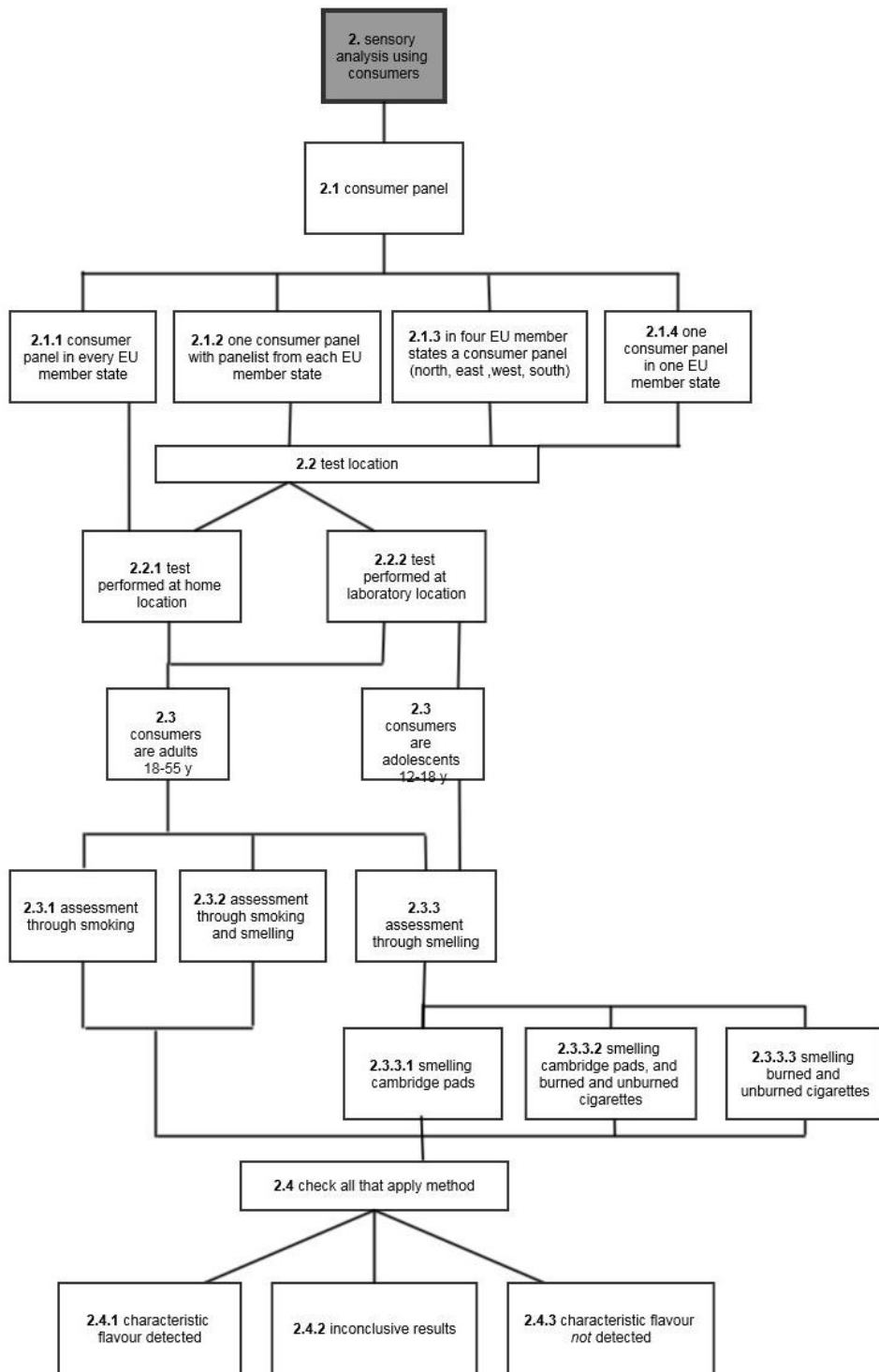
When the results of the expert panel lie within a "grey" area, a subsequent test session could be done with a consumer panel to see whether "regular" consumers perceive this flavour. If not, the product may stay on the market, as eventually consumer perception is most important.

#### **1.4.2.4 Characteristic flavour not detected**

Product falls within the confidence ellipse of the range of tobacco product without a characterising flavour, and can stay on the market.



### 3.2.2 Approach 2



### 3.2.2.1 Explanation with the flow chart approach 2

#### 2. Sensory analysis using consumers

##### 2.1 consumer panel

A consumer panel should consist of at least 100-120 panel members (depending on the population group and whether assessment is done through smelling or smoking) to have sufficient statistical power. The panel should be representative for the population and therefore both men and women should be included.

##### 2.1.1 Consumer panel in every European member state

This is not feasible in case adolescents are used and tests need to be performed at location, due to logistics. In case of adults smoking at home this might be an option. However, when adult consumer panels in every member state can be established possible cultural differences between all countries are taken into account.

##### 2.1.2 One consumer panel with panellists from each EU member state

As variability among panellists may be larger between countries, more panellists need to be included to have sufficient statistical power. However, when a consumer panel is set up including adolescents or adults from multiple countries this leads to results that are independent of cultural differences (or, are taken into account).

##### 2.1.3 A consumer panel in four EU member states

This might be feasible with adolescents at location, but is more easily to establish with adults who perform sensory tests at the home location. The four member states represent the east, west, north and southern part of Europe. Because of that it is thought to represent the possible cultural differences in a broad sense.

##### 2.1.4 One consumer panel in one EU member state

This option does not consider cultural differences. What might be considered to be a characterising flavour in one country might not be considered a characterising flavour in another country. Therefore, this option does not seem to be the most suitable one.

#### 2.2 test location

##### 2.2.1 Test performed at home location

Performing the tests at home location has the disadvantage that conditions are not standardized and the risk of unreliable data increases

##### 2.2.2 Test performed at laboratory location

Performing the tests at location has the advantage that conditions are standardized, which decreases the change of erroneous results.

#### 2.3 consumers are adults (18-55y) or consumers are adolescents (12-18y)

Young adolescents are most susceptible to start smoking with the use of flavoured tobacco and are therefore interesting to include. Therefore, the best option would be using consumers in the age category 12-18 years. However, this imparts smelling tobacco products instead of smoking. It is not allowed to use minors in a smoking experiment as determined by a medical ethical committee. Therefore, the trade-off is 1) using adult consumers that smell and/ or smoke 2) using adolescents that smell tobacco products; burned and unburned tobacco products, tobacco packages, and Cambridge pads.

### **2.3.1 Assessment through smoking**

Assessing the flavour of tobacco products through smoking would be a straightforward decision as this is the natural way of using the products. This might be the best strategy to base legislation on.

### **2.3.2 Assessment through smoking and smelling**

When assessment is done through smelling and smoking the tobacco products, we include flavour perception as perceived when first coming into contact with the product and when the tobacco product is used in the way it is used.

### **2.3.3 Assessment through smelling**

Assessing tobacco products through smelling would have the advantage that no special sensory smoking laboratory facility needs to be build and no ethical concerns are involved when no adolescents are used. From our small pilot study, participants perceived flavours more obvious than when smoking the tobacco product, as flavours were not masked due to the smoke.

Panel members could assess the tobacco products based on smelling packages, cigarettes (burned and unburned) and Cambridge pads (Annex I). Smelling tobacco packages and cigarettes is the first step in tobacco use. Furthermore, training the panel to assess flavours through smelling will take less time compared to smoking.

#### **2.3.3.1/2/3 smelling Cambridge pads and burned and unburned cigarettes**

Smelling a Cambridge pad is not a valid method, however when assessment is done through smelling this comes most close to actually smoking the cigarette. Volatile and non-volatile flavour compounds stick to the tar that will be captures on the Cambridge pads. While smelling un-burned tobacco products is the flavour perceived in the first step in tobacco use, burned tobacco products might contain different flavours (intensities) due to lighting the cigarette.

## **2.4 Check All That Apply (CATA) method**

From the method assessment as described in chapter one, we concluded that in case of a consumer panel Check All That Apply (CATA) is the best method to use for the determination of characterising flavours in tobacco products.

### **2.4.1 Characteristic flavour detected**

Product falls outside the confidence ellipse of the range of tobacco products without a characterising flavour. Product is deemed to have a characterising flavour.

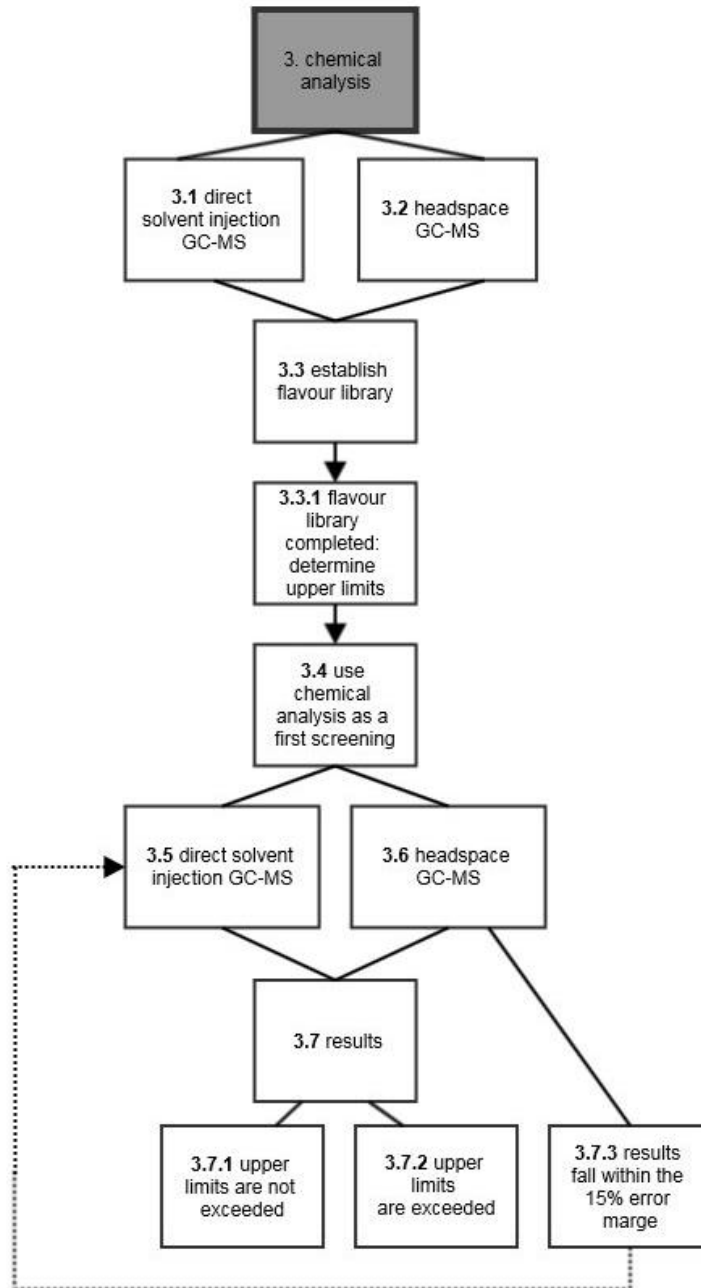
### **2.4.2 Inconclusive results**

When the results of the consumer panel lie within a “grey” area, a second test session could be done with an expert to see whether trained assessors perceive a characterising flavour. If not, the product should stay on the market.

### **2.4.3 Characteristic flavour not detected**

Product falls within the confidence ellipse of the range of tobacco product without a characterising flavour.

**3.2.3 Approach 3**



### 3.2.3.1 Explanation with the flow chart approach 3

## 3.0 chemical analyses

### 3.1 & 3.2 Headspace GC-MS and direct solvent injection GC-MS

Headspace and direct solvent injection both can be used to qualify and quantify chemical (flavour) compounds. Headspace is better suited for screening of products for all present volatile components, whereas with DS one looks for one or more specific components. Although headspace GC-MS has a larger error margin (10-15%) compared to DS GC-MS, the headspace GC-MS might be the best option for regulation purposes as this method is less labour intensive. The method that is chosen is dependent on time and budget.

### 3.3 establish flavour library

Combine data from the chemical analysis with sensory panel data to set upper limits for specific characterising flavours.

#### 3.3.1 Determine upper limits

### 3.4 Use chemical analysis as a first screening

Once the upper limits are set with use of the flavour library, chemical analysis (either headspace or direct solvent injection) can be used as a first screening, in case a tobacco product exceeds the upper limit for a characterising flavour, the product should be removed from the market. However, when the product does not exceed this limit the product still needs to be tested by a sensory panel to determine whether it does not contain a 'new' characterising flavour that is not yet in the flavour library. The screening can be performed with direct solvent injection or headspace.

### 3.5 direct solvent injection GC-MS

Once the upper limits are set direct solvent injection could be used to control tobacco product on characterising flavours. However, this method is labour-intensive and therefore more expensive.

### 3.6 Headspace GC-MS

Once the upper limits are set Headspace GC-MS could be used to determine whether the tobacco product imparts a characterising flavour that exceeds the upper limits. The headspace GC-MS measurements have a 10-15 % error margin. In case the amount of characterising flavour falls within the error margin, direct solvent injection could be used as a second control. Because DS GC-MS is more accurate in quantification compared to headspace GC-MS.

## 3.7 results

### 3.7.1 Upper limits are not exceeded

Perform a test with a sensory panel.

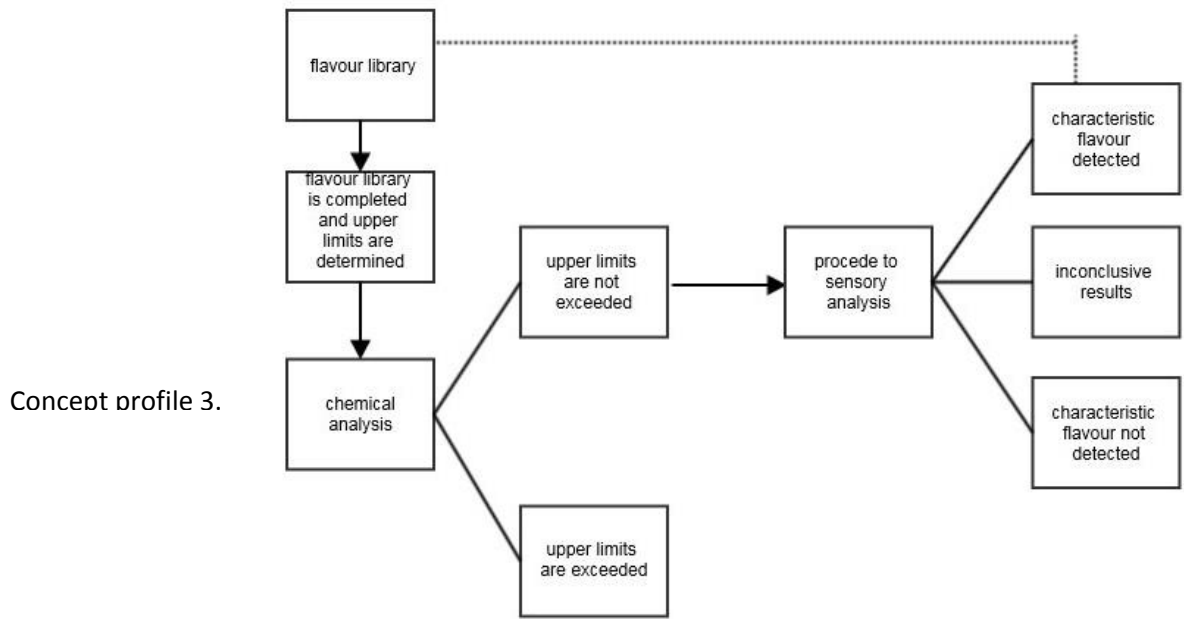
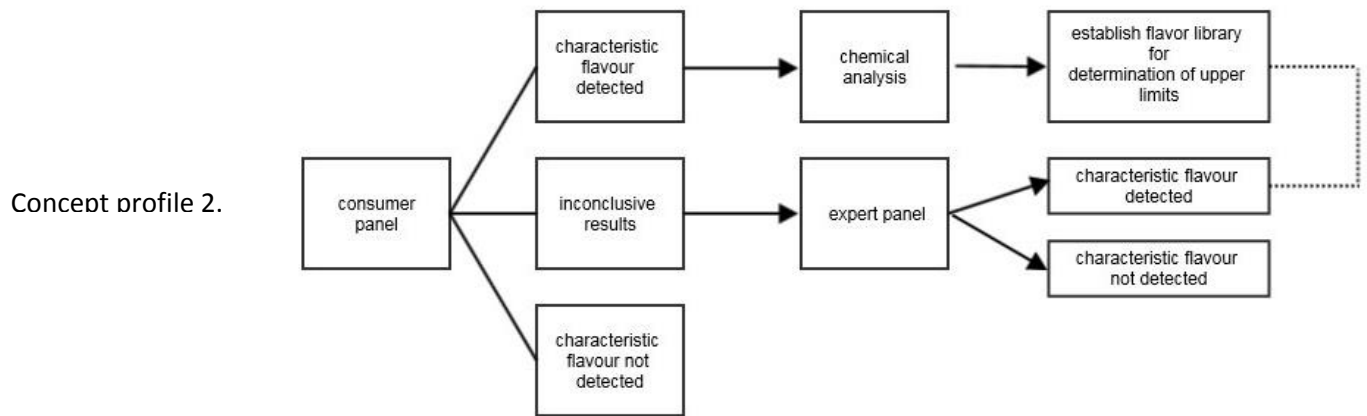
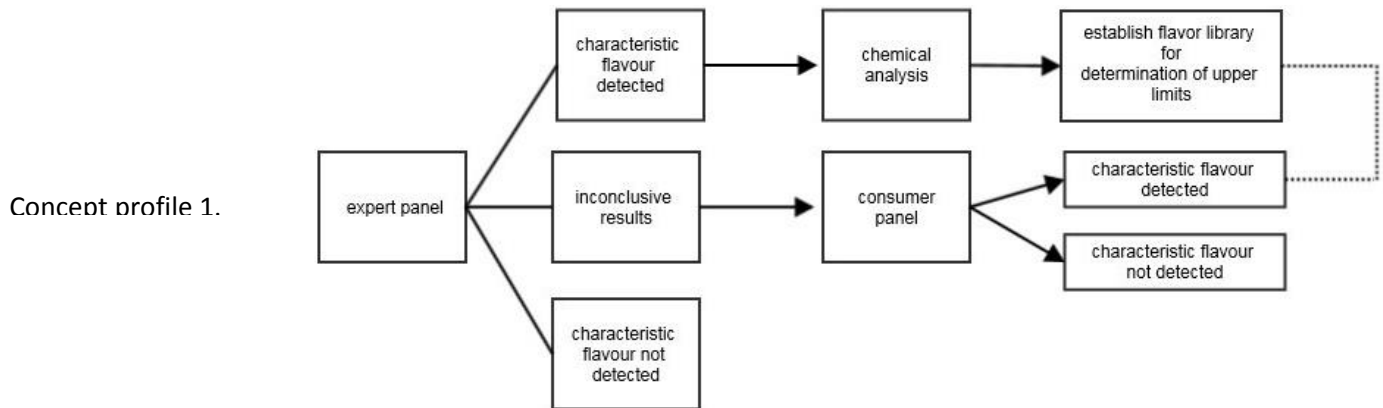
### 3.7.2 Upper limits are exceeded

The product has a characterising flavour.

### 3.7.3 Results fall within the 15% error margin

Perform direct solvent injection GC-MS to obtain a more precise measure.

### 4 Procedure; concept profiles and possible orders.



## **4.1 Explanation with the procedure flow charts**

### **4.1.1 Concept profile 1**

The expert panel will first test the tobacco product. With use of statistical analysis, it can be decided whether or not a product imparts a characterising flavour (see statistical analysis of detailed basic approaches §5 of Annex IV). When the tobacco product contains a characterising flavour; chemical analysis can be performed to obtain information about the flavour compounds present in tobacco products. With this combined information of sensory and chemical analysis a flavour library can be made. When this library is established upper limits could be determined for specific characterising flavours. The upper limits can be used as a first screening, which has the advantage that less product need to be tested by an expensive sensory panel (order 3).

However, when results from the expert panel are inconclusive (see statistical analysis of detailed basic approaches §5 of Annex IV). Verification of the findings could be done with use of sensory analysis by a consumer panel to see whether layman perceive a particular characterising flavour.

### **4.1.2 Concept profile 2**

As consumer perception is important considering the project aim, starting with a sensory method performed with a consumer panel would be logical. However, in this case, attributes and product space are not defined yet. Attributes could also be defined by an independent flavourist (flavour engineer) or, a company that provides the characterising flavours for tobacco products could be consulted. However no on beforehand determined product space is defined, it is unclear whether this would be possible.

When the consumer panel defined a product as having a characterising flavour, chemical analysis can be performed to obtain information about flavour compounds present in tobacco products. With the combined information of sensory and chemical analysis a flavour library can be made. However, when setting up a flavour library it would be better to use expert panel data as it gives you direct information about intensities and data are more clear (less variance).

When the consumer panel does not detect a characterising flavour the product can stay on the market, when the results are inconclusive flavour evaluation can be done with use of the expert panel to make a final decision.

### **4.1.3 Concept profile 3**

The order as proposed can only be performed when the flavour library is established and upper limits are determined.

With the use of upper limits chemical analysis can be performed as a first screening to determine whether the tobacco product imparts a characterising flavour. If upper limits are exceeded the product contains a characterising flavour and should be removed from the market.

However, if the upper limits are not exceeded this does not give a definite answer on whether the product contains a characterising flavour. A sensory panel should determine this, as new flavours can be introduced that are not yet present in the flavour library.

## 5 Preliminary detailed basic approaches

### 5.1 Approach 1

#### Quick overview

This methods' purpose is to determine whether a tobacco product imparts a characterising flavour. Characterising flavours and intensities will be determined by a group of tasters selected, trained and monitored as a panel.

#### Methodology

Sensory analysis performed with an expert panel using the quantitative descriptive analysis (QDA) method. Trained assessors will smell and/or smoke tobacco products and rate intensities of odours/flavours according to an attribute list that is generated through group consensus.

#### Procedure

A central expert panel with adults (18-55y) or adolescents (12-18) is established within the European Union. Prospective participants are screened on: odour threshold, identification and discrimination ability, taste threshold and identification ability, verbal skills and health. The panel should consist out of 12 panel members, both male and female can be included, as long as smell and taste ability falls within the normative range.

The panellist will assess the tobacco products by smelling the tobacco package, Cambridge pads, and unburned cigarettes and/or by smoking cigarettes and roll-your- own tobacco. Panellists are trained to come to consensus on an attribute list in English, and with the use of reference odours they are trained to identify odours in tobacco products. Panel and panellist performance will be checked; when satisfactory the panel will test the products on characterising flavours. The frequency of panel maintenance training sessions is dependent of panel performance. The Panel leader is responsible for monitoring panel performance on a regular basis and for scheduling sufficient training sessions.

#### Panel leader task

The panel leader should facilitate training sessions and keep track of panel performance. The panel leader should not be involved in decision making decisions as the method needs to be as objective as possible.

#### Technical requirements for necessary laboratory facilities

Precautions must be taken in order that the assessors in the test room are not influenced by external factors.

- The test room must be free from foreign odours and easy to clean.
- The test room and its lighting must be such that the properties of the products to be scored are not affected.
- In case of assessment through smoking, the room must be equipped with appropriate air ventilation to clear space from smoke.

#### Choice of samples for assessment in the panel session

Will be determined after the seminar

#### Sample preparation using specified standard procedures

Visual cues affect taste and smell perception; therefore, visual cues need to be eliminated as much as possible. The cigarettes will be offered to the panel members in a test tube, coded with 3 random numbers. In addition, visual cues on the cigarettes themselves will be removed as much as possible.



Tape straps will be placed over the brand name close after the filter of the cigarette. Obviously, not all visual cues can easily be eliminated (e.g., the length of the cigarette or the shape of the filter).

### **Validity, reliability and robustness**

The validity and reliability of a sensory method using an expert panel is high but depends on the panel training. Results can be controlled for discriminability, repeatability, and panel consensus by analysing data. Consequently the panel can be trained until there are no significant differences between ratings of attributes, between panellists over sessions. In that case, the sensory analysis performed by the panel can be seen as a calibrated measure.

### **Results and statistics**

#### Product space

With use of statistical analysis (principal component analysis) it can be decided whether or not a product falls within the cluster of non-characteristic tobacco products (this product cluster is determined on beforehand, by testing products representative for the tobacco market) or outside the cluster of non-characterising flavours. When the product falls inside the non-characteristic flavour cluster, the product does not contain a characterising flavour, when it falls outside the cluster the distance from the product to the cluster (in product space) should be considered with confidence ellipses.

When the product is clustered together with tobacco products that contain menthol (this product space is determined on beforehand, by testing products representative for the tobacco market), it is obvious that the product contains a characterising flavour (menthol, in this case). However, when a tobacco product forms a cluster on its own it is indecisive whether the product contains a characterising flavour.

#### Cut-off values

To be determined.

### **Considerations**

#### Costs

Training of the panel will take 30-40 sessions, estimated costs including incentives 40-60K. Costs per session (10 cigarettes incl. reference) are 2K. Maintenance of the trained panel will cost on a monthly basis 12K (one measurement session per week and one staff member for 8 hours per week (E15 per hour)).

#### Time

Training an expert panel can take up to a year before it is fully calibrated, depending on the difficulty of identifying a characterising flavour in tobacco products and is determined by panel performance.

#### Panel maintenance

The frequency of panel maintenance training sessions is dependent of panel performance.

#### Practicability

A trained panel is only cost efficient if it is in operation for two or more years (most expert panels in industry are in operation for many years).

#### Cultural differences

An expert panel is trained to identify flavours in a standardized manner independent on previous experiences and cultural background.

### **Advantages and disadvantages**

Advantage of an expert panel is that detailed information about flavours and their intensities is obtained that could be related to data of chemical analysis to establish thresholds. Furthermore, panel performance can be controlled, what can be seen as calibration of the panel. Disadvantages are the cost and time associated. The panel needs maintenance therefore regular training sessions should be scheduled. In addition, a sensory laboratory suitable for smoking should be developed (\* Assessing cigarettes through smoking can be done in a laboratory at Wageningen University until July 2015).

## **5.2 Approach 2**

### **Quick overview**

This methods' purpose is to determine whether a tobacco product imparts a characterising flavour. Characterising flavours will be determined by a group of consumers that assesses flavour through smelling or smoking tobacco products.

### **Methodology**

Sensory analysis performed with a consumer panel using the Check All That Apply (CATA) method. A group of adolescents or adults smell, or smoke tobacco products and indicate from a list with attributes which odours/ flavours are present in the tobacco product.

### **Procedure**

A Consumer panel with adolescents (12-18y) or adults is set up within one or multiple EU member states. In case of one central consumer panel, prospective participants are screened on health, smell identification and discrimination capability. After the screening participants will join two training sessions to recognize and identify different odours according to the attribute list developed by the expert panel or flavourist. For the test session participants come to the test location to assess the test samples in triplicate by smelling in a standardized manner under controlled conditions. Participants have to identify on a computer screen whether or not an odour is present, odour descriptors are presented one by one in a randomized order. When consumer are adults and tests are performed at home, no training sessions will be scheduled but the attributes and how to perform the test will be explained by sending participants an email/letter. In this letter we will also explain that people should perform the test in an odour- free room of the house or outside (i.e. not in the kitchen, or performing the test right after smoking their own cigarettes in the house).

The number of times participants need to perform test sessions is dependent of how the assessment is done, through smelling or smoking. The participants will assess the package of the tobacco product, Cambridge pads, burned and unburned cigarettes through smelling, and/or smoking cigarettes.

### **Vocabulary and assessment of a sensory profile**

The vocabulary used by the consumer panel can either be generated by an expert panel or determined by a flavourist (flavour engineer)

### **Technical requirements for necessary laboratory facilities**

Precautions must be taken in order that the assessors in the test room are not influenced by external factors.

- The test room must be free from foreign odours and easy to clean.
- The test room and its lighting must be such that the properties of the products to be scored are not affected.
- In case of assessment through smoking, the room must be equipped with appropriate air ventilation to clear space from smoke.

### **Requirements for performing tests at the home location**

When adult consumers perform tests at home, they should receive instructions on how to perform the tests in a standardized manner. In addition, the tests should be performed outside (when smoking) or in a neutralized room. For example, tests should not be performed in the kitchen close to spices and foods.

### **Choice of samples for assessment in the panel session**

Will be determined after the seminar

### **Sample preparation using specified standard procedures**

Samples are unidentifiable and presented in a randomized monadic order. The cigarettes will be offered to the panel members in a test tube, coded with 3 random numbers. In addition, visual cues on the cigarettes themselves will be removed as much as possible. Tape straps will be placed over the brand name close after the filter of the cigarette. Obviously, not all visual cues can easily be eliminated (e.g., the length of the cigarette or the shape of the filter).

### **Validity, reliability and robustness**

Comparison of CATA with other methods (mapping, QDA, ideal profile) shows high similarity (RV coefficients ranging from 0.90 to 0.97). These results suggest that the precision and reproducibility of sensory information obtained by consumers with CATA is comparable to that of traditional profiling.

### **Results and statistics**

#### Product space

With use of statistical analysis (correspondence analysis) it can be decided whether or not a product falls within the cluster of non-characteristic tobacco products (this product cluster is determined on beforehand, by testing products representative for the tobacco market) or outside the cluster of non-characteristic flavoured tobacco products. When the product falls inside the non-characteristic flavour cluster, the product does not contain a characterising flavour, when it falls outside the cluster the distance from the product to the cluster (in product space) should be considered with confidence ellipses.

When the product is clustered together with tobacco products that contain menthol (this product space is determined on beforehand, by testing products representative for the tobacco market), it is obvious that the product contains a characterising flavour (menthol, in this case). However, when a tobacco product forms a cluster on its own it is indecisive whether the product contains a characterising flavour.

#### Cut-off values

To be determined.

### **Considerations**

#### Costs

An untrained consumer panel costs E7000 per session of one hour, for triplicate measures of 10 cigarettes the total costs would be around 20K.

#### Time

Consumer panel selection and performing the test takes approximately 2 months.

#### Practicability

Consumer panel is not practical if weekly measurements need to be performed, or need to be often repeated.

### Cultural differences

Consumer panels from different European countries or one large consumer panel including participants from different nationalities could solve the problem of cultural differences (i.e. what is considered a characterising flavour in one country might not be considered a characterising flavour in another).

### **Advantages and disadvantage**

Advantage of a consumer panel is that it is faster and cheaper than a fully trained panel, while obtaining valid results. Consumer panels do not need maintenance, and can be recruited at any moment. However, this method is not suitable for weekly testing and no information is obtained about intensities. In addition, when the tobacco products' flavour is assessed through smelling Cambridge pads, there needs to be a GC-MS device at every consumer panel location to make the pads, as Cambridge pads lose odour over time.

## **5.3 Approach 3**

### **Quick overview**

This methods' purpose is to determine whether a tobacco product imparts a characterising flavour. Characterising flavours will be determined through chemical analysis.

### **Methodology**

Characterising flavours will be determined through chemical analysis.

Flavour compounds will be identified and quantified through Headspace-Gas Chromatography Mass Spectrometry (GC-MS) or by direct solvent injection GC-MS.

### **Procedure**

Sample cigarettes packs; remove the tobacco filler from 3 randomly chosen cigarettes of 1 package and homogenize the tobacco and weigh 0.2 gram in a headspace vial. Samples will be measured in triplicate after calibration of the GC-MS. When upper limits are determined by quantification with headspace, and that the results appear to fall within the error margin, DS GC-MS is performed for quantification that is more accurate.

### **Technical requirements for necessary laboratory facilities**

Headspace and DS GC-MS analysis will be performed on a gas chromatography single quadrupole mass spectrometer and a Multi-Purpose Sampler (for example an Agilent 7890B gas chromatography equipped with a 5977 single quadrupole mass spectrometer and a Multi-Purpose Sampler MPS-2, Gerstel), which is present at the laboratory in the RIVM.

### **Choice of samples for assessment in the panel session**

Will be determined after the seminar

### **Sample preparation using specified standard procedures**

From each tobacco type 0.2 g tobacco will be weighted in 20 ml headspace vials and measured using headspace gas chromatography – mass spectrometry.

The headspace vials containing tobacco samples are incubated for 30 minutes in an Agitator oven at a temperature of 140 °C. 1 ml of the formed vapour will be injected in the GC column and the vapour will be analysed.

### **Quality control**

Standard addition will be used for quantification and qualification of the characterising flavours. The method of standard addition is a type of quantitative analysis approach often used in analytical chemistry, where the standard is added directly to the aliquots of analysed sample.

Cigarettes with a known concentration of flavour will be included.

### **Procedure of the methods**

- The tobacco sample is heated in an oven of 140 °C causing all flavour components to become volatile.
- 1 ml of the formed vapour is injected on a GC column. In this column, all vapour components are separated and trapped.
- Temperature rising causes the vapour components to be released from the column. All components are released at a different compound specific temperature. The time after which a component is released from the column is called retention time.
- Components released from the column pass by a detector, which displays the released components in a chromatogram. Each peak represents one component, for example a flavour.
- The components pass by the detector; the components are fractionated in several different ion masses. These are measured, which is called mass spectrometry.
- A chromatogram is provided, with each peak consisting of unique masses, characteristic for a specific component. The masses and retention time are associated with a specific compound.
- Identification of the flavour compound, by searching a digital flavour library. The library gives presumable matches; it is required to compare the peak outcome with standards.
- Comparing with a standard: a known compound is analysed with headspace GC-MS, which provides information about the retention time and the mass spectrum of that specific compound.
- When the retention time and mass spectrum from the standard and the outcome of the flavour library are compared, the different flavour compounds of the tobacco that is analysed can be identified. This provides a list with all present components in the different types of tobacco products. For quantification, the flavour compounds will be compared to the standards, which provides a concentration of these compounds
- The flavour of all measured chemical components will be determined using the Leffingwell Flavour Base-9. This flavour data base will help us determine the flavour of the chemical compounds, which thereafter can be compared with flavours identified by a sensory panel.

### **Validity, reliability and robustness**

The GC-MS is calibrated with a standard mix, control measurements are performed, and results are compared with a known list of ingredients (Leffingwell Flavour Base-9) and flavour molecules of the standard mix. Results are robust as the gas chromatography is coupled with a mass spectrometry; this combines the high selectivity of the chromatographic separation with the sensitivity of a MS detector. GC-MS has been widely heralded as the “gold standard” for analytical substance identification. In case a product falls within the error margin of the headspace GC-MS, DS GC-MS method is performed, for quantification that is more accurate.

### **Results and statistics**

#### Flavour library

Components were identified by the FFNSC2 mass spectral library and/or NIST library and verified by analytical standards. Quantification will be done by a calibration line using calibration standards; the number of components that will be quantified is dependent of time and budget. The coefficient of variation will be around 15% using headspace.

## **Considerations**

### Costs

The estimated cost for the chemical analysis of measuring and analysing 10 cigarettes (including equipment material cost) is approximately 1K. This increases when quantification is required to determine upper limits and even more when DS GC-MS is performed.

### Time

It will take the technician about 45 min per cigarette to weigh, analyse, and report. Time of reporting depends on the quantity of cigarettes, the quantity of different flavours or cigarette brands, the experience of the technician, and finally the amount of components we want to quantify. For quantification, standards need to be tested separately for each component that needs to be quantified

### Practicability

Chemical analysis can easily be performed at any time as often as preferred.

### Cultural differences

Tobacco products from the same brand and the same edition may differ across countries. As noticed for Cola and probiotic drinks flavour.

## **Advantages and disadvantage**

GCMS is highly selective and sensitive compared to for example GC-FID. However, it is an expensive device and knowledge is necessary. Other injection techniques than headspace on the GCMS are liquid injection and SPME injection. Both techniques are usable, but are more labour intensive than the injection method we propose. Therefore, DS GC-MS is performed in case the quantification of characterising flavour falls within the error margin of the headspace GC-MS method. The main advantage of chemical analysis in general is that it is an objective method; however, the disadvantage is that human perception of flavours cannot be measured with this method. A solution for this problem is ultimately combining expert panel data with data of the GC-MS so threshold values can be determined. The drawback of chemical analysis is that flavours are often built up out of many different molecules, making it hard to conclude which combinations cause certain flavours.

## Annex III: WP3 - MAIN document

### 1 Purpose and scope

The purpose of the method described in this document is to assess odours of tobacco products and to determine whether a tobacco product imparts a characterising odour derived from an additive and/or combinations of additives. Conclusions following from the results of the method provide support for legislative purposes. The methods as described are applicable to cigarettes, and roll your own tobacco. The assessment of characterising odours is performed by a group of panellists operating as a sensory expert panel. The panellists have been selected, trained and monitored with respect to their odour assessment abilities. Chemical analysis is performed to complement the results of the sensory panel by identifying the odour components causing the odour and to build an odour library. Chemical analysis will also be used to determine whether the characterising odour is derived from a compound or combination of compounds that is not present in tobacco leaves (Burley, Virginia, Oriental), and therefore may be considered an additive.

This method has been peer-reviewed and comments have been taken into account.

### 2 General

For the general basic procedure steps of the methods and details we refer to the standard operating procedure documents (Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1), and Annex V: WP3 – Standard Operating Procedure 2 (SOP 2)).

### 3 Definitions

#### Advisory expert panel

A group of 18 panellists selected, trained and monitored as a sensory expert panel of which the 12 best performing panellists, identified as such during the preceding performance test-sessions, perform the test-sessions during which the tobacco products are assessed on characterising odours.

#### Characterising odour

Clearly noticeable smell other than one of tobacco, resulting from an additive or a combination of additives, including but not limited to, fruit, spice, herb, alcohol, candy, menthol or vanilla, which is noticeable before the consumption of the tobacco product (i.e. before smoking).

#### Chemical analysis

Identification of chemical components of tobacco products with use of headspace Gas Chromatography Mass Spectrometry- method (GC-MS).

#### Maintenance meeting

Occasion where panel leaders retrain panellists. The function of a maintenance meeting is to keep track of and maintain panel performance after the initial training of the panel. A maintenance meeting takes a whole day and consist of 4 sessions of 1.5 hour of which 3 could, when needed, be replaced by test sessions.

#### Meeting

The duration of a meeting is one day during which 4 sessions (of 1.5 hour) can take place. These sessions can have the purpose of maintaining the panel or testing products. The first of the 4 sessions is always a performance test-session to assess panel(lists) performance.

**Odour Additive**

Chemical compound or combination thereof, inducing an odour, that is not present in tobacco leaves (Virginia, Oriental, and Burley) and therefore may be considered an additive.

**Performance**

Ability of a panel or an assessor to make valid and reliable assessments of stimuli and stimulus attributes.

**Performance test-session**

Occasion during which panel(lists) performance is determined. A performance test-session needs to be scheduled as a first session of the maintenance meeting.

**Product space**

A multidimensional plot or graphical representation of tobacco products based on the variance of the intensity scores for their odour attributes. In this plot, products appear close together when they are similar, and further apart when they are different based on more than one variable (multivariable). In this case the variables are the odour attributes. The product space is needed to determine the reference space based on a cluster of products that are similar and are considered as products that do not impart a characterising flavour.

**Reference space**

A product space (multidimensional plot) consisting of reference products only (i.e. products that do not impart a characterising odour). Based on the initial product space including all products (+/- 50) a cluster of reference products (without characterising odours) is chosen from these 50 products. After that, a new product space (PCA-plot) will be determined with only the reference products; the reference space. With use of the reference space it can be determined whether a new test product is significantly different from the products in the reference space.

**Reliability**

Results of the method are reliable if the assessment is performed according to the method as describe in the SOP and according to the ISO standards associated with this method. Results are reliable if panel performance standards are met and statistics are performed by a statistician as described in the SOP.

**Repeatability**

Agreement in assessments of equivalent product samples under the same test conditions by the same assessor or panel.

**Robustness**

Robustness is the ability of this method to perform effectively when its variables or assumptions are altered. A robust method gives valid results under a variety of conditions. Robust statistical test provides insight in a problem despite altered assumptions.

**Session**

A session has a duration of 1.5 hour during which the panel can be trained, maintained or products can be tested.

**Tobacco**

Leaves and other natural processed or unprocessed parts of tobacco plants, including expanded and reconstituted tobacco.

**Test session**



Occasion during which twelve panellists assess products. The duration of a test session is 1.5 hour. During one test session 2 reference products and 2 test products can be tested.

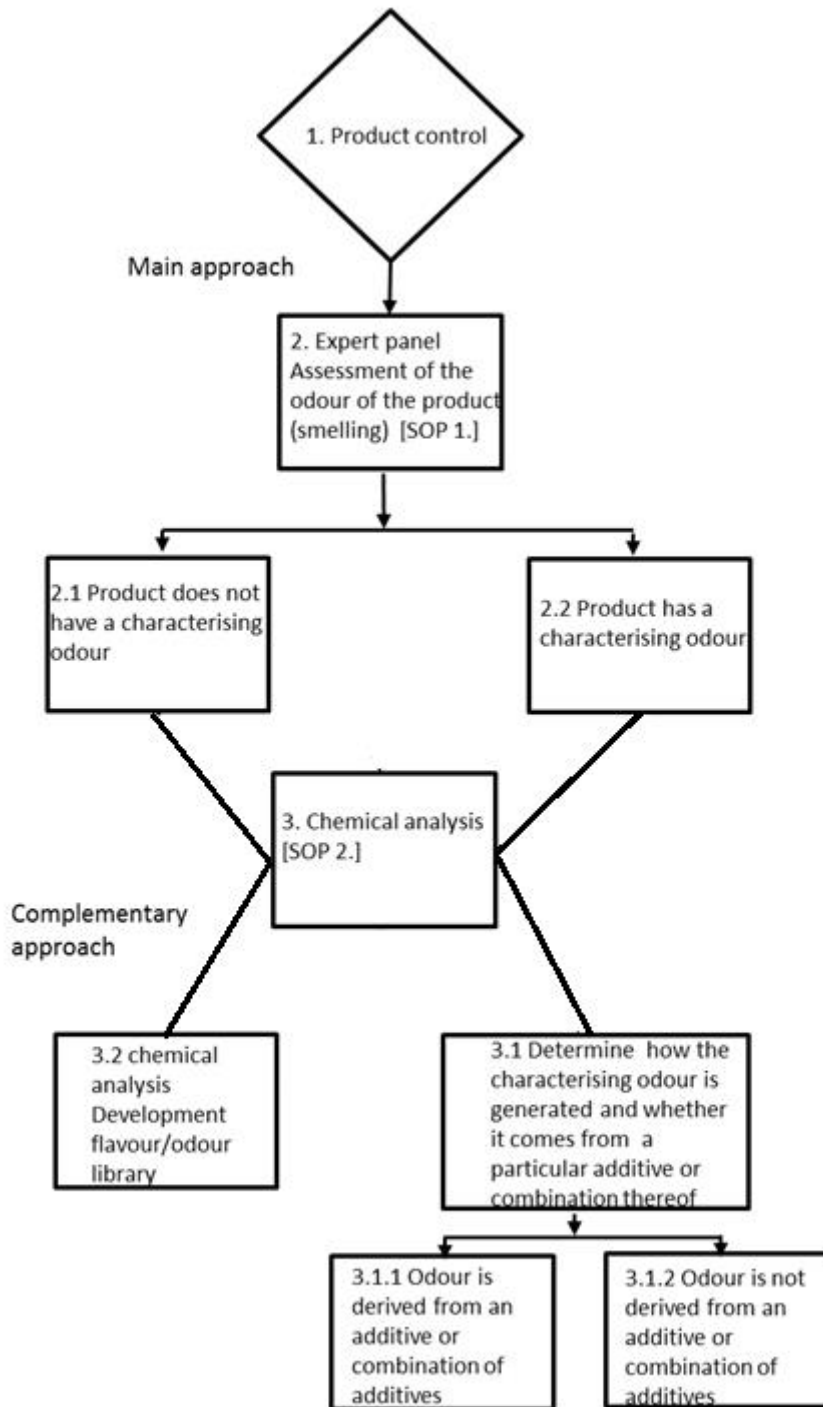
**Training session**

Occasion where panellists are trained by panel leaders to assess tobacco products. The duration of a training session is 1.5 hour.

**Validity**

The extent to which the method's outcome corresponds accurately with the truth.

## 4 Control and classification of the tobacco product



### **1. Product control: submission of a tobacco product to the panel**

EU Member States or the Commission sends a request to the independent advisory expert panel.

### **2. Expert panel; assessment of the odour of the product through smelling**

Main approach: all products requested will be assessed by the advisory expert panel through smelling. The procedures are described in standard operating procedure 1 (SOP 1).

#### **2.1 Product does not have a characterising odour**

The product does not fall outside the reference space, or is not significantly different ( $P > 0.01$ ), in terms of a higher intensity of at least one of the odour attributes, from the reference group (see SOP 1 §5).

#### **2.2 Product has a characterising odour**

The product falls outside the reference space and is significantly different ( $p \leq 0.01$ ), in terms of a higher intensity based on at least one of the odour attributes, from the reference group (see SOP 1 § 5).

### **3. Chemical analysis**

Apart from the sensory approach, all products reported by the EU Member States or the Commission will be chemically analysed with use of the headspace Gas Chromatography Mass Spectrometry-method (GC-MS). See standard operating procedure 2; chemical analysis method for the assessment of characterising odours in tobacco products.

#### **3.1 Characterising odour and additives**

##### **3.1.1 Characterising odour is derived from an additive**

If the odour components present in the tobacco product are not present in the tobacco leaves Burley, Virginia, Oriental, then the odour components are additives. Please keep in mind, that not every additive is an odour component, this can be examined using for instance the Leffingwell flavour database (See SOP2, § 9).

##### **3.1.2 Characterising odour is not derived from an additive**

If the odour components present in the tobacco products are also present in the tobacco leaves Burley, Virginia, Oriental, the odour components are no additives. These components were already present in the leaves and not added to the product.

#### **3.2 Flavour library**

Complementary approach: Chemical analysis will be performed where after sensory information will be combined with the data from the chemical analysis to develop a flavour library. The flavour library can be used to make predictions on whether or not a product may impart a characterising flavour or odour. However, with use of chemical analysis alone it cannot be concluded whether a product imparts a clearly noticeable characterising flavour/odour.

## **5 Advisory expert panel**

An independent advisory expert panel will perform the odour assessments of the tobacco products (Appendix I. Declaration of honour). The panel is trained in assessing tobacco products on odours based on smelling. The panel consists of two panel leaders and eighteen panellists (explained in more detail in §5.1 and §5.2 of this annex). Panellists and panel leaders are recruited and selected for their respective roles (SOP 1, §4.1) At least one of the panel leaders should be an expert in training a sensory panel, both panel leaders should have good communication and organisation skills (SOP 1, §4.1.1). The panel leaders are responsible for selecting the panellists, are in charge of training the

panel and monitoring panel(lists) performance in such a way that the panel meets the required standards as described in (SOP 1, §4.2.6). Two panel leaders are needed to monitor and lead conversations of a group of 18 panellists during the training phase.

The panellists must be selected and trained on account of their skills to identify and name odours, distinguish between odours, and rate odour intensity. The standards as described in the standard operating procedure should be maintained.

Prerequisites for sensory analysis by advisory expert panel

- The panel (lists) quality standards are met: panel consensus, good discrimination ability and repeatability (SOP 1, §4.2.6).
- The sensory test is performed under standardized conditions as described in §5.5 of this document and in (SOP 1, §4.3).
- As a control measure, at least 2 reference products (SOP 1, §4.3.8) should be measured within each test session.
- Statistics are performed by a qualified statistician, or a panel leader with experience and good knowledge of sensory related statistics (SOP 1, §4.1.1).

### **5.1 Panel leaders**

Two panel leaders are needed to be able to follow the group discussions that take place during the training phase of the study, and should therefore both be present at training sessions. Furthermore, in case of illness there will always be a panel leader available. For the respective roles of the two panel leaders see SOP 1, §4.1.2. The panel leaders do not take part but only facilitate the group discussion. The panel leaders should be passive in order not to bias the outcome of the panel towards a certain outcome.

### **5.2 Panel selection**

The advisory expert panel consists of eighteen panel members. Generally, for descriptive test, 20 or fewer panel members are used with an optimum of 12 and a minimum of 10 panel members [1]. Including more panel members does not improve product differentiation [1].

Twelve panel members is the optimal number of panellists as it balances power and the ability to reach consensus [2], however 6 additional panellists are needed to cover drop-outs. Recruiting and screening participants takes on average 8 hours per person.

The purpose of the screening tests is to select panel members on the basis of task specific qualities such as sensorial capacity and verbal- and social skills, and on the basis of general prerequisites such as availability, motivation and interest (SOP 1, §4.1.7).

Good olfactory function is needed to perceive the odours of the product. Verbal skills are assessed, as good vocabulary is needed, as describing odours is generally thought to be difficult. Social skills are important, as panellists should not be too dominant (as this can bias the panels' result towards the opinion of one panellist).

Most importantly, the participants should be motivated and interested to become a panel member. Other characteristics that are of importance are reliability and punctuality (panellists arrive on time when they are supposed to), ability to follow instructions (not wearing perfume etc.). All of the aforementioned qualities shall be assessed during the screening sessions and during personal interviews (similar to a job interview).

Based on experience, between 10-15% of the participants are expected to pass the screening tests. The starting pool of individuals should therefore be large, as the inclusion criteria are strict with respect to sensory capacities. For this reason, we propose that 180 people need to be screened to select eighteen panellists. Although the test-sessions should be performed by twelve panellists, to account for dropouts, six additional panel members are trained- and maintained (SOP 1, §4.1 and §4.2).

### **5.3 Panel training**

Before tobacco products can be tested on characterising odours, the panel needs to be set up and trained which will take at least several months (estimation: approximately half a year).

The panel training is needed for the panel members to learn to describe or categorize the perceived odours when smelling tobacco. In addition, panellists need to be in consensus with each other about the odour qualities (attributes) and intensities of a product. The panel members should be able to repeat product assessment in such a way that the ratings of the odour attributes are similar, furthermore they should be able to discriminate between the products (SOP 1, §4.2.6). During the initial start of the panel, the panellists generate odour attributes. The attributes should not be determined beforehand as the attributes are panel (as well as product) specific and should be chosen based on panel consensus. In this way, the panellists will understand the meaning of the attributes and agree on its use.

### **5.4 Panel training frequency**

For the panel to be operational the panellists should be extensively trained; standard times for each stage are 15-20 h for terminology development, 10-20 h for introduction to scaling, 15-40 h for initial practice, 10-15 h for small product differences and 15-40 h for final calibration [3, 4], thus 65-135 h of training is needed. Considering a 1.5-hour training session to be the maximum amount of time panellists can train due to fatigue, about 43-90 training sessions need to be scheduled. In general practice a training frequency of 2-3 times a week is often maintained when training a panel.

This frequency is maintained for practical reasons such as availability of panellists, but also for the panellist not to become too tired in case of having multiple sessions in a row. Because of that it is estimated that training an expert panel will take half a year. For the panel members to become accustomed to assessing the products, and to familiarize themselves with the odours present in the different tobacco products, frequent repeating of the tasks is important for the panel to become trained. Within these 50 sessions, panellists should be exposed to approximately 80 different brands to ensure they are exposed to a wide variety of odours varying in strength (see recommendations).

### **5.5 Test session frequency and panel maintenance**

Once the panel is trained, it is unknown how often the panel has to meet to remain well trained. There is no relevant empirical evidence and scientific basis for this, as expert panels are product and task (e.g. hedonic, descriptive or quality control) specific.

Based on sensory experts' opinions and considering the time frame in which an answer is needed on whether or not a product imparts a characteristic odour (6-8 weeks) and the number of products that can be tested a year (10-100) it is advisable to have maintenance meetings once a month that can be replaced by test sessions if needed.

During the maintenance meetings panellists' performance is assessed and panellists are retrained e.g. by practicing to rate products and odour attributes (SOP 1, §4.2.8). If panel performance of 12 panellists is sufficient, the maintenance meeting can also be used to update the reference space by

letting the panel assess the products of the reference space again. Depending on the frequency and distribution of test sessions, it may be possible to increase meeting intervals without affecting motivation and performance of the panel; this could be determined empirically. When having large time intervals for panel maintenance, at-home tasks could be given to the panellists that can be performed online. That way, the panel leader can monitor the progression of the panellists. Panellists whose performance has declined should be interviewed by the panel leader in an attempt to determine the cause. Wherever possible, assistance should be offered in an attempt to restore performance.

### **5.6 General test set-up**

The frequency of sample testing by the trained panel depends on the number of tobacco products to be tested on a yearly basis. For example, at a frequency of testing once a month, 1 performance test-session and 3 test sessions can be held during one whole-day meeting (see SOP 1, 4.2.8 and §4.3.2). During one session of 1.5 hour four products can be tested (of which two are reference products, randomly selected) in triplicate (4.3.10). Taken together,  $2 \times 3 = 6$  tobacco products can be tested on characterising odours during 3 test sessions. In case of a larger amount of products that need to be tested during one month, the maintenance sessions can be replaced for testing purposes and can be extended by one or two days (see SOP 1, §4.3.2).

The sensory test is performed in sensory booths under standardized conditions and conducted by the computer program EyeQuestion (Logic8, version 3.7.6) (see SOP 1, §4.3).

Products are given to the panellist in a blinded manner as visual cues may elicit expectations and perceptions due to the association with the package, brand or cigarette stick. Blindfolding the panellists is therefore the most objective way of testing the product. The filter and paper of the cigarette stick is removed and the tobacco is placed directly in the test bottle. The test bottles have to be prepared the day before the test session in order to have a stable odour headspace in the bottle (i.e. the tobacco odour has evaporated in the bottle and comes to equilibrium). It is the general experience that odours seem to be stronger when placed inside a bottle, compared with odours coming from the tobacco (stick) directly. However, the stronger percept of the odours will be counterbalanced by the fact that the reference products are treated in a similar way. This point is more extensively explained in the recommendations part of the pilot study results document (5.10, Product space and Reference products).

During the test-sessions, panellists judge the odour of each tobacco product individually by rating the odour attributes as generated by the panel (see Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1)). The duration of a test session should be no longer than 1.5 hour and no more than twelve product assessments can be done within this time due to panellist fatigue.

### **5.7 Product purchase**

For the testing, ten packages of each product to be tested need to be purchased, including for the products that are included to determine the reference space. The supplier or selling point is not of importance as the differences between individual cigarettes, packages, or batches of the same brand are negligible in comparison with the sensitivity of the panel as observed during the pilot testing, and their goal to assess characterising odours (e.g. differences from reference products, see §5.9). For the sensory analysis, three packs per brand are needed (one pack per test-session, products are tested in triplicate). For the chemical analysis, an additional package is needed. The remaining 6 packages serve as a backup in case the whole assessment procedure should be repeated. When collecting a sample for inspection, the procedure used for storage and handling of the sample should be done in such a way, that it does not affect the sensory properties of the sample. For the sample storing protocol see (SOP 1, §4.3.7)

## **5.8 Product profiling**

All panellists smell and evaluate the tobacco product(s) submitted for examination in triplicate. The classical Quantitative Descriptive Analysis QDA® suggests four to six replications; however, the general agreement among sensory scientists is that three replications are sufficient and give enough statistical power in case the panel is of a sufficient size ( $n \geq 10$ ) and sufficiently trained.

The panellists mark the intensity of all odour attributes, as generated by the panel.

The attributes are listed on one page by the computer program eye question (SOP 1, §4.3.9) in an order that is decided upon previously (i.e. during the training) by the whole panel (i.e. the order in which the *panel* perceives the odours or the order that is logical to them).

The products should be randomized as order effect may bias the results otherwise. For example, it is more difficult to assess a product after smelling a menthol flavoured tobacco product because of the typical intense menthol where after it is difficult to neutralize the nose-space. Intensity of the odour is marked on a 100 mm line scale (visual analogue scale) labelled with opposite terms at 10 and 90 mm (strong and weak). The anchors are needed to reduce the amount of error [5]. Furthermore, panellists should be instructed to use the whole scale (also 0 and 100) if applicable. More specifically they should be instructed to use the 0 in case the odour is not in the product at all, in case of an odour that is in almost every sample but not very distinct (for example hay odour) the panellists should score around 10 mm.

The upper part of the scale is for when a product is strong: when they cannot remember whether that product was indeed the strongest for a certain attribute, then they can score around 90mm, in order to have the additional 10 mm left in case another product they assess imparts the same odour even more intense.

## **5.9 Processing of data by the panel leader**

The profiling data, collected by EyeQuestion, is examined by the panel leader to control for missing data. The panel leader will feed the data into statistical software R (SensMineR and FactoMineR packages) to calculate the distance of the evaluated tobacco product (PCA hierarchical cluster analysis) from the reference space, and to calculate whether the product is significantly different from the products in the reference space based on the intensity of any odour attribute. The assessment must be carried out in triplicate, the average of the three measurements shall be used for the confirmation of a product having a characterising odour or not. Repeated tests must be carried out with the same panellists under the same conditions, but cigarettes or tobacco should be taken from different packages.

## **5.10 Product space, reference space and product testing**

### Product space

Principal component analysis is performed to obtain a product space based on the odour characteristics of the tobacco products (product- by attribute data).

Different product spaces should be determined for roll-your-own tobacco and cigarettes as these two types of product are assumed not to be part of the same cluster (see cluster analysis in reference space paragraph) as they constitute very different odours.

By estimation, the product space of cigarettes should consist of around 50 different tobacco brands (monitor cigarettes and tobacco leaves can also be included), to have a good overview of products currently sold on the EU market.

The product space of the roll-your-own tobacco should be constituted of approximately 25 brands. This is less than for cigarettes as it is thought that 1) there are less brands of RYO compared to cigarettes, 2) the differences between different brands of RYO is generally thought to be smaller compared to cigarettes as there are less RYO brands available on the market compared to cigarettes, and they are assumed to smell more similar. However, this should be confirmed when assessing the 25 RYO products.

The tobacco products, of which the product space is constituted, should be determined by the independent coordinating institution or authority that is in charge of the advisory expert panel. Probably, the most objective way of choosing these products would be to set up a list with all cigarette brands available on the EU market and randomly pick 50 brands of cigarettes, and do this in a similar way for selecting 25 roll-your own brands. It is advisable to have three kinds of tobacco leaves (cured Virginia, Oriental and Burley) among the 50 cigarette brands to be able to compare the deviation of commercially available products from the original product (leaf). The tobacco leaves could also be taken into account for the RYO space, these may deviate extensively from each other due to the 'Smokey' odour of some of the RYO therefore this should be determined when the space is constructed and various RYO brands have been assessed by the panel.

Ideally, the products space represents all existing tobacco taste styles. These are all commercially available types of blends from the three tobacco varieties (Virginia, Burley and Oriental), and their curing and processing variants.

Based on the variability between the products and the required statistical significance of the analysis it could be decided to analyse less or more samples to define the reference space during the set-up of the reference space.

Once panel performance is optimal, maintenance meetings are scheduled once a month. The first two months after the panel training can be used to assess the products to be represented in the product space. During one meeting, 4 sessions of 1.5 hour can take place. The first session should be dedicated to assess panel performance and is performed by all eighteen panellists. The best, available twelve performing panellists are then selected to test the products of the product space in the following 3 sessions. The other panellists will join 3 sessions or re-training during which they can practice assessing products and rating attributes they do not perform well on.

During a session of 1.5 hour we can assess 4 products in triplicate, therefore during one meeting 12 products of the product space can be assessed by the panel. Therefore, 7 maintenance meetings are needed to assess 50 cigarettes and 25 roll-your own tobaccos. To speed up the process of setting up the reference space one maintenance meeting can be extended by 2.5 day to be able to have 7 maintenance meetings during 2 months. During this extension only the panellist that were among the best 12 performing panellists should perform the assessment, the 6 remaining panellists will come back the next month. The following month the process of selecting the best 12 will start again.

#### Reference space

Before analysis can be done to determine whether a product imparts a characterising odour, a reference space should be determined, ideally involving all panellists. The reference space is a product space based on a cluster of products chosen from a wide variety of cigarette brands (50 different brands). This cluster consists of products that do not impart a characterising flavour or odour, based on hierarchical cluster analysis. From the hierarchical cluster analysis, clusters of



products that are most similar (within a cluster) and clusters are most different (between clusters) are obtained.

The reference space then reflects the average of the products on the market without characterising odour. This reference space is consequently used to determine whether a test product significantly differs from the products in the reference space based on one or more attributes with use of 99% CI ellipse and Hotellings' t test (see SOP 1, §5.2). If the test product contains an odour that was not an original attribute (i.e. is not part of the reference group), the product may automatically fall into the 'characterising' category, since it is different from the reference space products. However, as the odour should also be clearly noticeable and this is dependent of the intensity relatively to reference products, it should be checked whether other products on the market also constitute this odour. In case other products contain this odour, the reference space should be updated with products containing the relevant flavour that are not considered as products with a characterising flavour (see section 5.3.1.7.2 for other solutions). However, if it is expected that more products with this odour will appear or are already on the market, and products with a lower intensity of this specific odour should not be excluded, then it might be considered to update the reference space by testing many products with this particular odour (updated or new product space) and including the non-characterising odour products to the reference space. To define the reference space for roll-your-own tobacco, the approach is the same. A more detailed description of the statistical analysis can be found in Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1).

#### Significance

For a product to be designated as significantly different from the reference space a 99% confidence interval or alpha 0.01 is set. This alpha is chosen to minimize the change of a type I error, where the true null hypothesis is rejected incorrectly (false positive). We increased the strictness of the test (as generally 0.05 is taken as significance level) to account for the fact that trained experts are more sensitive compared to consumers in assessing odours. A significance level alone does not take into account the minimum relative distance between products that have to be set in order for a product to be designated as having a clearly noticeable odour. However, if a product is overall significantly different from the reference space/products, this means the product is perceived as 'different from tobacco', and this can be traced back to significant differences in specific attributes (noticeable smells, as generated by the panel by consensus). If the attribute is rated significantly higher in a specific test product than the reference space, thus higher than a 'noticeable smell', we conclude that this is a 'clearly noticeable smell'.

## 6 Chemical analysis

All products that are sensory analysed by a panel, will be chemically tested. Chemical analysis is performed to complement the results of the sensory panel by identifying the odour components causing the odour and to build an odour library. Chemical analysis will also be used to determine whether the characterising flavour or odour is derived from a compound or combination of compounds that is not present in cured tobacco leaves (Burley, Virginia, Oriental), and therefore may be considered an additive.

Components are identified based on the probability of the mass spectral library. This involves a certain uncertainty (to identify components with a 100% certainty, standards must be used for confirmation). The components that are present in flavoured tobacco products but absent in cured tobacco leaves (Virginia, Oriental, Burley), are, in this case considered to be flavour/odour-compound additives. However, the Leffingwell flavour database should be used to identify a component as a flavour- or odour compound as not all compounds impart an odour or flavour.

Chemical analysis is performed with headspace GC-MS, since this method is sensitive, reproducible, specific, sustainable, and not labour-intensive.

Testing is preferably performed under ISO/IEC 17025: General requirements for the competence of testing and calibration laboratories, ensuring good laboratory practices. These guidelines include qualifications needed for personnel.

### **6.1 Chemical analysis prerequisites**

- The method is performed according to the standard operating procedure (Annex V: WP3 – Standard Operating Procedure 2 (SOP 2));
- The GC-MS has been calibrated;
- Interpretation of the results is done by a qualified researcher.

### **6.2 General test set-up**

Chemical analysis is performed in a laboratory. Tobacco is isolated and pulverised to homogenise the samples. Approximately 200 mg is weighed accurately and measured with headspace GC-MS. Specifications of the chemical method are described in more detail in Annex V: WP3 – Standard Operating Procedure 2 (SOP 2).

### **6.3 Product profiling**

All tobacco products assessed by the sensory panel are also tested in triplicate chemically. To assess whether a product imparts a characterising odour departed from an additive we take the following cured tobacco leaves as reference: Virginia, Burley, and Oriental. If the sensory panel indicates a product as having a characterising odour and that odour is due to specific components (can be identified by comparing the sensory and chemical data) it is assessed whether these components are present/not present in tobacco leaves. In case the components are not present in tobacco leaves, the odour is designated to be derived from an additive or a combination thereof.

Based on the product space we can determine the group of products that do not impart a characterising odour, this will also be the reference group for the chemical analysis. Based on this, we can determine whether a new product is significantly different from this reference group (based on the chemical components). Furthermore, we can determine which compounds predict the sensory attributes as identified by the panel. Together, this information will be used for a flavour library.

### **6.4 Processing of data by the researcher**

The data is analysed by a chemical researcher. The chromatograms obtained by GC-MS are evaluated to identify the components that are present in flavoured tobacco products but absent in unflavoured products or leaves. These components are considered additives. Flavours of each additive is determined using the Leffingwell Flavour Base – Tobacco Version to compare the flavour as which the product is advertised to the flavour according to the chemical composition.

## **7 Risk assessment**

Below, details on risk management, including practical risks, and back-up solutions are provided.

### **Less than 12 panellists are available during one of the test-sessions**

- Low probability, High impact

- To cover dropouts or people being unavailable to join test sessions for various reasons such as illness and holidays, 6 additional panellists are trained simultaneously
- With high certainty, in most cases there will be 12 panellists available at one time point within 4 weeks.
- In case of more than 6 panellists drop out of the panel, new panellists need to be trained.

**One of the panel leaders is unavailable to be present at the training sessions.**

- Medium probability, low impact
- The training can be done by only one panel leader; if possible, secretariat can help making notes. In that case, the panel leader can focus on guiding the discussion. To decrease the risk, two panel leaders with similar qualities (i.e. same qualities of the panel head) could be hired.

**Product imparts an odour that is not included in the reference space**

- Low probability, High impact
- Because of this, the reference space should constitute of sufficient number of products, e.g. 50 cigarette and 25 RYO brands that are currently sold on the European market.
- In case the product imparts a characterising odour that is not included, the reference space should be updated with new reference products of which is expected that they also impart that particular odour to be able to see whether the intensity is high enough for the odour to be characterising.

**Panel drift occurs (The panel becomes more sensitive as side-effect of becoming more experienced after the initial training, long term)**

- Low probability, medium impact
- To counter this possibility, reference products are measured together with the newly tested products, and function as a calibration. In case the panel rates reference products significantly different from before (i.e. reference products no longer fall inside the reference space), the panel needs to renew the complete reference space.

**The training of the panellists takes longer than the initial estimation (i.e. approximately six months)**

- Low probability, medium impact
- In that case, it will take longer before products can be tested (i.e. before the panel performance is optimal). In case it takes too long for law reinforcement, the panel can perform the test sessions without having optimal panel performance. To account for low panel performance, the test could be often repeated to have high statistical power, or a consumer panel could be set up to determine the reference space. Once the panel is fully trained, the reference space should be reconstituted by the panel that is now fully trained.

**No characterising odour is detected, but product is still suspected to have a characterising flavour?**

- Low probability, medium impact
- This may result in still having characterising flavoured products sold on the EU market. However, especially in the first years this is not likely to occur. Based on pilot experience people perceive a wider variety of odours when smelling compared to smoking. However, on the long term, tobacco products may be developed that do not impart a characterising odour but characterising flavours are initiated once the product is lit.
- It needs to be evaluated whether, on the long term, an expert panel should, besides the smelling assessment, assess products through smoking.

**Industry removes all products with a flavour from the market before the panel is fully trained, or no products are available that can be tested over a longer time.**

- Medium probability, High impact
- Out of precaution, stores might not purchase any tobacco products that are suspected to have a characterising flavour, as they already know that these products will be banned. Because of that, the tobacco industry may already remove these products from the market (as they will not be purchased by stores anyway).
- A cost-time efficient decision should be made to conclude whether the panel should be maintained or whether a new panel should be trained once new products need to be tested.

## 8 References

1. Stone, H., R.N. Bleibaum, and H.A. Thomas, *Chapter 6 - Descriptive Analysis, in Sensory Evaluation Practices*.
2. Varela, P. and G. Ares, *Novel Techniques in Sensory Characterization and Consumer Profiling*. 2014: CRC Press Taylor & Francis Group.
3. Lawless, H.T. and H. Heymann, *Sensory evaluation of food principals and practices.*, ed. N.Y.C.a. Hall. 1998.
4. Meilgaard, M.C., G.V. Civille, and B.T. Carr, *Sensory evaluation techniques (2nd ed.)*, F.C.P. Boca Raton, Editor. 1991.
5. Gower, J.C. and G.B. Dijksterhuis, *Multivariate-Analysis of Coffee Images - a Study in the Simultaneous Display of Multivariate Quantitative and Qualitative Variables for Several Assessors*. *Quality & Quantity*, 1994. **28**(2): p. 165-184.

## Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1)

### 1 Purpose and scope

The purpose of this method is to assess the intensity of odours of tobacco products with the aim to determine whether a tobacco product imparts a characterising odour. Conclusions following from the results of the method provide support for legislation purposes. The method as described is applicable to cigarettes and roll-your-own tobacco only. Results are only valuable when the assessment is performed by a group of selected, trained and monitored panellists, operating as an expert panel.

### 2 General

This standard operating procedure (SOP) is number one of the two SOPs that are part of the test approach to detect characterising odours in tobacco product as described in the main document. This SOP describes the steps to detect characterising *odours* in tobacco products as perceived when smelling tobacco products. Standard operating procedure number two describes the steps for the *chemical analysis*. Chemical analysis is performed to complement the results of the sensory panel by identifying the odour components causing the odour and to build an odour library. Chemical analysis will also be used to determine whether the characterising odour is derived from a compound or combination of compounds that is not present in cured tobacco leaves (Burley, Virginia, Oriental), and therefore may be considered an additive.

§4 of this document describes all steps necessary to determine characterising odours in tobacco products. In the main document, the justifications behind the steps and choices made are described.

*This method has been peer-review and comments have been taken into account.*

### 3 Definitions

#### **Advisory expert panel**

A group of 18 panellists selected, trained and monitored as a sensory expert panel of which the 12 best performing panellists, identified as such during the preceding performance test-sessions, perform the test-sessions during which the tobacco products are assessed on characterising odours.

#### **Characterising odour**

Clearly noticeable smell other than one of tobacco, resulting from an additive or a combination of additives, including but not limited to, fruit, spice, herb, alcohol, candy, menthol or vanilla, which is noticeable before the consumption of the tobacco product (i.e. before smoking).

#### **Chemical analysis**

Identification of chemical components of tobacco products with use of headspace Gas Chromatography Mass Spectrometry- method (GC-MS).

#### **Maintenance meeting**

Occasion where panel leaders retrain panellists. The function of a maintenance meeting is to keep track of and maintain panel performance after the initial training of the panel. A maintenance

meeting takes a whole day and consist of 4 sessions of 1.5 hour of which 3 could, when needed, be replaced by test sessions.

### **Meeting**

The duration of a meeting is one day during which 4 sessions (of 1.5 hour) can take place. These sessions can have the purpose of maintaining the panel or testing products. The first of the 4 sessions is always a performance test-session to assess panel(lists') performance.

### **Performance**

Ability of a panel or an assessor to make valid and reliable assessments of stimuli and stimulus attributes.

### **Performance test-session**

Occasion during which panel(lists) performance is determined. A performance test-session needs to be scheduled as a first session of the maintenance meeting.

### **Product space**

A multidimensional plot or graphical representation based on variance. In this plot, products appear close together when they are similar, and further apart when they are different based on more than one variable (multivariable). In this case, the variables are the odour attributes. The product space is needed to determine the reference space based on a cluster of products that are similar and are considered as products that do not impart a characterising flavour.

### **Reference space**

A product space (multidimensional plot) consisting of reference products only (i.e. products that do not impart a characterising odour). Based on the initial product space including all products (+/- 50) a cluster of reference products (without characterising odours) is chosen from these 50 products. After that, a new product space (PCA-plot) will be determined with only the reference products: the reference space. With use of the reference space, it can be determined whether a new test product is significantly different from the products in the reference space.

### **Reliability**

Results of the method are reliable if the assessment is performed according to the method as describe in the SOP and according to the ISO standards associated with this method. Results are reliable if panel performance standards are met and statistics are performed by a statistician as described in the SOP.

### **Repeatability**

Agreement in assessments of equivalent product samples under the same test conditions by the same assessor or panel.

### **Robustness**

Robustness is the ability of this method to perform effectively while its variables or assumptions are altered. A robust method gives valid results under a variety of conditions. Robust statistical test provides insight in a problem despite altered assumptions.

### **Session**

A session has a duration of 1.5 hour during which the panel can be trained, maintained or products can be tested.

**Tobacco**

Leaves and other natural processed or unprocessed parts of tobacco plants, including expanded and reconstituted tobacco.

**Test session**

Occasion during which twelve panellists assess products. The duration of a test session is 1.5 hour. During one test session, 2 reference products and 2 new products can be tested.

**Training session**

Occasion where panellists are trained by panel leaders to assess tobacco products. The duration of a training session is 1.5 hour.

**Validity**

The extent to which the method’s outcome corresponds accurately with the truth.

**4 Standard operating procedure**

**4.1 Panel set up**

**4.1.1 Recruitment and hiring of panel leaders**

Two panel leaders need to be available to train the sensory panel, one head panel leader and one vice-panel leader. The panel leaders are trained experts and have good communication and organisation skills. For the workload of both panel leaders throughout the phases of the expert panel set up, training and testing, see cost-estimate document.

Preferred qualities head panel leader:

General:

- An MSc degree in sensory science or related field (food science, behavioural/social science) and courses in statistics and working knowledge of Quantitative Descriptive Analysis is preferred.
- At least 5 year(s) related work experience (R&D, market research, innovations, or consumer marketing).
- Experience with developing, training and leading a sensory panel (large groups).
- Working understanding of statistical analysis and research, and ability to translate data analysis into implications.

Leadership competence:

- Planning and execution
- Good communication skills
- Analytical and problem solving
- Group moderator, discussion facilitator

Minimum qualities vice-panel leader:

General:

- An MSc degree in sensory science or related field (food science, behavioural/social science).
- Interested in developing, training and leading a sensory panel (large groups).

Leadership competence:

- Planning and execution
- Good communication skills
- Analytical and problem solving
- Group moderator, discussion facilitator

#### **4.1.2 Role of the panel leaders**

The panel leaders are responsible for recruiting and selecting the panellists, are in charge of training the panel and monitoring of the panel(lists) performance in such a way that the panel meets the required standards as described in §4.2.6 of this annex. The panel leaders do not take part in the panel discussions but have the main task of facilitating the discussion. The panel leaders should not participate in the assessment of the products. The panel head is responsible for determining the training content, performing statistics (together with a statistician see §4.1.3 of this annex) and keeping track of and maintaining panel performance. The vice-panel leader has a supporting and organisational role. The vice-panel leader helps with the panel training set up, preparing samples and making training schedules in agreement with the panellists.

#### **4.1.3 Supporting personnel**

##### Administrator/secretary

A secretary is needed for administrative work, such as:

- contacting prospective panellists (i.e. recruitment)
- contracts and declaration of honour with respect to absence of conflict of interest
- payments including participants' reimbursements
- making schedules for panel training, maintenance sessions and test sessions
- coordinating and planning which products should be tested by the panel

For the workload of the secretary , see Annex IX. Cost estimate.

##### Statistician

Statistical analysis to determine whether a product imparts a characterising odour should be performed by a qualified statistician or panel leader with extensive knowledge of sensory data analysis. Panel performance statistics can be done by the panel leader, as this does not require extensive statistical knowledge.

For the workload of the statistician, see Annex IX. Cost estimate.

#### **4.1.4 Recruitment of participants**

The panel leaders recruit and select the panellists. To select eighteen panellists (see §4.1.7) at least a tenfold number of participants should be included in the screening. Potential candidates should receive written information on the in- and exclusion criteria, what is expected of panellists (i.e. commitment; number of training sessions per week/year) and that taking part in the screening does not guarantee that they will become panellists, as only a selected number of people with good smell function and verbal and social qualities will be included. The recruitment of the panellists should be in accordance with ISO standard 8586 [1].



#### 4.1.5 Information meeting for potential candidates/panellists

During the information meeting, the aim of the sensory panel is clearly explained. Furthermore, inclusion and exclusion criteria as well as the screening sessions should be explained. Thereafter, participants fill in a questionnaire for determining whether they are eligible to become panellist, according to the in- and exclusion criteria of the study. If participants are interested and eligible to become a panel member, they will be included in the screening process of the panel. After they pass the complete screening, panellists sign an employee contract from the independent coordinating institute that is in charge of the expert panel, together with a declaration of honour with respect to absence of conflict of interest (Appendix I. Declaration of honour of this annex)

#### 4.1.6 Selection of participants based on in and exclusion criteria

Inclusion criteria:

- Is fluent (preferably, a native speaker) in the language of the country where the panel is established [2, 3].
- Is between 18 and 55 years old at the day of the screening
- Is available to train 1.5 hour on two days a week, for the entire period of the training (estimated to last approximately 6 months)
- Is willing to join a whole-day maintenance meeting once a month
- Does not have problems with sense of smell, as indicated by a self-report
- Is in good health (as defined by a health questionnaire) (Appendix II. Health questionnaire of this annex)
- Does not use medication that influences smell
- Females: not pregnant or planning to get pregnant within the period of the panel training (i.e. 6 months) and are not breastfeeding at the moment of inclusion\*.

Exclusion criteria:

- Hay-fever
- Non-regulated or unstable diabetes
- Depression or other neurological or mental disorders

*\* In case of pregnancy at a later stage the panellist is replaced by one of the additional panellists during the time the panellist cannot perform test-sessions. After this period the panellist should first be retrained (during the maintenance sessions) until panellist performance is optimal.*

#### 4.1.7 Screening of participants on olfactory function, verbal and social skills.

The screening consists of a variety of tests (explained below) that need to be performed in a stepwise approach to end up with the eighteen best panellists. In the next parts of this paragraph, the tests to select panel members are described.

##### Motivation and availability interview

During this interview, the two panel leaders explain the aim and content of the panel training, panel maintenance and product testing. Motivation of the potential panellist should become clear and the participant should be available to train two times per week for a longer period.

##### Odour recognition test

Recognition skills are assessed with use of flavoured tobacco products. Ten tobacco products are offered to the participants in non-transparent glass bottles to eliminate visual cues. These ten tobacco products are Coresta monitor cigarettes (CM6) with different aromas - from electronic

cigarette liquids - added. The panel leader is free to choose odours that he/she thinks are useful to assess based on the odours that the panel leader expects to be frequently added to cigarettes or roll-your-own (e.g. vanilla, chocolate).

Participants smell the samples and indicate the odours they perceive from a multiple-choice list (check-all that apply CATA) either on paper or conducted by computer program EyeQuestion (Appendix III. Screening form tobacco odour recognition of this annex). Participants obtain points for correctly indicating the odour as described on the package of the electric cigarette filler package; in case of wrongly indicating an odour, points are subtracted.

The total score cannot go below zero, however as scores are subtracted in case participants indicate the wrong odour scores of 0 are expected. Participants who are among those with the highest scores and indicated at least 6 out of 10 odours correctly are selected.

#### Odour intensity test

Consistency or repeatability of rating products is an important quality that potential panellists should possess. Therefore, an odour intensity discrimination test will be repeated twice. The same stimuli will both times be given to the participants in the same order but with different sample codes. The participants rate the four samples according to their odour intensity on a line scale (100 mm visual analogue scale) with anchors going from 'weak' to 'strong' (at 10 and 90 mm) in duplicate. Rank order correlations will be used to quantify each panellist's repeatability. Panellists with the highest repeatability score and panellists that rate the product intensity in such a way that it corresponds with the rank order (i.e. the sample with the lowest concentration of the flavour/odour has been rated lowest on the line scale) are selected<sup>1</sup>. For the rank order, a minimum score of 5 out of 8 is needed for participants to be included in the next round.

#### Olfactory function

Olfactory function is assessed using the so-called 'Sniffin Sticks' developed by Hummel and colleagues. Screening document for this test can be found in Appendix IV. Screening forms sniffin stick test of this annex. The smell test contains three parts: an odour threshold test (THR), an odour discrimination (DIS) and an identification test (ID), which added together give a total TDI score.

To measure odour detection thresholds (THR), a standard series of pens with 16 dilutions of n-butanol will be used, starting from a 4% n-butanol solution. Three pens will be presented in a randomized order, of which one contains the odorant and two containing solvent. Then, participants have to identify the pen containing the odorant (three-alternative forced-choice method). The triplets will be presented at intervals of approximately 20-30 seconds, until the participant correctly identified the butanol-containing pen two times in a row at the same dilution level. This dilution is the starting point of the threshold measurements.

The odour discrimination (DIS) test consists of 16 triplets (two equal odorants and one different odorant). To prevent visual identification of the odorant-containing pens, participants will be blindfolded during the discrimination test. Before odour presentation, the cap of the pen will be removed by the investigator, and the pen's tip will be positioned approximately two centimetres in front of the participant's nostrils for  $\pm$  three seconds. The participant has to discriminate which of the three pens smells differently (three-alternative forced-choice method). The triplets will be presented at intervals of approximately 20 seconds in a randomized order (24, 27).

During the odour identification (ID) test, 16 common odours will be presented. Participants have to identify the odour using a multiple-choice task presented on a list of four different odorants (four-alternative forced-choice method). The different odours will be presented at intervals of

approximately 20-30 seconds (24, 27). The duration of the complete smell test is approximately 50 minutes.

Participants are selected if they pass the smell test, i.e. if they have a threshold of 9.0 or higher and at least score 36 TDI points (a threshold of 9.0 and a score of 36 is equal to the 50<sup>th</sup> percentile of the best performing age and gender group based on tests in 3000 people (Appendix V. Normative scores sniffin stick test of this annex) and are among the best performing panellists for this screening test<sup>1</sup>.

#### Verbal skills

During the verbal skill test, participants first need to come up with 10 odour descriptors they perceive in two tobacco products. Participants should be able to use odour descriptors that are not hedonic, but descriptive. Participants that are not able to come up with 10 odours are excluded.

#### Social skills

The social skills of the participants are assessed during a group discussion. During this discussion, the panel leader observes whether participants take part in discussions and are not too dominant. Participants that do not participate in the discussion or are too dominant are excluded.

### **4.1.8 Panellist selection**

Potential panellists declare any interests related to the panel activities. Based on the screening, and an evaluation of the interest declarations, b, the best eighteen participants are selected for the panel in a stepwise approach. These eighteen panellists are fully trained during the initial training period of at least half a year and will join the maintenance sessions. The test-sessions should be performed by the twelve best panellists, as twelve panel members is the optimum number of panellists as mentioned in literature[2]. Which panellists have the best performance will be determined during the performance test sessions scheduled before every test session. The additional 6 panellists are included in the panel to account for dropouts and for situations in which panellist are on holidays or sick on days of the test-sessions.

## **4.2 Panel training**

The aim of the panel training is to improve panel members' ability to identify odours in tobacco products and to rate them according to the intensity of the odour. Panel members have to be able to judge the products in an accurate, consistent and repeatable manner.

### **4.2.1 Determine training schedule**

Based on the availability of the panel members a training schedule should be made. All eighteen-panel members should train together in order to function as a panel. In general practice, training sessions are scheduled at least twice a week for the panel to become trained. Panel training is estimated to take approximately half a year. Once the panel is fully trained, a different training interval may be maintained (§4.2.8).

### **4.2.2 Instructions before the training**

Half an hour prior, the test panel members should not eat food with strong spices or herbs, candy, liquorice or gum. In addition, they should not brush their teeth or smoke a cigarette half an hour before the training/test. At the day of testing/training, panel members should not wear perfume, aftershave, deodorant with a persistent odour, or lipstick. Before the start of the test panel members have to wash their hands with neutral soap (unscented soap).

### 4.2.3 Attribute (generation) training sessions

Attributes are generated by the panellists and are therefore panel specific.

During the attribute training, panellists are given the task to determine, first individually and afterwards on a group level, which attributes discriminate between the samples (reversed grouping). During the first training, they are given five samples from the complete product set (the set of products that is used for determining the product space). The samples should represent a wide range of odour attributes (e.g. contain typical odours according to panel leader) to make the panellist feel comfortable in performing the tasks. During the attribute generation itself, the panel leader should choose the products in such a way that they are not all completely different. When generating the attributes, it is important to instruct the panellist to use descriptive words that refer to specific odour sources, such as tea, and avoid hedonic words such as 'nice'. Panellists are asked to exclude repeatedly one product from the group that is most different and write down based on which odour they think the product should be eliminated from the group. All individually generated words should be written down and should be grouped where possible. Words that seem to have the same meaning should be eliminated. This process takes approximately one hour. The second training has the same set up, but with different products that are more similar to one another compared with the first training. During the training panellists are also asked to come up with examples of references or to give a description of the odour. References or examples, such as foods, wood and hay are used to clarify the generated sensory attributes, this is mainly useful in case panellists are not in consensus or confused about the meaning of some sensory attributes. This type of session is repeated until all products are compared to ensure all potential attributes have been listed. The attributes should be generated in accordance with ISO standard 5492 [4] and ISO Standard 11035 [5].

### 4.2.4 Consensus training sessions

After the attribute (generation) training, the training focusses on shortening the list of attributes by asking if the panellists think some of the attributes have the same meaning or indicate the same odour; this can be done with use of references. The panellist should indicate the references that they find most appropriate for the attributes and preferably bring their own references. In general, the panel leader is non-directive in all discussions, however, in case no group consensus seems to be reached the panel leader should make decisions and ask the group whether they can agree with that. The main goal is to end up with an attribute list that covers all the differences among the samples but does not cause panellist fatigue. To minimize 'dumping' (i.e. rating an odour under a different descriptor due to lack of the correct descriptor or attribute) the attribute list will also have an "other" attribute where panellists can score the intensity of an odour that is not indicated on the attribute list.

### 4.2.5 Test training sessions

During the test training session, the panellists are shown how to use the computerized data acquisition system (EyeQuestion) and how to use the visual analogue scale (100 mm line scale) with anchors on 10 and 90 mm indicating 'weak' and 'strong' (Appendix VI. EyeQuestion example of this annex). Subsequently, the panellists receive a subset of the product set (in triplicate) that they have to judge based on the previously determined attribute list. From these data, the panellist performance can be checked and based on this information new training sessions should be focused on products and attributes that were difficult to assess (panel feedback training sessions). Once the panel performance is acceptable (§4.2.6 of this annex), the actual data collection starts.

#### 4.2.6 Monitoring panel performance

The panel(lists) should meet the quality standards: panel consensus, good discrimination ability and repeatability. The panel should be trained until there are no significant differences between ratings of different attributes, between panellists over sessions.

Throughout the training phase the panel members will be assessed on panel performance (§4.2.6 of this annex), this will also give additional information about the difficulty of rating specific tobacco odours. Panel performance should be monitored in accordance with ISO standard 11132 [6].

##### Performance of the panel

Performance of the panel is assessed per attribute by performing an ANOVA. The following model will be used:

$$Y = \mu + \text{product effect} + \text{panellist effect} + \text{session effect} + \text{product-panellist interaction} + \text{product-session interaction} + \text{panellist-session interaction} + \epsilon,$$

Where the intensity of the sensory attribute is the dependent variable Y, and the product, panellist and sessions are the independent variables.

For evaluating panel performance, the following parameters are of interested:

- If the product effect is significant (at a particular confidence level), it means that the intensity values of a specific attribute, evaluated by the panellists, are significantly different between products (at that level), therefore the panel has been able to discriminate the products with respect to that specific sensory attribute. (=Discrimination ability of the panel)
- If the product-panellist interaction is significant, it means that the intensity values of the attribute evaluated are significantly different depending on both, the type of product and the panellist, hence there is no consensus or agreement among panellists for that particular attribute. Panel members must have different perception of the products regarding that attribute.
- If the product-session interaction is significant, it means that the intensity values of the attribute evaluated are significantly different depending on the type of product, but also on the sessions when it was evaluated; therefore, the panel is not repeatable from one session to another.
- If the panellist effect is significant it means that the panellist does not use the scale identically; however, this is of less importance as we are interested in the relative difference between products. This is similar for a significant session effect, whether panellists use the scale consistently from one session to another.
- Panellist-session interaction indicates whether some panel members use the scale different from one session to the other. But this is also not of importance due to our interest in the relative differences between products [7].

To have a simple overview of the panel performance, a table will be built using the statistical program R (FactoMineR and SensoMineR packages). The dependent variables (intensity of sensory attributes) will be presented in the first column, and the independent variables (product, panellist, sessions, and their interactions) will be in the first row. The table values will represent the p values for each attribute at any specific effect or interaction. In this table the cells that are below significance level (0.05) are coloured to have a good overview of the panel performance [7].

##### Performance of the panellists

The individual performance of each panellist per tobacco product compared to the overall panel scores can be compared. To compare individual panellist's results to one another and the panel

mean, the same strategy as mentioned for panel performance can be applied. However, when interested in the performance of an individual panellist, an ANOVA model per panellist should be obtained. After that, results per panellist for all attributes separately for the product effects are obtained and this indicates the global performance (discriminability) of each panellist. To assess the consensus of one panel member with the panel as a whole, the product effect of the panellist can be compared with that of the complete panel. Repeatability of the panel member can be assessed by comparing the standard deviation of the residuals of the ANOVA panellist model with the residuals of the attributes (one by one). Over the course of the panel training phase and the panel maintenance meetings the panel leaders should keep track of the panellists performance of each individual panellists [7].

#### **4.2.7 Panel feedback**

The training sessions following the performance training sessions should focus on the attributes and products that were:

- Not used to differentiate between the products
- Not assessed in consensus among the panellists.
- Not assessed in a repeatable manner by the panellists

During these training sessions, the panellists practice differentiating products for these attributes. If the panel is not able to differentiate products or come to consensus on a certain attribute the panel needs to discuss a better name or reference for this odour, if this remains difficult elimination of the attribute should be discussed. When the panel is in consensus of eliminating the attribute, it can be excluded from the attribute list.

#### **4.2.8 Panel maintenance, robustness and sustainability control**

To ensure sustainability of a panel in terms of attendance over a long period of time the following guidelines are recommended once a panel is fully trained. Once a month all eighteen panellists join a maintenance meeting for a whole day. During this meeting, there should be four sessions of 1.5 hour, during which the panel is trained actively. The panel should have enough breaks in-between for the panellists to remain focused.

- During the first 1.5 hour sessions panel(lists) performance will be assessed, based on this ad-hoc decisions need to be made by the panel leaders on which attributes and which products the focus should be during the following 3 sessions. During the following three sessions the panel performs tasks such as ranking or rating the products after which panel discussion takes place to come to consensus on the ranking/rating of the products.
- During holidays, maintenance meetings can be skipped, however additional training might be necessary in the month thereafter (depending on the panel performance) but this is left to the panel leader to decide.
- If 12 or more panellists perform well, the maintenance meeting may also be used to re-assess the products of the reference space to update the space.

## 4.3 Test phase

### 4.3.1 Setting up the reference space

Once panel performance is satisfactory, maintenance meetings are scheduled once a month. The first two months after the panel training can be used to assess the products in order to establish the reference space. Two reference spaces should be determined one for cigarettes and one for roll your own tobacco (§5.9 of Annex III: WP3 - MAIN document). During one meeting, 4 sessions of 1.5 hour can take place. The first session should be dedicated to assess panel performance. During first two months after the panel is trained, it is thought that panel performance is still optimal. Therefore, preferably, all eighteen panellists (but at least 12 panellists) perform the assessment of the products to set up the reference space. The head panel leader will guide these sessions. The panellists, of which the performance is not sufficient, will join 3 sessions or re-training during which they can practice assessing products and rating attributes they do not perform well on. The vice panel leader will be in responsible for these training sessions.

During a session of 1.5 hour 4 products in triplicate can be tested, during one meeting 12 products can be assessed by the panel. Therefore, 7 maintenance meetings are needed to assess 50 cigarettes and 25 roll-your own tobaccos. To speed up the process of setting up the reference space one maintenance meeting can be extended by 2.5 day to be able to have the 7 maintenance meetings during 2 months. During this *extension* only the panellists that had sufficient panel performance will stay to assess the products, those whose performance was not sufficient can come back the next month. The following month the process of selecting only the panellists whose performance is sufficient starts again.

### 4.3.2 Test sessions

The monthly maintenance meetings can be (partly) replaced by test sessions in case products need to be tested on characterising odours.

In case of a test-session, one or more of the 1.5-hour sessions of the maintenance meeting can be replaced by a test session. However, panel performance needs to be assessed before the test session to verify whether it meets the standard (§4.2.6 of this annex). Therefore, one performance test-session of 1.5 hour needs to be held prior to the test session to see if panel performance is sufficient.

- During the performance test-session of 1.5 hour, the panel assesses reference products (in the same way as a real test session), where after the panel leader checks panel(lists) performance.
- In case panel performance does not meet the standard, maintenance sessions should be held, during which the panel is re-trained until panel performance is sufficient (assessed through another performance test-session). Only when panel performance is sufficient test sessions can be performed by the panellists.
- In case more than 6 products need to be tested during one month, the meeting can be extended with one or two days.

### 4.3.3 Materials

- Tobacco products (cigarettes, roll-your-own)
- Glass test bottles (non-transparent)
- Storage boxes
- Refrigerator
- Weighing scale (in mg)
- Rubber gloves
- Dishwasher
- Odourless soap
- Tin foil and labels

Equipment must be selected to ensure that there is no transfer of volatiles between these (e.g. containers) and the tobacco. The tobacco should be kept in closed packages and rolled in tin foil labelled with the date of the purchase. The package should be kept in closed containers in a refrigerator. The panel leader should wear gloves when transferring the tobacco from the cigarettes or from the roll-your-own package into the test bottles and take new gloves for every new tobacco brand. The test bottles should be prepared a day before the test or training session and kept in the fridge. The test bottles should be taken out of the fridge at least 4 hours before the start of the test session.

### 4.3.4 Test room

The test room must be free from odours free (neutral) and easy to clean and should have a natural colour, white/light grey. The room must be equipped with appropriate air ventilation and temperature must be controlled. The room should contain sensory booths each with a computer screen. In addition, the room should be quiet for panellists to stay focused [8]. The test room should be in accordance with ISO standard 8589 [9].

### 4.3.5 Number of samples

The number of samples that can be assessed within one session is dependent on the panellist fatigue when judging the products. Experience has shown that it is feasible to test twelve tobacco products within 1.5 hours given that panellists have at least 30 seconds of resting/neutralizing between samples and have a 5-minute break after evaluating five samples.

### 4.3.6 Sample selection

Samples used for the training should be chosen based on the performance of the panel; meaning that the products or attributes that the panel finds difficult to assess should be used more often compared to the brands that the panel is able to assess well.

For the sensory test sessions, three fresh packs per brand are needed (1 additional for the chemical analysis and 6 reserve). For each test session a new package should be used and the 3 samples tested of a similar brand should be from 3 different packages to encounter difference in the products coming from different packages.

### 4.3.7 Sample preparation

The test samples should be prepared a day before the test session (Annex III: WP3 - MAIN document). The paper of the cigarette is removed and the tobacco is put in test bottles (semi-transparent brown glass bottles with a plastic lid). Glass bottles are preferred as they do not take up the odour of the tobacco and are easy to clean. However, glass bottles are usually semi-transparent. In these cases, the bottles should be covered (e.g. with tinfoil) to cover the tobacco inside completely. In case of roll-your-own products, the amount of tobacco used is 0.7 gram, which is in line with the average weight of cigarettes.



#### 4.3.8 Randomization

The order in which the panellists receive the samples should be randomized for the actual test session. However, with the randomization both order and carry-over effect should be taken into consideration. Meaning, that each sample is tested the same number of time at each position, and is following each of the other samples equally. However, for the performance test sessions the samples should not be randomized as to determine whether panellists find it difficult for example to assess a product after a menthol product. In that case the menthol product should always come last or the neutralization period between samples should be longer.

#### 4.3.9 Test set-up- product assessment

The sensory test sessions are performed in sensory booths under standardized conditions as described in §4.3.10 and conducted by the computer program EyeQuestion (Logic8, version 3.7.6).

In the sensory booths, a glass of water and tissues are placed for the panellist to use when needed. Three-digit coded tobacco test samples are offered to the panellists in a blinded manner (i.e. semi-transparent odourless dark-glass test bottles). Panellists can re-smell the sample as often as preferred and assess the attributes at their own pace. Panellists receive the next sample after evaluating all attributes for the previous sample. After evaluating all attributes, there will be an open question. This allows the possibility to indicate the intensity of an odour that is not included in the original attribute list.

Between samples, there is a 30-second break and panellists are instructed to 'neutralize' their sense of smell by smelling their own arm/clothing. After the assessment of six tobacco products, there is a 5 minutes break in which participants can play a short computer game to regain their focus before evaluating the remaining six products. The duration of a test session should be no longer than 1.5 hour, and no more than twelve tobacco products can be tested within this time frame.

#### 4.3.10 Test procedure

The assessment must be carried out in triplicate with the same panellists under the same conditions in order to designate a product as having, or not- having a characterising odour. During one 1.5-hour test session, 4 products (of which 2 should be reference products) can be tested in triplicate. As one meeting consist maximum of three 1.5-hour sessions (originally maintenance sessions) 6 new products can be tested. The two reference products (per 1.5-hour session) serve as a control measure, to control for possible deviations from the place these products originally have in the product space. In case the products are significantly different from the products in the reference space:

1. Panel performance should be checked after the test-sessions (just as before the test-session).
2. Panel drifting (see section 0) may have occurred, meaning that the panel has become more sensitive due to increase in experience and therefore uses the scale in a different manner. This could, though highly unlikely, lead to reference products that are re-tested having characterising odours compared to the reference products as tested during the initial set up of the reference space. If this occurs, all products of the reference space should be measured again to adapt the reference space according the sensitivity of the panel.

Half an hour preceding the test panel members should not eat food with strong spices or herbs, candy, liquorice or gum. In addition, they should not brush their teeth or smoke a cigarette half an

hour before the training/test. At the day of testing/training panel members should not put on perfume, aftershave, deodorant with a persistent odour, or lipstick. Before the start of the test panel members have to wash their hands with neutral soap (unscented soap).

#### **4.3.11 Product profiling**

All panellists smell and evaluate the tobacco product(s) submitted for examination, marking the intensity of all odour attributes, as generated by the panel, in randomized order. Intensity is marked on a 100 mm line scale (visual analogue scale) with opposite terms at 10 and 90 mm (strong and weak). An example of the profiling display as provided by EyeQuestion can be found in Appendix VI. EyeQuestion example.

## 5 Statistics

### 5.1 Data processing procedure

The profiling data collected by EyeQuestion is examined by the panel leader. The three replicates will be combined and used for statistical analysis. The panel leader will feed the data into the statistical software programme R (FactoMineR and SensoMineR packages). Principal component analysis will be performed to visualize the distance between the products based on the odour profile and variance explained of each product. Consequently, hierarchical cluster analysis will be performed to calculate groups of products that cluster together to determine the reference space. Furthermore, 99% confidence ellipses will be calculated around the mean odour profile for each product. To test whether a product is significantly different from the reference product group the “tested” product can be plotted into the reference space (i.e. PC space of products not considered to have a characterising flavour). The statistical analysis will be explained in more detail in §5.2 in this annex. For reasoning behind the significance level see §5.9 in Annex III: WP3 - MAIN document.

### 5.2 Data analysis

#### Multi-Analysis of variance

When the panel is fully trained, attributes that differentiate the products are identified with use of a two-way mixed model analysis of variance performed for every single attribute (similar to the panel performance statistics) provided the following assumption is met:

- The dependent variable  $Y_s$  (the attributes) should be normally distributed.

Normal distribution should be checked with use of ANOVA by looking at whether the residuals are normal distributed ( $N(0, \sigma^2)$ ).

In case the attributes are not normally distributed they should be log-transformed.

Mixed model analysis is performed as odour attributes are rated on a continuous scale, but the other two variables (i.e. panellist and session) are not continuous. The odour attribute is the dependent variable, independent variables in the model are the product, panellist and session variables. The following statistical model is used:

$$Y = \mu + \text{product effect} + \text{panellist effect} + \text{session effect} + \text{product-panellist interaction} + \text{product-session interaction} + \text{panellist-session interaction} + \epsilon$$

In this model the product effect is fixed and the panellist and sessions are random.

For panel performance evaluation, the interaction between panellist and session and product variables in their effect on the intensity ratings of the sensory attribute is of most interest.

Once the panel has been trained, there should be no significant interaction between panellist and session. Because of that, it can be assessed whether the product variable significantly influences the intensity of sensory attributes evaluated. To test the assumption of the panel being well trained, the other variables (panellist and session) are evaluated and their interaction to verify that the panel has consensus and is repeatable before the start of the test session, by doing a test-training session. If there is no interaction between panellist and session the test session can be performed.

For example, to explain the attribute menthol with respect to the type of product, the question that should be answered is; 'Is menthol an attribute that panellists can distinguish between products?' If the effect of this variable (product effect) is statistically significant, this means that the menthol attribute is significantly different among products, and therefore it plays an important role in defining the sensory profile of each product.

The statistical model should be applied to all the dependent variables (intensity of sensory attributes) to point out the variables that are the most characteristic according to the set of products in its whole, and to each of the products in particular. To do that, the function *decat* (description of categories) of the *SensoMineR* package in R can be used. This function is designed to test the main effect of a variable (product) and the significance for a set of dependent variables (intense of sensory attributes) at a given ANOVA model.

Once the sensory attributes that differentiate the products are identified, the products that are specific for those attributes (e.g. which products are differentiated based on menthol) can be defined. In that case, answer is given to the following question 'for the sensory attribute menthol, which product can be considered significantly different from some kind of average products (all other products of your sensory test)?' This can be done with a simple T-test, not comparing two products but 1 product comparing with the average product. In this case, it would be comparing 1 'new' product in a sensory space defined by products from which is assumed that they do not impart a characteristic flavour or odour.

When the T-test is significantly different ( $p \leq 0.01$ , 99% CI) for products we can conclude that these products are different from the average product (all the other products in the product space) regarding the sensory attribute menthol (having either less or more menthol flavour/odour compared to the average).

Again with use of the *decat* function we can determine the sensory attributes that are the most characteristic of a set of products as well as for product by product

Principal component analysis and product/reference space

Principal component analysis (PCA) is applied when products are described by quantitative variables. PCA calculates the distance between the products and the distance between variables by means of Variance. With this analysis the correlation coefficient (or covariance) are calculated over all individuals. PCA depicts a product map or space visualising all products in a 2D subspace reflecting as well as possible (showing the optimum amount of variance in the data) the differences among the products. Before PCA is applied the data needs to be centred by using the estimated means for each variable of each product.

$$\forall i,j, x_{ij} \leftarrow x_{ij} - x_j$$

Where  $x_{ij}$  is replaced by  $x_{ij} - \bar{x}_j$ , where  $\bar{x}_j$  is the mean of variable j calculated over the n individuals.

Furthermore, for each variable to applied with similar weight, the data has to be scaled to unit variance.

$$\forall i,j, x_{ij} \leftarrow \frac{x_{ij} - \bar{x}_j}{s_j}$$

Where  $x_{ij} - \bar{x}_j$  is divided by  $s_j$ , where  $s_j$  is the standard deviation of variable j calculated over the n individuals.

Inertia determines the distance between variables. Inertia is a general notion of variance in case of multiple variables or some kind of 3D standard deviation, meaning that Inertia of  $N_i$  is equal to the sum of the variances over the variables.

$$I(N_i) = \sum \text{Var}(j)$$

Principal component analysis is used to obtain an overview of products depicting differences and similarities in case of multivariate data. The main objective of PCA is to summarize the information contained in a multivariate data set into graphical representation of individuals and variables. PCA represents the data in a low dimensional subspace (usually two dimensions) depicting the variance in the data in the best possible manner. This is done by using vector space formed of vectors of  $J$  variables (the attributes) and  $N$  products. PCA produces a map, called product space, in which the similarities and differences between products are depicted in a low-dimensional space rather than the multivariate space. If the products are close in this product map, it means the products are similar, and in the opposite, if products are far apart it means products are different. Thus, PCA simply changes the frame of reference of a high-dimensional space into a low-dimensional space

Principal component analysis is used to obtain two product spaces, one for cigarettes ( $n=50$ ) and one for roll your own tobacco ( $n=25$ ). A product space is a map based on the principal components obtained from the analysis, called the principal dimensions ( $x,y$ ) of the product space.

#### Hierarchical clustering

The distances between the products obtained by performing PCA can be biased, as products may seem close to each other (and so similar) while actually they are far apart. This may happen in case the difference between the products is not only explained by the first two principal components, but for example by the third as well. Hierarchical ascendant classification calculates the distance between products, whereas PCA only visualizes this. To indicate whether products are similar or different, hierarchical clustering on principal components (HCPC) is used. HCPC takes into account all principal components. The HCPC requires the analyst to set a number of clusters by clicking on a hierarchical tree or dendrogram. The number of clusters chosen should be based on common sense and is subjective. HCPC does suggest a natural cut of the hierarchical tree, based on the ratio of variance within clusters, and the total variance of the scatter plot. The analysis aims at balancing the definition of clusters that are as homogenous as possible (with similar individuals), and as different as possible from each other. The variance is of course not only dependent of the variance of the products, but also of the variance due to the panellists. It is therefore important that the panel performance is optimal before performing final tests. Once the number of clusters is chosen, the individuals can be plotted in a product space, with a colour code depending on their cluster. Furthermore, for each cluster a list of significant attributes can be obtained. The list of significant attributes is sorted according to the sign of the difference between the average score of that cluster, the overall mean for each attribute, and the significance of the p-value of the test that compares the mean over the individuals of the cluster to the overall mean. If a cluster does not indicate any sensory attributes, it means that the cluster is not characterised by any sensory attribute, and that that cluster is perceived as some kind of average product. This can be seen as the reference group that does not imparts a characterising odour. When depicted on the product space, this cluster of products will lie close to the inner centre of the product space.

#### Reference space

50 cigarettes brands and 25 roll your own tobaccos need to be assessed and principal component analysis need to be done to obtain the product space (see main document §5.9). Based on the product space, a cluster of products that are similar to each other should be determined by means of hierarchical cluster analysis in combination with the ellipse analysis. This cluster of products will then

be plotted (with use of PCA, see previous page) in a new map called the reference space (as it consists of reference products).

It may be possible that a product comes on the market that contains a 'new' flavour, and this flavour did not come out as significant attribute before or has been not been listed as attribute during the panel training, and is thus not part of the product space. If the test product contains an odour that was not an original attribute (i.e. is not part of the reference group), the product may automatically fall into the 'characterising' category, since it is different from the reference space products. However, as the odour should also be clearly noticeable and this is dependent of the intensity relatively to reference products, it should be checked whether more products on the market constitute this odour. In case more products contain this odour the reference space should be updated (see recommendations 5.3.1.7.2, or §7 of Annex III: WP3 - MAIN document).

#### Confidence ellipses

Virtual panels are generated using Bootstrap techniques (random sampling to estimate properties of the estimator) based on the variance in the data in order to display confidence ellipses around products. When calculating the confidence ellipses we choose to repeat the experiment 100 times (99% confidence interval or ellipses in this case) with a panel size of 12 [7].

From this analysis we obtain a product map (similar to PCA) where each product is displayed, showing with a square the original product place (i.e. the result from the panel, without the virtual panel bootstrapping technique) in the PCA map. Furthermore, each product is circled by its 99% confidence ellipse generated by the virtual panel. If the circles have no overlap it means that the products are significantly different. By performing multi factor analysis we can also obtain the associated p-values.

#### Hotelling's T-test

Hotelling's T-test or Hotelling's t-squared is a multivariate counterpart of Student's T-test and is based on the Hotelling's T<sup>2</sup> distribution. The Hotelling's T-test, tests the differences between the (multivariate) means of different populations, or in this case different products. We will use this test to see whether the product to be tested significantly differs from our reference group, taken into account all attributes.

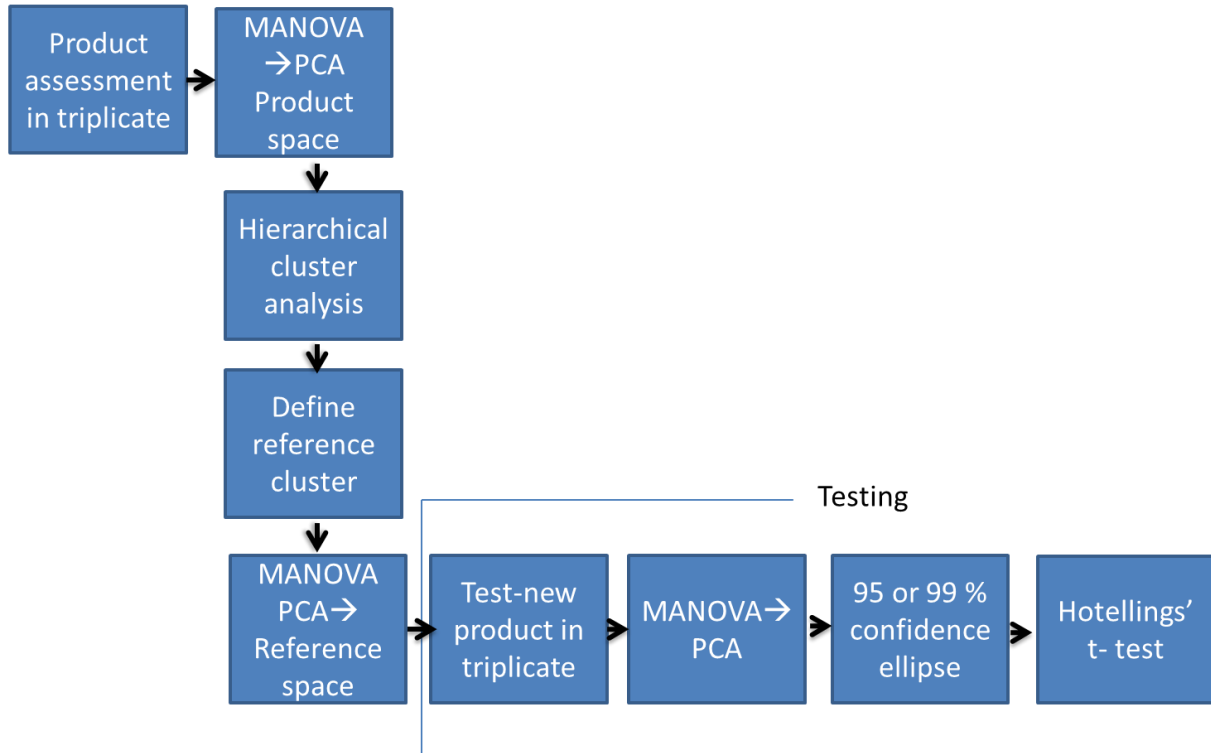
#### Determination of whether or not a product imparts a characteristic flavour

When a new product needs to be tested (or one that was among the previous sixty to eighty products that were needed to set up the reference space), the new product can be plotted into the reference space and determine with use of the Hotelling's T-test whether this product is in overall significantly different from the reference group and based on which variables the product is different. If the product is significantly different from the reference group based on one or more attributes (odours or flavours) the product is defined as having a characterising flavour.

Once it has been determined with Hotelling's T-test whether a product is significantly different from the products in the reference space overall, it should be determined whether the product is different based on *one* odour attribute. The reason for this is that products may be slightly different on every odour attribute, which when added up makes the product significantly different; however, the product should have at least one clearly noticeable characterising odour. Therefore, besides the overall Hotelling's T-test, the product should be significantly different on at least one single sensory odour attribute. This can be done with LSD post-hoc test (least significant difference, Bonferroni) assessing the odour attributes per product one by one (see Case study section of the results document). If the attribute is rated significantly higher in a test product than the mean ratings of the

reference space, it can be concluded that this is a 'clearly noticeable smell' (i.e. more intense than a 'noticeable' smell).

In this way the method is able to 1) determine characterising odours in products that are 2) 'clearly noticeable' as stated in the definition of 'a product with a characterising odour'.



**Figure 26. Statistical approach to 1) determine the reference space and 2) to determine whether a product imparts a characterising odour**

## 6 References

1. ISO, ISO standard 8586-1 Sensory analysis-General guidance for the selection, training and monitoring of assessors. .
2. Lawless, H. and H. Hildegarde, Sensory Evaluation of Food principles and practices. Second edition ed. 2010, New York, Dordrecht, Heidelberg, London: Springer.
3. Chapter 15 “Geographic, cross-cultural, and individual variation in human olfaction” Smell and Taste in Health and Disease.
4. ISO, ISO standard 5492-Sensory analysis- Vocabulary
5. ISO, ISO standard 11035-Sensory analysis- Identification and selection of descriptors for establishing a sensory profile by multidimensional approach. .
6. ISO, ISO standard 11132-Sensory analysis-Guidelines for monitoring the performance of a quantitative sensory panel. . 2013.
7. Le, S. and T. Worch, Analyzing Sensory Data with R, ed. C. Press. 2014, Croydon (UK): Taylor & Francis Group,LLC.
8. Lawless, H.T. and H. Heymann, Sensory evaluation of food principals and practices. 1998, New York: Chapman and Hall.
9. ISO ISO standard 8589-Sensory analysis-general guidance for the design of test rooms. . 2007.



## 7 Appendix I. Declaration of honour

### Declaration

of absence of conflict of interests and of confidentiality

I the undersigned, ..... , declare that there is no conflict of interest for the impartial and objective exercise of the above.

I am neither currently involved, nor have been in the past five years, in activities in the field of tobacco which could be subject to a conflict of interests.

I understand there is a conflict of interest where the impartial and objective exercise of the review is compromised for reasons involving family, emotional life, political or national affinity, economic interest or any other shared interests in the particular field of tobacco products. Such a conflict of interests could arise in particular as a result of economic, professional, or personal interests (such as collaboration with the tobacco industry, memberships in boards, consultancies, employment, payments, ownership of investments/shares etc.), political or national affinities, family or emotional ties, or any other relevant connection or shared interest;

I confirm that, if I discover during the time I am part of the tobacco-panel, that such a conflict exists, I will declare it immediately.

I also confirm that I will personally carry out any task assigned to me. I will neither communicate nor disseminate any confidential information that is revealed to me or that I have discovered. I will not make any adverse use of information given to me.

Signed: .....

Done at.....

Date...../...../.....

## 8 Appendix II. Health questionnaire

This questionnaire is used to assess the medical fitness before participation in the tobacco expert panel.

In case you want to participate in the above-mentioned study, we ask you to fill in this questionnaire according to the truth with a pen (not with pencil). This information will be kept private and, if needed, will only be discussed with a medical examiner or nurse.

### 1. Personal information

#### Participant

Male/Female

Surname: \_\_\_\_\_

First name: \_\_\_\_\_

Street and house number: \_\_\_\_\_

Zip code: |\_\_|\_\_|\_\_|\_\_| |\_\_|\_\_|

City or town: \_\_\_\_\_

Place of birth: \_\_\_\_\_

Telephone: \_\_\_\_\_(private) \_\_\_\_\_(mobile)

### 2. General

E-mail address: \_\_\_\_\_

#### 2.1 What is your gender?

Male

Female

#### 2.2 What is your date of birth? |\_\_|\_\_|-|\_\_|\_\_|-|\_\_|\_\_|\_\_|\_\_| (dd-mm-yyyy)

#### 2.3 Do you speak XXX as a mother tongue?

yes

no

**2.4 Are you willing to take part in the expert panel and did you consider the time schedule as described in the information brochure?**

yes

no

**2.5 Are you currently taking part in another panel?**

yes                                      please specify: \_\_\_\_\_

no

### **3 Smoking**

**3.1 Do you smoke?**

no, I was never a smoker

no, I quit smoking more than 3 months ago

yes,  
I smoke cigarettes/cigars/pipe or combination.

---

*If you smoke, please answer the following 3 questions*

**3.2 What is your standard brand of cigarettes/roll-your-own/pipe tobacco?**

---

**3.3 For how many years have you been smoking?**

- Shorter than 5 years
- Longer than 5 years
- Between 5-10 years
- Longer than 10 years

**3.4 How many cigarettes do you smoke on average?**

- Less than 1 a day
- 1-5 a day
- 6-10 a day
- 11-15 a day
- More than 15 a day

**4 Senses**

**Do you have problems with:**

- |     | yes                      | no                       |  |
|-----|--------------------------|--------------------------|--|
| 4.1 | <input type="checkbox"/> | <input type="checkbox"/> | Sight (other than glasses or contact lenses) |
| 4.2 | <input type="checkbox"/> | <input type="checkbox"/> | Hearing (exclusive hearing aid)              |
| 4.3 | <input type="checkbox"/> | <input type="checkbox"/> | Smell  |

Did you answer yes on one of the above questions than please explain below:

**5 Allergy**

**5.1 Do you have hay-fever?**

- yes
- no

**If yes, please indicate below the last period of symptoms**

the last time I had symptoms was \_\_\_\_: years ago

**6 Chronic diseases**

**6.1 Do you have any neurological or mental diseases/disorders (including depression)?**

- yes    namely: \_\_\_\_\_
- no

**6.2 Do you have a chronic disease that affects your daily life**

- yes    namely: \_\_\_\_\_
- no

**6.3 Do you have diabetes**

- yes
  - blood glucose levels are stable
  - blood glucose levels are not stable
- no

**7 Medication**

**7.1 Do you use medicine or have you used medicine on a daily basis in the last 3 months?**

- no
- yes, namely?
  - a. name medicine \_\_\_\_\_
  - how often            \_\_\_\_\_
  - how many            \_\_\_\_\_
  - for what             \_\_\_\_\_
  - b. name medicine \_\_\_\_\_
  - how often            \_\_\_\_\_
  - how many            \_\_\_\_\_
  - for what             \_\_\_\_\_
  - c. name medicine \_\_\_\_\_
  - how often            \_\_\_\_\_
  - how many            \_\_\_\_\_
  - for what             \_\_\_\_\_

**8 The next questions only apply to woman**

**8.1 Are you pregnant at the moment?**

- yes  
For \_\_\_months.  
 no

**8.2 Are you planning to get pregnant within the next six months ?**

- yes  
 no  
 I do not want or are not able to give answer to this question

**8.3 Are you currently giving breastfeeding?**

- yes  
 no

**9. To be signed by the participant**

The undersigned declares that he/she filled in this form truthfully

Signature: \_\_\_\_\_

Date (dd-mm-yyyy): |\_|\_|-|\_|\_|-|\_|\_|\_|\_|

## 9 Appendix III. Screening form tobacco odour recognition

### CHECK ALL THAT APPLY

#### Instructions:

- For each question it is stated with sample you should assess, please make sure the sample number matches the number on the questionnaire
- During this test you will receive 10 samples which you will assess one by one.
- For every sample you indicate on a list which odours you perceive, you can indicate multiple odours per sample
- You are allowed to smell the sample as often as you like to be able to assess the odour correctly
- This test will take approximately 20 minutes
- For this test you will receive points when you indicate the correct odour(s) however points will be subtracted if you indicate the wrong odour.



Kruis alle geuren aan die u waarneemt als u aan het sample ruikt.

- |                                      |                                       |  |
|--------------------------------------|---------------------------------------|--|
| <input type="checkbox"/> Cassis      | <input type="checkbox"/> Perzik       | <input type="checkbox"/> Vanille         |
| <input type="checkbox"/> Koffie      | <input type="checkbox"/> Thee         | <input type="checkbox"/> Bessen          |
| <input type="checkbox"/> Aardbei     | <input type="checkbox"/> Hooi         | <input type="checkbox"/> Drop            |
| <input type="checkbox"/> Kaneel      | <input type="checkbox"/> Mint/menthol | <input type="checkbox"/> Banaan          |
| <input type="checkbox"/> Kersen      | <input type="checkbox"/> Almond       | <input type="checkbox"/> Chocolade       |
| <input type="checkbox"/> Bloemen     | <input type="checkbox"/> Tabak        | <input type="checkbox"/> Chocolade/Cacao |
| <input type="checkbox"/> Sinaasappel | <input type="checkbox"/> Rook-geur    | <input type="checkbox"/> Noten           |
| <input type="checkbox"/> Anijs       | <input type="checkbox"/> Snoep        | <input type="checkbox"/> Gras            |
| <input type="checkbox"/> Kruidnagel  | <input type="checkbox"/> Appel        | <input type="checkbox"/> Citroen/Limoen  |
| <input type="checkbox"/> Caramel     | <input type="checkbox"/> Honing       | <input type="checkbox"/> Toffee          |
| <input type="checkbox"/> Kokos(noot) | <input type="checkbox"/> Mango        | <input type="checkbox"/> Peer            |

**Volgende >**

© Logic8 BV 2001-2014

**Figure 27. Eye-question CATA- task screen**

Above the odour list it is stated that the participant should indicate all odours as perceived in this sample. Participants can press next (volgende) to go to the next sample.

## 10 Appendix IV. Screening forms sniffin stick test

Screening document

Screening No.

Subject No.



# 1. "Sniffin stick": Threshold test

General instructions:

- 20-30 seconds interval between sets, Red 16=lowest, R1=highest
- Use pen 1 to familiarize the subject with the butanol odour.
- The participants should put on the sleeping mask before starting with the test
- Clean hands with odourless soap
- Tell the participant:
  - You receive 3 odour pens in 1 round. Of which 1 has an odour. In some cases this odour is in a low concentration while in others it might be stronger. --> let the participant smell pen 1 red to indicate that they need to indicate that odour when smelling a set of pens.
  - Breath normally
  - Do not move your head
  - concentrate
  - tell me if you would like to pause
- Start with 16 and skip 1 set in-between (16,14,12) until they correctly identify the butanol odour 2 times in a row at the same level. The threshold is the average row number of the last 4 turning points.
- Take 3-5 minutes break in between tests

X: Correctly identified - : not correctly identified

++ - ++ - ++ -

Dilution step	1								
	2								
	3	↑	↓	↑	↓	↑	↓	↑	↓
	4								
	5								
	6								
	7								
	8								
	9								
	10								
	11								
	12								
	13								
	14								
	15								
	16		↓	↑	↓	↑	↓	↑	↓

Score:

Total score within normal range:

Yes

No

## 2. "Sniffin stick": differentiation test

### Green is different

*Please circle the answer given by the participant, afterwards indicate if correct.*

SET	ORDER			CORRECT ( tick, if correct, when green is identified)
1.	RED	GREEN	BLUE	<input type="checkbox"/>
2.	BLUE	RED	GREEN	<input type="checkbox"/>
3.	GREEN	BLUE	RED	<input type="checkbox"/>
4.	RED	GREEN	BLUE	<input type="checkbox"/>
5.	BLUE	RED	GREEN	<input type="checkbox"/>
6.	GREEN	BLUE	RED	<input type="checkbox"/>
7.	RED	GREEN	BLUE	<input type="checkbox"/>
8.	BLUE	RED	GREEN	<input type="checkbox"/>
9.	GREEN	BLUE	RED	<input type="checkbox"/>
10.	RED	GREEN	BLUE	<input type="checkbox"/>
11.	BLUE	RED	GREEN	<input type="checkbox"/>
12.	GREEN	BLUE	RED	<input type="checkbox"/>
13.	RED	GREEN	BLUE	<input type="checkbox"/>
14.	BLUE	RED	GREEN	<input type="checkbox"/>
15.	GREEN	BLUE	RED	<input type="checkbox"/>
16.	RED	GREEN	BLUE	<input type="checkbox"/>

Score:

Total score within normal range:

- Yes
- No

**3-5 min break**

### 3. "Sniffin stick": identification test

Please circle the right answer, afterwards indicate if correct.

STICK No.	ANSWER				CORRECT ( tick, if correct, when grey is identified)
	1	2	3	4	
1.	Orange	Blackberry	Strawberry	Pineapple	<input type="checkbox"/>
2.	Smoke	Glue	Leather	Grass	<input type="checkbox"/>
3.	Honey	Vanilla	Chocolate	Cinnamon	<input type="checkbox"/>
4.	Chive	Peppermint	Fir	onion	<input type="checkbox"/>
5.	Coconut	Banana	Walnut	Cherry	<input type="checkbox"/>
6.	Peach	Apple	Lemon	Grapefruit	<input type="checkbox"/>
7.	Liquorice	Cherry	Spearmint	Cookies	<input type="checkbox"/>
8.	Mustard	Gum	Menthol	Turpentine	<input type="checkbox"/>
9.	Onion	Sauerkraut	Garlic	Carrots	<input type="checkbox"/>
10.	Cigarettes	Coffee	Wine	Smoke	<input type="checkbox"/>
11.	Melon	Peach	Orange	Apple	<input type="checkbox"/>
12.	Cloves	Peppermint	Cinnamon	Mustard	<input type="checkbox"/>
13.	Pear	Plum	Peach	Pineapple	<input type="checkbox"/>
14.	Camomile	Raspberry	Rose	Cherry	<input type="checkbox"/>
15.	Anise	Rum	Honey	Fir	<input type="checkbox"/>
16.	Bread	Fish	Cheese	Ham	<input type="checkbox"/>

--	--	--	--	--	--

**Score:**

**Total score within**  **normal range:**

- Yes
- No

## 11 Appendix V. Normative scores sniffin stick test

**Table 2** Descriptive statistics of normative values obtained in healthy subjects

	Female subjects				Male subjects			
	THR	DIS	ID	TDI	THR	DIS	ID	TDI
<b>Age group A 5–15 years</b>								
<i>N</i>	25	25	59	25	17	17	51	17
Mean	6.59	12.32	11.75	30.67	7.22	11.71	12.41	30.87
SD	2.23	1.70	1.77	3.60	2.59	1.57	1.77	4.79
Minimum	2.75	10	6	24.50	4.00	9	8	23.00
Maximum	13.50	16	15	36.50	12.00	14	16	40.00
Percentiles								
5	3.13	10	8	24.58	4.00	9	8	23.00
10	4.30	10	9	24.90	4.00	9.8	10	23.80
25	5.00	11	11	27.75	5.00	10	12	27.88
50	6.00	12	12	31.00	6.75	12	13	31.50
75	8.00	14	13	33.88	9.00	13	14	32.88
90	9.35	15	14	35.60	12.00	14	14	40.00
95	12.30	15.7	14	36.28	12.00	14	15	40.00
<b>Age group B 16–35 years</b>								
<i>N</i>	760	741	827	704	579	587	672	552
Mean	9.39	12.91	13.68	36.06	9.24	12.61	13.48	35.31
SD	2.56	1.92	1.62	4.17	2.99	1.95	1.73	4.73
Minimum	1.75	5	8	23.00	1.00	5	6	18.00
Maximum	16.00	16	16	46.75	16.00	16	16	47.00
Percentiles								
5	5.51	9	11	29.50	4.75	9	10	27.91
10	6.50	10	11	30.50	6.00	10	11	29.50
25	7.50	12	13	33.50	7.00	11	12	32.00
50	9.00	13	14	36.00	8.75	13	14	35.00
75	11.25	14	15	39.00	11.50	14	15	38.60
90	12.50	15	16	41.50	13.75	15	16	41.50
95	14.00	16	16	43.00	14.80	15	16	43.00
<b>Age group C 36–55 years</b>								
<i>N</i>	295	291	586	288	208	207	491	207
Mean	9.08	12.46	13.49	35.16	8.43	11.94	13.10	33.20
SD	3.09	1.96	1.56	4.52	3.47	2.24	1.88	6.05
Minimum	1.00	6	4	22.50	1.00	5	4	15.00
Maximum	16.00	16	16	45.75	16.00	16	16	44.25
Percentiles								
5	4.25	9	11	26.86	2.75	7	10	20.60
10	5.50	10	12	28.75	3.75	9	11	24.95
25	6.75	11	13	32.50	6.25	10	12	29.50
50	8.75	13	14	35.50	8.50	12	13	34.50
75	11.00	14	15	38.00	10.50	14	14	37.50
90	13.60	15	15	40.50	13.02	15	15	39.55
95	15.30	15	16	42.89	14.91	15	16	42.48
<b>Age group D &gt;55 years</b>								
<i>N</i>	147	143	251	143	142	139	238	139
Mean	7.44	10.66	12.06	29.83	7.15	10.69	12.20	29.81
SD	3.51	2.50	2.31	6.77	3.59	2.77	2.57	7.17
Minimum	1.00	4	4	11.00	1.00	4	3	9.00
Maximum	16.00	16	16	43.00	16.00	16	16	44.00
Percentiles								
5	1.55	5.2	7	17.25	1.04	5	6.95	14.50
10	2.75	7.4	9	19.05	2.25	7	9	19.75
25	5.50	9	11	25.75	4.44	9	11	26.25
50	7.25	11	12	30.50	7.50	11	13	31.00
75	9.00	13	14	34.25	9.25	13	14	34.50
90	12.60	13.6	14.8	37.65	11.68	14	15	37.75
95	14.70	14	15	40.20	14.35	15	16	40.50

Reference: Normative data for the “Sniffin Sticks” including tests of odour identification, odour discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. T. Hummel · G. Kobal · H. Gudziol · A. Mackay-Sim

## 12 Appendix VI. EyeQuestion example

How intense do you perceive the odour **hay** in this sample

English: Weak

Strong

zwak

Sterk

--	--	--

zwak

Sterk

--	--	--

How intense do you perceive the odour **menthol** in this sample

English: Weak

Strong

*All attributes are listed on one page, panel members do not have to follow the order the way the attributes are listed. Panellists are instructed to first rate the attributes that are most intense when they open the bottle, thereafter they can rate the other attributes*

## **Annex V: WP3 – Standard Operating Procedure 2 (SOP 2)**

### **CHEMICAL ANALYTICAL METHOD FOR THE ASSESSMENT OF CHARACTERISING ODOURS IN TOBACCO PRODUCTS**

#### **1 Purpose and scope**

All products that are sensory analysed by the advisory expert panel, will be chemically analysed. Chemical analysis is performed to complement the results of the sensory panel by identifying the odour components causing the odour and to build an odour library. Chemical analysis will also be used to determine whether the characterising flavour or odour is derived from a compound or combination of compounds that is not present in cured tobacco leaves (Burley, Virginia, Oriental), and therefore may be considered an additive.

Flavoured as well as unflavoured tobacco products and tobacco leaves are chemically analysed. Components are identified based on the probability of the mass spectral library. This involves a certain uncertainty (to identify components with a 100 per cent certainty, standards must be used for confirmation). The components that are present in flavoured tobacco products but absent in the cured tobacco leaves are considered additives. However, it should be considered that not every additive is an odour/flavour component. The Leffingwell flavour database can be used to determine whether an additive is an odour/flavour component. Chemical analysis is performed with headspace GC-MS, since this method is sensitive, reproducible, specific, sustainable, and not labour-intensive.

The method as described is applicable to cigarettes and roll-your-own tobacco only. For legislation purposes the method should be performed by an independent institute designated by the European Commission.

#### **2 General**

This standard operating procedure (SOP) is number two of the two SOPs that are part of the test approach to detect characterising flavours in tobacco products. This SOP describes the steps for chemical analysis to identify components or additives that contribute to a characterising odour. Standard operating procedure number one describes the steps to detect characterising odours in tobacco products.

Paragraph 3 of this document describes all steps necessary to identify the chemical components that contribute to characterising odours in tobacco products. The main document describes the general procedure and procedures for identification of flavours/odours in more detail and justifies the steps and choices made to come to the standard operating procedure.

#### **3 Terms and definitions**

##### **Flavoured product**

Tobacco product suspected to have a characterising flavour.

##### **Unflavoured product**

Tobacco product not suspected to have a characterising flavour

##### **Odour components**

Chemical compounds that contribute to the characterising odour of a tobacco product.

## 4 Standard Operation Procedure

### 4.1 Preferred qualities chemical analyst

General:

- A bachelor's degree in chemical analysis or related field (biochemistry, analytical chemistry)
- Working knowledge of GC-MS with headspace
- At least a few year(s) related work experience (GC-MS).
- Working understanding of chemical analysis and research, and ability to analyse and interpret GC-MS data.

Competences:

- Planning skills
- Good communication skills
- Analytical and problem solving
- Accurate
- Quality-oriented

### 4.2 Requirements of the lab facility

- Fully equipped testing or calibration laboratory capable of performing tests or calibrations, with access to a GCMS-headspace.
- Performing under ISO 17025

#### 1. Method summary

- 1.1 Tobacco is isolated and pulverised to homogenise the samples. Approximately 200 mg are weighed accurately and measured with headspace GC-MS.
- 1.2 The method is a headspace (HS) gas chromatographic (GC) method using a HP-5ms Ultra Inert capillary column (30 m x 0.25 mm ID x 0.25  $\mu$ m). A Single Quadrupole MS is used for mass spectrometry (MS). Approximately 200 milligrams of homogenised tobacco in a headspace vial is incubated for 30 minutes in an Agitator oven at a temperature of 140°C. A volume of 1 ml of the formed vapour is injected on the GC column using a 1.5 ml-HS syringe with a temperature of 140°C.
- 1.3 HS-GC-MS analysis results in chromatograms, from which the peaks are identified using a mass spectral library.

#### 2. Safety and environmental precautions

- 2.1 Follow routine safety and environmental precautions as in dealing with any chemical laboratory activity.
- 2.2 The testing and evaluation of certain products against this test method may require the use of materials and or equipment that could potentially be hazardous or harmful to the environment. This document does not purport to address all the safety aspects associated with its use. Anyone using this test method has the responsibility to consult with the appropriate authorities and to establish health and safety practices as well as environmental precautions in conjunction with any existing applicable regulatory requirements prior to its use.

#### 3. Apparatus, equipment, and software

Common laboratory apparatus are used, in particular:

- 3.1 Analytical balance capable of measuring to at least four decimal places.
- 3.2 A Gas Chromatograph – Mass Spectrometer with headspace injection possibility.
- 3.3 A HP-5ms Ultra Inert capillary column, 30 m x 0.25 mm ID x 0.25  $\mu$ m (or equivalent).



- 3.4 The following software programs for component identification:
  - 3.4.1 Flavours and Fragrances of Natural and Synthetic Compounds (FFNSC) library
  - 3.4.2 NIST Mass Spectral Search Program
  - 3.4.3 Automated Mass Spectral Deconvolution and Identification System (AMDIS)
- 3.5 Leffingwell Flavour Base 9 – Tobacco Version

#### 4. Materials

- 4.1 Cigarettes and roll-your-own tobacco products advertised as being flavoured.
- 4.2 Cigarettes and roll-your-own tobacco products advertised as being unflavoured.
- 4.3 Tobacco leaves.
- 4.4 Mortar and pestle.
- 4.5 20 mL headspace vials.
- 4.6 Suitable closable cups for each tobacco type separately.

#### 5. Preparation of tobacco samples

- 5.1 Isolate the tobacco filling from the cigarettes of one package or use at least 2 grams of roll-your-own tobacco and tobacco leaves.
- 5.2 Pulverise the tobacco of each product using a mortar and pestle and collect the tobacco of each package in a separate suitable cup.
- 5.3 Shake the cups for at least 10 seconds for homogenisation of the tobacco sample.
- 5.4 Accurately weigh 200 mg of the homogenised test sample into a 20 mL headspace vial.
- 5.5 Analyse the headspace vials on the GC-MS.

#### 6. Sample analysis

This method utilises headspace GC-MS analysis to identify chemical components that contribute to a characterising odour in tobacco products.

##### 6.1 GC Operating Conditions: example

GC Column: HP-5ms Ultra Inert, 30 m x 0.25 mm ID x 0.25 µm (or equivalent)

Carrier Gas: Helium at a flow rate of 1.5 ml/min

Injection volume: 1 mL

Injection mode: Split 1:20

##### Temperature Program

Syringe temperature: 140°C

Injector temperature: 250°C

Agitator temperature: 140°C

Agitator time: 30 minutes

GC oven start temperature: 50°

Heating rate: 10°C/second to 200°C hold for 5 minutes

Run time: 20 minutes

##### 6.2 MS Operating Conditions: example

Transfer line temperature: 280°C

Ionisation mode: Electrospray ionisation (EI)

Ionisation voltage: 70 eV

Ion source temperature: 230°C

Quadrupole temperature: 150°C

Solvent delay: 3 min

Mode: Full scan from m/z 40 to 500

### 6.3 Expected retention times

- 6.3.1 Differences in temperatures, gas flow, rate, age of the column, etc. may alter retention time.
- 6.3.2 Retention times must be verified before analysis is begun.
- 6.3.3 For the conditions described here, the expected sequence of examples of important peaks in tobacco samples with a characterising flavour will be propylene glycol, menthone, l-menthol, menthyl acetate, nicotine, vanillin, ethyl vanillin.

## 7. Data analysis

- 7.1 Perform three agitation runs of the tobacco samples under identical conditions.
- 7.2 Use AMDIS software to generate reports form each measurement in which all chemical components present are listed. Make sure that the AMDIS software uses the FFNSC and the NIST library to generate the reports.
- 7.3 Exclude chemical components with a signal to noise ratio below 55 and a 'weighted' or 'reverse' score below 70.
- 7.4 Determine which odour components are present in flavoured tobacco products and absent in cured tobacco leaves and by removing the components present in cured tobacco leaves from the component lists of flavoured tobacco products. The remaining components are considered additives. Many additives will be flavour additives that contribute to the characterising odour of a flavoured tobacco product.
- 7.5 Evaluate the remaining chemical components using the Leffingwell Flavour Base 9 – Tobacco Version to determine the flavour of each component separately.

## 8. Statistical analysis

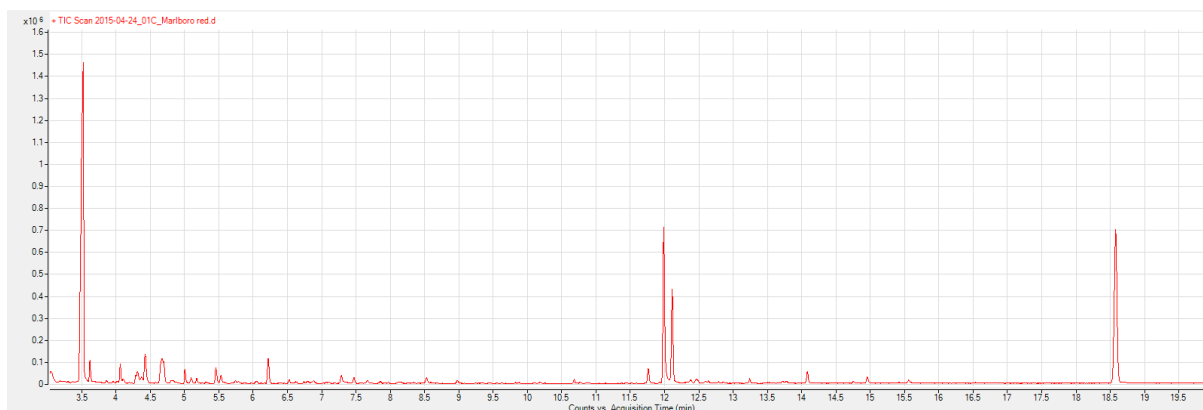
Perform Principal Component Analysis (PCA) on the signal-to-noise ratio of the peaks remaining after step 8.4. The PCA product space will depict which chemical components are specific for which products and how the products relate to each other (i.e. similar products will be more close together).

## 9. Data reporting

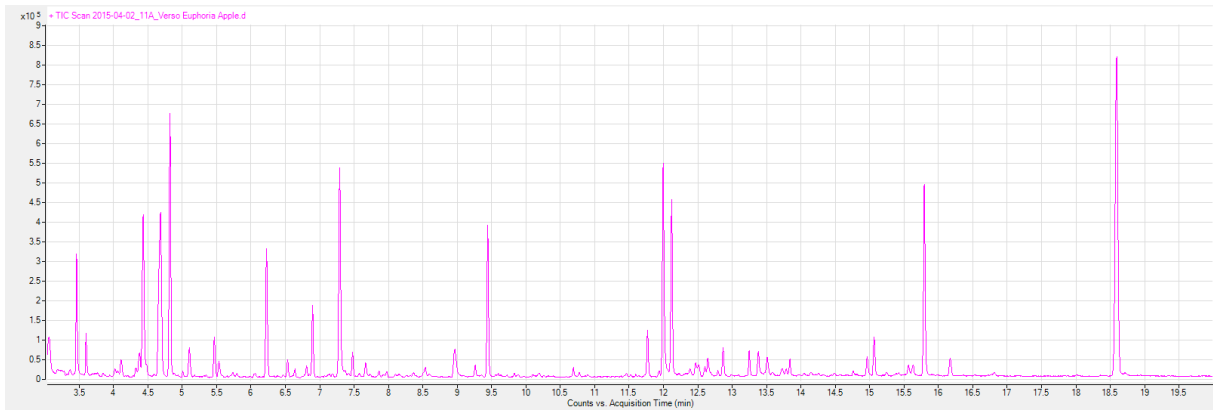
- 9.1 Report the odour additives that are present in the flavoured tobacco products separately (and absent in the cured tobacco leaves), combined with the odour percept they produce according to the Leffingwell flavour database.
- 9.2 Report the results of the Principal Component Analysis.

## 10. Appendix 1: Example of a chromatogram

Two chromatograms are shown from a tobacco product advertised as being unflavoured (1) and from a product advertised as being flavoured (2).



1. Chromatogram of Cig-RP1 tobacco.



2. Chromatogram of Cig-TP9 Apple tobacco.

## Annex VI. WP4: Questionnaire sent to industry

### 1 Questionnaire on costs aspects involved in the analysis of tobacco products

The new EU Tobacco Product Directive (TPD) prohibits the placing on the market of cigarettes and roll-your-own tobacco with a characterising flavour other than tobacco. The Consumers, Health and Food Executive Agency (CHAFAEA) has contracted the HETOC Consortium to develop a method based on sensory profiling and chemical-analytical measurements to decide whether a tobacco product imparts a characterising flavour other than tobacco.

To gather information on various types of costs involved in sensory analysis the attached questionnaire is addressed to you to ask for your expert knowledge/input in this field. The questionnaire involves several aspects regarding sensory analysis of tobacco products such as costs involved in setting up, training, operating and maintaining a sensory panel.

We kindly ask you to complete the questionnaire. Any information you provide will be used exclusively for the purpose for which it is collected and aims at enabling evidence-based regulation. The information can be shared with the European Commission and will be treated in a confidential manner.

We are looking forward to receiving your input at your earliest convenience, but not later than **13 May 2015**. Please return your questionnaire by replying to HETOC@rivm.nl. This email address is made available only for this purpose.

Thank you for your time. Kind regards, the HETOC Consortium.

National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands  
 Wageningen University (WUR), Wageningen, the Netherlands  
 OP&P Product Research (OP&P), Utrecht, the Netherlands

*This questionnaire was produced under the Health Programme (2008-13) in the frame of service contract n° 2014 62 02 Concerning the Request for Specific Services N° Chafea/2014/health/19 for the implementation of Framework Contract N° EAHC/2013/Health/23 – Concerning mapping of best practices and development of testing methods and procedures for identification of characterising flavours in tobacco products with the Consumers, Health and Food Executive Agency (Chafea) acting under the mandate from the European Commission. The content of this report represents the views of HETOC CONSORTIUM and is its sole responsibility; it can in no way be taken to reflect the views of the European Commission and/or Chafea or any other body of the European Union. The European Commission and/or Chafea do not guarantee the accuracy of the data included in this report, nor do they accept responsibility for any use made by third parties thereof.*

Contact details of the person who completed this questionnaire
Country

Name:
Function:
Name of company:
Address:
Telephone:
Email:

**1 General questions regarding sensory analysis of tobacco products carried out by your company**

1.1 Does your company use <b>sensory panels for product assessment</b> or have you used such a panel in the last five years, including outside the EU?
Answer:

*Please describe in general terms the sensory panel(s) used, in particular the reason for setting it/them up, but also the main ways of operation and associated costs. If multiple types of panels (e.g. expert and consumer panels, smoking v. smelling, laboratory environment v. home testing etc.) are/were used by your company in the last five years and the panels were not created for an adhoc/short term need, please describe each of the main types identified and reply to the subsequent questions separately for all main types identified. Please submit each type using a separate sheet/questionnaire.*

*Please address the following aspects in the description of the panel:*

1.2 <b>Type of panel</b> (experts/ trained consumers/untrained consumers):
Answer:
1.3 <b>Product assessment</b>

- Does the panel assess products through smoking the product and/or smelling the product?
- In case of smelling: Does the panel assess products through smelling the burnt or unburnt tobacco? Please explain the underlying reasons for the approach.

Answer:

**1.4 Purpose of the panel** (please explain the underlying reasons for choosing this panel type for this particular purpose)

Answer:

**1.5 Operation of the panel:**

- Location of testing (which country?)
- Operator of the panel (e.g. your company, a dedicated subsidiary of your company, a third party) and the main reasons for your choice. If only part of the work is outsourced, explain what and why. Please indicate to whom you have outsourced (short description of the company)

Answer:

**1.6 Panel composition** (e.g. number of panellists, age, gender, nationality)

Answer:

**1.7 Screening:**

- What type(s) of test(s) are used for screening the sensory capacities of potential panel members?
- Required qualification of panel members

Answer:

**1.8 Training needs:**

- What kind of training(s) is used before a panel member is allowed to participate in the sensory testing? Please describe including duration and main steps/methods.
- Do you use any standardized training methods and if so which and why?

Answer:

**1.9 Panel leader(s)**

- Number of panel leader(s)
- What are the qualification requirements of the panel leader(s)?
- Does the panel leader follow a specific training? Do you use any standardized training methods and if so which and why?

Answer:

**1.10 Standard methodology applied (e.g. ISO-standards)**

- Does the panel perform the sensory research under ISO norms or under other quality norms?  
(ISO: please name ISO type / other quality norms: please indicate which)
- Have you developed internal standards? (if yes, please send a copy of the main documentation)

Answer:

<p><b>1.11 Quality assurance</b></p> <ul style="list-style-type: none"> <li>- How is the quality of the output generated by the panel maintained over time?</li> <li>- How is the performance of panel members assessed?</li> </ul>
<p>Answer:</p>
<p><b>1.12 Setting of testing:</b> Laboratory environment v. home testing et alia. If testing is carried out in a laboratory please explain the laboratory setup (e.g. dedicated smoking lab, internal/external lab)</p>
<p>Answer:</p>
<p><b>1.13 Duration of the panel</b> (For how many years is the panel established?)</p>
<p>Answer:</p>
<p><b>1.14 Methodology</b>, including data collection and analysis</p> <ul style="list-style-type: none"> <li>- Are sensory attributes defined beforehand or generated by the panel?</li> <li>- Which tool/method is being used to generate data / which scales are being used?</li> <li>- Does the panel assess products once or are products tested in duplo or triplo?</li> <li>- Which type of statistical analysis is performed on the data generated by the panel (e.g. model, procedure, software package)?</li> </ul>
<p>Answer:</p>



<p><b>1.15 Workload/Frequency of meetings</b> (How frequently is the panel consulted? How many samples, on average, are assessed by the panel on a yearly basis?)</p>
<p>Answer:</p>
<p><b>1.16 Human resources</b> – Number, qualifications and role of personnel involved in setting up and running the panel (e.g. administration, statistician etc.)</p>
<p>Answer:</p>
<p><b>1.17 Maintenance of panel over time</b></p> <ul style="list-style-type: none"> <li>- What are the mechanisms to replace panel members? (please explain)</li> <li>- In which way is loyalty of panel members maintained (e.g. routine panel activities, extracurricular activities)? (Please specify and define frequency of such activities)</li> </ul>
<p>Answer:</p>
<p><b>1.18 Do you use a comparison between products</b> in your assessment? If yes, what type of product(s) do you choose as a reference? Please specify your choice of the reference product(s)</p>

Answer:

*The subsequent questions should be answered for each panel type that was identified in the reply to section 1*

## 2 Questions on the procedure and costs for setting-up a sensory panel

2.1 What are the costs involved in selecting and training the panel (before it starts being operational)?

Please specify main types of costs (e.g. panel leader, panel member, other staff, other) and amounts in €

Answer:

2.2 What are the other costs involved in setting up the panel (before it starts being operational)?

Please specify main types of costs and amounts in €

Answer:

2.3 What are the **total costs** involved in setting up the panel (CAPEX)?

Please indicate main types of costs and amounts in €

Answer:

### 3 Questions regarding maintaining the panel and testing products by sensory analysis

3.1 What are the costs involved for maintaining/running the panel (after it has become operational) – please limit to staff (i.e. panel leader, panel members, administrative, statisticians, other)?

Please specify main types of costs and amounts in €, ideally on an annual basis (if that is not readily available, please provide information you have and estimates which could form the basis of extrapolations to annual costs). Please distinguish between fixed (e.g. general maintenance of panel) and variable (related to the testing of products) costs.

Answer:

3.2 What are the other costs for maintaining/running the panel (e.g. overheads, materials, renting of facilities)?

Please specify main types of costs and amounts in €, ideally on an annual basis (see above). Please distinguish between fixed (e.g. general maintenance of panel) and variable (related to the testing of products) costs.

Answer:

3.3 What are the total costs for maintaining/running the panel?

Please indicate amounts in €, ideally on an annual basis (see above). Please distinguish between OPEX with and without depreciation. Please explain any difference between the total amount and the sum of the amounts listed under 4.1 and 4.2 (apart from depreciation). Please distinguish between fixed (e.g. general maintenance of panel) and variable (related to the testing of products) costs.

Answer:

3.4 How many samples are tested by the panel on a yearly basis?
Answer:

**Thank you for your time.**

## **Annex VII. Minutes of seminar**

### **1 Minutes of characterising flavours in tobacco products seminar**

**Jan. 12-13 2015, RIVM, Bilthoven, the Netherlands**

*Seminar of independent experts to discuss the three concept profiles for testing the presence of characterising flavours in tobacco.*

Contract n° 2014 62 02

Concerning the Request for Specific Services N° Chafea/2014/health/19 for the implementation of Framework Contract N° EAHC/2013/Health/23 –

Concerning mapping of best practices and development of testing methods and procedures for identification of characterising flavours in tobacco products

Meeting room: V134

National Institute for Public Health and the Environment (RIVM)/Centre for Health Protection (GZB)

Antonie van Leeuwenhoeklaan 9

3721 MA Bilthoven, the Netherlands

#### **Deliverable of meeting:**

Input of the participants on the three most suitable profiles for testing characterising flavours in tobacco.

#### **Introduction to the seminar**

At the start of the seminar several introductory talks were given by both DG SANTE and HETOC members introducing the background and aim of the project.

A short summary was given on the new Tobacco Product Directive (TPD) and the aim of this project to develop a defensible and solid procedure to determine cigarettes and roll-your-own tobacco *with* characterising flavours before and/or during consumption. The project aims and conclusions of the literature review on the most suitable methods (WP1) were summarized, as well as the transfer of this knowledge into the development of the basic approaches as described in WP2. A brief introduction was given to the three different concept profiles, which consist of the following approaches: expert panel, consumer panel, and chemical analysis. The different possibilities to combine these approaches in different orders were mentioned, as well as key decisions to be made on the panel composition, product assessment and type of analytical method. In WP3 pilot testing is conducted to obtain information about the procedure, logistics and requirements of the final method.

It was mentioned that the literature review and concept profiles have been sent to the tobacco industry. Comments given by the industry on WP1 and WP2 were presented and were to be considered by the participants.

#### **Discussion to identify choices to be made for WP3 (pilot testing)**

At the first day of the seminar, a round table discussion was held about the concept profiles as presented in WP2 and decisions to be made for WP3. The different options were briefly introduced by three HETOC members and three external speakers.

At the second day the discussion continued using set discussion points, taking into account all arguments mentioned during the brainstorm at the first day of the seminar. Informal voting was performed at this point to quantify the views of all participants of the seminar and to help in defining choices for WP3.

**Discussion point 1— ‘what are non-characterising flavours in tobacco products: How to determine a baseline or reference to define the product space’**

The definition of characterising flavours according to the TPD, "a clearly noticeable flavour, other than tobacco, resulting from an additive or a combination of additives, including but not limited to fruit, spices, herbs, alcohol, candy, menthol or vanilla, which is noticeable before or during the consumption of tobacco products", is not providing a quantitative threshold for the distinction between products with and without characterising flavours.

A test (authenticity test) was proposed by the moderator, in which smokers are asked to assess a product for differences to their own tobacco product. Seminar participants felt that this method would bring along difficulties in panel selection and products and might also be too sensitive for the current aim. The method proposed would not be appropriate for differentiating a characterising flavour due to the large variety in baseline products. It was concluded that for the aim of the project it is crucial to define the reference product space.

Four different options for reference products were proposed:

1. products stated to be additive free;
2. natural tobacco;
3. ask large group of consumers: ‘Is this normal tobacco or flavoured?’;
4. around 10 products with the largest market share / representing the European market that have no **suspected** characterising flavour (statistical clustering will be performed to determine ‘product space’).

The different options for reference products were discussed. It was mentioned that because of its harshness completely natural tobacco is rarely smoked by consumers. Several types of natural tobacco exist that can differ between batches. It was discussed that a baseline with commercial tobacco products not suspected to have a characterising flavour would probably lead to a higher cut off point compared to natural tobacco. This is because such commercial tobacco products in general already contain many flavours. When comparing products to be tested to a reference group consisting of these commercial products, we can expect that fewer products would be considered different from such a reference (whereas more products to be tested would be considered different from a reference consisting of natural tobacco).

Participants agreed that the reference space should contain different tobacco blends. It is not expected that these products will differ largely, but this remains to be determined. Instead of taking the top 10 market share products, a different option could be to ask the tobacco industry to define a top 10. Regulators think that given tobacco industry’s interests, such a list could only be considered as advisory. It was mentioned that batch differences within products do not have to be taken into account since the tobacco industry minimizes these in order to keep taste constant.

As roll-your-own and cigarettes are different product types, separate reference spaces should be established. Also the fact that the TPD refers to products with characterising flavour noticeable before or during the consumption needs to be taken into consideration when determining the product spaces. Separate product spaces would need to be determined for both smelling and smoking tobacco products.

**Voting:** *unanimous agreement for point 4, because it is found to be the best representation of the European market.*

**Discussion point 2 — ‘The role of an expert versus a consumer panel: Discuss training and implementation, added value and (dis) advantages, cultural and legislative validity’**

The method must be defensible, meaning that the procedure should be based on uniform rules, produce reliable, and reproducible/repeatable and robust results and it must lead to binary outcome (yes or no). During the discussion, it was agreed that adults, not adolescents, should participate in the sensory panel. The nationality of panel members is not expected to be a critical parameter. Relative product differences are expected to be the same, although different attributes may be used to describe the product depending on the panel’s cultural background.

On the second day, it was discussed whether consumer or expert testing would be most suitable to the purpose, and whether QDA (quantitative descriptive analysis) or CATA (check all that apply) was the preferred method. The main point of discussion was the advantages and disadvantages of expert- vs consumer panels and in particular if there would be differences in the results provided by different panels. In this context, it was concluded (also taking into account the results from a preliminary smelling experiment carried out among the participants at the start of the meeting) that consumers can detect differences in flavour of different cigarettes and often even correctly identify the flavour. Consumer panels can differentiate between products. Although the assessments of individual panellists have a larger variability than those of expert panellists, this is compensated for by the large number of consumers that assess the products. Regarding the exact amount of participants in a consumer panel, it was mentioned that this depends on the reference space and the amount of tobacco products to be analysed. A few hundred consumers in a panel should be sufficient for this purpose and participants should be randomly selected for each testing session. As regards a consumer panel, it was also emphasized that the detection of flavours per se and the ability to differentiate between products is not sufficient to take decisions on tobacco products with characterising flavours. Also the intensity of the flavour must be determined. It was mentioned that the proportion of consumers detecting a flavour is related to the intensity of a flavour, but there was no consensus that this would be an appropriate proxy for assessing intensity.

An expert panel on the other hand is extensively trained and therefore more consistent, meaning that the assessments of panellists have a smaller intra- and inter-observer variability. Furthermore, an expert panel could be trained to be able to quantify the strength of particular flavours, e.g. via the quantitative descriptive analysis (QDA) method, which is relevant for the regulatory decision at stake here, i.e. to decide whether a product has a characterising flavour or not. To perform expert panel tests, ISO norms for reproducibility should be met. This requires an estimated 6 months for panel training, and the panel needs to operate on a regular basis in order to be sustainable.

It remained undecided whether an expert panel or a consumer panel would give more robust results for court cases.

Due to time constraints within the project/pilot to fully train an expert panel, it was discussed whether some preparatory work, such as establishing a preliminary product space could be done by consumers and that aspects regarding setting up an expert panel could be described as part of the project.

**Vote:** *the majority of the participants (12 to 5) voted for an expert panel and QDA. Some participants were undecided and mentioned that both panels could ‘do the job’ and therefore financial, time, legal and political aspects also should be considered.*

**Discussion point 3— ‘Smelling and/or smoking of tobacco products’**

The TPD refers to tobacco products with a characterising flavour present before and/or during consumption of the tobacco product.

Unburnt tobacco smelling experiments can be performed using non-smokers and/or smokers, as they have negligible health effects, are easy to execute, and therefore even adolescents could be included. Smelling side stream smoke, on the other hand, poses health risks to the participants, as ETS exposure is not safe. However, flavours resulting from combustion processes cannot be assessed by smelling.

The relationship between the perceptibility of added flavours in the unburnt product and the smoke of the burnt product is not known. Unburnt tobacco can possibly have a flavour that is not detectable in burnt form, and *vice versa*. Some flavours, e.g. resulting from combustion, or from inventions such as microcapsules in tobacco products, may only be released upon burning the tobacco. When a characterising flavour is not smelled it may possibly still be detected by smoking the product.

It was mentioned that for smelling and smoking, a different reference space would need to be defined, and therefore two different experiments would need to be carried out.

As smelling experiments are practically and ethically easier to perform, they are deemed to be the most sensible first step in product assessment. Smoking protocols would also need to be tested by performing pilot experiments. Important practical conditions such as (provisional) smoking facilities and approval of medical ethical committee have been arranged by the HETOC consortium.

***Vote:*** unanimous agreement on first smelling unburnt tobacco product. When a characterising flavour is not detected, proceed as appropriate to assessment through smoking.

**Discussion point 4 — ‘Statistics’**

The data will be analysed by statistical methods in order to help define boundaries for the definition of a characterising flavour. An introduction was followed by a discussion on how to draw statistic conclusions and define cut-offs from QDA or CATA data. Several statistical methods were described, and it was mentioned that to test for differences, ANOVA is preferred. To check for similarity, TOST is often used (two-one sided tests).

A threshold value for defining a characterising flavour is not determined in the TPD. It is stated that a characterising flavour is a flavour ‘clearly noticeable’. All experts agree that perception and identification of a flavour is dependent on its intensity and concentration. These factors are related; the probability to detect a flavour is increased when the concentration of the flavour in the product is higher.

Participants agreed that products falling outside the reference space are not necessarily defined as having a characterising flavour; this rather depends on the difference between the reference space, and the tested products, as well as how the threshold is set. Depending on the test method used, it will not result in a ‘yes or no’ answer, but a distance to the reference value. The participants discussed how to define whether a product has a characterising flavour. In, the observed sample result that is used for testing a hypothesis is usually using a significance level of 5% (corresponding to a 95% confidence interval), but stricter/less strict significance levels could be chosen. As for other



test methods, just-noticeable differences (JND) may be used, defined as a value at which 50% of a population can detect a difference; so clearly noticeable should be more than 50% of panel members that should identify a characterising flavour in the product.

It was discussed that it is difficult to define a threshold value, since no practical experience/data exist on when a flavour should be considered "clearly noticeable" in line with the definition used in the Tobacco Products Directive. Some participants mention that it is important to consider what would legally hold during a court case. First, pilot experiments to define the reference point need to be performed to obtain information on the conditions and the outcomes of the sensory task. Following this data a decision on which test and statistical method to use for analysis of the data should be taken. It is preferable to choose an approach that has been used in court cases. Information on the variability of the reference space will define a boundary or threshold.

***No voting performed:*** test and statistical method to be used and cut-off will be decided based on results of pilot testing or baseline product calculations and additional arguments given by legislators.

#### **Discussion point 5 — ‘Chemical analysis’**

Headspace GC-MS is discussed as a fast, simple, and accurate technique with which volatile compounds can be analysed, and therefore a suitable technique to measure odours in tobacco that will be perceived when smelling tobacco.

Direct solvent extraction GC-MS allows for detection of non-volatile flavours. This technique could be performed in addition or in combination with headspace GC-MS. A combination with LC-MS-MS techniques was also suggested. It was discussed that headspace GC-MS can very well be used to generate a characteristic fingerprint of flavours present in a product. Quantification of the flavour components may be difficult in some cases, since some flavours are a complex mixture of multiple components including flavour enhancers.

Participants agree that a combination of sensory and chemical analysis has added value, as it allows for identification and quantification of flavours, which is also useful in defining the reference space and the flavours present in products in this space. Chemical analysis can help in defining if an additive or combination of additives is responsible for a characterising flavour. The presence of most flavours will be relatively easy to demonstrate using chemical analysis, although flavours mimicking natural tobacco components (e.g. hay or tobacco flavour) will be more difficult to demonstrate. Complex mixtures of multiple components can be defined as a characteristic fingerprint; this type of analysis is relatively easy to perform in combination with sensory analysis.

***Vote:*** agreement to start with headspace GC-MS as a first screening tool to analyse flavour components in products and ‘reference products’ and being complementary to sensory analysis.

#### **Final conclusions**

- As a baseline a representation of the European market – about 10 brands - consisting of different tobacco blends should be analysed
- The majority of seminar participants was in favour of expert testing, but agreed that the development of a fully trained expert panel was not feasible within the current timeframe of the pilot study.
- Smoking or smelling: first smelling, followed as appropriate by smoking when a characterising flavour is not detected.

- Test method/Statistics: Different methods are available and depend also on the data which is generated. Whatever approach is chosen, a statistical method for testing is available and a cut-off can be determined.
- The type of test/statistical method that ultimately will be used, needs to be decided based on the results of the pilot experiment
- Check whether court examples exist in which evidence based on sensory research has been used.
- Chemical analysis: headspace GC-MS is considered the most efficient method, but this is not an exclusive decision, as other methods are also suitable.

**Proposal for pilot testing by HETOC:**

- The conclusions of the seminar should be considered in launching the pilot, but not every goal discussed during the seminar can be achieved in the 3 months available for pilot testing.
- According to the technical offer, a minimum of 5 products need to be tested. This is feasible, but it does not generate a sufficient amount of data to result in a method already sufficiently robust for product regulation.
- To gain realistic data for a representative set of cigarettes and attributes, decisions are to be made on:
  - how many brands are needed to define a product space;
  - how many and which type of attributes need to be selected;
  - how many and which type of persons need to be in the panel;
  - what type of data analysis needs to be performed?

It is not possible to fully train an expert panel nor do smoking experiments before summer 2015, as this will take at least 6 months. However, alternative options such as testing based on a semi-trained panel or experiments with consumers had been proposed by the consortium and would be further discussed with the Commission following the seminar.

## Annex VIII. Medical Ethical Committee application

### Time line

<i>Start writing proposal:</i>	<i>8 September 2014</i>
<i>First submission:</i>	<i>6 October 2014</i>
<i>First comments METC:</i>	<i>22 October 2014</i>
<i>Second submission:</i>	<i>7 November 2014</i>
<i>Second comments METC:</i>	<i>28 November 2014</i>
<i>Third submission:</i>	<i>3 December 2014</i>
<i>Approval by the METC:</i>	<i>3 December 2014</i>

### Reason for submission:

To obtain ethical approval for performing research, in which panel members are asked to smoke cigarettes. We specifically choose to get approval for an expert smoking panel as this poses the highest burden for participants and therefore it is most difficult to get ethical approval.

Note: In January possible changes will be made to the existing protocol based on input from the expert seminar. In order to do so we have to hand in an amendment to get ethical approval for the changes. For the approval of the amendment, additional time ( $\pm 1$  month) should be reserved in WP3

### General content research proposal:

The research proposal describes a research study whereby an expert panel is trained to assess tobacco products through smoking cigarettes. As a method we choose Quantitative descriptive analysis (QDA), and the study consists of 3 phases: phase I: screening of potential candidates, phase II: training of the selected panel members and phase III: testing of different tobacco products. The screening consists of 3 tests to evaluate gustatory-, olfactory- function and verbal skills. The duration of this screening is approximately 1.5 hour. The training phase consist of 30 training sessions, 3 sessions per week of 1 ½ hour for 10 consecutive weeks. The last phase is the testing phase and consists of 3 sessions of 1 ½ hour that take place during 1 week in which five tobacco brands including roll-your-own tobacco and filter cigarettes and one reference (cigarette without additive) will be tested. These six tobacco products will be tested in a randomized complete block design and the tobacco products are evaluated in triplicate. The study population consist of adults (both male and female) whose native language is Dutch (*this will be changed in the amendment when we will train experts, training will be done in English*), aged between 18-55 years old. Participants have to be smokers in the past five years, and currently smoke at least 10 cigarettes per day. For the training phase of this study we will need 30 eligible participants, from which 12 will be needed for the final test phase.

### Comments by the METC and changes to the research proposal:

- Furthermore, the METC concludes that the studies as written in the research proposal can only be done with adults that are smokers. – We already emphasized on that with the research proposal, only adults who smoke 10 or more cigarettes per day will be included.
- Furthermore people who know that they want to quit smoking needs to be excluded- Added to the exclusion criteria “participant is planning to quit smoking during the upcoming 6 months or has tried to quit smoking in the past 3 months”.
- Another comment was on the safety of personal and research stuff, how they will be protected from cigarette smoke. - This will be prevented with the use of separate smoking boots with good ventilation.
- In addition, the METC was curious whether people should keep the same smoking pattern over the time span of the study. This is not the case, although researches agree that increasing the number of cigarettes smoked can lead to decreased taste and smell performance, it is not a necessity for participants to keep the same pattern. Panel

performance is regularly controlled for, if a panel member does not perform well he/she will be excluded for participation in the test sessions.

- Who is going to use the panel once trained, and what is the publication policy. This study functions as a pilot study and therefore we are not sure whether or not the panel will be further trained to use as an expert panel. The panel could be used by the European commission for tobacco control.

## Annex IX. Cost estimate

### 1 Estimation of costs for implementing the methodology as described in WP3

#### 1.1 *Cost-estimation for the identification of tobacco products with a characterising flavour by smelling the product using a sensory expert panel*

*The cost estimations presented here are separated in costs involved in setting up the panel (1.1.1), determining the reference space (1.1.2), and costs involved in the operation and maintenance of the panel (1.1.3 and 1.1.4). The cost calculation is based on the assumption that the panel is set up to assess tobacco products for a period of 4 years following the initial set-up period of about half a year. In the first half year, costs will involve setting-up the facilities, screening participants and training the panel. Once the panel is trained, it is operational to determine the reference space and to test product. For the set-up of the reference space 50 cigarette brands (including 3 tobacco leafs) and 25 RYO tobacco products are needed. We assumed that during the following time period around 10-100 products need to be assessed yearly. Assessment of products and/or panel maintenance will take place by having regular meetings (e.g. once a month). During these meetings at first panel performance is monitored (in a performance test-session). Based on these results the panel leader decides if the panel performance meets the standards and if additional maintenance meetings are needed.*

Definitions:

#### **Session**

A session has a duration of 1.5 hour during which the panel (up to 18 members) is trained, maintained or products can be tested.

#### **Product test session**

During one session up to 12 products can be assessed meaning 4 products in triplicate. For assessment of products in triplicate, three different packs of tobacco products of one brand are purchased by one or several distributors or selling points. During the testing phase 1-2 random selected references and 2-3 products are assessed during one session. For establishing the reference space 4 products can be assessed during one session.

#### **Product**

A product is one cigarette or 0.7 gram of a roll-your-own tobacco product from a particular brand and batch.

#### **Performance test-session**

Occasion during which panel(lists) performance is determined. A performance test-session needs to be scheduled before every test session or maintenance meeting (i.e. is always the session of the maintenance meeting).

**Meeting**

The duration of a meeting is usually one day during which up to 4 sessions take place. These sessions can have the purpose of maintaining the panel, testing products or assessing products to establish the reference space. The first of the sessions is usually a performance test-session to assess panel(lists) performance. The panel should have enough breaks in-between for the panellist to remain focused.

**Maintenance meeting**

Occasion where panellists are (re)trained by panel leaders. The function of a maintenance meeting is to keep track of and maintain panel performance after the initial training of the panel. A maintenance meeting takes a whole day and consists of 4 sessions of which 3 could, when needed, be replaced by test sessions.

**Personnel costs**

These costs are based on a standard personnel costs calculation, based on RIVM data. General costs arising from having the available capacity in both people and resources are included. It should be noted personnel needs to be hired on a contract for a fixed term. Yearly personnel cost of a panel leader (1 fte): €63.450. Yearly personnel costs of administrative personnel (1 fte): €48.600. Yearly personnel cost per research scientist (1 fte): €63.450.

**Facility costs**

Renting lab-facilities and office space, business operations, costs for infrastructure costs: €62.100 per year

**Product costs**

Pack of cigarettes or a pack of roll-your-own tobacco: €7

**Compensatory costs**

Compensation of €45 per 1.5h session per panellist.

**Travel costs**

€40 travel costs per panel member. €200 travel costs/hotel allowance per panel member.

### 1.1.1 Cost estimation setting-up the panel (estimated timeframe: 6 months)

For additional information on panel selection and training see §5.2-5.3 of Annex III: WP3 - MAIN document and SOP 1, §4.1-4.22

Facilities and equipment: start-up costs (CAPEX)		Description
setting up a test facility for smelling (multiple usage)	€ 100.000	Included in these cost are requirements as high quality air ventilation systems, testing rooms, discussion/training rooms, office space and an area dedicated to sample preparation. Indicated start-up costs are not part of the total cost estimation when test facilities and office settings for smelling experiments are already present at the testing facility. Rental of testing facility from €350/m <sup>2</sup> /yr. An estimated 150m <sup>2</sup> is needed.
purchase of equipment	€ 10.000	tablets and computers, depreciation over a period of 5 years
software (EyeQuestion)	€ 16.000	Purchase Intranet €12.000 – 20.000 average system, 22 concurrent users or Internet lease: several options e.g. €500/month plus €2,50/panellist meeting.
METC research protocol approval (provisional item, may not be necessary)	€ 2.000	
<b>total costs for a five year period</b>	<b>€ 128.000</b>	

#### Setting-up a sensory expert panel (screening and training)

personnel		Description
panel leaders	€ 38.070	2 panel leaders, for half a year 3 days a week (employed 0.6 fte each).
administrative personnel	€ 14.580	For half a year 3 days a week employed (0.6 fte).
costs personnel	€ 52.650	
<b>screening phase</b>		

Compensation for 180 persons participating in screening	€ 18.000	Screening takes about 6h with a compensation of €100 per person. Human resources for recruiting and screening indicated under personnel (panel leaders). 180 persons traveling once. Sniffin Sticks extended test complete by Burghart, and Taste Strips by Burghart.
traveling costs panel members	€ 7.200	
sample costs	€ 2.000	
<b>costs screening phase</b>	<b>€ 27.200</b>	
<b>training phase</b>		
Training, costs, e.g. two times a week for 1.5h over a period of 26 weeks	€ 42.120	52 training session with 18 panel members. 2x a week for 26 wks
320 packs of tobacco products	€ 2.240	80 different brands of tobacco products are used, 4 packs each brand.
traveling costs panel members	€37.440	travel frequency of 2x a wk for 26 wks
<b>costs training phase</b>	<b>€ 81.800</b>	
<b>total costs for setting-up a panel</b>	<b>€ 161.650</b>	

### 1.1.2 Cost estimation for establishing the reference space for cigarettes and roll-your-own tobacco products in the first year (timeframe: 1 month)

*Depending on the variability between the products and the required statistical significance of the analysis it could be decided to analyse less or more samples to define the reference space. Here we made a cost-calculation based on a reference space of 75 products (25 roll-your own products and 50 cigarettes).*

*For additional information on setting up the reference space see §4.3.1 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1)*

Product testing:

**75 products:** 18 panellists come together **every week during one month** to assess products to determine the reference space. In the first and third week they meet for 2 working days. In the second and fourth week they meet for 1 working day.

		<b>Variable costs</b>
		analysis 75 products
<b>personnel costs</b>		<b>Description</b>
panel leaders	€ 6.345	2 panel leaders, employed for 1 month 3 days a week, i.e. one twelfth of 0.6 fte.
statistician	€ 3.172	For 1 month 3 days a week employed (0.6 fte).
administrative personnel	€ 2.430	For 1 month 3 days a week employed (0.6 fte).
<b>sampling costs</b>		
product costs	€ 6.000	Collecting samples from the market and transporting them. Included logistical expenses and travel time when necessary.



<b>testing costs</b>		
per month: 6 days to assess reference products	€ 15.390	19 test sessions with 18 persons.
225 packs of tobacco products	€ 1.575	3 packs of tobacco products in triplicate
300 packs of tobacco products (back-up)	€ 2.100	4 packs of tobacco products per sample/brand needed as a back up for maintenance meetings.
traveling costs panel members	€ 10.080	Hotel allowance 2x for 18 panel members. Traveling costs 4x for 18 panel members traveling in one month.
<b>total costs per month</b>	<b>€ 47.092</b>	
<b>costs per product</b>	<b>€ 627</b>	

### 1.1.3 Cost estimation for maintaining the panel

To ensure sustainability of a panel in terms of attendance over a long period of time panellists should join a maintenance meeting once a month. During this meeting there should be, under normal circumstances, four sessions of 1.5 hour, one performance test-sessions and three maintenance sessions.

For information on test sessions and panel maintenance see §4.3.2 and §4.2.8 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1).

For additional information on the role of the panel leader see main doc. §5.1 of Annex III: WP3 - MAIN document and §4.1.1 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1). For the role of the statistician see §5.2 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1).

<b>personnel costs</b>		<b>Description</b>
personnel costs	€ 66.960	2 panel leaders, 0.4 fte each
statistician	€ 15.862	0.25 fte
administrative personnel	€ 12.150	0.25 fte
<b>training costs to add new panel members</b>		
52 training sessions to add 2 novel panel members	€ 4.680	2x a wk for 26 wks.
<b>costs to maintain the expert panel well trained</b>		
per year: 12 performance test sessions	€ 9.720	1 performance test session per month
per year: (up to) 36 maintenance sessions	€ 29.160	3 maintenance sessions per month
240 packs of tobacco products	€ 1.680	80 different samples are assessed in triplo.
320 packs of tobacco products (back-up)	€ 2.240	4 packs of tobacco products per sample/brand.
<b>traveling and subsistence costs</b>		
traveling costs panel members	€ 8.640	18 panel members traveling for 12x a year
<b>total costs per year</b>	<b>€ 151.092</b>	

### 1.1.4 Cost estimation for assessment of products in subsequent 4 years (2016-2020) by monthly meetings of the expert panel

For additional information on the role of the panel leader see main doc. §5.1 of Annex III: WP3 - MAIN document and §4.1.1 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1). For the role of the statistician see §5.2 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1).

For information on test sessions and panel maintenance see §4.3.2 and §4.2.8 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1). For information on the test phase see §4.3 of Annex IV: WP3 – Standard Operating Procedure 1 (SOP 1).

Product testing and panel maintenance training:

Product testing will normally coincide with panel maintenance whenever practicable (unless there is a peak period). The actual costs depend on number of samples submitted for regulatory testing and the way the testing sessions can be incorporated within the normal schedule of the maintenance meetings. For the purpose of cost estimate two distinct scenarios are considered. It is expected that the actual costs will lay between these two specific scenarios:

**10 products:** All 18 panel members come together **once a month for a whole day**. At first a maintenance test-session is performed to check panel performance than during one test session up to 3 products are tested in the remaining time maintenance training is performed.

**100 products:** All 18 panel members come together **once a month for one and a half day**. At first a maintenance test-session is performed to check panel performance than during three test sessions up to 9 products are tested. No maintenance training is needed unless panel performance reduces.

All cost estimates below include maintenance costs.

	Variable costs		Description
	analysis of 10 products yearly	analysis of 100 products yearly	
<b>personnel costs</b>			
personnel costs	€ 66.960	€ 66.960	2 panel leaders, 0.4 fte each
statistician	€ 15.862	€ 15.862	0.25 fte
administrative personnel	€ 12.150	€ 12.150	0.25 fte
<b>training costs to add new panel members</b>			
52 training sessions to add 2 novel panel members	€ 4.680	€ 4.680	2x a wk for 26 wks.
<b>testing costs</b>			
method development	€ 2.000	€ 2.000	Optimising the method and regular quality control
Per year: 50 product test sessions	x	€ 40.500	
300 packs of tobacco products	x	€ 2.100	Analysis 100 samples in triplicate.
4 packs of tobacco products per	€ 280	€ 2.800	400/40 packs of tobacco

EUROPEAN COMMISSION

sample/brand (back-up)			products
Per year: 5 product test sessions	€ 4.050	x	
30 packs of tobacco products	€ 210	x	Analysis 10 samples in triplicate
<b>costs to maintain the expert panel well trained</b>			
per year: 12 maintenance test sessions	€ 9.720	€ 9.720	
per year: 31 maintenance sessions	€ 25.110	x	31 maintenance sessions a year for training are needed for 18 panel members (analysis 10 samples/yr.) or no maintenance sessions are necessary (100 samples/yr.).
120 packs of tobacco products	€ 840	€ 840	Analysis 40 samples in triplicate (analysis 10 samples/year)
<b>traveling and subsistence costs</b>			
traveling costs panel members	€ 8.640	€ 43.200	18 panel members traveling for 12x a year (10 products).Travel costs/hotel allowance for 18 panel members traveling 12x a year (100 products)
<b>total costs per year</b>	€ 150.502	€ 200.812	
<b>costs per sample product on a yearly basis</b>	€ 15.050	€ 2.008	

### 1.1.5 Total cost estimate for setting up and running the panel

	10 products/yr. + reference products	100 products/yr. +reference products
CAPEX	€ 128.000	€ 128.000
Costs for setting up a panel	€ 161.650	€ 161.650
Costs for setting up the reference space	€ 47.092	€ 47.092
Running costs for subsequent 4 years (incl. compliance costs)	€ 618.208	€ 803.248
<b>Total costs over 4 years</b>	€ 954.950	€ 1.139.990
<b>Yearly costs</b>	€ 238.737	€ 284.997
<b>Cost per product</b>	€ 23.873	€ 2.849

## 1.2 Cost estimation for chemical analysis of tobacco products to identify products with a characterising flavour

### 1.2.1 Cost estimation for assessment of reference products in the first year (timeframe: ± 1 month)

Costs are based on outsourcing the chemical analytical work to a laboratory equipped with a GC-MS quadrupole met headspace/liquid injection system. Analysis, statistical calculations and reporting of the data are performed by a trained research scientist experienced in performing this type of analysis. For information on chemical analysis see §6 of Annex III: WP3 - MAIN document.

		Variable costs	
		analysis of 75 products	
personnel costs		Description	
1 trained scientist	€ 4.230	For 1 month 4 days a week employed 1/12 of 0.8 fte. Preparation of 75 samples, analysis samples in triplo, data analysis.	
Lab-facilities	€ 7.762		
testing costs			
consumables	€ 750	Laboratory needs (GC columns, standards etc.)	
<b>total costs per month</b>	<b>€ 12.742</b>		

### 1.2.2 Cost estimation for assessment of products over a period of 5 years (2015-2020)

Using chemical analysis, it can be determined how a characterising flavour as detected by sensory analysis is generated and whether it may derive from a particular additive or combination thereof.

In order to be able to define thresholds for a specific flavour the chemical components that form the basis of the flavour should not only be identified but also quantified. In this way a detection threshold or threshold above which a flavour is considered characterising (based on results of prior sensory analysis) can be determined. Costs estimated here represent the analysis for detecting one chemical component (e.g. menthol).

		Variable costs		
		analysis of 10 products yearly	analysis of 100 products yearly	
personnel costs				Description
1 trained scientist (0.3 fte)	€ 3.807	€ 19.035	Scientist will work at 0.3 fte.	
Lab-facilities	€ 12.420	€ 62.100		
testing costs				

EUROPEAN COMMISSION

packs of tobacco products	€ 70	€ 700	One pack of tobacco products per analysis Optimising the method and regular quality control Laboratory needs (GC columns, standards etc.)
method development	€ 2.000	€ 2.000	
consumables	€ 6.000	€ 6.000	
<b>total costs per year</b>	€ 24.297	€ 89.835	
<b>costs per product on a yearly basis</b>	€ 2.429	€ 898	

## **Annex X. Pilot proposal**

Based on our own expertise, and input from the participants of the seminar, we proposed three options for methodologies suitable to assess characterising flavours in tobacco: consumer panels, semi-trained panels, and expert panels. In addition, we proposed to perform chemical analysis. Only consumer panel and semi-trained panel tests were feasible to be carried out in the timeframe of the project. These proposals, together with their timeframes and advantages/disadvantages, are presented in Annex XI. Course of the panel training.

In line with the conclusions of WP2 and the discussion at the seminar, indicating that an expert panel better fit to the regulatory need, i.e. deciding whether or not a tobacco product has a characterising flavour, it was decided to have the procedure for an expert panel tested with a pilot experiment using a semi-trained panel. This means that certain limitations have to be taken into account. It may not be possible to obtain valid test results and a representative reference space following a 10 weeks training, therefore only limited information regarding cut-off values can be obtained.

The reason that a trained panel can perform reliable tests with only 12 panel members instead of several 100 as in a consumer panel is because the variability in the data is smaller; keeping statistical power equal to as when having a large number of consumers rating the products.

In case of 12 semi- trained panellists for test sessions, there will still be larger variability among panel members as compared to a fully trained panel (i.e. the panel may not fulfil all the requirements of panel performance; consensus, repeatability and discriminability) and because of that the statistical power may be low and it may not be possible to obtain valid results or cut of values of a reference space.

To determine whether or not a product imparts a characterising odour, a reference space needs to be defined. When a reference space is determined it can be concluded that if a product falls within this space the product is not significantly different from the reference space, however when the product falls outside this range the product may impart a characterising odour.

In addition to sensory analysis, chemical analysis will be pilot tested.

## **1 Sensory tests**

### **1.1 Study aim**

The purpose of the pilot is to obtain relevant information about the procedure to assess products with characterising flavours, including test set up, facilities and logistics.

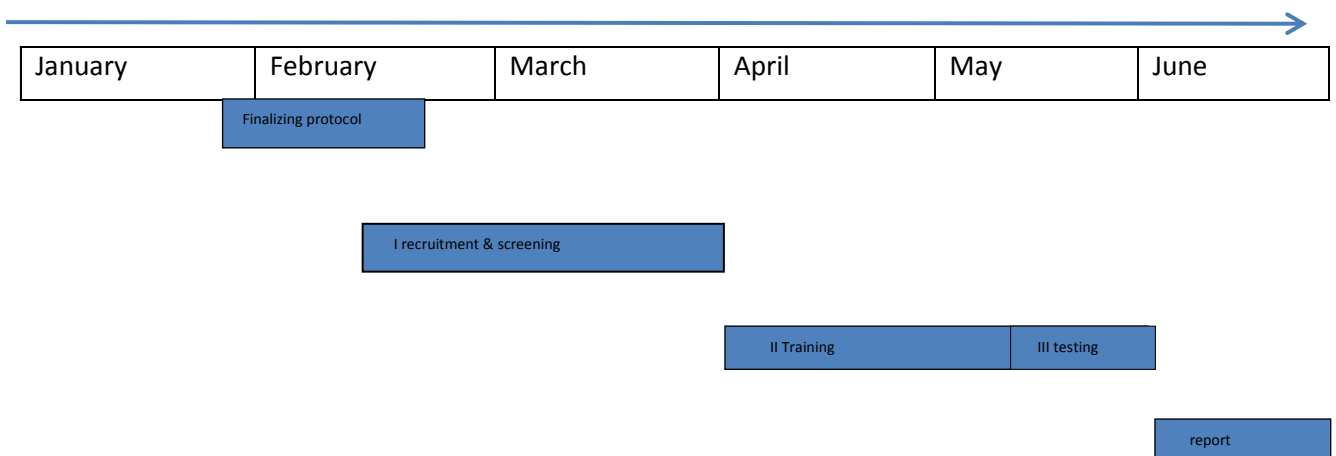
The final report will include a detailed description of participant selection and screening, experience gained during the pilot including the training sessions, information on the assessment of the panel performance and recommendations for future operation of such a panel including additional considerations (e.g. tobacco product assessment via smoking). This pilot test is limited in terms of panel training time (7 weeks) and the number of tobacco products that will be assessed (20 products). Because of this, we will obtain a preliminary list of potential attributes and a preliminary product space as final study results.

### 1.2 Study design

The study consists of 3 phases, phase I: screening of potential candidates, phase II: training of the selected panel members and phase III: testing of 20 different tobacco products. A sensory panel (trained for 7 weeks) will assess odours of unburned tobacco products by rating odour intensity according to an attribute list (generated by the panel). The method as described is a Quantitative descriptive analysis sensory method, QDA.

### 1.3 Time line

Within the time frame of the project (February- May) 18 panellists will be trained for 7 weeks. The training period will be followed by a test period (3 weeks) in which 20 different tobacco products will be assessed on odour intensity in triplicate.



## **1.4 Study population**

The study population consist of eighteen healthy adults (aged between 18-55 years old) whose native language is Dutch, both male and female will be included. Included panel members are required to have a good smell- function and be able to recognize and rate odour (intensity).

### In- and exclusion criteria Phase I Screening

In order to be eligible to participate in this study, a subject must meet all of the following criteria:

- Participant should speak Dutch as mother tongue
- 18-55 years (born between 01-06-1959 and 17-02-1997)
- Healthy as defined by a health questionnaire
- Available for a minimum period of 3 months for the pilot test

A potential subject will be excluded from participation in this study if he/she is:

- Using medications that are known to affect smell perception
- Pregnant or having plans to get pregnant
- Is lactating
- Having problems with smelling, hearing or sight

### Inclusion criteria phase II Training and III Testing

After inclusion and exclusion criteria have been applied, people are invited for the screening. Participants with the highest scores on the screening tests will be included in the training phase. Eighteen panel members need to be trained, they have to participate in all training sessions and should have sufficient discrimination ability, show repeatability of results and be in agreement with the group in order to be included in the final test sessions. The final test sessions should be performed with at least twelve panel members but preferably with all eighteen.

### **Sample size requirements**

This pilot study aims to describe how to recruit, select and train a panel and to use this panel as a sensory instrument to determine the (characteristic) odours of different tobacco products. A fully trained Quantitative Descriptive Analysis (QDA) panel typically consists on average of twelve panel members as it is considered to provide the optimal balance between ensuring a sufficient statistical power on the one hand and the need to achieve consensus between panellists on the other hand (13). However, considering the relatively short training period of this pilot study eighteen panellist is considered to be the optimum number of panellists: a larger number of panellists would lead to an increase in power, but also makes it more difficult to come to consensus (thereby decreasing power). Taking into consideration drop-outs, eighteen eligible participants for the training phase is sufficient as for the final test phase we need at least twelve panel members, although ideally, all eighteen panellist perform the final tests to account for higher variability due to the short training period.

#### **1.4.1.1 Method**

The pilot study consists of 3 phases, phase I: screening of potential candidates, phase II: training of the selected panel members and phase III: testing of different tobacco products. The screening consists of two sessions and 3 tests. During the first screening participants perform a tobacco odour recognition test and an intensity-rating test. This takes approximately 1 hour. Participants with highest scores on both test combined will be invited for the second screening. During this screening olfactory function will be assessed with use of the Sniffin' Sticks test (14). This test takes approximately 1.5 hour.



The training phase consists of 14 training sessions, 2 sessions per week of 1 ½ hour for seven consecutive weeks. The last phase is the testing phase and consists of 6 sessions of 1 ½ hour that take place during 3 weeks.

For the screening session participants will be asked not to eat or drink anything besides water, not smoke or use chewing gum or brush their teeth, half an hour prior to the screening session (14).

Before the start of the test and training sessions panel members have to wash their hands with neutral soap (unscented soap). It is important to inform panel members about the importance of concentration and that they should take enough time to perform the tests. Half an hour preceding the training or test panel members should not eat food with strong spices or herbs, candy, liquorice or gum. In addition, they should not brush their teeth or smoke a cigarette half an hour before the training/test. At the day of testing/training panel members should not put on perfume, aftershave, deodorant with a persistent odour, or lipstick. Panel members have to inform the panel leader in case they do not feel well, have a cold or allergic reaction (hay fever), or use medication (13).

The screening, training and final sensory tests will all take place at Wageningen University, Biotechnion.

### Phase I: Screening

For the fully trained expert panel, potential panellists need to be screened with a variety of tests in a step wise approach to end up with the best 18 panellists. In general, the screening tests aim to select panel members on the basis of general prerequisites such as availability, motivation and interest, and on the basis of task specific qualities such as sensorial capacity and verbal and social skills.

Because of limited time we will only use a sub sample of the task specific tests to screen participants for the pilot study (15). We will screen potential participants on their ability to recognise odours that are present in tobacco products, their ability to indicate odour intensity and on their olfactory function.

#### *Screening 1*

Odour recognition skills will be assessed with use of flavoured tobacco products. Ten tobacco products will be offered to the participants in non-transparent cups to eliminate visual cues. Participants have to smell the sample and indicate the main odour they perceive from a multiple-choice list (in line with the check-all-that-apply (CATA) method). Participants with the highest scores (indicating the odour as described on the package of the tobacco product) will be selected (15).

Consistency or repeatability of rating products is important to determine when selecting panel members. This test has to be performed at two time points with the same stimuli but given to the participant in the same order but samples should be coded differently. For this test we will add almond aroma in 4 concentrations to (0.10 mg) RYO-RP2 roll-your-own tobacco. The participant has to rate the 4 samples according to the odour intensity on a line scale with anchors going from 'weak' to 'strong'. The participants that are best able to rate the products in intensity corresponding with the rank order (i.e. the sample with the lowest concentration of the flavour/odour has been rated lowest on the line scale) will be selected for the panel training phase (15).

#### *Screening 2*

Olfactory function will be assessed using an odour stimuli test, the so-called 'Sniffin' Sticks' developed by Hummel and colleagues(16). This test consists of pen-like odour dispensing devices with odours considered to be familiar. The smell test contains three parts: an odour discrimination (DIS) and identification test (ID) and odour threshold test (THR). Participants are selected if they pass the smell test, i.e. if they meet the normative standards, and are among the 18 best performing panellists (14). In the unlikely case that we end up with less than eighteen panellists who meet the

normative standards for the Sniffin' Sticks test we re-invite participants from the first screening (those with the highest scores) until we are able to include eighteen eligible panellists for the training phase.

### Phase II Training

The aim of the panel training is to improve panel members' ability to identify odours in tobacco products and to rate them according to the odours intensity. Panel members have to be able to judge the products in an accurate, consistent and repeatable manner(15).

The training phase consist of fourteen training sessions in which participants need to be committed to two, one and a half hour training sessions a week, for seven consecutive weeks. The eighteen panellists will be divided in two groups (a morning and an afternoon group). These groups are not fixed, panellists can switch between groups as long as they join two training sessions per week, provided that they are not on the same day. Non-fixed group will increase panel consensus, as this way, we hope to prevent the emergence of two distinct groups, each with a different way of evaluating the products.

During the first two training sessions, panellists will be provided with 7 products per sessions (14 different brands in total). During the sessions, each panellist has to smell all 7 samples and exclude the product they think should not be in the group, or is most outstanding (reversed grouping). Every time they exclude one product, they have to write down based upon which descriptors or attributes they exclude the product from the group. Panellists do this individually, and afterwards a group discussion will follow to come to an attribute list. To make this task easier, a list of attributes (appendix I) will be provided that the panellists can use in case they have difficulties describing the odour. However, panellists can also come up with their own descriptors/attributes.

During the second weeks' training sessions, the panellist will use the attribute list as developed in the first week to rate the products on a five point scale (from very weak, to very strong). After this task a group discussion will follow to discuss which products have the highest and lowest score for each attribute and to discuss whether some attributes on the list have the same meaning or are redundant. During the third week, the revised attribute list will be tested in a real test manner, where participants will perform QDA in sensory booths for 10 tobacco products.

During the training sessions in week four, panellists will receive the individual and group results of the test performed in week 3. Based upon the results of this test, the attributes and products that panellist find most difficult to rate (no consensus) are determined. During the fourth week, special attention will be paid to these attributes and products as panellist will score these products according to the attributes on a five point scale (similar to week two), which again will be followed by a group discussion to increase consensus. The fifth weeks' training sessions will be similar to the training sessions of the third week but for the remaining ten tobacco test-products.

During the training sessions in week six and seven, panellists will practice with the attributes and products that were most difficult to rate according to the training sessions in week five. During these training sessions, panellists will rate each product on a five points scale for each attribute individually followed by a group discussion.

**Table 26. Panel training schedule and content**

Week	Training session number	Content
1	1,2	Attribute generation through reversed grouping
2	3,4	Practicing with attribute list
3	5,6	Real test exercise, assessment of 10 products in duplicate
4	7,8	Discussion results week 3. Panellist practice rating the difficult products and attributes.
5	9,10	Real test exercise, assessment of 10 products in duplicate (different from products used in week 3)
6	11,12	Discussion results week 5. Panellist practice rating difficult products and attributes.
7	13,14	Panellist practice rating difficult products and attributes

### Performance of the panel

Throughout the training phase the panel members will be assessed on panel performance, this will also give us additional information about the difficulty of rating specific tobacco odours. Discriminatory power, panel consensus and repeatability are to be considered and monitored throughout the training sessions (17).

Performance of the panel is assessed per attribute by performing an analysis of variance (ANOVA). The following model will be used:  $Y = \mu + \text{product effect} + \text{panellist effect} + \text{session effect} + \text{product-panellist interaction} + \text{product-session interaction} + \text{panellist-session interaction} + \epsilon$

Intensity of the sensory attribute (on 100mm VAS) is the dependent variable; product, panellist and sessions are independent variables.

This model is similar to the model described below ('statistical analyses') to evaluate the difference of attribute intensity between products. For evaluating panel performance, we are interested in the following parameters:

- If the product effect is significant, it means that the intensity values of a specific attribute, evaluated by the panellists, are significantly different between products, therefore the panel has been able to discriminate the products with respect to that specific sensory attribute. = Discrimination ability of the panel
- If the product-panellist interaction is significant, it means that the intensity values of the attribute evaluated are significantly different depending on both the type of product and the panellist, hence there is no consensus or agreement among panellists for that particular attribute. Panel members must have different perception of the products regarding that attribute.
- -If the product-session interaction is significant, it means that the intensity values of the attribute evaluated are significantly different depending on the type of product, but also on the sessions when it was evaluated, therefore the panel is not repeatable from one session to another.
- If the panellist effect is significant it means that the panellist does not use the scale identically, however this is of less importance as we are interested in the relative difference

between products. This is similar for a significant session effect; whether panellists use the scale consistently from one session to another.

- Panellist-session interaction indicates whether some panel members use the scale different from one session to the other. But this is also not of importance due to our interest in the relative differences between products (17).

To have a simple overview of the panel performance, a table will be built using the statistical program R. The dependent variables (intensity of sensory attributes) will be presented in the first column, and the independent variables (product, panellist, sessions, and their interactions) will be in the first row. The table values will represent the  $p$ -values for each attribute at any specific effect or interaction. In this table we highlight the cells that are below significance level (0.05) to have a good overview of the panel performance (17).

#### Performance of the panellists

We can compare the individual performance of each panellist per tobacco product compared to the overall panel scores. To compare individual assessor's results to one another and the panel mean, the same strategy as mentioned for panel performance can be applied. However, as we look into the individual panellist we apply an ANOVA model per panellist. We then obtain results per panellist for all attributes separate for the product effect and this indicate the global performance (discriminability) of each panellist. We can compare the product effect of the panellist with the product effect of the complete panel to assess the consensus of one panel member with the panel as a whole. By comparing the standard deviation of the residuals of the ANOVA panellist model with the residuals of the attributes (one-by one) we can say something about the repeatability of the panel member (17). To have a simple overview of the results of the ANOVA for all panellists we will make 3 tables showing the discriminability, consensus and repeatability results for each of the attributes per panellist.

#### Phase III Testing

In a single session the panel members can evaluate 10 tobacco products and the 20 tobacco products should be evaluated in triplicate. Therefore, panel members must attend six sessions to evaluate all the samples. The total duration of a test session will be 1.5 hour.

#### Test procedure

The sensory test will be performed in sensory booths under standardized conditions (odour free, good lightning). The tests will be conducted by the computer program EyeQuestion (Logic8, version 3.7.6), and will provide the introduction, some questions about the individual (Panel member number) and the odour-intensity questions about the attributes for the tobacco test products. During the testing phase panel members will evaluate tobacco in the same manner as practiced during the training phase (week 3 and 5). Panel members will receive a sample and rate the intensity of +/- 30 odour attributes in a randomized order. Participants are allowed to re-smell the sample as often as they like and rate the attributes at their own pace. Between smelling the different tobaccos, the participants have a 20 seconds break to neutralize their smell. The questions in EyeQuestion will be answered via a 100 mm visual analogue scale. On the left and right hand of the scale at 10 and 90 mm, opposite terms will be used, weak and strong (in Dutch: zwak en sterk). The tobacco products are offered to the participant in a randomized order. After assessing 5 tobacco products there will be a 2 minutes break in which participants can play a short computer game to regain their focus before evaluating the remaining 5 products.

#### Study products

During the three final test sessions twenty tobacco products will be assessed. We will include six original filter cigarettes of brands with a large market share, two popular original roll-your-own tobacco brands, three types of tobacco leaves and one tobacco product without additives. We

predict this group to cluster and to constitute the reference group, and we will test empirically how far apart natural tobacco leaves and brand cigarettes will be. We will include eight tobacco products that we expect to have a characterising flavour according to the package or ingredient list (Test products). Two are menthol cigarettes made from different tobacco leaves (Cig-TP2 is Virginia mixture), furthermore, we include two vanilla-flavoured cigarettes, high and low in vanilla according to the ingredient list, and one roll your own. Three products of this group are thought to be (other) flavoured cigarettes. The package of the Cig-TP5 clearly indicates Cherry flavour and the package of the Cig-TP7 cigarette does not indicate a specific flavour, however we included this product, as it is suspicious due to the pink package and cigarettes. We suspect the Cig-TP6 (red) cigarettes to be flavoured based on the ingredient list of the brand (high in cocoa).

Tobacco product		Number	Brands
<b>Products to determine the reference space (reference products)</b>	Original filter cigarettes	6	Cig-RP1 (American blend), Cig-RP2, Cig-RP3 , Cig-RP4, Cig-RP5, Cig-RP6
	Roll your own (original)	2	RYO-RP1 (half zware), RYO-RP2 (zware)
	Tobacco leave	3	Virginia, burley, oriental
	Filter cigarette without additive	1	Coresta Monitor 6 (CM6)
<b>Test products</b>	Menthol cigarettes	2	Cig-TP1, Cig-TP2
	Vanilla flavoured cigarettes	2	Cig-TP3 Hp 19 (expected low), Cig-TP4 (expected high)
	(Other) potentially flavoured cigarettes	3	Cig-TP5, Cig-TP6 (red), Cig-TP7
	Roll your own (flavoured)	1	RYO-TP1 hand rolling vanilla
		Total = 20	

### Randomization and blinding

Each brand will be presented three times for each panel member in randomized order over the 6 sessions (complete block design). In this study the blocks are 2 sessions and the 20 tobacco products are randomized across those panellists within each block. The study will be single-blinded, when it comes to brand recognition, as samples will be coded and the panel members will not have knowledge of which code belongs to which roll-your-own or filter cigarette product brand. To prevent brand recognition, the cigarettes offered to the panel members in a non-transparent test cup to eliminate all visual cues. During data analysis, products will be 3-digit coded, in that way the researcher performing the analysis will be blinded and conclusions can be drawn without prejudice.

### Product storage

Samples are stored under conditions that do not impact the sensorial properties. All samples will be stored in their original package in boxes with closed plugs in the refrigerator.

## 1.5 Statistical analysis

This section describes an initial outline of the statistical analysis foreseen for the pilot study. More detail will be provided at a later stage during the pilot study. Moreover, the methods to be used in the approach ultimately proposed will be re-assessed and further refined within WP3 based on the experience gained during the pilot.

Statistical analysis will be performed with use of the statistical software R and the packages FactoMineR and SensoMineR (17).

The main aim of performing descriptive analysis is to obtain an accurate description of the products regarding the sensory attributes: a sensory profile of a product. This sensory profile can be used to identify the sensory attributes that are the most important for defining the product space and to show relationships between products. The product space can consequently be used to determine the reference space that in turn can be used to determine whether a product is different from the average reference product on specific attributes (see \* below). The product space is defined by the (intensity of the) sensory attributes. To have an overall or complete product space, products need to be included that are the most extreme or typical versions of every sensory attribute. Based on the product clusters of the product space, a reference space can be determined including products that are considered not to have a characterising flavour based on the corresponding attributes for that particular cluster(17). Furthermore, we can compare a 'new' product with an 'average product' (the reference space) when we included the product in the overall product space. If the new product falls outside the cluster of reference products we can designate the product as having a characterising flavour.

### More detailed explanation of the statistics

When the panel is fully trained, attributes that differentiate the products are identified with use of Multivariate Analysis of Variance (MANOVA), performed for every single attribute (similar to the panel performance statistics). We choose to perform the MANOVA as odour attributes are rated on a continuous scale and we want to explain the odour attribute (dependent variable) by the product, panellist and session effects (independent variables). We use the same statistical model as to evaluate panel performance:

$$Y = \mu + \text{product effect} + \text{panellist effect} + \text{session effect} + \text{product-panellist interaction} + \text{product-session interaction} + \text{panellist-session interaction} + \varepsilon$$

For panel performance evaluation we were interested in the interaction between panellist and session variables with the product variable that influenced the intensity ratings of the sensory attribute. Once the panel has been trained, we assume that there is no significant interaction between panellist and session. Because of that, we can now assess whether the product variable significantly influences the intensity of sensory attributes evaluated. To test our assumption of the panel being well trained, we still evaluate the other variables (panellist and session) and their interaction to verify that the panel has consensus and is repeatable.

For example, we want to explain the attribute menthol with respect to the type of product. The question you could answer would be 'Is menthol an attribute that panellists can distinguish between products?'. If the effect of this variable (product effect) were statistically significant, this would mean that the menthol attribute is significantly different among products, and therefore it plays an important role in defining the sensory profile of each product.

We will apply the statistical model to all the dependent variables (intensity of sensory attributes) to point out the variables that are the most characteristic according to the set of products in its whole, and to each of the products in particular. To do that, we will use the function *decat* (description of categories) of the SensoMineR package in R. This function is designed to test the main effect of a variable (product) and the significance for a set of dependent variables (intense of sensory attributes) at a given ANOVA model

Once we have identified the sensory attributes that differentiate the products we can define which products are specific for those attributes (i.e. which products are differentiated based on menthol).

In this case we give answer to the question ‘for the sensory attribute menthol, which product can we consider as significantly different from some kind of average products (all other products of your sensory test). This can be done with a simple T-test not comparing two products but 1 product comparing with the average product. In our case it would be comparing 1 ‘new’ product in a sensory space defined by products from which we assume they do not impart a characteristic flavour/odour.

When the T-test is significantly different ( $p \leq 0.05$ , 95% CI) for products we can conclude that these products are different from the average product (all the other products in the product space) regarding the sensory attribute menthol (having either less or more menthol flavour/odour compared to the average).

Again with use of the *decat* function we can determine the sensory attributes that are the most characteristic of a set of products as well as for product by product (17).

ANOVA is also used in sensory data analysis to obtain adjusted means that hence can be used for Principal Component Analysis (PCA). PCA is a multivariate method, which you can use to explain one variable or product considering multiple quantitative variables (attributes). Using this method, we can see the similarities and differences between products in a product space. In this product space products are located close to each other if their sensory profiles are similar and are further apart if they are different.

By performing PCA we can obtain a factor map representing all products according to two dimensions. Simultaneously we obtain a factor map representing the sensory attributes. With use of the FactoMineR function *dimdesc* we can obtain a list of variables that are significantly linked to each dimension. Remember that these dimensions or this product space is only build up out of the attributes that were significantly different for products (17).

After performing the PCA we can perform clustering methods to test the distance between products instead of only visualizing these distances as with PCA. An example of a cluster method is Hierarchical clustering on principal components (HCPC). This clustering method only takes into account the first 5 main components that define the product space. HCPC shows you a hierarchical tree and a cut of value based on the ratio of variance within the clusters and the total variation and then shows the PCA map with products in a similar colour belonging to the same cluster. Furthermore we can obtain for each cluster a list of significant attributes (17).

\* It is possible that a product is placed on the market that contains a ‘new’ flavour, which did not come out as a relevant attribute before or has not been listed as an attribute during the panel training, and is thus not part of the product space. In that case, it is recommended to ‘update’ the product and reference space, however, that approach falls outside the scope of this pilot experiment.

## 2 Chemical analysis

### 2.1 Aim

The 20 tobacco products tested by the sensory panel will be chemically analysed with headspace GC-MS as well.

Our aim is to objectively confirm the flavours determined by sensory analysis. For instance, when the sensory panel detects cherry flavour, we analyse the product and check for components that cause a cherry flavour to confirm the sensory result.

The components that cause the flavour are identified and compared to sensory results. We identify the components that are present in the 3 types of tobacco leaves and in the 17 tobacco products with and without a characterising flavour. After performing this experiment, we provide lists with all present identified components and the flavour they cause according to the Leffingwell Flavour Base 9 – Tobacco version to compare these with the sensory results.

We now cannot predict how the flavour levels of the flavoured products will relate to the products not advertised as being flavoured, since the reference products still need to be determined. Also, one flavour (i.e. menthol) can consist of a different combination of chemical components (i.e. Cig-TP1 consists of L-menthol, menthone, and menthyl acetate, while another menthol brand consists of another combination of components). Therefore, we can compare relative amounts of a flavour between two tobacco products *if* this flavour consists of *the same combination* of chemical components.

Therefore, we can compare the composition of the tested tobacco products. Whether we can say something about relative amounts of flavours, can only be predicted when the experiment is finished.

For now, we focus on identification of the flavour components to objectively confirm the flavours present according to the sensory analysis.

When the experiment is finished, we will send you the results to discuss how to interpret this data and what this means for future legislation.

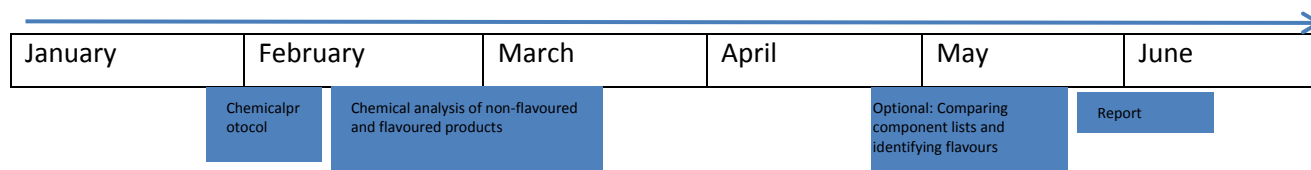
### 2.2 Experimental design

We will measure the 12 brands (listed in 'study products') from which we assume not to have a characterising flavour (leaves and non-flavoured products) with headspace GC-MS. This method will provide a list with all chemical components present in these tobacco products. We will compare the component lists of the non-characterising products, also using Principal Correspondence Analysis (PCoA, qualitative data) and/or Principal Component Analysis (PCA, quantitative data), to determine differences between leaves and non-flavoured products. This eventually can help in determining which brands are suitable as reference when we know how they differ from each other and from the flavoured products.

We will measure 8 brands that presumably have a characterising flavour, these brands will be the same as tested by the sensory panel (see 'study products'). We will compare the component lists of the non-flavoured brands to the lists of the brands with a characterising flavour. The components that are present in the characterising flavour products and not in the non-flavoured brands, are assumed to be the components that are added to evoke a characterising flavour. We will use the Leffingwell Flavour-base to identify the flavour that every of these chemical components cause. When at least one component causes the specific flavour that is determined by sensory analysis, we can confirm with chemical analysis that the characterising flavour is present.



### 2.2.1 Time line



### 2.2.2 Products

See products used for sensory tests.

### 2.2.3 Method

Headspace GC-MS.

## 3 Budget (pilot)

This section describes the budgetary planning for the pilot study. A cost-estimate including maintenance fee for the final panel and setting up test facilities will be described in WP4.

### Sensory analysis

#### Incentives participants

After participation in the one hour screening session (Phase I), participants not eligible to participate in the study will receive a financial compensation of €10. Participants that performed both screening sessions but were not included in the panel (3 hours in total) receive €25. Panel members will receive financial compensation of a total of €230 for the complete training and test phase (Participants will receive the compensation by rate if they drop out of the study.)

Phase	Number of participants	compensation	total
I screening	+/- 100	€35	€3500
II training + III test	18	€230	€4140
Round up total:			€8000

#### Material Costs

Material	Number	Price	total
Cigarettes			
I screening	75	€7	€525
II training	80	€7	€560
III test	120	€7	€840
Reference odours	150	+/- € 10	€1500
Sniffin sticks	-	-	€820
Other materials (test cups etc.)	-	-	€1000
Round up total:			€5500

#### Other costs

	Costs
METC research protocol approval	€2000
Preparation smoking lab	€1100
Round total:	€ 3100

**Chemical analysis**

Based on headspace GC-MS analysis the estimated costs:

	Costs
Cost cigarettes	€800
Other costs	€200
Round total:	€1000

\*Based on 20 brands (19 cigarettes per package, costs €7)

**Total cost pilot study**

€17600
+ 10% unexpected costs = € 19360 <b>Round up total: € 20.000</b>

**4 References**

1. Hummel, T., et al., *Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects*. European Archives of Oto-Rhino-Laryngology, 2007. **264**(3): p. 237-243.
2. Meilgaard, M.C., Vance Civile, G., and Carr, T.B, *Sensory evaluation techniques*. 4th ed. 2007, New York: CRC Press.
3. OPP Research, *Workshop setting op a sensory panel*. 2001.
4. Hummel, T., et al., '*Sniffin' Sticks*': *Olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold*. Chemical Senses, 1997. **22**(1): p. 39-52.
5. Le, S. and Worch, T., *Analyzing sensory data with R*. The R series. 2014, Boca Ration, FL: Taylor & Francis Group.

## Annex XI. Course of the panel training

### Week 1: training 1

Introduction of the aims of the tobacco panel and panel members introduce themselves. Participants were offered 7 products from which they had to exclude a product that was most different compared to the others and write down based on which odour they thought it was different. They continued doing this until only one product was left. Panel members were given 45 minutes to perform the task.

Products that were used during this training sessions were Cig-RP1, RYO-RP2, Burley, Coresta monitor 6, Cig-TP2, Cig-TP4, CIG-TP6.

Products were chosen in such a way that one of each category was included, for example menthol and a vanilla product (see proposal pilot study).

Odour descriptors that were generated by group one were:

Vanilla, Caramel, Sweet, Smokey, Burned, Chocolate, Marmite, Salty, Old wet hay, earth, plum, fruity tea, nutty-odour, fish, peppermint, menthol, liquorice, 'pottertjes' (Dutch kind of liquorice), tutti-frutti, black tea, tobacco, wood, hay, cocoa, toffee, burned wood, mocha

Odour descriptors that were generated by group two were:

Tobacco, smokey, burned wood, roasted, vanilla, burned sugar, caramel, toffee, coconut-crème, fruit, chocolate, coffee, leather, rooibos-tea, hay, stable, grass, mint, toothpaste, eucalyptus, menthol, green tea, honey, sweet, raisins, mocha, liquorice, spices

### Week1: training 2

During the second training 7 other products were given to the panellist, similar to the first training. Odour attribute that were not yet on the list (the attributes of the first and second group of the first training combined) were written down to have a complete overview. After that, the group of panel members started to decrease the number of attributes, by excluding attributes that were not relevant, had the same meaning or that could be merged.

Products that were used for the second training were: Cig-RP4, Cig-RP6 , RYO-RP1, Virginia, Oriental, Cig-TP5, Cig-TP7.

Products were chosen in such a way that one of each category was included (see pilot proposal).

In bold the attributes that were included in the first version of the attribute list:

Group one: **Smoke, Smoked, Tobacco, burned**, roasted, **vanilla, burned sugar, caramel**, toffee, **coconut**, fruit, **chocolate (sweet), cocoa (bitter), coffee/mocha, leather, rooibos-tea, liquorice, marmite (salty), nutty**, fish, **tutti-frutti**, salt-ammonic, **red fruit**, glühwein, **clove**, orange, **cinnamon, hay**, stable, grass, **mint/peppermint**, toothpaste, Vicks, **menthol, eucalyptus, tea**, green tea, fruity tea, **honey**, sweet, **raisins, spices, old wet hay, earth, plum**, 'pottertjes' (Dutch kind of liquorice), **wood**, melon, **flowers**, sangria, peach, **syrup, dried fruit, vinegar (sour odours),honey**.

### Week 2: training 3

During the third training the panel assessed 5 products on all attributes as previously determined (see training 2). They had to indicate the intensity of the attribute for each product by using + and – signs, see below.

- ++ = most strong
- + = strong
- +/- = not weak not strong
- = weak
- = most weak

Products that were used during this training were: Cig-RP3 (#1), RYO-RP2 (zware) (#3), Coresta Monitor 6(#4), Cig-TP1 (#5), and Cig-TP2 (#2).

Both menthol products were included for the panellist to be able to compare the strength of the menthol odour for these products.

Panellists were instructed to assess the 5 products on the first odour and continue with the second odour to be able to make comparisons in odour strength for all products.

After 45 minutes there was a group discussion in which the following conclusions were made:

#### Group 1:

- Smoke and smoked should be merged together, in everyday life these are different odours but the panel is not able to indicate the difference within the tobacco products.
- Chocolate and cocoa are clearly different but not everyone has the same idea about this so references are needed to clarify the meaning of these two attributes.
- Eucalyptus is a stronger version of menthol and therefore these can be merged together because other than the strength it is difficult to differentiate between these two attributes.
- Product number 1 and 3 are typically for tobacco odour, number 1 is stronger for tobacco odour compared to number 3.
- Product number 3 is most strong for burned, smoke and smoked odour and leather.
- Product number 4 is most strong for tea and hay.
- Product number one has a little of chocolate but not cocoa.
- Product number 2 and 5 are both menthol/mint but product number 5 is most strong.

#### Group 2:

- Smoke and smoked is different from tobacco odour, as smoked is more the odour that comes from the ashes of a cigarette.
- Tobacco odour is difficult to describe, but number 3 is the best example together with number 1. But number one is less tobacco than number 3 and number 1 is more smoked odour.
- Chocolate and cocoa should be merged together, after smelling real cocoa it seems that that is sweet too.
- Menthol and eucalyptus can be merged together, difficult to distinguish.
- Need examples of different kinds of chocolate, old wet hay, hay, tea, honey and vanilla odour.
- Product number 1 is typical for hay and old wet hay together with a little bit of leather odour.
- Product number 2 and 5 are both menthol but 5 is more strong compared to 2.
- Product number 3 is strong for burned/smoked odour.
- Product number 3 is also leather according to one half of the panel although the other part thinks is more the smoked odour that is present.
- Product number 4 is typical tea, and a little bit of a coffee odour.
- Product number 5 smelled a little bit like vanilla or honey besides the menthol odour.

Week 2: training 4

Similar to the third training but now with 5 other products: Cig-RP2 (#7), Cig-TP3 Hp 19 (#6), Cig-TP4 (#8), Cig-TP5 (#10), Cig-TP7 (#9).

The Panellists were again instructed to first assess the 5 products on the first odour and then continue to the second odour to be able to make comparisons in strength for certain odours.

After 45 minutes there was a group discussion in which the following conclusions were made:

Group 1:

- Vanilla is unclear how this odour in tobacco is different from caramel, need examples of these odours to clarify them.
- To describe tobacco odour panellist use hay, tea, wood, honey, chocolate.
- Product number 7 is typical for dried fruit; product number 10 does not have this odour at all.
- Product 6 and 7 both are strong for hay, but 6 are more strong compared to 7.
- Product number 8 when just opening the bottle, vanilla is most intense but when the lid is off the bottle for a little while you also smell tobacco odour.
- The attribute syrup is unclear and not necessary to describe the odour of the products (i.e. does not add to the other attributes)
- When the lid is a little off from the bottle, different odours can be perceived as when just opening the bottle
- Chocolate and cocoa are the same odour

Group 2:

- Product number 6 is chocolate, and dried fruit
- Product number 7 is mainly tobacco and dried fruit
- Product number 7 and 9 are both wood, but product number 9 is more tea, and sour like odour that should fall under old wet hay
- Product number 10 is almond and red fruit, and clove

The preliminary attribute list that was used for the training test sessions of week 3 can be found in Annex XII.

Week 3, training 5 and 6

Test-training session. During these training sessions panellists assessed the products individually in the sensory booths similar to a real-test situation. Panellists assessed 10 products per training session. Testing was done in duplicates.

Products used were:

1. Cig-RP4
2. Cig-TP6
3. Cig-RP1
4. Coresta Monitor 6
5. Cig-RP6
6. Cig-TP3
7. Cig-TP1
8. Burley
9. RYO-RP2
10. Cig-TP5

Strong vanilla tobacco products were not included as we wanted to focus on certain types of products, to be able to train the panel on the attributes belonging to these products in the week

thereafter. It was not possible to train on all attributes simultaneously and another test-training session was planned in week 5 hereafter we planned on focussing on the other attributes and products such as vanilla.

The following comments were made by the panellists:

- 1.5 hour to assess 10 products on all the attributes is sufficient for all panellists
- The breaks (30 seconds) in-between products are sufficient to neutralize the nose-space.
- The 5 minute break where panellists could play Tetris was good to regain focus.
- Some smell different odours depending on how long the bottle remained open.
- One panellist said that there were too many attributes that he had to assess for each product, not all of them were relevant
- Sometimes panellists smell odour coming from my hands instead of the tobacco sample.

What was noticed by the panel leader:

- Some panellists assess the odours of each product fast, however after looking at panel performance the panellists who assessed the odours fast were not the panellist who performed badly
- Panellist should be instructed to smell the sample before assessing each attribute, besides the odours that are most strongly perceived when opening the bottle
- Panellist should not shake the bottle before opening

Results of panel performance can be found in Annex XII. For the test training sessions, products were numbered between 1 and 10 as this makes discussion of the products in the following training sessions easier (panellist were allowed to make notes). The order in which the products were given to the panellist was not randomized as in this case we were not interested in the final results but only in panel performance. The other reason we did not randomize the order is to be able to ask panellist whether it was difficult to assess some of the products after the other, for example if it was difficult to perceive other odours besides 'smokey/burned' after smelling RYO-RP2 tobacco. This was not the case, but neutralization time between samples could have prolonged in case this was not sufficient.

General conclusions following from the test training panel performance and results:

- The panel discriminated and was in consensus for most attributes but not for the odours: rooibos, liquorice, coconut, honey, flowers, leather, salty, dried fruit, old wet hay, almond and coffee. Possible reasons are 1) they did not understand the odour attribute 2) the odour is not present in any of the tested tobacco products
- Panellists limit their ratings to the extremes of the scale rather than using the whole scale, as instructed. Therefore, in the next weeks we used scales with anchors on the outer sides (0 mm, 100mm) of the line-scale.
- Typical odours of some of the products as indicated by the panel:
  - Burned and smoked: RYO-RP2 RYO
  - Tobacco: RYO-RP2 RYO and Cig-RP1
  - Vanilla: Cig-RP4
  - Caramel: Cig-RP4, Cig-RP1
  - Chocolate: Cig-RP4, Cig-RP1, Cig-TP6
  - Almond: Cig-TP5
  - Old wet hay: Burley
  - Clove: Cig-TP5
  - Red fruit: Cig-TP5
  - Cinnamon: Cig-TP5



**Table 27. Rank order for the odour attribute vanilla**

<b>Odour: Vanilla</b>		
<b>Rank</b>	<b>Group 1</b>	<b>Group 2</b>
<b>A (strongest)</b>	Cig-TP6	Cig-RP4
<b>B</b>	Cig-RP4	Cig-RP6
<b>C</b>	Cig-RP6	Cig-RP1
<b>D(weakest)</b>	Cig-RP1	Cig-TP6

**Table 28. Rank order for the odour attribute Caramel**

<b>Odour: Caramel</b>	
<b>Rank</b>	<b>Group 1</b>
<b>A (strongest)</b>	Cig-TP3
<b>B</b>	Cig-RP4
<b>C (most weak)</b>	Cig-TP6, Cig-RP6 , Cig-RP1

**Table 29. Rank order for the odour attribute Chocolate**

<b>Odour: Chocolate</b>	
<b>Rank</b>	<b>Group 1 and group 2</b>
<b>A (most strong)</b>	Cig-RP1
<b>B</b>	Cig-RP4
<b>C</b>	Cig-TP3
<b>D</b>	Cig-RP6
<b>E(most weak)</b>	Cig-TP6

**Table 30. Rank order for the odour attribute Dried fruit**

<b>Odour: Dried fruit</b>	
<b>Rank</b>	<b>Group 1 and group 2</b>
<b>A (most strong)</b>	Cig-RP1
<b>B</b>	Cig-RP4
<b>C</b>	Cig-TP3
<b>D</b>	Cig-RP6
<b>E(most weak)</b>	Cig-TP6

Note, these tables reflect (preliminary) results during training, and are not final results.

The panel did not perceive Coconut odour in any of the products, therefore this should be eliminated in case for the other ten products this attribute is not relevant either.

Week 4: training 8

This training was similar to training 7; however different products and different odour attributes were used. From the previous training we concluded that ranking for more than 3 products is too difficult therefore, for this training we only used 3 products per attribute. Products were chosen based on having high, medium and low intensity ratings as indicated during the test-training session by the panel.

Attributes discussed were: coffee, hay, wood, liquorice, marmite, leather, smoke.



Group 1: Coffee: only strong for Cig-RP1

Group 2: Coffee is not an odour attribute that can be perceived in Cig-RP1, Cig-RP4, and Coresta Monitor 7.

Group 1 and 2: Hay odour is strongly perceived in the burley sample, and not at all perceived in Cm6 and Cig-RP6 .

Group 1 and 2: Wood is most strongly in the burley sample, second rank is Cig-RP1. Wood is not perceived in the Cig-RP6 sample

Group 1: liquorice odour is not present in Cig-RP6 , Cig-RP4 and Burley

Group 2: liquorice is present in Cig-RP6 and Cig-RP4, but not in Burley

Group 1: marmite/salty is perceived in RYO-RP2 sample but not in Coresta monitor 6 and Burley

Group 2: marmite/salty not perceived in RYO-RP2, Coresta monitor 6 and Burley

Group 1 and 2: Burley is the best example for leather odour, most intense not perceived in RYO-RP2 and Cig-RP1.

Comments:

- Cig-RP6 is chocolate and hay
- Burley is more wood than smokey
- RYO-RP2 is more smokey than wood
- Coresta monitor is tea and hay
- Burley smells like forest wood
- Cig-RP1 is a little bit wood-odour
- RYO-RP2 is most salty, Burley is a little bit salty

#### Week 5: training 9 and 10

Results of panel performance can be found in Annex XII. Note that some of the attributes are not used to describe the products (product column) i.e. are not relevant. However, keep in mind that this is not only dependent of the panel performance but also on the products used. For example, red fruit odour has not been used, which makes sense because Cig-TP5 is most typical for this odour and that product was not assessed during these training test sessions but was assessed during week 3.

Attributes that were not used to describe the products (not relevant) were:

Coffee

Rooibos

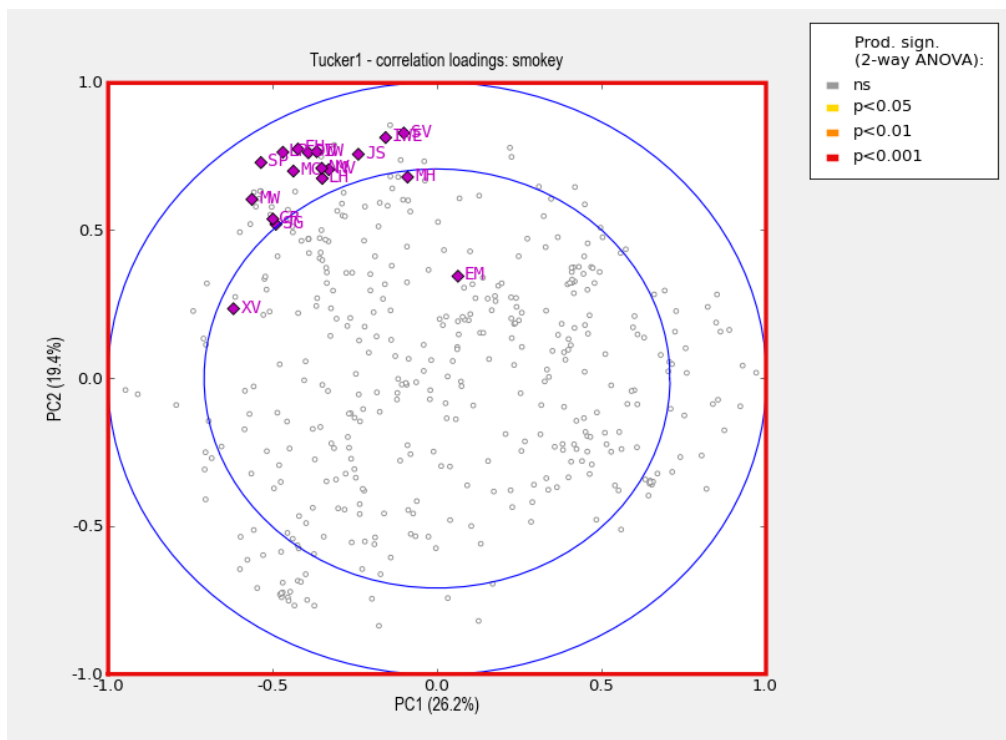
Liquorice

The panel consensus for attributes that were used to describe the products can be found in table 31. In consensus means that all panellists assessed the odour in a similar way for the same products, see figure 28. If all panellist (purple dots) are clustered together this indicates that they are in consensus if one or two falls outside the cluster, it means that these two panellists are not in agreement with the panel or do not understand the attribute. If none of the panellists cluster together it means there is no consensus and either the attribute is not used to describe any of the samples because the odour is not perceived, or all of the panellist have a different idea in mind when rating this attribute and examples of the odour may help to have panellists think and rate in a similar way.

**Table 31. Panel consensus for each attribute that was used (significantly) by the panel to describe the differences between products**

Attribute	Consensus (based on visual inspection of tucker-plots*)	Number of people that were not in agreement
Smokey-burned	Good consensus	2
Tobacco	Not in consensus	11
Vanilla	Most panellists are in consensus	6
Caramel	Good consensus	1
Coconut	Not in consensus	10
Chocolate	Most panellists are in consensus	5
Almond	In consensus	3
Old-wet hay	Not in consensus	18
hay	Not in consensus	18
Menthol/mint	Good consensus	2
Tea	Not in consensus	18
Salty	Most panellists are in consensus	6
Honey	Not in consensus	18
Wood	Not in consensus	18
leather	Not in consensus	18

Example of a tucker plot:



**Figure 28. Tucker-plot for attribute Smokey. Grey dots are all measurements, purple dots are the averages (over all products) for the individual panellists. The inner and the outer ellipse represent 50% and 100% explained variance, respectively**

A Tucker-plot (Figure 28) is a correlation loading plot, a principal component analysis based on the correlation between the panellists. It is used in situations where the panellists interpret the attributes differently, the corresponding correlation loadings (the purple dots) will be located closer to the centre (origin of the plot). This way, variables can be identified that are weak in relation to the underlying data or products. The two blue ellipses in the plot represent 50% (inner ellipse) and 100%

(outer ellipse) explained variance, which is the variance that can be explained by the correlation between the panellists. For a well-trained and calibrated panel the correlation loadings of the attributes should be close to the outer ellipse with all panellists clustered closely together (4).

From this plot (Figure 28) we can see that the attribute smokey has been used to describe differences between products. The red border of the plot indicates that this attribute was significantly different in the two-way ANOVA ( $p < 0.001$ ). Most of the panellist cluster together (are in agreement), except for panellist EM and XV.

#### Week 6: training 11

Based on the panel performance of week 5 this training's focus was on the attributes: tobacco, hay, old wet hay, tea, honey and wood as least consensus was found for these attributes. References or examples of odours (real products) were used for each attribute besides tobacco.

Products used during this training were: RYO-RP1, Cig-RP5, Cig-RP3, Cig-TP7, Virginia, Oriental, RYO-TP1, Cig-RP2. These products were chosen as they were typical for at least one of the attributes discussed during this training.

#### Tobacco

Group 1: This attribute is not needed to describe the products as it is a combination of all other attributes together. This attribute is only used to indicate if a product does not smell like tobacco (such as with the strong vanilla or menthol products). Therefore, this attribute can be removed from the attribute list.

Group 2: This attribute is a specific odour and is not the same as combining all other attributes. RYO-RP1 is the best example of tobacco, second is Cig-RP3 and RYO-TP1 has the tobacco odour too but the vanilla odour that is mostly present.

However, when the panellists were asked to indicate how tobacco was different from smokey or burned they were not able to answer this question.

#### Hay and old wet hay

Group 1: Difference remains difficult; samples used during this training do not contain a clear hay or old wet hay odour therefore these attributes should be discussed during training where all 20 products can be smelled.

Group 2: Difference is not difficult, oriental is the best example of old wet hay odour, Virginia also smells like old-wet-hay but is stronger for hay. Cig-RP3 is weak hay; Cig-RP2 is strongly hay but not old wet hay.

Old wet hay is a farm or stable like odour, Hay is a drier or dusty odour.

#### Tea

Group 1 and group 2: Cig-RP2 is best example for this attribute, Virginia also smells like tea, and Cig-TP7 does not have a tea odour.

#### Honey

Group 1 and group 2: Cig-TP7 and Cig-RP5 are both good examples for the honey attribute. Cig-RP2 does not contain a honey odour.

#### Wood

Group 1: this attribute is difficult as different wood varieties have different odours, difficult to have all panellist think of the same type of odour (i.e. wood from furniture or forest wood). Cig-RP2 is the tobacco with most wood odour, second is RYO-RP1 and Cig-TP7 does not smell like wood.

This attribute may be excluded from the attribute list as it is too complicated and it is not a very distinct odour.

Group 2: Cig-RP2 is most tobacco odour, second Cig-TP7 and third RYO-RP1. There should be a distinction between different types of wood.

Week 6: training 12

During this training panellist assessed 4 products on all attributes. They indicated the strength of each attribute with + and – (see Figure 28). They first assessed all odours for one product before continuing to the next product. Once all panellists finished the assessment of the first product a group discussion was held.

Sample Code:

Tabak-studie	-- (1)	- (2)	+/(3)	+(4)	++(5)
<b>Smokey/burned</b>		X			
<b>Tobacco</b>			X		
<b>Vanilla</b>		X			
<b>Honey</b>					

Figure 29. Example of task-form to be filled in by the panellist

Group 1:

The following conclusions were made concerning the attribute list

- Raisins are dried fruit
- Rooibos is sweet-like odour
- Honey is a sweet/sour odour like flowers/nectar
- Tobacco should not be in the attribute list
- Salty is rated equally to liquorice
- Almond is not present in these four products
- Caramel and Vanilla are always rated together, even though they are different odours
- Raisins or dried fruit are always rated together
- Flower-like odours should be rated under honey
- Rooibos is rated similar to tea

Group 2:

- Tobacco should not be on the attribute list
- Caramel is sweet and is difficult to distinguish from chocolate
- Flowers should be excluded from the attribute list
- Wood and coffee are never used to really describe a product is difficult to get everybody to think about these odours in a similar way.

Week 7: training 13 and training 14

During these training sessions all products were assessed. Panellist were instructed to first individually rate all products for one odour attribute by placing the number of the product on a line

scale. To clarify, on one 100 mm line scale products were placed based on their intensity for that particular odour (from weak to strong).

After the individual ratings the odour attribute and products that imparted this odour were discussed and placed on the right place on the scale based on a group discussion.

Conclusions concerning the attribute list:

- Salty and smoked/burned can be merged together as the same products were rated equally high for both attributes.
- Vanilla and Caramel can be merged together; it is too difficult to distinguish between the two odours in this panel that was trained for a limited number of hours.
- A better descriptor for almond is nutty
- Leather odour is not strongly perceived in any of the products. It is too difficult for the panel to distinguish products based on this attribute
- Old wet hay and hay can be merged together, difference is very small and panel is not in consensus about the difference between these two attributes
- Spices should be Clove, this is more specific and the panel did not perceive other spiced besides clove.
- Rooibos is not an odour that is clearly present in tobacco products therefore it should be merged together with tea.

**Final attribute list:**

<i>Smokey/burned</i>	<i>Honey</i>	<i>Clove</i>
<i>Vanilla/caramel</i>	<i>Liquorice</i>	
<i>Coconut</i>	<i>Hay</i>	
<i>Chocolate/cocoa</i>	<i>Red fruit</i>	
<i>Nutty</i>	<i>Menthol/mint</i>	
<i>Raisins</i>	<i>Tea</i>	

## Annex XII. Extended results of pilot study

### 1 CATA results screening I

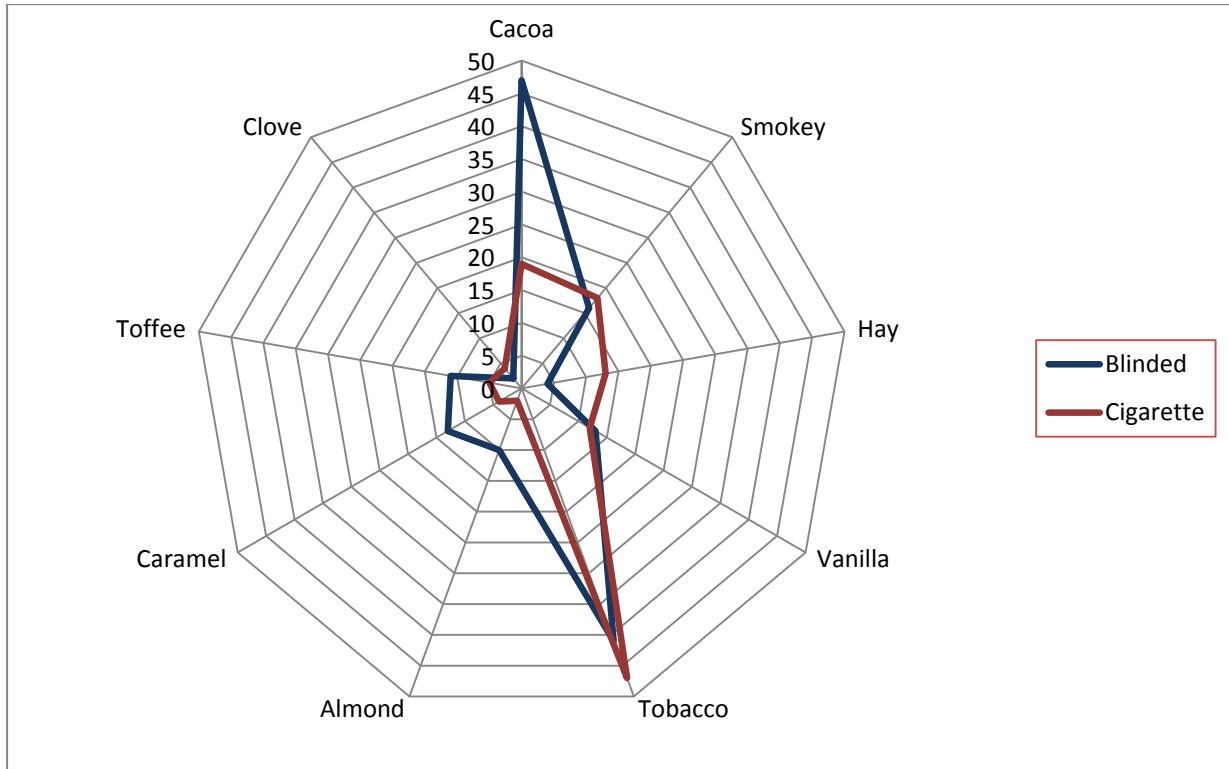


Figure 30. Cig-TP8- Chocolate

Spider plot depicting mean VAS (100mm) intensity scores for the different attributes.

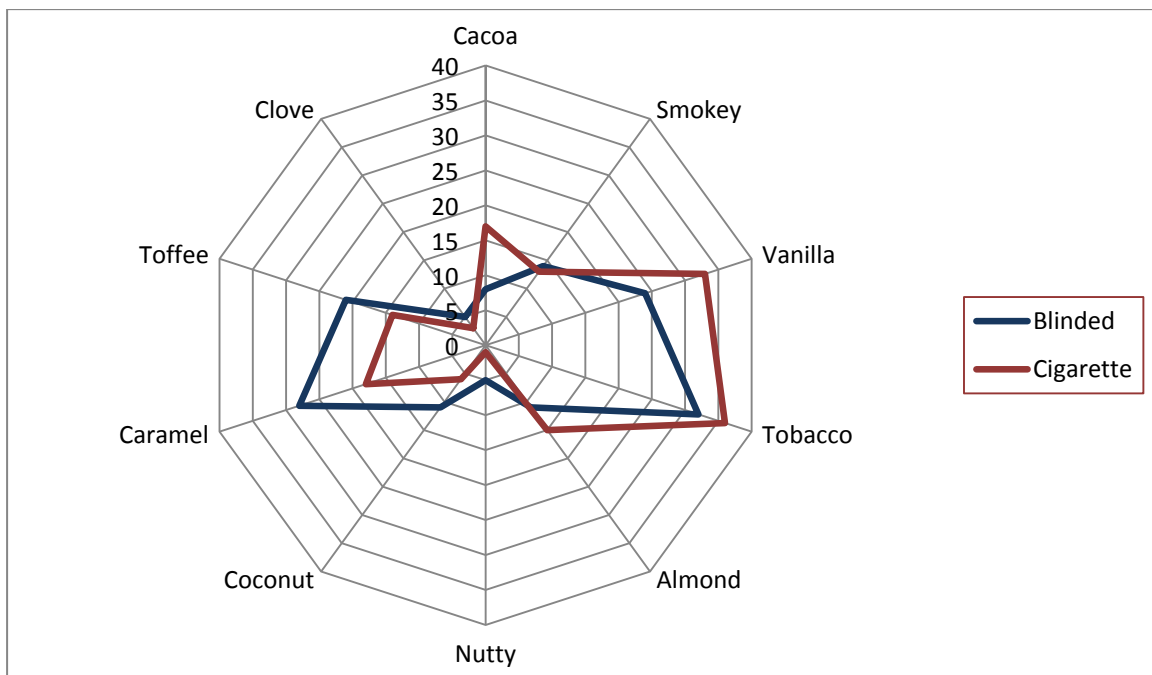


Figure 31. Cig-TP4- Vanilla

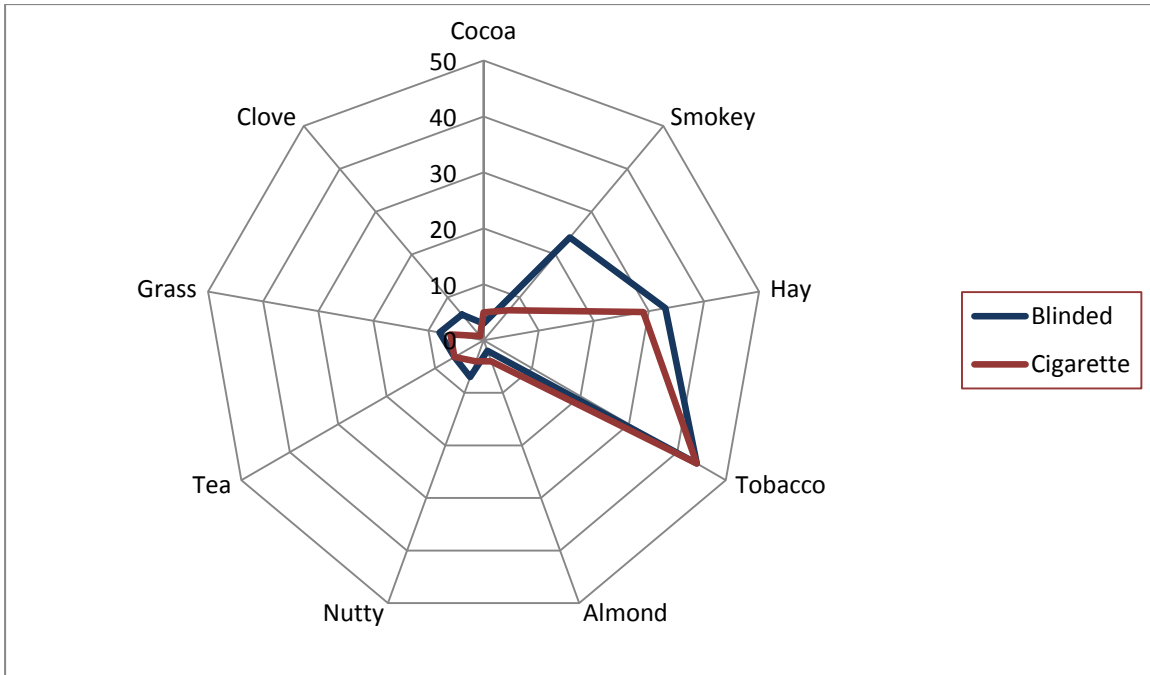


Figure 32. Cig-RP6-regular

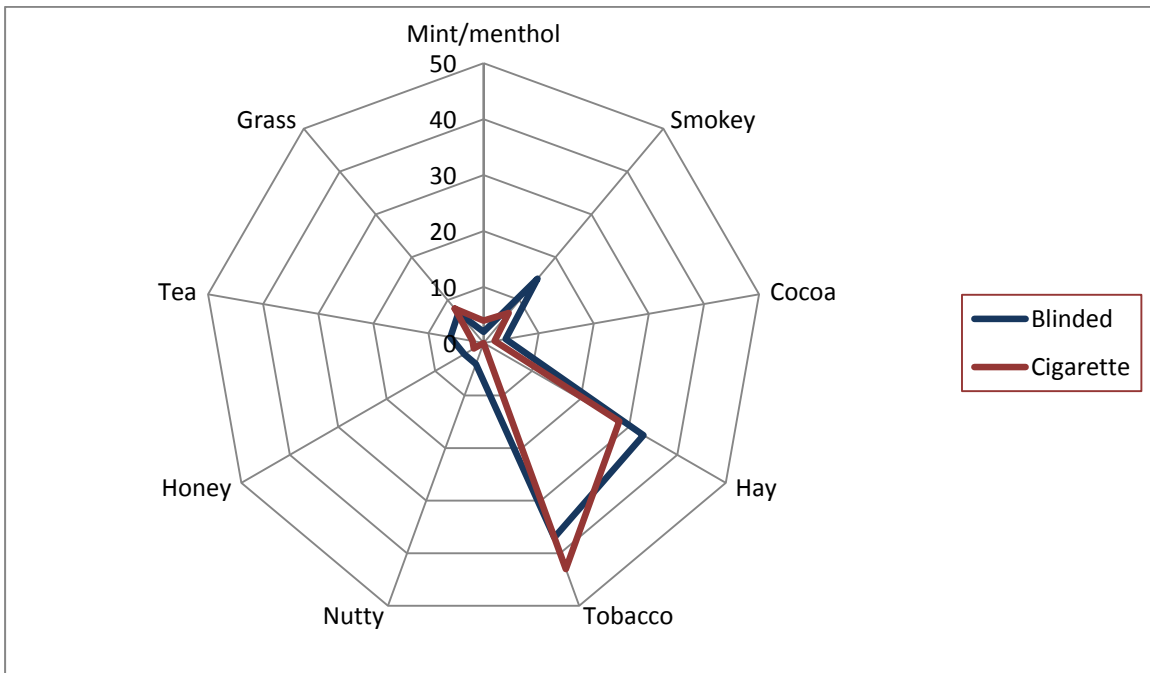


Figure 33. Cig-RP2-regular

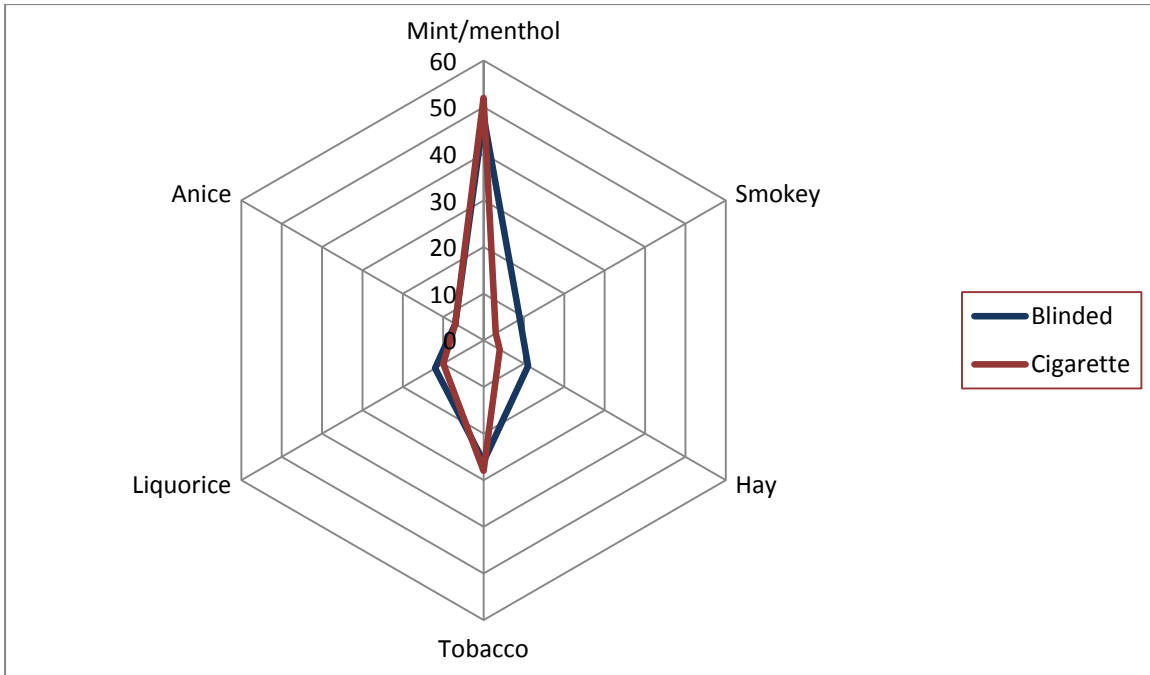


Figure 34. Cig-TP1

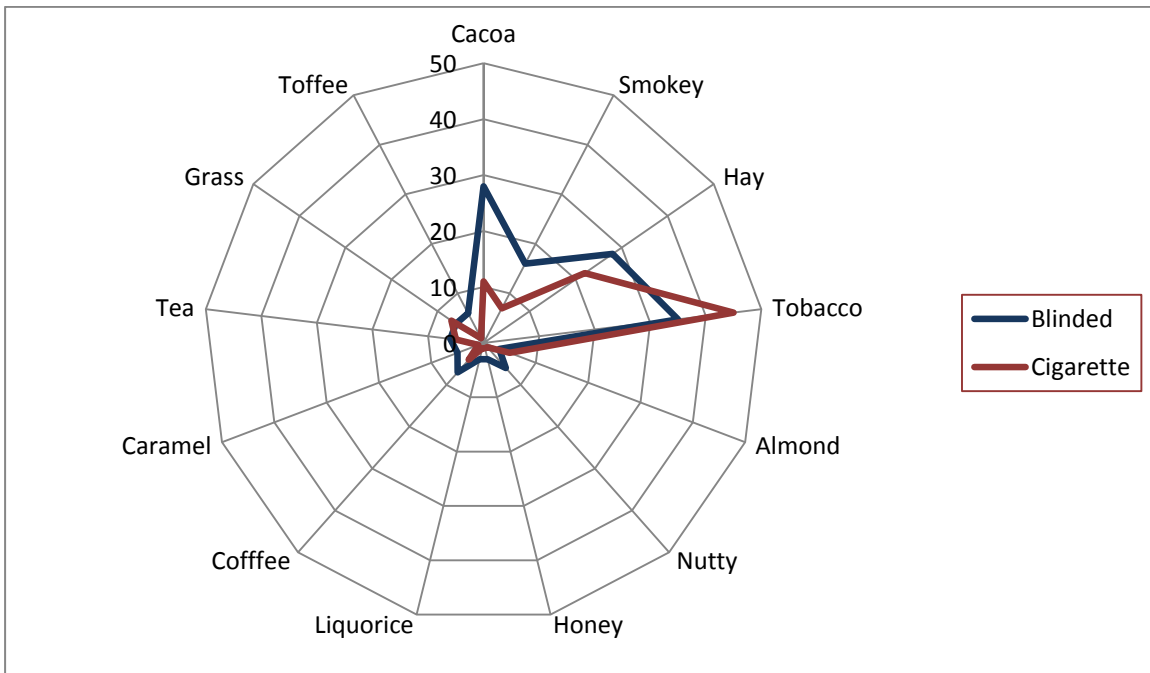


Figure 35. Cig-RP1



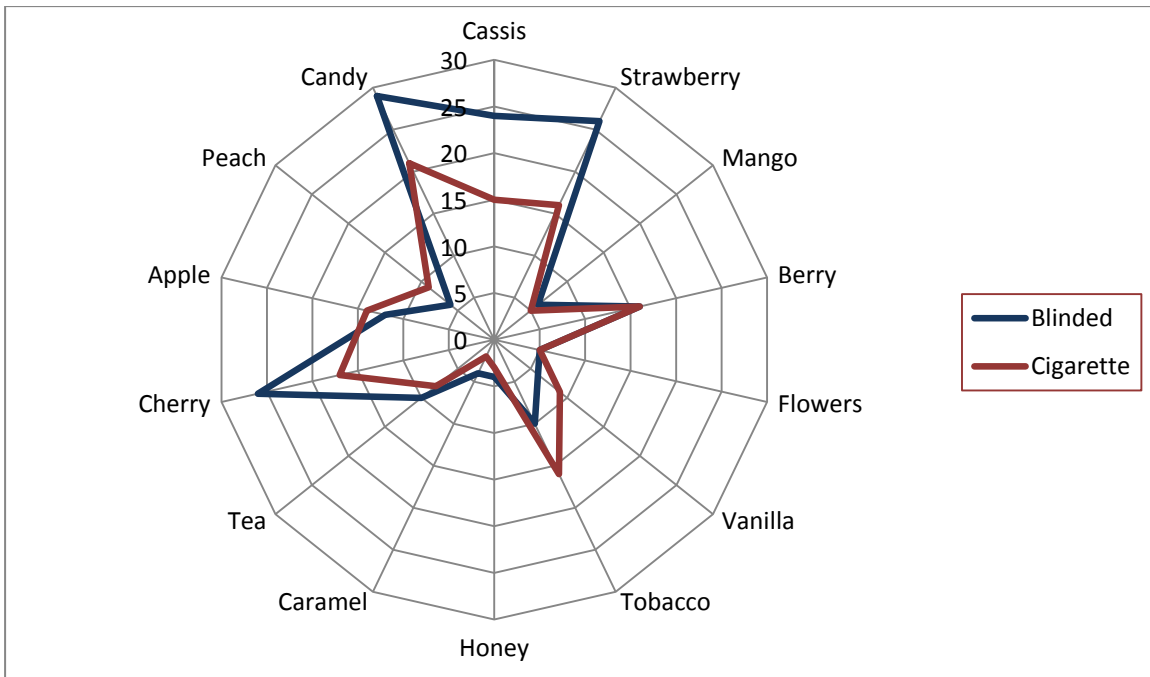


Figure 36. Cig-TP9- Apple

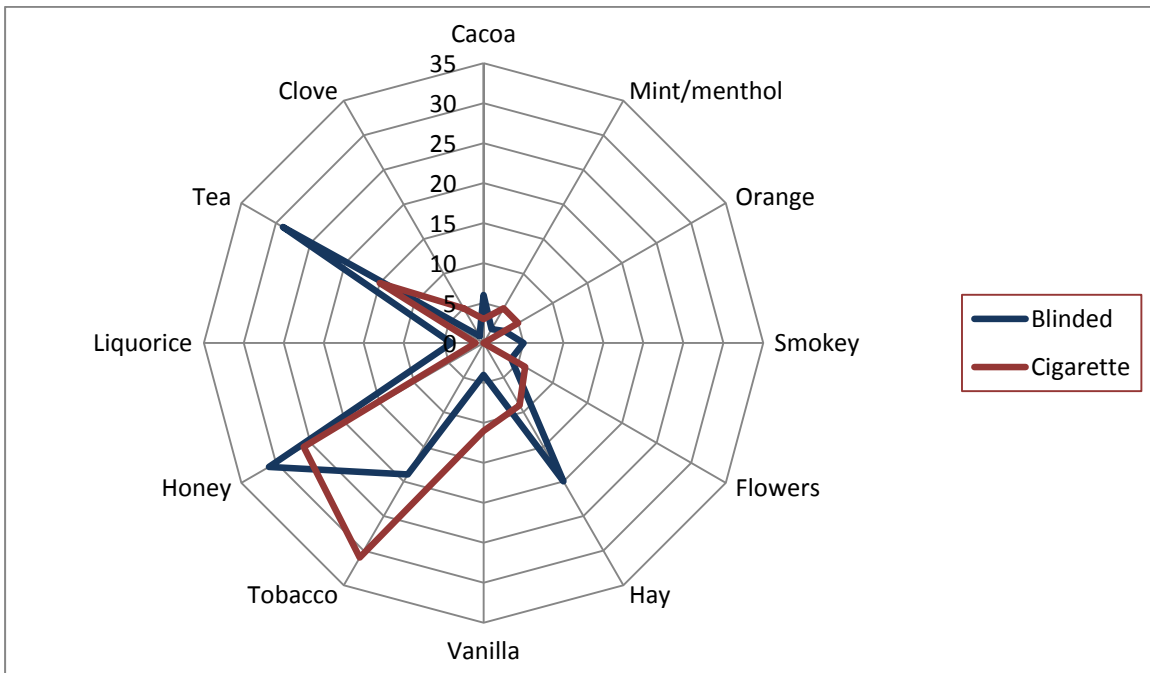


Figure 37. Cig-TP10- Honey

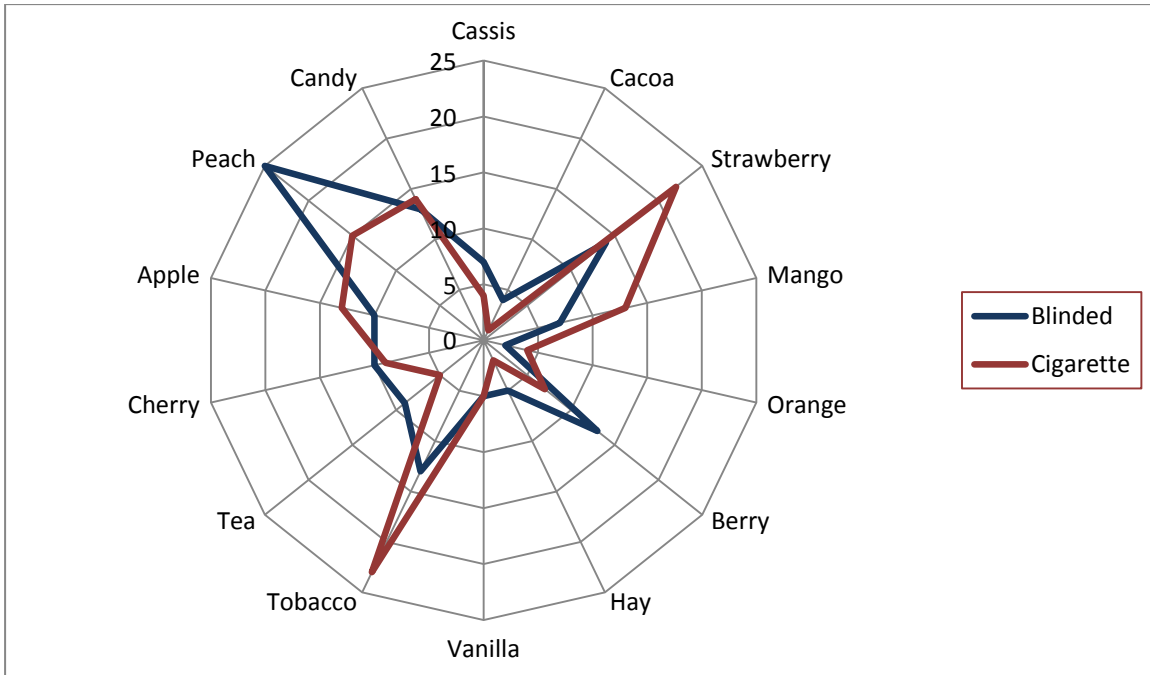


Fig 9. Cig-TP11- Mango

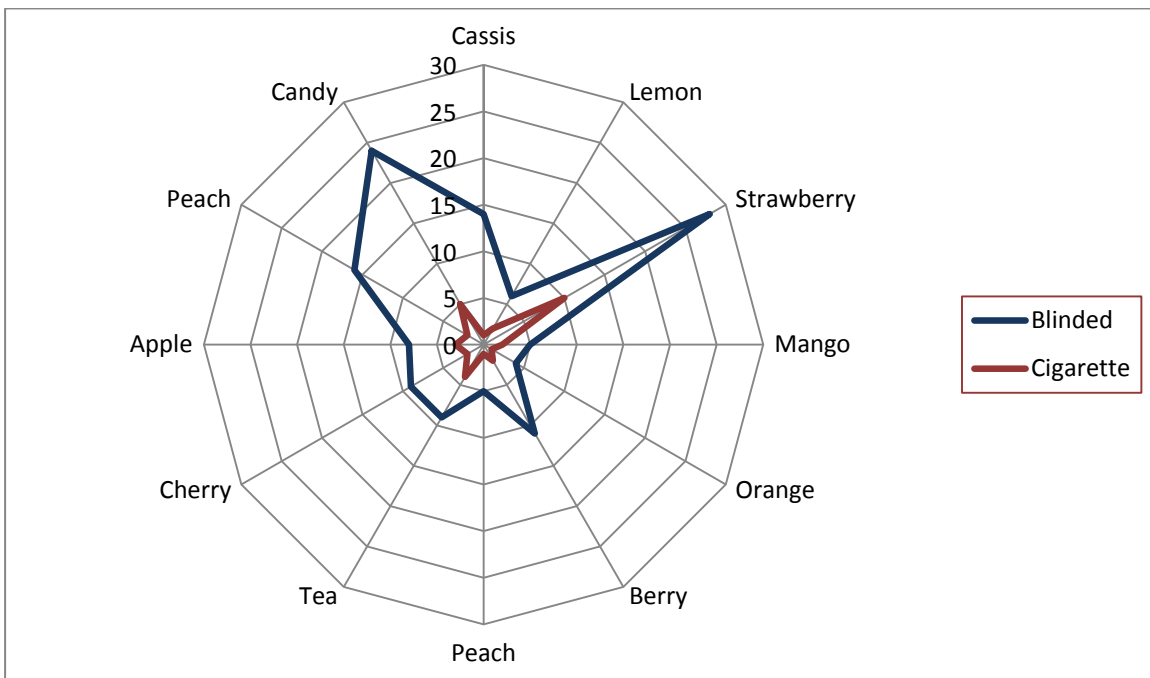


Fig. 10 Cig-TP12- Strawberry

## 2 Preliminary list of attributes generated by the panel

This list of attributes and their descriptors as found below are generated by the pilot panel.

Preliminary:

*Smoky/smoked odour*: burned and Smoky odour, including cigarette smoke

*Burned*: burned wood and sugar, 'smell of a campfire'

*Vanilla*: vanilla pod, sweet odour

*Caramel*: toffee-like odour

*Coconut*: bounty-like, sweet creamy odour

*Chocolate/cocoa*: both sweet and bitter chocolate odours

*Coffee*: including mocha

*Leather*: including shoe wax

*Rooibos*: sweet tea- like odour

*Liquorice*: All sweet liquorice, including liquorice, laurel-liquorice, and 'pottertjes'

*Salty- odour*: including marmite and salt- ammoniac

*Dried fruit*: including tutti-frutti, plum, dried apple, raisins

*Almond*: including marzipan, macaroons, amaretto

*Old- wet- hay*: stable/cowshed odour, sour-like, including vinegar

*Red fruit*: including strawberry, cherry, berries and raspberry

*Clove*: Spices

*Cinnamon*: Spices

*Hay*: dry, dusty and musty

*Menthol/mint*: including eucalyptus, toothpaste, chewing gum, Vicks and peppermint

*Tea*: all kinds including black-, green- and fruity tea

*Honey*: sweet-odour

*Wood*: mild odour, 'Ikea wood'- like (not forest wood)

*Spices*: all spices (examples; chi-tea and laurel) except for cinnamon and clove

### 3 Results first test training session week 3 panel performance

**Table 32. Results of the ANOVA Model: attribute intensity ~ Product+panellist+ Session + Product: panellist + Product: Session +panellist: Session showing the significances of the F-tests for attribute (Y), product, panellist session and their interaction. In dark grey the significant ( $p<0.05$ ) results are indicated.**

	product	panellist	session	productpanellist	product:session	panellist:session
menthol	8.424e-58	3.265e-12	0.7703	1.525e-06	0.2887	6.805e-10
smokey	9.735e-44	1.826e-36	0.05309	0.000994	4.233e-06	0.001255
tea	6.072e-21	5.457e-17	0.003538	0.1108	0.1252	0.1374
caramel	3.434e-16	2.589e-33	0.007148	0.0001358	2.145e-05	0.4175
cacao	2.678e-15	3.75e-20	0.3481	2.309e-05	0.07034	0.04807
tobacco	6.518e-15	4.223e-18	0.9785	0.4446	1.668e-05	0.757
OWhay	6.136e-14	1.468e-14	0.361	0.7178	0.3019	0.01032
vanilla	7.184e-10	1.333e-26	0.08658	0.000298	0.3848	0.003236
wood	6.068e-09	8.337e-34	0.3368	0.1396	0.4726	0.6507
honey	8.038e-07	3.477e-29	0.7231	0.00311	0.002682	0.0002803
salty	1.492e-06	1.834e-31	0.8869	0.0006425	0.9102	0.001052
dryfruit	2.463e-06	2.281e-25	0.3124	0.005282	0.01861	0.0001954
hay	6.29e-06	5.357e-21	0.4011	6.255e-12	0.002742	0.01971
leather	0.001493	2.45e-28	0.6124	0.002263	0.005156	0.1238
coffee	0.01177	1.199e-22	0.03705	0.2939	0.2772	0.7159
coco	0.02073	2.641e-19	0.1885	0.0006808	0.8628	0.02579
rbos	0.07206	2.085e-31	0.004358	1.955e-08	0.04778	0.04302
flowers	0.1227	1.222e-23	0.04495	0.004077	0.3897	0.9658
liquorice	0.1413	3.157e-45	0.5988	0.0007728	0.3931	0.009186
cinnam	0.2601	6.766e-46	0.4162	0.08318	0.2995	0.001013
clove	0.4302	1.707e-58	0.04387	7.372e-06	0.151	0.0005846
almond	0.5285	2.411e-52	0.05415	0.0001296	0.4173	8.099e-17
rfruit	0.6783	1.334e-18	0.1042	4.744e-05	0.2173	0.7669
spices	0.8777	8.645e-24	0.08926	0.6653	0.2345	0.03826

## 4 Results test-training session week 5 panel performance

**Table 33. Results of the ANOVA Model: attribute intensity ~ Product+panellist+ Session + Product: panellist + Product: Session +panellist: Session showing the significances of the F-tests for attribute (Y), product, panellist session and their interaction. In dark grey the significant (p<0.05) results are indicated**

	product	panelist	session	product:panelist	product:session	panelist:session
menthol	7.963e-25	0.02329	0.1542	0.1042	3.303e-31	0.05744
coco	5.24e-19	0.02977	0.3405	0.7541	7.619e-23	0.9869
caramel	6.348e-15	1.405e-12	0.6342	0.8988	2.786e-11	0.9856
vanilla	9.182e-15	4.348e-12	0.2888	0.5785	4.565e-18	0.9995
tobacco	2.197e-06	4.517e-15	0.1116	0.4833	2.964e-07	0.8765
OWhay	1.041e-05	1.242e-09	0.04487	0.4575	9.832e-10	0.783
hay	1.671e-05	1.374e-05	0.02978	0.8308	9.734e-12	0.9325
tea	0.00023	2.824e-10	0.1017	0.8157	1.983e-07	0.8293
almond	0.0002895	0.0003116	0.1811	0.8688	2.326e-05	0.9069
cacao	0.0003261	3.259e-13	0.7595	0.2228	0.0004962	0.4282
honey	0.01333	3.648e-12	0.04177	0.9918	0.000843	0.05976
salty	0.01541	4.09e-11	0.7081	0.3011	2.539e-06	0.07209
wood	0.02662	3.202e-17	0.2627	0.6909	0.01739	0.6087
coffee	0.07457	3.613e-15	0.3586	0.706	0.2235	0.9352
rbos	0.07465	1.474e-05	0.08749	0.9397	0.04645	0.1214
leather	0.07756	3.52e-05	0.1286	0.4782	0.009115	0.9673
flowers	0.09933	0.000243	0.4364	0.6705	0.2068	0.1466
rfruit	0.1705	2.156e-05	0.04941	0.7578	0.002535	0.9497
liquorice	0.2734	3.498e-10	0.7607	0.9565	0.1384	0.9899
dryfruit	0.3322	0.000378	0.5899	0.6508	1.832e-05	0.6213
cinnam	0.4021	2.076e-05	0.5317	0.9831	0.4909	1
clove	0.4302	4.865e-07	0.6056	0.6493	0.4302	0.5507
spices	0.4893	1.461e-10	0.1371	0.4375	0.5947	0.8322

## 5 R-Scripts

```

library (FactoMineR)
library(SensoMineR)
setwd("M:/My Documents/final test results")

tobacco_qda_experts <-read.table(file.choose(),header=T,sep="\t",dec=".")

charflav=tobacco_qda_experts
charflav
charflav$panelist<-as.factor(charflav$panelist)
charflav$product<-as.factor(charflav$product)
charflav$session<-as.factor(charflav$session)
charflav$rank<-as.factor(charflav$rank)

summary(charflav)

##panel performance
res.panelperf=panelperf(charflav,firstvar=5,formul="~product+panelist+session+product:panelist+product:session+panelist:session")
names(res.panelperf)
res.panelperf$p.value
coltable(res.panelperf$p.value[order(res.panelperf$p.value[,1]),],col.lower="gray")

##panellist performance

#Repeatability
res.paneliperf<-paneliperf(charflav, formul = "~product+panelist+session+product:panelist+product:session+panelist:session",formul.j = "~product+session", col.j = 1, firstvar = 5, synthesis = TRUE)
names(res.paneliperf)
res.magicsort<-magicsort(res.paneliperf$res.ind,method="median")
round(res.magicsort,3)
coltable(round(res.magicsort,2),level.lower=1,level.upper=2,col.lower="gainsboro",col.upper="gray")

#consensus between each panellist and the whole panel
res.magicsort<-magicsort(res.paneliperf$agree.ind,method="median")
round(res.magicsort,3)
coltable(round(res.magicsort,2),level.lower2=0.299999,level.lower=0.68,level.upper=0.69,col.lower="red",col.lower2="orange", col.upper="chartreuse3")

#discriminability
res.magicsort<-magicsort(res.paneliperf$prob.ind,method="median")
round(res.magicsort,3)
coltable(round(res.magicsort,2),level.lower=1,level.upper=2,col.lower="gainsboro",col.upper="gray")

##for sensoryattribute x which product can I consider as significantly different from some kind of average product?
Clove.lm <-
lm(formula=Clove~product+panelist+session+product:panelist+product:session+panelist:session,
data=charflav)

```

```

anova (Clove.lm)
# product effect significant; products have been differentiated regarding the Clove attribute

## are my products different regarding the sensory attributes
res.decat<- decat(charflav,formul=~product+panelist", firstvar=5, lastvar=ncol(charflav),
graph=FALSE)
names(res.decat)
res.decat$resF

# if significant products have been differentiated and some attributes are specific to some products.

# which attributes are specific for a particular product
options(contrasts=c("contr.sum","contr.sum"))
summary.lm(fruity.lm)
levels(charflav$product)
res.decat$resT$tuttifrutti

#product space
round(res.decat$adjmean,3,)

#pca, dim 1 and dim 2
res.pca <- PCA(res.decat$adjmean, axes = c(1, 2))
names(res.pca)
res.pca$eig[1:5,]
res.pca$ind$coord
res.pca$ind$contrib

#pca, dim 2 and dim 3
res.pca <- PCA(res.decat$adjmean, axes = c(2, 3))
names(res.pca)
res.pca$eig[1:5,]
res.pca$ind$coord
res.pca$ind$contrib

# clusters
res.hcpc <- HCPC(res.pca)
res.hcpc$desc.var$quanti

# ellipse 95%
res.panellipse <- panellipse (charflav, col.p=4, col.j=1, firstvar=5,level.conf = 0.95,
level.search.desc=1)

#ellipse 99%
res.panellipse <- panellipse (charflav, col.p=4, col.j=1, firstvar=5,level.conf = 0.99 level.search.desc=1)

res.panellipse$hotelling

coltable(res.panellipse$hotelling, main.title = "P-values for the Hotelling's T2 tests")

```

## HOW TO OBTAIN EU PUBLICATIONS

### Free publications:

- one copy:  
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:  
from the European Union's representations ([http://ec.europa.eu/represent\\_en.htm](http://ec.europa.eu/represent_en.htm));  
from the delegations in non-EU countries  
([http://eeas.europa.eu/delegations/index\\_en.htm](http://eeas.europa.eu/delegations/index_en.htm));  
by contacting the Europe Direct service ([http://europa.eu/eurodirect/index\\_en.htm](http://europa.eu/eurodirect/index_en.htm))  
or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (\*).

(\* ) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

### Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

### Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union  
([http://publications.europa.eu/others/agents/index\\_en.htm](http://publications.europa.eu/others/agents/index_en.htm)).





Publications Office