The Influence of Olfactory Concept on the Probability of Detecting Sub- and Peri-threshold Components in a Mixture of Odorants

Johannes H.F. Bult[1], Hendrik N.J. Schifferstein[2], Jacques P. Roozen[1,4], Alphons G.J. Voragen[1] and Jan H.A. Kroeze[3,4]

[1]Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, Wageningen University and Research Centre, Wageningen, [2]Department of Social Sciences, Marketing and Consumer Behaviour Group, Wageningen University and Research Centre, Wageningen, [3]Psychological Laboratory, Utrecht University, Utrecht and [4]Wageningen Taste and Smell Centre, Wageningen, The Netherlands

Correspondence to be sent to: Jacques Roozen, Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, Wageningen University and Research Centre, Biotechnion, PO Box 8129, 6700 EV Wageningen, The Netherlands. e-mail: jacques.roozen@chem.fdsci.wau.nl

Abstract

The headspace of apple juice was analysed to obtain an ecologically relevant stimulus model mixture of apple volatiles. Two sets of volatiles were made up: a set of eight supra-threshold volatiles (MIX) and a set of three sub-threshold volatiles. These sets were used to test the hypothesis that sub-threshold components can change the quality of a familiar smelling mixture of odorants when added to this mixture. In order to test this hypothesis, three successive dilutions of the sub-threshold volatiles were prepared in such a way that the strongest was at the threshold concentration and the two lower concentrations were below the threshold. The detection probabilities of the sub-threshold components in a blank stimulus were compared with the detectabilities in MIX. The sub- and peri-threshold volatiles were detected no better in MIX than in a blank. On the contrary, sub- and peri-threshold volatiles were better detected alone than when added to MIX. However, when the group of subjects was split into two sub-groups, employing either a rough or a detailed concept definition of the target stimulus, respectively, the subjects with highly refined concepts were better able to detect the presence of sub-threshold volatiles in MIX than those with poorly refined stimulus concepts. The effect of stimulus concept definition occurred independently of the proportions of correct detections of sub-threshold volatiles in a blank.

Introduction

In food science it is a common practice to formulate complex food aromas from their odorous constituents as identified in isolation. A widely used method in this field is gas chromatography olfactometry (GCO). This method, developed during the 1960s, is still being applied according to the design first described by Dravnieks and O'Donnell (Dravnieks and O'Donnell, 1971). In typical GCO experiments human subjects detect odorous mixture constituents by sniffing components that elute sequentially from the gas chromatograph at a sniffing port (SP). These detected constituents are assumed to be the relevant contributors to the aroma of the mixture. A general finding, however, is that the thus recomposed aroma, albeit similar, is usually not identical to the original aroma. So far, satisfactory reconstruction of a complex food aroma by means of GCO/SP identification of constituents has not been accomplished. The finding that the perceived smell of GCO reconstructed mixtures differs from the original smell while the con-centrations of the odorous constituents are identical in both mixtures is referred to here as the 'reconstruction discrepancy'.

One explanation for reconstruction discrepancy could be that components not detected by GCO/SP play a role in perception of the original mixture. However, since these components have sub-threshold intensities at the SP they are not selected for construction of the mixture. The contribution of sub-threshold components to overall perception can be understood using concept formation theory (Miller and Johnson-Laird, 1976; O'Mahony, 1991). According to this theory the extent to which a subject can discriminate between stimuli depends on refinement of the subject's conceptual representation of that particular stimulus. This is supported by the fact that the ability to discriminate between instances of a set of qualitatively different odours is positively related to the subject's familiarity with these stimuli (Rabin and Cain, 1984; Rabin, 1988; Jehl *et al.*,

1995). Therefore, we hypothesize that manipulating a familiar smelling mixture of odorants (MIX) by adding sub-threshold components will result in a mixture (MIX+) that is more easily discriminated from MIX than the sub-threshold components (BLANK+) can be discriminated from a blank stimulus (BLANK). One should note that if this hypothesis holds, the most familiar smelling mixture would be MIX+, since this is the mixture that optimally approximates the composition of the original food aroma.

Most of the available studies on olfactory mixture perception can be characterized by two features: a limited number of mixture constituents eliciting unfamiliar odour qualities (Schiet and Frijters, 1988; Laing and Glemarec, 1992; Berglund and Olsson, 1993; Olsson, 1994). In recent years, however, the importance of studying complex, familiar smelling and ecologically relevant mixtures was recognized. Livermore and Laing studied subjects' capacity to identify mixture constituents when these constituents themselves were familiar smelling, complex mixtures of odorants (Livermore and Laing, 1998). Others studied the effect of changing the concentrations of odorants in complex mixtures modelled after food aromas (Blank *et al.*, 1992; Guth and Grosch, 1994; Guth, 1997; Schieberle and Hofmann, 1997; Czerny *et al.*, 1999).

Likewise, in the present study we made up a mixture that reflects the complexity of a natural food aroma. In contrast to previous studies, however, this mixture was used to investigate the influence of adding sub-threshold components. Few cases of sub-threshold components affecting olfactory mixture perception have been reported (Guadagni *et al.*, 1963; Laska *et al.*, 1990; Laska and Hudson, 1991; Patterson *et al.*, 1993). These studies reported additive or even synergistic effects under conditions where all mixture constituents were at peri- or sub-threshold concentrations. In the present study, however, the sub-threshold components were added to a supra-threshold mixture.

A sample taken from the headspace of an apple juice dilution was used to select the constituent components for a model mixture. The components in the headspace sample were identified sensorially by GCO/SP. In parallel, the components were instrumentally identified by mass spectrometry (MS) analysis. Components identified in both analyses were used to reconstruct the original food aroma as a mixture of odorants (MIX). In addition, several components with concentration levels similar to those of the selected odorants but not detected sensorially at the SP were selected (Experiment 1). The latter components (BLANK+) constituted the sub-threshold mixture. Using the components selected in Experiment 1 we investigated the effect of adding sub-threshold components on the perceived odour quality of MIX. To test the hypothesis that the degree of odour concept refinement influences component detectability, we also investigated whether identifying the target stimulus as an apple aroma affected the probability with which MIX was discriminated from MIX+ (Experiment 2).

Experiment 1

A number of quantification methods relating the amount of an odorous component to its intensity in the mixture have been proposed in the literature. Several methods are based on subjects' direct intensity judgements. Dilution methods derive a measure of intensity from the number of dilution steps a component is above its threshold level. The detection frequency method relates the number of panellists' co-incident responses to the amount of an eluting component at the SP. Aroma extract dilution analysis (AEDA), introduced by Ullrich and Grosch (Ullrich and Grosch, 1987), is an example of the dilution method, while Pollien *et al.* and Van Ruth and Roozen used the detection frequency method (Pollien *et al.*, 1997; Van Ruth and Roozen, 1994). A hybrid method is CHARM analysis, proposed by Acree *et al.* (Acree *et al.*, 1984). Essentially a dilution method, CHARM analysis also encompasses the use of the number of coincident responses to infer a component's odour intensity. Although these methods do not psychophysically quantify odour intensities, reliable relationships between the stimulus concentrations and the number of coincident respondents have been reported (Van Ruth *et al.*, 1996; Pollien *et al.*, 1999). In this experiment we employed the detection frequency method presented by Van Ruth and Roozen (Van Ruth and Roozen, 1994).

Materials and methods

Subjects

Sixteen paid volunteers, 10 female and six male, ranging in age from 18 to 43 years, participated in the experiment. They were recruited from the local Wageningen community. All were non-smokers and none had any history of olfactory dysfunction. Participants were selected according to their performance on odour recognition and attribute generation tests, designed especially for this purpose. Subjects were naive with respect to both the nature of the stimuli and the objectives of the experiment. During the experimental sessions none of the subjects suffered from colds, allergic reactions or other adverse conditions of the respiratory tract.

Preparation of stimulus material (GCO)

Commercial quality 'Jonagold' apples, taken from a batch picked in France in October 1997 and subsequently stored for 6 months under controlled atmosphere conditions, were processed sequentially during a 3 week period. The apples were peeled and their cores were removed fresh from storage. Subsequently they were homogenized using a food processor (AEG) that yielded filtered apple juice. Three parts of this juice were diluted with two parts of distilled water. The complete process took no more than 3 min. Immediately after preparation 15 ml of the diluted sample was poured into the container of a 'purge and trap' device (Van Ruth *et al.*, 1995) and heated to 30°C. The solution

was then purged for 10 min with purified nitrogen gas (30 ml/min) while being stirred constantly at a rate of 250 r.p.m. Volatile components thus extracted from the dilution were trapped on granulated organic adsorbent material (Tenax TA, 35/60 mesh; Alltech Nederland, Zwijndrecht, The Netherlands).

Instrumental analysis

Volatiles were thermally desorbed from Tenax at 260°C for 300 s and trapped at –120°C by a cold trap/thermal desorption device (Carlo Erba TDAS 5000; Interscience, Breda, The Netherlands). Subsequently the volatiles were analysed by GC (Carlo Erba MEGA 5300; Interscience) equipped with a Supelcowax 10 capillary column with a 0.25 mm inner diameter and a length of 60 m. The oven temperature was initially 40°C for 4 min, after which it was increased to 92°C at a rate of 2°C/min, followed by an increase to 272°C at a rate of 6°C/min. The total running time was 65 min. Column effluents were split: the total flow was divided over a flame ionization detector (FID) and two sniffing ports (SP) in a 1:2:2 ratio, respectively. Desorbed volatile components were identified using combined GC/FID and SP. The chemical identities of the components were determined additionally by MS analysis with a VG MM 7070 F (Fisons Instruments, Weesp, The Netherlands) on duplicate samples.

Calibration curves for a number of identified components were determined on this GC system by transferring a series of 10 linearly incrementing amounts of every pure component dissolved in hexane to Tenax. Desorption and subsequent analysis of the components was executed using the same GC system settings as used for the apple aroma samples. The pure components used are listed in Table 2.

Sensory analysis

Subjects participated in two identical GCO sessions. The first session was regarded as a training session and, therefore, the results were discarded. During a session two subjects were simultaneously involved in analysing the effluents at a SP. A screen prevented the subjects from seeing each other. Subjects were not allowed to interact in any way. Room temperature was kept at 21°C by air conditioning. Effluents from the SP were humidified prior to presentation.

The subjects were positioned behind the SP. Keystroke responses were recorded on a laptop computer that was placed in front of them. These recordings were synchronized in time to the FID registrations. The subjects were instructed to strike a key (any key) whenever they noticed an odour at the SP. The moment they stopped perceiving the odour they had to strike a key again. The computer provided visual feedback on whether a key strike was being registered. After they indicated that no odour was being perceived anymore subjects gave a precise description of the odour smelled. Total session time was 60 min.

Data analysis

Subjects' responses during SP sessions were recorded at 1 ms intervals. Before aggregation they were corrected for GC retention time differences. Since the SP sessions were repeated with identically prepared samples, the FID profiles should be similar. Retention times, however, are subject to variation due to fluctuations in GC performance. Therefore, all chromatograms, along with the accompanying sensory time events, were matched according to the retention times of nine selected reference components in one of the FID chromatograms. Time scores of events that occurred between two reference peaks were interpolated linearly.

The number of subjects responding simultaneously at a specific time interval (1 ms) was calculated and plotted as a coincident response chromatogram (Van Ruth and Roozen, 1994). If this number was below four it was considered to be noise. The noise level was determined according to the method of J.H.F. Bult *et al.* (in preparation). With this method noise levels are estimated using the response time distributions derived from 'stimulus present' data. Thus, in order to be selected in the model mixture the volatiles had to be identified by at least four of the 16 subjects. In addition, the descriptions of the lower scores had to be similar (for instance 'sweet fruity' and 'strawberry' would be considered similar descriptors). Components not detected by the subjects but still detected by FID were selected for the set of sub-threshold components. Concentration levels of these non-perceived components were in the same range as the levels of the sensorially detected components.

Results

A typical FID chromatogram of the Jonagold aroma mixture is shown in Figure 1, along with the time-corrected, accumulated subject responses in the corresponding response chromatogram. Ten sniffing peaks scored above noise level. However, not all of these sniffing peaks coincided unambiguously with FID peaks. The first sniffing peak, located at 12 min, could not be related to any specific component. The accompanying descriptions (see Table 1) did not give any further indication for identification either.

Clear matches with respect to both response timing and consistency of odour descriptions were found for propyl acetate (2), butyl acetate (5), hexanal (6), 2-methyl-1-butyl acetate (7) and *trans*-2-hexenal (9). The numbers in parentheses refer to the indexes in Table 1 and Figure 1. The fifth SP peak, with an onset time close to 20 min, was immediately preceded by FID readings for propyl propanoate (3) and isobutyl acetate (4). Therefore, and because of the typical descriptors that accompanied these two components, they were both selected, assuming that the ensemble had been responsible for the predominantly sweet and fruity descriptions (see Table 1). Hexanol (10) was clearly present in the FID chromatogram, but was not detected in the sniffing sessions. Therefore, hexanol was
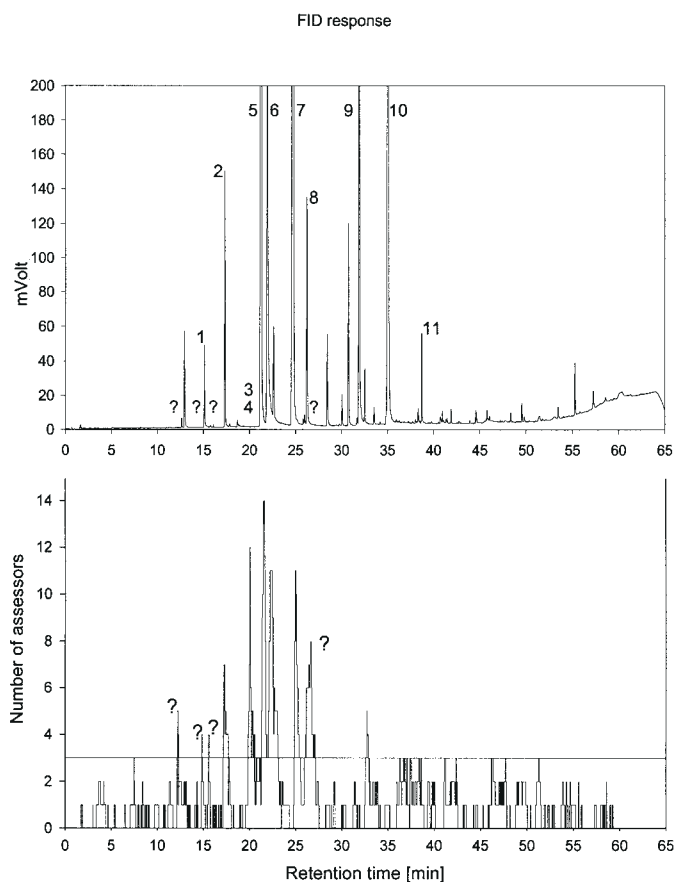
Figure 1 Combined FID-chromatogram (top) and olfactogram (bottom) of volatile components obtained from purge-and-trap apple-juice samples. The olfactogram is composed of the results from 10 different sniffing-port sessions in which 16 assessors were involved.

selected for the list of sub-threshold components. Ethanol (1) and 1-butanol (8) seem to coincide with sniffing peaks. However, the detection threshold of ethanol, 100 p.p.m. (Flath *et al.*, 1967), is considerably higher than the concentration found in this experiment, while the accompanying odour descriptions for 1-butanol are far from consistent with typical descriptions reported for this component, e.g. 'alcohol like', 'chemical' and 'paint like' (Dravnieks, 1985). Possibly ethanol and 1-butanol eluted close to low-threshold components not detected by FID and MS. Therefore, ethanol and 1-butanol were included in the list of non-detected components. Hexyl acetate (11) was also included in this list since it was not detected at the SP, although it was clearly present in the FID chromatogram. The two sniffing peaks that nearly coincided with the elution of ethanol and the sniffing peak that coincided with the elution of butanol were ascribed to the presence of components not identified by FID/MS. This selection procedure resulted in an initial set of seven components for the supra-threshold mixture and a set of four components constituting the sub-threshold mixture.

With the exception of ethanol, calibration curves for the

identified components allowed for reliable estimates of their masses at the SP (Table 1). Due to its hydrophilic nature, however, ethanol has a lower affinity for Tenax (Novák *et al.*, 1981). Even the use of a syringe to transfer an ethanol solution to the central region of the Tenax tube did not result in a precision index higher than $r^2 = 0.837$.

Estimated masses for both SP-detected and non-SP-detected components are also shown in Table 1, along with the number of GC analyses on which these estimates are based and the corresponding descriptions for SP-detected components.

Experiment 2

Materials and methods

Subjects

The panel of 16 subjects that participated in Experiment 1 was extended by seven subjects to form a panel of 23 paid volunteers, 17 female and six male (age range 18–51 years). Selection and specific requirements with respect to health and habits equalled those of Experiment 1. Subjects were naive with respect to both the nature of the stimuli and the objectives of the experiment.

Determination of partition coefficients

Partition coefficients of the 11 components dissolved in distilled water were determined for equilibrated static headspaces at 30°C. Because unexpected matrix interactions might alter partition coefficients, components were dissolved as one mixture. The headspaces were sampled from vials of 12.25 ml containing 3.0 ml of solution. They were loaded using an automated sampling unit (Fisons HS800) and subsequently injected onto a HRGC 5300 Mega Series gas chromatograph (Carlo Erba Interscience, Breda, The Netherlands) equipped with a DB-wax column (30 m × 0.542 mm). The oven temperature was initially 40°C for 10 min and was then raised to 220°C at 15°C/min, where it remained for another 3 min. Component detection was by FID. The concentrations used to determine the partition coefficients were 1, 5 and 25 p.p.m. (v/v) with duplicate samples for every concentration.

Calibration curves on this GC system were determined by manually injecting a series of 10 linearly incrementing amounts of every pure component in hexane on the DB-wax column. The temperature programme and system settings were similar to those used for the partition coefficient determinations.

Formulating the mixtures from the identified constituents

Volatile concentration levels at the SP cannot be derived directly from the amount of eluting volatiles. Concentration levels depend on the dilution of volatiles in the air immediately after their release at the SP. This dilution depends heavily on the respiratory capacity of the subject and the exact positioning of the nose. Therefore, an attempt

Table 1 Mean component masses at the SP, r^2 quality indices of the calibration curves for the GC/SP system, number of analyses on which these were based and the corresponding odour descriptions

Component	Corresponding peak in Figure 1	r^2 value of the GC-SP/FID calibration curve	Mean[a] mass (ng) at GC/SP[b]	No. of analyses	Descriptors generated during SP sessions
SP-detected components					
Unknown	0?				'sweet' 'pungent sweet' 'sweet warm musty wood' 'sulphide' 'apple sour'
Unknown	0?				'flowery spicy' 'sweet plums' 'strawberry odour' 'light sweet blossom'
Unknown	0?				'sweet dough warm vanilla' 'butter' 'butter sour' 'unpleasant'
Propyl acetate	02	0.997	48.0	15	'fresh lemon like' 'sour candy' (2×) 'sweet' 'chemical alcohol sweet' 'petrol' 'sweet strawberry candy'
Propyl propanoate[c]	03	0.996	4.7	13	'sweets' (4×) 'sweet' (2×) 'sweet sour' 'raspberry strawberry' 'strawberry lollipop' 'chewing gum' 'sweet fruity' 'flowery'
Isobutyl acetate[c]	04	0.994	23.0	15	see propyl propanoate
Butyl acetate	05	0.995	341.0	14	'glue' (5×) 'paint' 'glue paint' (3×) 'paint acetone' 'lacquer' (2×) 'sour'
Hexanal	06	0.999	286.0	14	'grass plant' 'grass' (5×) 'plant hedge' 'green grass hedge' (2×) 'apple' 'green plant'
2-Methyl-1-butyl acetate	07	0.994	335.0	15	'sweet green' 'glue' (5×) 'sweet apple-like' (2×) 'sweet sour' 'sour candy' 'lacquer paint'
Unknown	0?				'plant' (2×) 'sour apple' (2×) 'grass' 'green sweet' 'apple' 'banana'
trans-2-Hexenal	09	0.995	231.0	15	'heavy sweet' 'spicy plant' 'spicy musty' 'apple-like pricking marzipan' 'pine needles' 'rum'
Non-SP-detected components					
Ethanol[d]	01	0.837	5.9	14	
1-Butanol[d]	08	0.997	16.0	14	
Hexyl acetate	10	0.997	426.0	14	
1-Hexanol	11	0.997	21.0	14	

[a]Coefficients of variation ranged between 0.19 and 0.65.

[b]FID masses were 50% of the masses that are given here for each single SP (see text).

[c]Propyl propanoate and isobutyl acetate eluted simultaneously.

[d]Descriptors that coincided with ethanol and butanol are attributed to co-eluting non-detected, low threshold components.

was made to adjust the concentrations of the volatiles in a way that all components previously perceived at the SP were also perceivable in the static headspace of the model mixture. This implied diluting the model mixture using distilled water. In doing so we kept the concentration ratios in the mixture identical to the mass ratios in the extract.

Initially all derived aqueous dilutions were calculated as if the SP masses from Experiment 1 were present in 1 ml of air, whereas the final diluted aqueous model was obtained after a subsequent 10-fold dilution in water. At this dilution rate presumed sub-threshold components were not discriminat-

able, whereas supra-threshold components were, with the exception of hexyl acetate (10). This component, not perceived by any of the subjects at the SP, was clearly perceivable when presented in the static headspace. For this reason hexyl acetate was transferred to the set of supra-threshold components (MIX). The concentrations of the aqueous solutions of the components in the mixture are listed in Table 2.

Stimuli

For construction of the model mixture 11 components were

Table 2 Components used for the model mixture, their derived partition coefficients, aqueous dilutions based on the presumption that SP-detected masses were present in 1 ml of air and the final aqueous dilutions which were chosen in such a way that headspace detectabilities maximally mirrored the SP detectabilities

Component	Nominal purity (%)	Mean mass (ng) at GC/SP	Partition coefficient[a] (30°C, mixture)	Derived aqueous dilutions (mg/l)	Model mixture (μl/l)
Supra-threshold components					
Propyl acetate[b]	>99	48.0	1.3×10^{-2}	3.90	0.50
Propyl propanoate[c]	>99	4.7	2.1×10^{-2}	0.23	0.10
Butyl acetate[c]	>99	341.0	1.8×10^{-2}	19.00	2.50
Hexanal[c]	>98	286.0	1.3×10^{-2}	22.00	2.50
2-Methyl-1-butyl acetate[d]	>94	335.0	2.9×10^{-2}	12.00	1.50
trans-2-Hexenal[b]	>99	231.0	3.3×10^{-3}	70.00	8.00
Hexyl acetate[d]	>99	426.0	3.5×10^{-2}	12.00	1.50
Isobutyl acetate[d]	>99	23.0	2.3×10^{-2}	0.99	0.15
Sub-threshold components					
Ethanol[c]	>99	5.9	3.0×10^{-4}	22.00	2.00
1-Butanol[c]	>99	16.0	5.4×10^{-4}	30.00	3.00
1-Hexanol[c]	>98	21.0	9.5×10^{-4}	22.00	2.00

Calibration curves for all components on headspace GC/FID had precision indices ranging from $r^2 = 0.995$ to $r^2 = 1.000$.

[a]Determined for components when in mixture; partition coefficients are averages over three determinations at aqueous concentrations of 25, 5 and 1 p.p.m., respectively.

Obtained from: [b]Janssen Chimica; [c]Merck; [d]Alldrich.

used (Table 2). Solutions were made in distilled water. Four stimuli were prepared: BLANK, consisting of distilled water only; BLANK+, the sub-threshold components dissolved in water; MIX, the supra-threshold components dissolved in water; MIX+, all 11 sub- and supra-threshold components dissolved in water. These four stimuli were prepared at three concentration levels: high, the initial solution, as in the last column of Table 2; medium, 1:4 dilution of initial solution; low, 1:16 dilution of initial solution. In these dilutions the relative concentrations of the various components remained constant.

Pre-testing assured that the combination of the components in the BLANK+ mixture was at detection threshold level for the highest concentration whereas the two lower concentrations were at sub-threshold levels. The MIX mixture was of supra-threshold concentration at all three concentration levels.

Design

A 2 (complexity) × 3 (concentration) × 2 (concept) full factorial design was used. The three variables were defined in the following way.

The detectability of three sub-threshold components was studied under two different complexity conditions that varied within subjects: a simple condition in which the three sub-threshold components had to be discriminated from distilled water (BLANK+ versus BLANK) and a complex condition in which the sub-threshold components were contained in a complex mixture of eight supra-threshold components. This complex mixture had to be discriminated from the supra-threshold mixture alone (MIX+ versus

MIX). The concentration factor also varied within subjects: the comparison tasks were carried out at three concentration levels. The concept factor was a between subjects variable. Subjects were assigned to one of two groups, depending on whether they used a poorly or highly refined definition of the target stimulus (an apple aroma).

Sensory analysis (discrimination test)

Since it was not possible to predetermine any sensory attributes on which MIX and MIX+ could be distinguished, the duo–trio method was applied to measure perceived qualitative differences (Ennis, 1990, 1993). A typical duo–trio discrimination trial comprises the presentation of three stimuli: one designated a standard and, subsequently, two stimuli designated comparison stimuli. One of the comparison stimuli is identical to the standard. The subject has to decide which of the two comparison stimuli differs from the standard.

The discrimination threshold concentration for BLANK+ was defined as the concentration at which the group proportion of correct responses in discriminating BLANK+ from BLANK equalled 0.75. Using the duo–trio method a 0.75 probability of correct scores corresponds to a 0.5 probability of correct detections. The highest concentration of the BLANK+ mixture was chosen according to this criterion from pre-test results.

Stimuli were presented in 200 ml glass jars closed by a low-odour plastic screw lid that could be opened by one simple twist. Every jar contained 10 ml of solution. To minimize the possible migration of odorous components from the lids to the headspace, the lids were separated from

the jar by a sheet of aluminium foil. The stimuli in the discrimination test were prepared at least 2 h before presentation. Presentation was at ambient temperature. Every sample was used only once.

In each session a subject was presented with all stimuli from the full factorial design: concentration (3) × complexity (2). Four possible duo–trio presentation sequences were used in a randomiszed order within each of the six design cells. For each concentration level all eight duo–trio trials were carried out in succession. The subject could either receive the four MIX with MIX+ comparisons first or the four BLANK+ with BLANK comparisons. The ordering of concentration levels and the order of complexity within concentration levels was randomized over subjects and sessions.

For each concentration level a sequence of eight sets (24 jars) was presented on a tray. Subjects had to twist off the caps from the jars and remove the aluminium foil while holding the jar close to their nose. They were instructed to sniff the headspace in small sniffs after opening the jar. Subjects could proceed in a free paced manner within each trial. Between trials, however, the subjects had to rest for 60 s. A complete session for all three 'concentration' levels (24 trials) lasted no more than 1 h.

The subjects were trained for 2 h on the experimental procedure. Results from these sessions were discarded. Subsequently four experimental sessions were completed. This resulted in 16 responses per cell per subject (four sessions, four duo–trio replications each). So every subject completed 96 trials (six cells in the design, 16 replications each). The time interval between two sessions was approximately 2 weeks for every subject.

Before the first training session subjects were instructed to individually verbalize the odour qualities of the samples. Although the subjects were not informed of the nature of the stimuli, they were encouraged to form some kind of mental representation of the odour quality throughout the experiment. Following the last session subjects were again presented with a high concentration apple odour stimulus (MIX). They were then asked to verbally describe the odour impression they had. According to their responses subjects were assigned to one of two categories: the 'highly refined' category was used for respondents referring to a clear conceptual representation of the apple aroma; the 'poorly refined' category was used for non-specific or blurred qualifications. Instances of the 'highly refined' category are 'apple' or more refined definitions of this category, like 'ripened apple'. 'Poorly refined' are qualifications like 'sweet', 'fruity', 'pungent' and the like.

Data analysis

The subjects in this experiment had to discriminate between nearly identical aromas. Therefore, the aromas used can be considered a complex background against which one is asked to detect a weak stimulus (i.e. the sub-threshold components). Signal detection theory (SDT) provides a psychophysical framework to interpret results from such an experiment (Swets, 1961). SDT is built upon the Thurstonian point of view that stimulus magnitudes are projected on a psychological continuum by means of representation processes that introduce variability into that projection. The probability distribution of the resulting psychological representation is hypothesized to be Gaussian (Thurstone, 1927). The probability that a subject will discriminate between two or more stimuli relates to the probability density functions of the stimulus representations. As a result, SDT provides an index for sensory difference. This index, known as δ, is expressed in the number of standard deviations of the Gaussian distribution. It can be calculated for a specific discrimination task, depending on the sensory comparisons that the subject is assumed to make. In a duo–trio discrimination task the probability of a correct response depends on δ according to the model (David and Trivedi, 1962; Ennis, 1993)

$$P_{\text{correct}} = F(\Phi) = 1 - \Phi(\delta/\sqrt{2}) - \Phi(\delta/\sqrt{6}) + 2 \times \Phi(\delta/\sqrt{2}) \times \Phi(\delta/\sqrt{6}) \qquad (1)$$

where Φ denotes the accumulated standard normal distribution.

We calculated d' scores, empirical estimates of the perceived sensory difference, from proportions of correct responses per cell and per subject. Unfortunately, this conversion compresses all proportions below 0.5 to the exact score of 0.5, since lower proportions are not allowed in this deterministic model. However, if one wishes to analyse empirical responses in a repeated measures design, a proportion of correct responses below 0.5 can very well occur due to random variation. Therefore, the proportions of correct responses were also transformed using the logit transformation (McCullagh and Nelder, 1983), which does not compress individual scores below 0.5. Like d' conversion, this conversion yields metrically comparable data suitable for analyses of variance. The logit conversion, however, lacks the psychophysical relevance of the d' conversion. We considered both sets of transformed data in parallel.

Transformed scores were subjected to repeated measures analysis of variance (ANOVA) using SPSS 7.5 software (SPSS Inc., 1997) with concentration and complexity as within-subjects variables. To study the modulating influence of concept refinement, the transformed proportions of correct MIX versus MIX+ discriminations, irrespective of mixture concentration, were plotted as a function of the corresponding BLANK versus BLANK+ discriminations for the 'highly refined' and 'poorly refined' concept groups separately. Because all BLANK+ versus BLANK correct discrimination proportions higher than 0.75 are above threshold by definition, we discarded these observations from the analysis, together with the corresponding MIX

versus MIX+ responses. The remaining BLANK+ versus BLANK correct discrimination proportions were allocated to three categories (low, medium and high discriminability for BLANK+ versus BLANK) so that the numbers of proportion correct scores in each category were approximately equal. We tested whether the BLANK+ versus BLANK discriminability category and the concept factor had an effect on transformed proportions of correct MIX versus MIX+ discriminations by means of ANOVA. Independent variables were BLANK+ versus BLANK difference (three categories, within subjects) and concept (two categories, between subjects).

Throughout this paper $P < 0.05$ was used as the level of significance.

Results

Proportions of correct discriminations in comparing MIX with MIX+ or BLANK with BLANK+ did not vary significantly between experimental sessions for the three concentration levels. Therefore, data were aggregated over the four sessions. Mean proportions of correct discriminations between stimuli with or without sub-threshold components are plotted as a function of concentration in Figure 2. The mean proportion of correct discriminations between the highest concentration of the BLANK+ versus BLANK (i.e. 0.78) is not significantly higher than the expected proportion for threshold concentrations (i.e. 0.75) [1-tailed $Z = 1.50$, $P = 0.07$], so all concentrations are on (or below) the detection threshold level.

Mean proportions of correct responses for the BLANK+ versus BLANK comparisons are higher than mean proportions of correct responses for the MIX versus MIX+ comparisons. Furthermore, the mean correct response proportions increase with concentration for both the BLANK+ versus BLANK comparison and the MIX versus MIX+ comparison (Figure 2). A repeated measures ANOVA on the logit-transformed proportions confirmed these outcomes by significant effects for concentration [$F(2,44) = 11.6$, $P < 0.001$], complexity [$F(1,22) = 116.2$, $P < 0.001$] and their interaction [$F(2,44) = 4.2$, $P < 0.05$], which can be ascribed to the weaker effect that concentration has on discriminability in the MIX+ versus MIX comparisons compared with the BLANK+ versus BLANK comparisons. An identical test for d'-converted data revealed similar results for the effect of concentration [$F(2,44) = 12.7$, $P < 0.001$], complexity [$F(1,22) = 106.6$, $P < 0.001$] and their interaction [$F(2,44) = 3.3$, $P < 0.05$].

Figure 3 shows the proportions of correct MIX versus MIX+ detections plotted as a function of correct response proportions on BLANK+ versus BLANK comparisons for the two concept groups. ANOVA on logit-transformed proportions did not reveal an effect of BLANK+ versus BLANK proportion correct on MIX versus MIX+ proportion correct [$F(2,44) = 0.8$, $P > 0.1$]. In other words, the
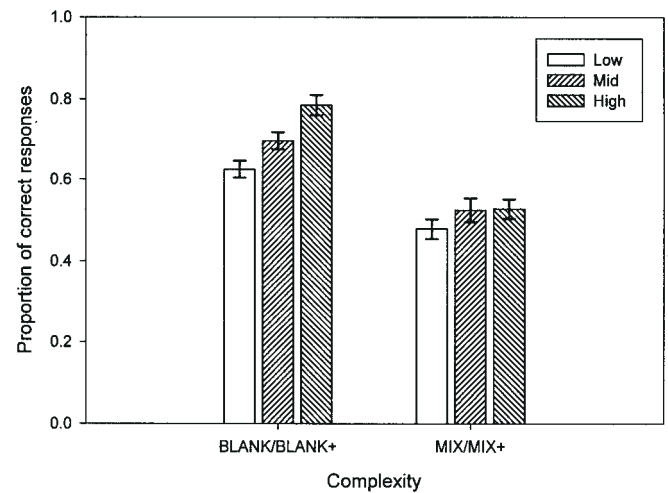

Figure 2 Proportions of correct discriminations (means ± SEM) between mixtures with either a sub- or peri-threshold set of components present or not. These proportions are shown for three different concentrations (i.e. low, mid and high) and grouped for the two paired complexity manipulations.
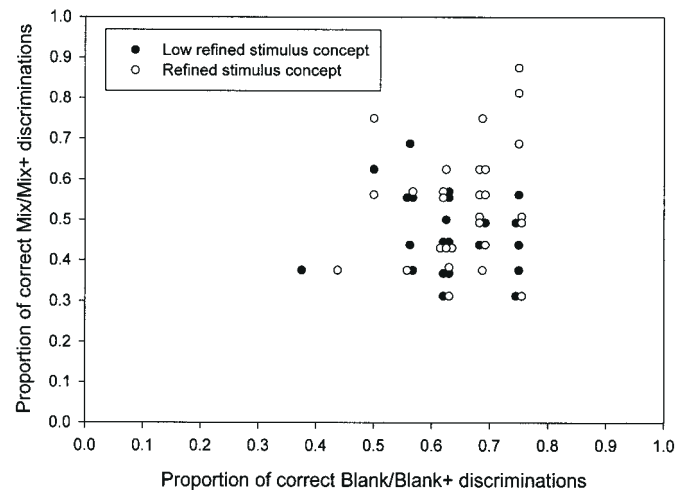

Figure 3 Proportions of correct MIX/MIX+ discriminations as a function of the proportion of correct BLANK/BLANK+ discriminations and marked for the two levels of refinement of the stimulus concept. The clustered dots represent identical proportions. The chance level for both axes is 0.5.

responses on BLANK+ versus BLANK comparisons do not predict the responses on the paired MIX versus MIX+ comparisons. However, the level of concept refinement does. Subjects having highly refined stimulus concepts (13 of 23) scored higher proportions correct on the MIX versus MIX+ comparisons than the subjects having poorly refined stimulus concepts (10 of 23) [$F(1,44) = 4.6$, $P < 0.05$]. Concept refinement was not found to interact with BLANK+ versus BLANK discriminabilities [$F(2,44) = 1.5$, $P > 0.1$].

Discussion

The first objective of the present study was to compose a mixture of odorants that reflects a natural food aroma with respect to its complexity and its perceived familiarity. Although four SP-detected peaks could not be identified by FID, we obtained a mixture generating descriptions related to apple aroma in more than half of the subjects. Since neither visual nor verbal clues were given, identification depended exclusively on the mixture's aroma. This suggests that the mixture's aroma reflected the natural apple aroma rather well. Because we failed to identify several SP-detected components, the model mixture may be expected to differ in aroma quality from the original apple aroma. Nonetheless, we consider it acceptable to use the present mixture aroma in studying possible causes for reconstruction discrepancy, since only the total absence of a distinct apple quality would interfere with the objective of relating stimulus concept refinement to the discriminability of minor aroma changes.

The probability of detecting peri- or sub-threshold components did not increase in the presence of a mixture of odorants representing an apple aroma. On the contrary, the apple aroma (MIX) decreased detectability of peri- and sub-threshold components. Although this contradicts our hypothesis that sub-threshold components could cause the reconstruction discrepancy, this outcome is not surprising if we relate it to studies that focus on sensory suppressive effects in mixtures. In both the olfactory and the gustatory domain the effect of suppression of certain odour and taste qualities by other agents is well documented (Schiet and Frijters, 1988; Berglund and Olsson, 1993; Olsson, 1994; Cain et al., 1995; Schifferstein and Kleykers, 1996). Moreover, Stevens and Traverzo observed that a multi-component mixture of tastants was more effective in masking a dissimilar tasting component than was each of the mixtures' components alone (Stevens and Traverzo, 1997). In fact, their data suggest complete masking additivity of the two masking agents used. When applied to the present study we would expect the complex apple base mixture to act as a powerful masker for the, already barely discriminatable, sub- or peri-threshold components. The observed suppression perfectly fits this expectation.

Regardless of the masking effect of MIX, the availability of a well-defined stimulus concept was found to improve the ability to discriminate between similar stimuli. This effect, however, was too small to counteract the suppressive influence of the apple aroma mixture. Note that the 'concept facilitation effect' and the 'mixture suppression effect' are unlikely to be located at the same level of stimulus processing. Mixture suppression has been attributed to various levels of interaction, ranging from the peri-receptor level (Ennis, 1996) to central processing levels (Algom and Cain, 1991; Rouby and Holley, 1995). Most likely, mixture suppression is due to concurrent peripheral and central interactions (Cain, 1975; Laing and Willcox, 1987). This suppression takes place during bottom-up information processing, i.e. the integration of information from physical stimuli resulting in a cognitive representation that allows an adequate response or further cognitive processing. In contrast, the concept facilitation effect depends on the influence of memorized sensory representations on processing of the incoming stimulus information. This is an example of top-down processing. Top-down processing is known to facilitate stimulus recognition in a number of sensory domains. For example, the positive relation between an odorants' familiarity and its discriminant ability (Rabin and Cain, 1984; Rabin, 1988; Jehl et al., 1995) may be attributed to top-down processing. Therefore, the top-down processing involved in concept facilitation is more central to and intrinsically different from the bottom-up processing involved in mixture suppression.

In this study concept formation theory was proposed as the theoretical framework by which reconstruction discrepancy can be explained. We hypothesized that sub- and peri-threshold components can change aroma quality under the condition that the subject has a well-refined stimulus concept. This hypothesis was supported in part by the results: subjects who employed a refined stimulus concept showed improved discrimination ability. Since food aromas are familiar to many subjects, we did not choose to manipulate the formation of stimulus concepts in a randomized experimental design. Instead, a quasi-experiment was designed: subjects were assigned post hoc to either of the two 'concept' groups according to their existing concept refinement. Since the sequence of effects was not controlled for, the design permits an alternative conclusion on the direction of the causal relation: good discriminators develop well-refined stimulus concepts. Nevertheless, we argue that the hypothesized causal direction, well-refined concepts make good discriminators, is the most plausible one. To compare a set of stimuli on a sensory property a subject has to employ (meta-)knowledge of stimulus properties and, therefore, discriminating processes cannot occur without the involvement of cognition. Since sensory discriminating ability starts from cognitive processing, sheer discriminators do not exist.

In the present case a general conclusion should be that peri- and sub-threshold components were not detected better in an apple-like mixture than in isolation. However, stimulus concept refinement did affect discriminability between aromas. For the case of the apple model presented here peri- and sub-threshold components could account for reconstruction discrepancy, albeit with stimulus concept as a confounding factor. The results of this study indicate that semantics are an important factor in odour mixture research because the ecological relevance of the aroma affects its sensory evaluation.

Acknowledgements

References

Acree, T.E., Barnard, J. and **Cunningham, D.G.** (1984) *A procedure for the sensory analysis of gas chromatographic effluents*. Food Chem., 14, 273–286.

Algom, D. and **Cain, W.S.** (1991) *Remembered odors and mental mixtures: tapping reservoirs of olfactory knowledge*. J. Exp. Psychol. Hum. Percept. Perform., 17, 1104–1119.

Berglund, B. and **Olsson, M.J.** (1993) *Odor–intensity interaction in binary mixtures*. J. Exp. Psychol. Hum. Percept. Perform., 19, 302–314.

Blank, I., Sen, A. and **Grosch, W.** (1992) *Potent odorants of the roasted powder and brew of arabica coffee*. Z. Lebensm. Unters. Forsch., 195, 239–245.

Cain, W.S. (1975) *Odor intensity: mixtures and masking*. Chem. Senses Flav., 1, 339–352.

Cain, W.S., Schiet, F.T., Olsson, M.J. and **De Wijk, R.A.** (1995) *Comparison of models of odor interaction*. Chem. Senses, 20, 625–637.

Czerny, M., Mayer, F. and **Grosch, W.** (1999) *Sensory study on the character impact odorants of roasted arabica coffee*. J. Agric. Food Chem., 47, 695–699.

David, H.A. and **Trivedi, M.C.** (1962) Technical Report 55. Department of Statistics, Virginia Polytechnic Institute, Blacksburg, VA.

Dravnieks, A. (1985) Atlas of Odor Character Profiles. American Society for Testing and Materials, Baltimore, MD.

Dravnieks, A. and **O'Donnell, A.** (1971) *Principles and some techniques of high-resolution headspace analysis*. J. Agric. Food Chem., 19, 1049–1056.

Ennis, D.M. (1990) *Relative power of difference testing methods in sensory evaluation*. Food Technol., 44, 114–117.

Ennis, D.M. (1993) *The power of sensory discrimination methods*. J. Sens. Stud., 8, 353–370.

Ennis, D.M. (1996) *A general molecular model for the effect of multi-component mixtures on biological systems*. Food Chem., 56, 329–335.

Flath, R.A., Black, D.R., Guadagni, D.G., McFadden, W.H. and **Schultz, T.H.** (1967) *Identification and organoleptic evaluation of compounds in delicious apple essence*. J. Agric. Food Chem., 15, 29–35.

Guadagni, D.G., Buttery, R.G., Okano, S. and **Burr, H.K.** (1963) *Additive effect of sub-threshold concentrations of some organic compounds associated with food aromas*. Nature, 200, 1288–1289.

Guth, H. (1997) *Quantitation and sensory studies of character impact odorants of different white wine varieties*. J. Agric. Food Chem., 45, 3027–3032.

Guth, H. and **Grosch, W.** (1994) *Identification of the character impact odorants of stewed beef juice by instrumental analyses and sensory studies*. J. Agric. Food Chem., 42, 2862–2866.

Jehl, C., Royet, J.P. and **Holley, A.** (1995) *Odor discrimination and recognition memory as a function of familiarization*. Percept. Psychophys., 57, 1002–1011.

Laing, D.G. and **Glemarec, A.** (1992) *Selective attention and the perceptual analysis of odor mixtures*. Physiol. Behav., 52, 1047–1053.

Laing, D.G. and **Willcox, M.E.** (1987) *An investigation of the mechanisms of odor suppression using physical and dichorhinic mixtures*. Behav. Brain Res., 26, 79–87.

Laska, M. and **Hudson, R.** (1991) *A comparison of the detection thresholds of odour mixtures and their components*. Chem. Senses, 16, 651–662.

Laska, M., Hudson, R. and **Distel, H.** (1990) *Sensitivity to biologically relevant odours may exceed the sum of component thresholds*. Chemoecology, 1, 139–141.

Livermore, A. and **Laing, D.G.** (1998) *The influence of chemical complexity on the perception of multicomponent odor mixtures*. Percept. Psychophys., 60, 650–661.

McCullagh, P. and **Nelder, J.A.** (1983) Generalized Linear Models, 1st edn. Chapman and Hall, London, UK.

Miller, G.A. and **Johnson-Laird, P.N.** (1976) Language and Perception. Cambridge University Press, London, UK.

Novák, J., Goliás, J. and **Drozd, J.** (1981) *Determination of hydrophilic volatiles in gas-aqueous liquid systems by Grob's closed-loop strip/trap method and standard-addition calibration*. J. Chromatogr., 204, 421–428.

O'Mahony, M. (1991) *Descriptive analysis and concept alignment*. In Lawless, H.T. and Klein, B.P. (eds), Sensory Science Theory and Applications in Foods. Marcel Dekker, New York, pp. 223–268.

Olsson, M.J. (1994) *An interaction model for odor quality and intensity*. Percept. Psychophys., 55, 363–372.

Patterson, M.Q., Stevens, J.C., Cain, W.S. and **Cometto-Muñiz, J.E.** (1993) *Detection thresholds for an olfactory mixture and its three constituent compounds*. Chem. Senses, 18, 723–734.

Pollien, P., Ott, A., Montigon, F., Baumgartner, M., Muñoz-Box, R. and **Chaintreau, A.** (1997) *Hyphenated headspace-gas chromatography–sniffing technique: screening of impact odorants and quantitative aromagram*. J. Agric. Food Chem., 45, 2630–2637.

Pollien, P., Fay, L.B., Baumgartner, M. and **Chaintreau, A.** (1999) *First attempt of odorant quantitation using gas chromatography–olfactometry*. Anal. Chem., 71, 5391–5397.

Rabin, M.D. (1988) *Experience facilitates olfactory quality discrimination*. Percept. Psychophys., 44, 532–540.

Rabin, M.D. and **Cain, W.S.** (1984) *Odor recognition: familiarity, identifiablility, and encoding consistency*. J. Exp. Psychol. Learn. Mem. Cogn., 10, 316–325.

Rouby, C. and **Holley, A.** (1995) *Temporal competition between odorants: effect of different time intervals on the perception of monorhinic and dichorhinic binary mixtures*. Perception, 24, 1083–1097.

Schieberle, P. and **Hofmann, T.** (1997) *Evaluation of the character impact odorants in fresh strawberry juice by quantitative measurements and sensory studies on model mixtures*. J. Agric. Food Chem., 45, 227–232.

Schiet, F.T. and **Frijters, J.E.R.** (1988) *An investigation of the equiratio-mixture model in olfactory psychophysics: a case study*. Percept. Psychophys., 44, 304–308.

Schifferstein, H.N.J. and **Kleykers, R.W.G.** (1996) *An empirical test of Olsson's interaction model using mixtures of tastants*. Chem. Senses, 21, 283–291.

SPSS Inc. (1997) SPSS Version 7.5.2. SPSS Inc., Chicago, IL.

Stevens, J.C. and **Traverzo, A.** (1997) *Detection of a target taste in a complex masker*. Chem. Senses, 22, 529–534.

Swets, J.A. (1961) *Detection theory and psychophysics: a review*. Psychometrika, 26, 49–63.

Thurstone, L.L. (1927) *Psychophysical analysis*. Am. J. Psychol., 38, 368–389.

Ullrich, F. and **Grosch, W.** (1987) *Identification of most intense volatile flavor compounds formed during autoauxidation of linoleic acid*. Z. Lebensm. Unters. Forsch., 184, 277–282.

Van Ruth, S.M. and **Roozen, J.P.** (1994) *Gas chromatography/sniffing port analysis and sensory evaluation of commercially dried bell peppers (*Capsicum annuum*) after rehydration*. Food Chem., 51, 165–170.

Van Ruth, S.M., Roozen, J.P. and **Cozijnsen, J.L.** (1995) *Volatile compounds of rehydrated French beans, bell peppers and leeks. I. Flavour release in the mouth and in three mouth model systems*. Food Chem., 53, 15–22.

Van Ruth, S.M., Roozen, J.P. and **Cozijnsen, J.L.** (1996) *Gas chromatograpy/sniffing port analysis evaluated for aroma release from rehydrated French beans (*Phaseolus vulgaris*)*. Food Chem., 56, 343–346.