A visual interface for augmented human olfactory perception in the context of monitoring air quality. - Issue 1.2.0

Winterboer, A.; Evers, V.; Groen, F.; Pavlin, G.

Publication date
2011

Document Version
Final published version

A visual interface for augmented human olfactory perception in the context of monitoring air quality

Contributors

Andi Winterboer A.K.Winterboer@uva.nl, University of Amsterdam (UvA)
Vanessa Evers v.evers@uva.nl, University of Amsterdam (UvA)
Frans Groen F.C.A.Groen@uva.nl, University of Amsterdam (UvA)
Gregor Pavlin Gregor.Pavlin@icis.decis.nl, Thales Research (Thales)

Table of Contents

Abstract

This report presents the experiments that were carried out to investigate ways in which an intelligent adaptive interface could support inhabitants in providing accurate smell descriptions.

We investigated the effect of multi-modal odor cues on human smell identification performance to inform the development of an adaptive interface for a mobile application. This involved a data elicitation study (N=429) to collect people's olfactory associations when exposed to nine sample odors. Based on these associations, we then developed a multimodal interface that offered textual, image or combined cues to augment subjects' odor perception, and 190 new subjects used the interface to identify odors. We found that participants' smell identification performance increased when the interface offered visual (image and/or text) cues for odor identification. Furthermore, participants experienced the combination of visual and textual cues as most useful and enjoyable. The results of this experiment show that human smell perception can be successfully enhanced with the help of an adaptive odor cue interface.

We have used the results of this study to develop a first prototype of an intelligent interface that automatically generates cues to assist human smell identification. This prototype is based on causal models (Bayesian Networks). We extracted these observation models for a few relevant chemicals. Due to the lack of data, all types of chemicals could not be covered. Nevertheless, we have shown that construction of models supporting detection and localization using human reports is possible.

1. Augmented human olfactory perception in the context of monitoring air quality

1.1 Introduction

Humans are very good at detecting and discerning smells. However, we are bad at identifying smells by name. In this paper, we report on an experiment investigating computer-supported human olfactory perception and classification using an adaptive mobile interface. The work is carried out in the context of crowdsourcing for environmental monitoring purposes. A distributed chemical sensor system based on Bayesian network techniques detects anomalies in the air quality and calculates hypothesized gasses in the air based on knowledge about the location and type of chemical processing factories in the area, harbor traffic data, and meteorological data. Smell descriptions from humans who live and work in the area are then gathered through an interactive mobile phone app to reduce the hypotheses and determine the location, severity, nature, and development of the pollution.

Olfactory perception is an interesting, yet underrepresented research area in human-computer interaction (HCI). Although humans are very good at odor detection (Yeshurun and Sobel, 2010 [36]), there has to date been little research on how to create interfaces making use of or supporting the human sense of smell. The research on using smell as output in human computer interaction is hindered by technological limitations. In this paper, we focus on computer-assisted human smell perception, and in particular the recognition and identification of smells, rather than using smell as system output to interact with users.

In the context of crowdsourcing for environmental monitoring, we are developing a system that collects location-based smell descriptions from human users to detect environmental pollution in the air (Winterboer, Cramer, Groen, et al. [54]). In urban-industrial areas, where people live and work near chemical factories or harbors, pollution complaints are a common occurrence (e.g., there were around 4.600 stench-related complaints in the area around the harbor of Rotterdam, The Netherlands, in 2009 alone). Concerned citizens living or working in such areas usually call the local environmental monitoring agency where they file a complaint.

In the EU FP7 project DIADEM [42] we are developing a distributed chemical detection network that will detect anomalies regarding air quality in a specific area. Whereas currently available chemical sensors ("e-noses") only detect certain chemicals, often deliver noisy data, are expensive and therefore only sparsely distributed, the system reported in this paper exploits information from chemical sensors, chemical experts and inhabitants and subsequently fuses the data obtained. By analyzing and taking into account data from static pollution measurement sensors, meteorological data (e.g. wind direction), and location based information (e.g. locations of factories and chemicals they process as well as harbor traffic data) the system calculates a number of hypotheses concerning potential chemicals detected in the air. A dedicated mobile application then contacts inhabitants on their mobile phones to receive location-based smell perceptions. The system will use these smell descriptions to reduce and re-rank the number of hypotheses in order to identify the substance that is causing the smell. In addition, the system will allow people to send smell descriptions to the system at any time. The underlying monitoring system then uses this data to create a dynamic real-time model of the geographic area to assess pollution levels of different gases at any given time.

The processed location-specific pollution data will be used by the regional environmental monitoring agency to detect pollution and act upon hazardous situations. The data will also be presented on a public map layer, such as a Google Maps overlay, that highlights (micro) areas of chemical pollution. The map could be used by inhabitants to decide which route to take to work, where to recreate, buy a house, and so on.

In the case of a serious chemical spill, users can be informed via their mobile phone and safety instructions can be issued. However, the development of such a system is highly dependent on reliable smell descriptions from members of the general public. Our focus in this paper is on providing a mobile interface that assists users in identifying the smells they perceive to increase the accuracy of the overall environmental monitoring system.

1.2 Theory of Olfactory perception

1.2.1 Olfactory perception in HCI research

Previous HCI research has explored the olfaction modality as a potential alternative to visual and auditory modalities for providing messaging notifications. Bodnar, Corbett, & Nekrasovski [17] carried out an experiment comparing these modalities as secondary display mechanisms used to deliver notifications to users working on a cognitively engaging primary task. The olfactory modality was shown to be less effective in delivering notifications than the other modalities, but produced a less disruptive effect on user engagement in the primary task. Other research has used olfactory perception to assist information retrieval. Brewster, McGookin, & Miller [18] created an olfactory interface for photo searching by allowing users to tag photos with particular smells and then use those smells to help recall the photos later. They identified some potential for smell-based tagging but also found lower performance than text tagging and challenges in the use of smell delivery devices. Ghinea and Ademoye [26] looked at the impact of enhancing multimedia applications with olfaction for information recall. They presented videos with and without matching scent stimuli and compared recall performance. They found that olfaction had a negative impact on user information recall. However, Ghinea and Ademoye used directed scents for the individual user in comparison to the ambient olfactory data used by Brewster et al, for instance.

In another line of research, environmentally appropriate scents were presented as an additional sensory modality consistent with other aspects of a virtual environment (Tortell, Luigi, Dozois, et al. [50]). Subjects were randomly assigned to receive scent during the virtual environment, and/or afterward during a task of recall of the environment. It was hypothesized that scent presentation during the virtual environment would significantly improve recall, and that subjects who were presented with scent during the recall task, in addition to experiencing the scented virtual environment, would perform the best on the recall task. They found that scent presentation during exploration of the virtual environment had a positive effect on subjects' recall of the environment, in support of the hypothesis that additional sensory modalities would in fact result in a more immersive experience.

One of the main obstacles in research on olfactory interfaces (discussed in Kaye [34,35]) concerns the lack of proper classification or description schemes for smells (other than those related to wine and perfume). The lack of a common smell 'vocabulary' makes communicating about smells difficult for both system and user.

Although experiments with olfactory interfaces have shown promising results, the lack of commonly used classification schemes and technical limitations of existing smell delivery devices have limited scientific contributions to olfactory perception in HCI research. In this paper we report research that does not concern smell as a system's output, but on augmenting human smell perception by assisting humans in what is difficult for them: identifying an odor. In order to do this we need to gain a thorough understanding of human olfactory perception. In the following chapter, we provide a brief overview of the human olfactory perception system.

1.2.2 Human Olfactory Perception

The olfactory perception system of humans is both fascinating and impressive. Odorants are volatile compounds with low molecular weight, typically organic, hydrophobic and polar (see Schiffman and Pearce [47]). When odorant molecules reach the olfactory epithelium, a small patch of tissue located at the roof of each nasal cavity, they stimulate a large population (100M) of olfactory receptor neurons (ORNs). This initiates a chain of biochemical and electrical signals that results in the sensation that we know as an odor [31].

The olfactory pathway can be divided into three general subsystems: (1) olfactory epithelium, where primary reception takes place, (2) olfactory bulb, where an organized olfactory image is created and, (3) olfactory cortex, where odor associations are stored [47]. Olfactory perception involves three basic tasks: intensity estimation, qualitative description, and hedonic tone [32]. The relationship between odorant concentration and perceived intensity is well understood, and follows a logarithmic law common to other sensory systems. When compared to intensity estimation, qualitative description of an odorant is a very difficult task. It is estimated that humans have the ability to discriminate up to 10,000 different odorants [43], though most of us consciously perceive only a fraction of these in our lifetime. Various schemes, such as Henning's odor prism (flowery, putrid, fruity, spicy, burned and resinous), have been proposed in the past in an attempt to classify odors into a small number of dimensions (see [34], for a review). Due to the lack of success of these efforts, current approaches employ odor profiling techniques, in which a large number of verbal descriptors are used to describe individual odors (e.g., [32, 43].

The smell of an odor is objective; it cannot be manipulated or influenced. However, the challenge of olfactory data is the fact that odor perception is extremely subjective, has a habit of changing, as well as being greatly influenced by age, sex, pregnancy, social and cultural factors, in addition to emotions, memory, experience and input from other sensory modalities [15, 20, 36, 45]. The hedonic tone is a qualitative property related to the pleasantness of an odorant. The hedonic quality is highly subjective, and is influenced by cultural factors and emotional associations (Wilson and Stevenson [53]). The hedonic tone of an odor is not hardwired in the brain, but shaped through experience; whether a subject likes or dislikes an odor depends largely on the associations made throughout life. Thus, the same odor may be perceived differently by separate groups of people or individuals. While one smell may be pleasant to some, it may be unpleasant to another, and neither pleasant nor unpleasant to others.

1.2.3 Odor identification for environmental monitoring

One of the main difficulties that human olfaction research faces is that although humans can detect and discriminate countless odorants, they can identify few by name (Yeshurun and Sobel, [56]). For a public environmental monitoring system based on human/social sensing this is problematic because the system is crucially dependent on accurate smell descriptions from humans. Therefore, we are investigating whether there are suitable ways to facilitate the communication of sensory information, such as olfactory perceptions.

Previous research has shown that the identification of odors (i.e. free recall of the name of an odor) without the presence of other relevant semantic information is a difficult task for humans [19, 41]. Other studies have revealed that the odor identification process works considerably better when participants can choose between possible labels, rather than free recall (e.g., De Wijk and Cain [23]).

The literature also reports strong ties between olfactory perceptions and other stimuli. For example, some studies found that olfactory perception is heavily dependent on learning, and especially personal memories (e.g., [53, 54]). According to Chu and Downes [21], for instance, there is at least preliminary evidence that olfactory stimuli can cue autobiographical memories more effectively than cues from other sensory modalities.

Yeshurun and Sobel [56] argue that the primary function of olfaction can be viewed as to signal approach or withdrawal and that this signal is best represented by pleasantness. Approach is the proper response to an edible food, a safe environment, or a fertile mate, and they all generally smell pleasant. Withdrawal is the proper response to a poison or a predator, and they generally smell unpleasant. In other words, because approach and withdrawal is the realm of olfaction, the language of olfaction is pleasantness, and an olfactory object is a given pleasantness. Yeshurun and Sobel [56] summarize further research, which indicated that humans consistently and rapidly describe an odor by its perceived pleasantness. Odorant pleasantness was the primary perceptual aspect of odor spontaneously used by subjects in olfactory discrimination tasks and was found both the most salient psychological dimension in the perception of odors and the primary perceptual aspect humans use to discriminate odorants [28, 46], or combine them into groups (Schiffman, Robinson, & Erickson, [47]). When using large numbers of verbal descriptors to describe odorants, pleasantness repeatedly emerged as the primary dimension in multidimensional analyses of the resultant descriptor space [37]. Yeshurun and Sobel conclude that findings indicate that humans are good at detecting and discriminating odorants but bad at naming odorants, and the one label they readily apply to an odor is its pleasantness.

Gottfried and Dolan [29] demonstrate that although human olfaction is unreliable, it benefits substantially from visual cues. Participants were faster and more accurate at recognizing odors when these odors appeared in the context of semantically congruent visual cues. In a study by Lyman and McDaniel [39], both odor recognition and recall of odor names were tested. The results show that subjects exposed to an odor, with no additional stimuli (free recall), achieved the lowest scores in correctly identifying odors, compared to subjects presented with relevant semantic context, such as odor-names, odor-pictures or odor-name-picture combinations, where the latter led to the most accurate answers. However, in this study participants were exposed to picture and smell stimuli for all odors involved before having to recall a single smell so that priming effects could have lead to an advantage in terms of identification for the picture and smell stimuli condition.

According to Gottfried and Dolan [29], these results are consistent with the notion that using multiple counts of stimuli, such as odor-name-picture, results in improved odor recognition, because unique features are encoded for each stimulus and multiple retrieval paths for olfactory information are formed.

Morrot et al. [40] found strong connections between olfactory perception and visual information, such as images, but also colors. Similarly, according to a study by Demattè et al. [24], humans often agree on the same color label when asked to associate an odor with a color. Gilbert et al. [27] also found strong associations between certain odors and colors and that such color-smell associations are most likely acquired and may be subject to variation between cultures. Other research, partly summarized in Köster [38], found strong links between (either positive or negative) emotions with an odor (see also [30, 33]).

1.3 Hypotheses

In the section above, different types of stimuli have been shown to assist odor recognition. The literature suggests that odors are subjectively associated with images, words, colors and personal memories. Furthermore, people tend to discriminate between smells based on how pleasant or unpleasant they find the smell. Based on these findings, we expect that cues provided by an adaptive interface (the mobile interface will display cues related to substances the system hypothesizes to exist at the users' geographical location) will facilitate accurate odor classification and lead to more accurate smell identifications in comparison to free recall (identification of an odor without stimuli). Therefore, we hypothesize that users will more accurately identify odors when selecting a cue presented by the interface rather than through free recall. This leads to hypothesis 1:

Hypothesis 1*: On-screen stimuli will improve the accuracy of user provided, real-time odor identifications compared with free recall.*

Even though people seem to readily describe odors in terms of pleasantness, pleasantness ratings may not sufficiently discernible and distinguishable to allow reliable classification by Bayesian decision networks among the multiple pollutants that could possibly cause the smell. Also, the combination of image and name associations yielded most accurate odor identification in experiments by Lyman and McDaniel [39]. The adaptive interface that is connected to the environmental monitoring system described in this paper automatically selects for presentation those cues that are relevant to the gases the system identifies as possible pollutants after an anomaly has been detected. We therefore expect that a combination of image and name stimuli will lead to more accurate identification of the smell that is perceived. This leads to the following hypothesis:

Hypothesis 2*: Users will identify odors more accurately when interacting with an interface that offers image and name stimuli to trigger odor associations compared with other stimuli.*

To test the hypotheses above, we carried out a study in two phases. In phase one, we elicited odor descriptions and associations in order to extract meaningful stimuli for human smell identification as well as to see which stimuli provide discernible rankings for automatic smell classification. In phase two, we carried out an experiment to assess which smell-related stimuli generated by the interface would lead to more accurate odor identification.

1.4. Olfactory Interface Cues Study: Phase 1 – Eliciting odor associations

In this study, we exposed participants to nine distinct smells (between subjects variable: odor) and asked them to provide multiple types of associations (within subjects variable: association; memory, color, image, and name associations) that the smell evoked.

1.4.1 Participants

51 students of a course on research methods for human-computer interaction at a large Dutch university performed the data collection in exchange for course credit. Participants were drawn

from the personal social and professional networks of these 51 students. The final sample consisted of 429 participants (180 males and 249 females). Ages ranged from 18-65 (M = 33.1, SD = 12.46). 31.6% of participants had completed high school education and 18.1% a university degree. 46.5% of participants indicated that their fathers had either college (33.9%) or university (17.7%) degrees. Also, 23.6% of mothers had a college and 10.1% a university degree. 89.2% of participants were of Dutch nationality and 91.0% indicated that their mother tongue was Dutch. We also asked where people grew up because we hypothesized that people growing up in the countryside may have different smell experiences than people growing up in the city. 40.6% of participants grew up in a suburban area in between countryside and cities, 30.6% grew up in a city, 21.6% grew up in the countryside, and, finally, 7.6% indicated that they grew up in an urban-industrial area. 79.6% of participants were non-smoking, and those who smoked indicated that they smoked an average of 10 cigarettes per week (range 1-22 cigarettes per week, M = 10.5, SD = 4.5). 3.2% of participants told us that they had a smell dysfunction. 1.6% of participants indicated they were colorblind. These subjects' responses were excluded from analysis.

1.4.2 Manipulations

The between-subjects independent variable 'odor' was manipulated by exposing each participant to one of nine odors with 'scratch & sniff' cards specifically developed by a professional manufacturer for the purpose of the study. In these cards the odorant is encapsulated and printed on paper that participants sniff after scratching. In contrast to other methods of stimuli presentation, for example specifically designed smell booths equipped with air ventilation or fans, scratch & sniff cards can be used in less restricted contexts that more closely resemble the conditions in which users would experience the smell in their day-to-day activities. By avoiding these tightly controlled laboratory environments, we could present participants with smell materials in their everyday environment. We used the following odors: rotten eggs, mercaptan (natural gas), gasoline, mildew, garlic, (freshly) cut grass, fishy, smoked wood, and cheesy.

Measures

Each participant (for one distinct smell only) was requested to provide four association types (counterbalanced to avoid order effects). Each participant was asked to provide

1. **a textual association**: 'Which word or term best describes the current smell?';
2. **a personal memory association**: 'Describe a memory you associate with the current smell';
3. **a color association**: a choice from black, white, red, green, yellow, blue, brown, purple, pink, orange, gray colors provided on the screen (based on the eleven basic colors in [16].
4. **a visual mental image the odor evoked**: 'Describe a visual image that comes to mind when you smell the odor.'.

For each association method, we measured how confident participants' felt that their provided association accurately described the odor. Moreover, because we thought people may find it more difficult to produce one association compared to another, we wanted to know about the perceived effort of providing the particular type of association. Each time participants provided an association, they were therefore asked to answer two questions:

Confidence/closeness of match: Participants rated on a seven point Likert-type scale how confident they were that the selected association best matched the odor they perceived (e.g., 'How confident are you that the image accurately describes the odor?'; scale from 1-7, 1 = 'not a close description at all' to 7 = 'very close description').

Effort: Participants rated on a seven point Likert-type scale the difficulty of providing the association ('How difficult was this task for you?').

Participants then reported their perception of intensity and pleasantness of the smell.

Intensity: Participants rated on a seven point Likert-type scale the perceived strength of the odor ('How intense was the odor?').

Pleasantness: Participants rated on a seven point Likert-type scale the perceived pleasantness of the odor ('How pleasant did you find the odor?').

In a short post-test interview, participants were asked which association method they had found easiest to provide and for what reason. They were also asked which association method they thought best described the smell and why.

1.4.3 Procedure

Participants were involved in individual sessions with one of the 51 experimenters. The sessions were highly protocolized and practiced beforehand to ensure that each of the experimenters used the exact same protocol. Verbal protocols included a brief introduction, answers to questions that could arise, a brief post-survey interview and debriefing. All of these were minimized in length and practiced prior to the study. Each experimenter was instructed to use the exact same wording. After receiving a brief standard introduction from the experimenter, the remainder of the instructions and the experiment session was guided by an online data collection tool developed in Surveymonkey (surveymonkey.com). Upon prompting from the online survey, participants received a scratch & sniff card from the experimenter with one odorant, completed the online questionnaire, and were asked the four remaining post-test interview questions. The online survey tool consisted of a welcome screen with a brief explanation of the study, a consent form and a 45-item survey. Participants could complete the survey in private. The experimenter stayed nearby should the participant have any questions, but with their back turned to the subject and they did not monitor the survey completion. Participants were told the purpose of the experiment was to examine how people recognized smells. Each participant evaluated one odor only. The order of the odor association questions was counterbalanced in the survey with the help of a script. After the data collection session was completed, the experimenter would explain in the debriefing that the goal of the study had been to collect smell associations.

Preparation for analysis

After data collection, some of the data needed coding before analysis. Coding was carried out by two separate groups of students for each type of association. An intraclass correlation coefficient of .79 to .84 confirmed the coding schemes' reliability. Image descriptions were coded as follows: First, for each association the one main object/concept that was referred to in the image description was identified (e.g., cheese). Consequently, for each odor, the frequency of the main object occurring was calculated. For instance, seven participants when confronted with the natural gas card reported to smell 'gas' and a further four reported 'chemical' to be the most suitable word association (see Table 1).

The name associations for the odors were coded with the help of the St. Croix odor descriptors wheel [50]. The St. Croix odor descriptor wheel provides eight recognized odor descriptor categories to classify odors (floral, fruity, vegetable, earthy, offensive, fishy, chemical, and medicinal). Specific odor descriptors within each category (between three and 20 per category) are listed on the outside of the wheel. The most frequent descriptions (e.g., cheese) were noted down. With help of the St. Croix odor wheel, similar descriptions were normalized to the terms provided in the odor wheel (for instance 'cheesy' became 'cheese'). Finally, frequencies were analyzed for the normalized word descriptions resulting in one or a few most common textual descriptions.

Memory associations were coded based on whether human judges thought they were positive or negative. This was done based on previous research by Grabbe, McCarthy, Brown, et al. [30] indicating that odors are strongly connected to the valence of memories. Two independent coders coded memory associations to determine negative and positive memories.

1.4.4 Phase 1 - Results

Because the data was not normally distributed, Kruskal-Wallis tests were used for analysis. The results showed that there were indeed significant differences between how intense a smell was perceived depending on which odorant participants were rating ($H(8)=49.45$, $p < 0.001$). Natural gas was perceived as the most intense smell ($M=5.7$, $SD=1.3$) and Mildew as the least intense one ($M=4.2$, $SD=4.2$).

The results of another Kruskal-Wallis test showed that there were also significant differences between pleasantness ratings ($H(8) = 101.89$, $p < 0.0001$) depending on which odorant participants were rating. According to participants, the most unpleasant odor was natural gas ($Mn=1.91$, $SD=1.12$), followed by fishy ($Mn=1.93$, $SD=1.06$), garlic ($Mn=2.74$, $SD=1.59$), rotten egg ($Mn=2.81$, $SD=1.51$), cheesy ($Mn=3.10$, $SD=1.16$), and mildew ($Mn=3.29$, $SD=1.55$). Participants perceived smoked wood ($Mn=3.39$, $SD=1.34$) and gasoline ($Mn=3.53$, $SD=1.49$) as more pleasant and freshly cut grass ($Mn=4.43$, $SD=1.56$) as most pleasant.

We then analyzed whether there were (significantly) different odor associations for each association method. If odor associations differed significantly across odors for a particular method (e.g., image associations) then that association method would be best to implement as stimuli in our system. The results of a Kruskal-Wallis test indicated that there was a significant difference between color associations ($H(8)=19.69$, $p=0.012$) depending on the odorant participants were confronted with. Thus, color associations could be used for providing feedback regarding olfactory perceptions.

Odor (frequency)	Top 3 word associations (frequency)
Cheesy (52)	Cheese (20), Kitchen waste (3), Garbage (2)
Garlic (50)	Garlic (5), Rubber/Plastic (4), Gas (2)
Gas (50)	Gas (7), Chemical (4), Gasoline (4)
Rotten eggs (47)	Burned plastic (3), Chemical (2), Herbs (2)
Fishy (47)	Feces (4), Puke (2), Industry (2)
Cut grass (46)	Cleaning supplies (9), Flowers (4), Soap (3)
Gasoline (47)	Gasoline (6), Oil (5), Turpentine (3)
Mildew (52)	Mildew (6), Plastic/Rubber (4), Rotten (2)
Smoked wood (52)	Burnt smell (11), BBQ (3), Smoke (2)

Table 1: Textual associations

Odor (frequency)	**Top 3 visual associations (frequency)**
Cheesy (52)	Cheese (24), Milk products (3), Garbage (2)
Garlic (50)	Food (9), Garlic (6), Rotten (4)
Gas (50)	Stove (11), Rotten eggs (4), Exhaust f. (3)
Rotten eggs (47)	(Food) Waste (13), Herbs (3), Industry (2)
Fishy (47)	Garbage (4), Feces (4), Fishy (3)
Cut grass (46)	Flowers (19), Fruit (7), Waste (3)
Gasoline (47)	Gas station (8), Paint (7), Engine (6)
Mildew (52)	Smoke (5), Garden (4), Farm (2)
Smoked wood (52)	Campfire (8), Burnt smell (6), BBQ (3)

Table 2: Visual associations

However, looking at the frequencies of provided color associations, we found that although some of the provided color associations fitted well with the odorant (38.5% of participants thought 'yellow' would be the best color match for 'cheesy'), 'brown' was the most often selected color for six of the nine smells, which makes color cues much less meaningful. Even though the image and name descriptions were very different for the different smells, there was too much variation in the answers to yield significance (see tables 1 and 2). We did not find significant differences between odors in terms of memory valence.

To assess the quality of a particular association method, we compared participants' confidence rating (how well the participant felt the association matched the odor) and found that they felt most confident about their color association (Mn=4.95, SD=1.39), followed by word (Mn=4.69, SD=1.61), image (Mn=4.42, SD=1.1.60) and memory associations (Mn=4.26, SD=1.65). We wanted to find out whether there is one method that people felt most confident about for all odorants, and we tested whether there was an interaction effect between association methods and smells on variables, such as confidence and effort level. To this end, we performed a factorial mixed repeated-measure ANOVA where we treated participants' confidence ratings for each association method as repeated within-subject measure and the odors as between-participant measures. Mauchly's test indicated that the assumption of sphericity had been violated (chi-square=42.96, $p<.05$), therefore degrees of freedom were corrected using Huyn-Feldt estimates of sphericity (epsilon=0.98). The results show that there was a significant interaction effect between participants' confidence ratings on the association methods and the different odors ($F(31.26, 0.98) = 1.70$, $p=.009$), meaning that confidence levels differed significantly for each method for a particular odor. Since we want to offer stimuli that will yield most accurate smell descriptions, it is important to us that users feel confident about the description/association they provide. Unfortunately, we were not able to identify an association method that people are most confident about across all smells. There was a main effect of the type of association, $F(3.91, 9)=27.99$. $p=.001$. Chi-square tests revealed that for each odor a different method was seen as yielding most accurate associations ($\chi^2(32)=47.016$, $p=.042$). Overall, participants reported from most to least confident: memory (29.9%), image (26.9%), text (25.1%), and color (18%). Similarly, from the interview responses we found that participants were most confident about their memory associations (31.9%), followed by image associations (27.9%), textual associations (25.1%), and least confident about color associations (18.0%).

We also compared how effortful participants thought each association method was and found no significant differences between methods. According to participants, the most difficult task was to find an appropriate image for an odorant (Mn=4.23, SD=1.61). The second most difficult task was the word association task (Mn=4.41, SD=1.65). The easiest task was the memory association (Mn=4.43, SD=1.69), followed by the color association task (Mn=4.92, SD=1.45).

Similarly, we carried out a mixed repeated-measure ANOVA for perceived effort. This time, Mauchly's test revealed no violation of the sphericity assumption ($p > .05$). The results show a significant interaction effect between participants' effort ratings on the association methods and the different odors, $F(32, 9)=1.50$, $p=.035$. There was also a main effect of the type of association, $F(4, 9)=18.88$. $p=.001$. Chi-square tests revealed that for each odor a different method was seen as least effortful ($\chi^2(32)=52.644$, $p=.012$). Over 45% and over 50% of participants who received the gasoline or the smoked wood odorants respectively, perceived the color association method as least effortful, whereas 29% of participants receiving a fishy odorant thought textual associations where the easiest one to provide. The general order of effort from easy to difficult: color (36%), memory (27.1%), image (20.9%), and text (15%). An evaluation of participants' responses to the interview questions revealed a similar response: 35% of participants thought providing colour associations was easiest, 27.1% thought that memory, 21.9% thought that image, 15% thought that textual associations were the easiest to provide.

To summarize, we found that participants rated the odors differently in terms of intensities and pleasantness again demonstrating that human noses are able to distinguish between and characteristics of different odors. Both survey and interview data showed that participants were most confident with their memory, image and text associations. Memory associations are individually very different and particularly difficult to present. Therefore, we decided to focus on image, text, and a combination of image and text cues, as well as pleasantness ratings in a second study to test whether and which of these stimuli types would lead to the highest accuracy and user satisfaction.

Determining cues for stimuli

We then determined which textual and image stimuli to use in phase two based on the data collected in phase one. Textual stimuli (odor names) were derived as follows: a) the actual name of the odor served as a name stimulus (e.g. natural gas), b) the textual descriptions most frequently mentioned by participants that were not the odor name (e.g. stove, rotten eggs) were also selected as name stimuli.

Image stimuli were derived by selecting images from an online database (www.gettyimages.com) for each of the main objects identified in the image coding effort. In addition, the most often used textual descriptions were gathered and images resembling these descriptions were selected from internet databases. Three researchers decided independently which of the gathered alternatives were most suitable. The alternative with more votes was chosen.

1.5 Olfactory Interface Cues Study Phase 2 – Interface generated cues to assist olfactory perception

In phase 2 of the study, we carried out an experiment to assess whether the association stimuli elicited from phase 1, automatically generated by an adaptive interface, would lead to more accurate odor identification.

1.5.1 Participants

The same 51 experimenters also assisted in data collection for this second study in exchange for course credits. Participants were again drawn from the social networks of these students but differed from the previous sample. The final sample consisted of 190 participants (106 males and 84 females). Ages ranged from 18-65 (M=32.0, SD=10.60). 35.2% of participants had completed high school education and 19.2% possessed a university degree. While 87.2% of participants were of Dutch nationality, 88.7% indicated that Dutch was their mother tongue. 85.6% of participants were non-smoking, and those who did smoke reported to smoke around 9 cigarettes per week (Range=2-25, M=9.4, SD=5.6). 2.1% of participants told us that they had a smell dysfunction. 1.2% of participants indicated they were colorblind. These participants' data was again excluded from analysis.

1.5.2 Manipulations

In the second phase of the study, the between-subjects independent variable 'odor' was manipulated by exposing each participant to one of three odors with 'scratch & sniff' cards. We focused on the three smells we considered to be most common and realistic in the environmental monitoring context: gasoline, rotten eggs, and natural gas. Also, there was considerable overlap in the descriptions received from participants in phase one for these smells (see table 1 and 2) and we therefore expected these smells to benefit most from additional stimuli to aid recognition. For the odor association stimuli, we arranged the cues derived from phase 1 (and related to the three smells) alphabetically on one screen of the online survey for easy visual access (see Table 3 and Figure 1).

Odor	**Association**	**'Similars'**
Natural gas	stove	e.g., rotten eggs, exhaust fumes
Gasoline	petrol station	e.g., paint, oil
Rotten eggs	rotten eggs	e.g., rotten vegetables, sweat

Table 3: Study 2 odor name stimuli

Each participant (for one smell only) was exposed to only one of five on-screen stimulus conditions:

- Textual: The participant could select a name that best matched the odor from a list of 25 words.
- Image: The participant could select an image that best matched the odor from a list of 25 images.
- Text/image: The participant could select a text/image that best matched the odor from a list of 25 text/image associations.
- Pleasantness: The participant could provide feedback on a 7-point Likert-type pleasantness scale: 'How pleasant do you think this odor smells?'.
- Free recall: Participants were asked to fill in the name of the odor in a textbox. They were not provided with any additional stimuli.

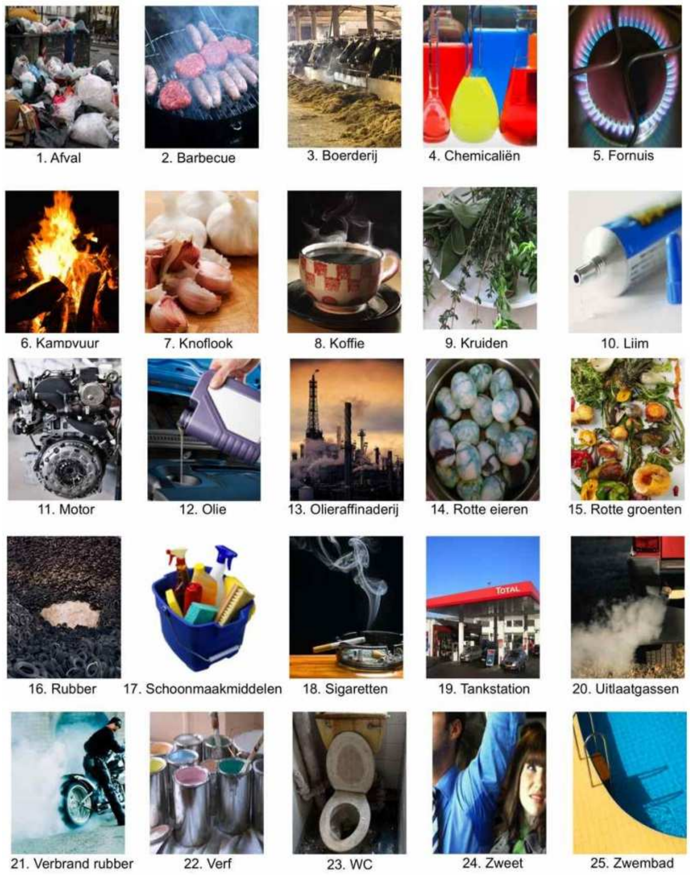

Figure 1: Visual cues presented in 2nd study

Confidence on accuracy: Participants rated on a 7-point Likert-type scale how confident they were that the selected association matched the odor they perceived (M=3.5, SD=1.6).

Satisfaction with choice: Participants rated on a 7-point Likert-type scale how satisfied they were with their choice (M=2.9, SD=1.4).

Enjoyment (based on [22]): 3 items, 7 point Likert-type scale, α=.73, M=3.5, SD=.43

Enjoy: *Did you find it enjoyable to pick the odor with this method?*

Pleasant: *Selecting an odor with this method was pleasant.*

Fun: *It was not fun to choose an odor with this method.*

Ease-of-use (based on Venkatesh, [52]): 3 items, 7-point Likert-type scale, α=.72, M=4.1, SD=.37 Ease: *It was easy to identify the odor with this method.* Problematic: *It was problematic to identify the odor with this method.* Difficult: *The correct identification of the odor was difficult.*
Usefulness (also based on Venkatesh, [52]): 3 items, 7-point Likert-type scale, α=.76, M=3.7, SD=.81 Practical: *To identify an odor this way was practical.* Efficient: *This way of identifying the correct odor went quickly.* Usefulness: *This way was useful for selecting the correct odor.*

1.5.3 Procedure

The procedure was again highly protocolized and practiced to ensure consistency across the experimenters. Verbal protocols included a brief introduction, answers to questions that could arise, and debriefing. Each experimenter was instructed to use the exact same wording. After receiving a brief standard introduction from the experimenter, the remainder of the instructions and the experiment session was again guided by the online data collection tool developed in Surveymonkey (www.surveymonkey.com). Upon prompting from the online survey, participants received a scratch & sniff card from the experimenter with one odorant and completed the online questionnaire. The online survey consisted of a welcome screen with a brief explanation of the study, a consent form, the interface with stimuli and a 20-item survey. Participants could complete the survey in private. The experimenter stayed close by should the participant have any questions, but did not monitor the survey completion. Participants were told that the purpose of the experiment was to examine how people describe smells. Each participant evaluated one odor only. Which odor association method each participant was presented with was randomized with the help of a script so that we received the same number of responses for each association method. After the data collection session was completed, participants were debriefed and thanked. Please note that participants were exposed to all the cues for each of the odors. For instance, a subject exposed to the smell 'natural gas' would see all the images that are presented in figure 1 and would click on the figure that best represented the smell they perceived. In the environmental monitoring system that is developed, the system would generate hypotheses about potential gases in the environment, each associated with a calculated probability. The users' odor identifications would help to reduce the number of hypotheses and determine the gas that is in the air. This study is designed to closely replicate such a situation.

1.5.4 Phase 2 Results

A new variable was calculated to measure user accuracy: two researchers coded all user responses. An 'accurate' score was assigned to every selection that directly represented the odorant (e.g., 'Rotten eggs' for rotten eggs) or through selection of 'similar descriptors' as identified from phase 1 (e.g., 'Oil' for 'Gasoline'). An 'inaccurate' score was assigned when the odor was not identified.

In support of Hypothesis 1, results of a Mann-Whitney test showed a significant difference in accuracy between free recall and the stimulus conditions combined ($U=1906.00$, $Z=-1.982$, $p=0.048$). When comparing accuracy means (1=direct; 2='similar', and 3=wrong), participants were most accurate in the image stimuli condition ($M=1.35$, $SD=.49$), the image and word combined condition (M=1.42, SD=.50), and then the word only condition (M=1.54, SD=.50). They were least accurate in the free recall condition without any stimuli (M=1.63, SD=.49).

Hypothesis 2 states that odors would be identified with more accuracy when word/image stimuli were provided by the interface. However, no significant difference in accuracy was found for the 4 stimulus conditions.

In partial support of hypothesis 2, perceived Enjoyment and perceived Usefulness did differ significantly between the conditions. The data for the three scales Enjoyment, Ease-of-use and Usability was not normally distributed. Kruskal-Wallis tests revealed significant differences in terms of Usefulness (H(4)=17.99, p=.001), and Enjoyment (H(5)=15.51, p=.004). No difference was found for Ease of Use. Tables 4 and 5 show how pleasant and useful participants found the stimulus conditions in order of pleasantness/usefulness. Both tables show that the image and word combined conditions was perceived as most pleasant and useful. The reason for this preference could be that participants thought they could make a well-informed choice and that selecting one of the offered image and text cues was a ‘fun’ thing to do.

Method	**rating (1=v. pleasant – 7=v. unpleasant)**
Word+picture	M=3.1, SD=1.2, range 1-6
Picture	M=3.2, SD=0.9, range 2-6
Word	M=3.6, SD=1.2, range 2-7
Pleasantness	M=3.9, SD=1.2, range 1-7
Recall	M=4.0, SD=1.5, range 1-7

Table 4. Pleasure of methods

Method	**rating (1=v. useful – 7= not at all useful)**
Word/picture	M=2.8, SD=1.1, range 1-5
Picture	M=3.1, SD=1.1, range 1-6
Word	M=3.3, SD=1.2, range 1-6
Pleasantness	M=3.7, SD=1.1, range 1-7
Recall	M=4.0, SD=1.4, range 1-7

Table 5. Usefulness of methods

The results of two Kruskal-Wallis tests revealed that there were no significant differences for satisfaction or confidence between the conditions.

1.6 Prototype of the intelligent odor identification interface agent

The results from the two previous studies indicate that interactive stimuli indeed improve human smell identification performance. The results of the second phase study show that participants thought that providing image and word stimuli was most pleasant and useful. Moreover, we found a significant difference in accuracy between free recall and the stimulus conditions combined. The smell identification accuracy was higher with provided stimuli.

1.7 Limitations

There are some limitations to this study. In selecting the images to use as stimuli in the image conditions, the experiment would have benefited from an additional step where participants rate the suitability of images to describe a particular smell. Even though we formalized verbal protocol and procedures, the use of multiple experimenters and data collection 'in the wild' will have introduced extra variables. However, the use context of the final environmental monitoring application will also differ for each participant. As future work we aim to collect responses from users without the physical presence of experimenters.

We also recognize that data collection performed by 51 students, even though we took many precautions and used standardized verbal protocols, standard procedures, and thoroughly practiced data collection, brings with it certain issues related to the consistency of experimentation. However, we believe that the large sample size should counterbalance the potential problems in terms of inter-experimenter variability so that the overall experiment outcome should not be affected.

We admit that most of the chosen smells are not representative of smells that could be experienced in urban-industrial areas. However, we believe the allowed for sufficient variation and the results of the experiments demonstrate significant differences regarding the perception of those smells. This shows that for this proof-of-concept experiment the chosen stimuli worked well enough. Nevertheless, we plan to perform a similar experiment with actually collected air in neutral smelling plastic containers.

1.8 Discussion

The results of our studies show that people will more accurately identify smells with the help of an interface that offers relevant stimuli. Participants that were presented with additional stimuli generally provided more accurate odor descriptions than those that had to provide a description without a cue. The findings also show that participants found the combination of word and picture stimuli most enjoyable and useful. We have used the results of this study to develop a first prototype of an intelligent interface that automatically generates such cues to assist human smell identification (see chapter 2). Computer-assisted human perception is an interesting field of research and the work described in this paper shows the potentially powerful ways in which human olfactory perception can be enhanced. We believe this merits further research and future work will continue to evaluate and develop interactive tools to assist human smell.

To our knowledge, the study reported here was the first large-scale user study to inform the development of a system to facilitate human smell perception and labeling in a mobile application. We have shown that computer-supported odor classification benefits from particular types of graphical user interface stimuli. Both perceived enjoyment and usefulness were highest for the picture plus word stimuli condition, which is a crucial finding for mobile applications designed for environmental monitoring that heavily relies on participation from volunteers.

Interestingly, we also found that pleasantness is indeed a differentiating indicator of smell perception. In our studies pleasantness ratings differed significantly across smells. Thus, our experiments show that both a combination of text and picture descriptors and individual pleasantness ratings are useful to accurately identify smells when the number of smells is limited (as it is the case in an industrial area) and the sample size large enough.

2. Development of a visual interface supporting human olfactory perception.

We have used the results of the previously presented study to develop a first prototype of an intelligent interface that automatically generates cues to assist human smell identification. The goal of the interface for users is to allow easy and efficient location-based reporting of stench anomalies without time-consuming calls to call centers of an environmental monitoring agency. For the Diadem system this interface could eventually mean that simultaneously a large group of volunteers could be contacted who would be able to submit smell perceptions at a specific location in a timely manner. This information then could be used for information fusion, i.e. for a combination of human-based smell data with data from sensors, meteorological data, ship traffic information and so on. Taking these sources together very reliable source detection should be feasible.

The Diadem system uses Bayesian networks to describe the causal relation between, e.g., a leakage and certain observations that can be made when a leak occurs. An example of such a Bayesian network can be seen in Figure 2. Here, the top node "Leak" contains states corresponding to situations where a certain gas is leaking from a factory. In this case these gases are H2S, mercaptan and oil/gasoline, the last states corresponds to the situation where no gas is leaking. The child node "Chem1" contains the conditional probability that one of these gases is observed given that the gas is actually present. Each node has all the necessary conditional probabilities stored in a conditional probability table (CPT). An example for the Chem1 node can be found in Table 6.

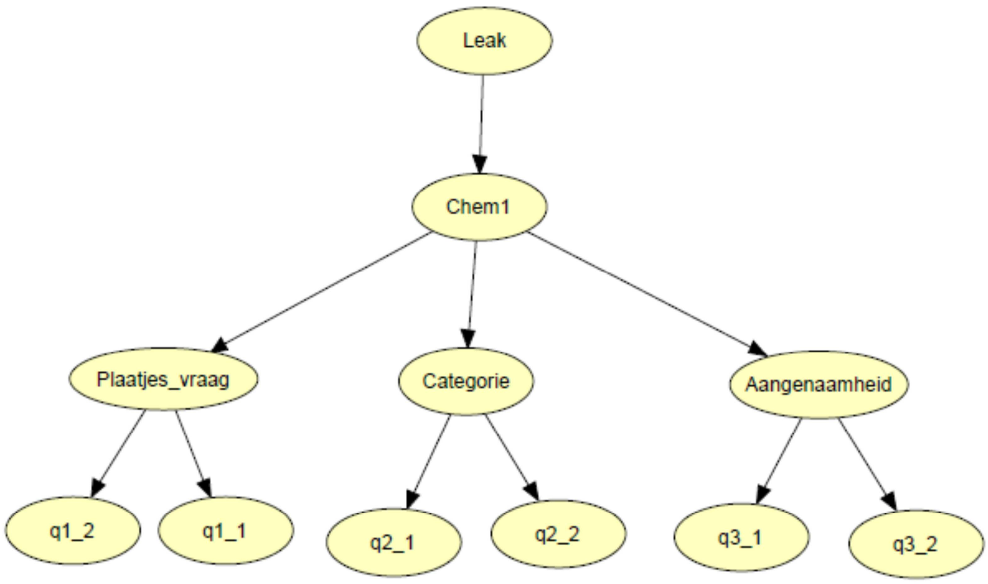

Figure 2: Example of Bayesian network

	H2S	Mercaptan	Oil/Gasoline	None
H2S	0.9	0.033	0.033	0.033
Mercaptan	0.033	0.9	0.033	0.033
Oil/Gasoline	0.033	0.033	0.9	0.033
None	0.033	0.033	0.033	0.9

Table 6: CPT for Chem 1

As can be seen from this table, the individual probabilities for each of the conditional probabilities are never 1. This means that there is always a small chance that one gas causes the same smell as another gas, because of circumstances that change the smell of the gas. All the nodes in the network contain CPTs which contain conditional probabilities, where the chances of a certain observation depend on the values of their parent nodes. The CPTs thus describe a certain level of uncertainty between the relation of observing a certain phenomenon (a human smell description) and the presence of another phenomenon (the presence of a certain gas in the air). By capturing this uncertainty in the network, the system is able to reason about an incident, while it might get some incorrect information. Note that the presented table has not been derived from the database. It serves purely for the illustration of the principles and for the evaluation of the mechanisms supporting adaptive querying. The tables that were extracted from the DCMR database can be found in chapter 3.

For our application we have developed a new Bayesian network which contains the relations between the presence of the three previously mentioned gases and the observation of answers to three types of question (see Figure 2). The questions we use for classification make use of a word/image combination, ratings of the smell on a pleasantness scale and classifying the gas into a category (see Figure 3). The first two types of questions are based on research into how people best communicate smell perceptions (see chapter 1). As stated before, people have a very acute sense of smell, but have difficulties identifying smells by name.

Based on the results of the experiments introduced in chapter 1 we were able to develop an adaptive interface of a mobile application for the Android operating system. This application is connected to the distributed environmental monitoring and decision-making system. The environmental monitoring system detects anomalies in the air quality through chemical sensors or complaints from inhabitants. It consequently calculates hypotheses concerning the gasses that are the most likely pollutants. Inhabitants in the area are contacted through their mobile phones and requested about potentially perceived smells in order to inform the detection system and to eventually support or reject hypothesis based on this newly received evidence. The adaptive interface on participating volunteers' mobile phones dynamically generates and displays visual and textual cues related to the hypothesized gasses. The user can then select the image/text combination that most accurately represents the smell they perceive.

Within the context of the overarching environmental monitoring project the goals of the mobile application are twofold: First, concerned citizens can file a complaint and inform responsible environmental monitoring agencies about unusual and unfavorable smells at their current location. Second, the system can inquire about users' smell perceptions in a potentially affected area, which are then communicated to the central detection system, in order to determine the likelihood, severity, and location of an incident.

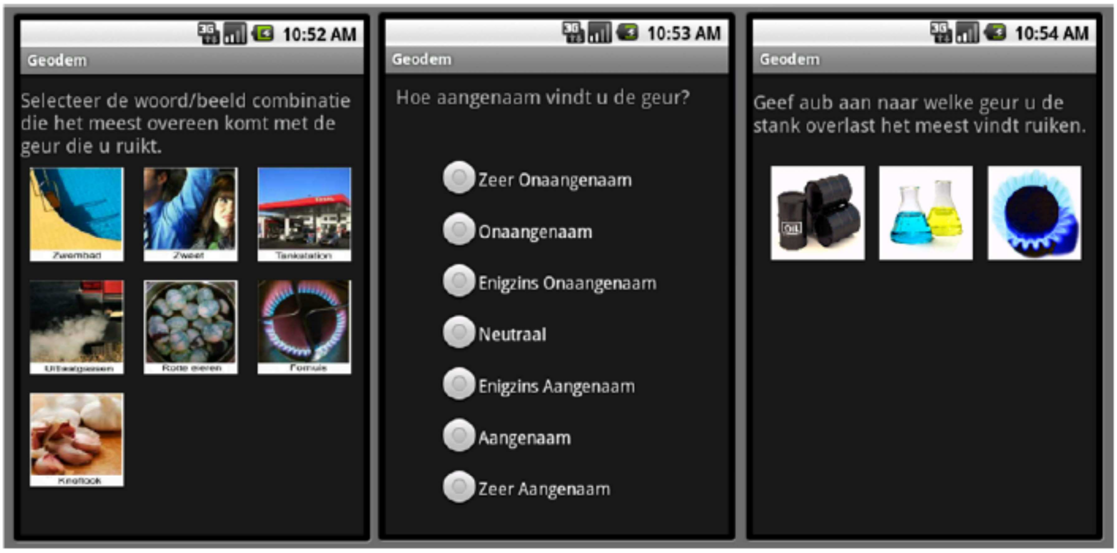

Figure 3: Screenshots of the prototype app (in Dutch)

2.1 Method

The interface offers users one of three stimulus screens (see Figure 3) based on which kind of feedback promises the highest information gain for the underlying Bayesian network-based gas detection system. (1) A 7-point pleasantness scale to indicate the subjective pleasantness of the perceived odor, (2) a screen allowing to select whether the user perceives the smell as oily, chemical, or gaseous (which allows the gas detection system to make critical computations concerning the likelihood and criticality of a gas pollution), and (3) a screen which shows visual and textual cues, as tested in this study. The selection of stimuli is based on Bayesian reasoning processes concerning possible substances derived from already available information from stationary sensors or earlier reports from concerned citizens.

As a proof of context implementation, the application is based on a simple HTTP client-server architecture implemented with the Web framework Django (www.djangoproject.com) and a trial version of the Bayesian decision-network tool HUGIN (www.hugin.com) on the server side, and Android (www.android.com) on the client side.

In Hugin we have implemented an extremely simplified form of the networks used in the Diadem project. There are two main reasons for this simplification. The first is that the network being used now is far too big and computationally intensive to use in our experiment systems. The second is that the currently used free Hugin version only allows 50 states to be used, which is reached very quickly. This also restricted us in the number of possible user interactions, the number of gases we could use as well as the possible number of users. However, as this will only be a proof of concept, adding questions later should not be problematic. The Bayesian network essentially models a factory and its potential leakages. Its topmost node contains states representing the gases it produces. The next node is a time slice node; it represents a subset of all users that live in a certain range and direction of the factory. Due to the previously mentioned restrictions we could only implement one time slice. The time slice consists of three questions which in turn consist of two users. In the Diadem system this is all done dynamically, so more users can be added later. This, however, was not feasible for our purposes, but can be implemented easily later.

The conditional probabilities entered in the Bayesian network are mostly approximations, whereas some stem from unpublished research. By iterating over all the leaf nodes of our Bayesian network, which represent (possible) answers given by our users, we can determine which question is going to give us the most discriminative information. This is done by 'clamping' a node, which means selecting a possible state/answer for that node, and observing the difference in probability for the states in the topmost node. The leaf node that shows the biggest difference and thus is the most informative is then sent back to the users in the form of a question. By answering that question, the user effectively permanently 'clamps' this node and all the probabilities in the network are updated. When asked for the most informative node again, the system will return a new (unclamped) node which, given the state the network is in, is now the most informative. This process starts up every time a person indicates that there has been an incident and via this question selection scheme speeds up the gathering of useful information. This is even more the case when many different questions are added. Also the concept of using tailor-made questions for users seems powerful. For instance different word picture combinations can be presented to a user when the system requires information about some particular examples.

2.2 Discussion

Based on these findings, we built an initial prototype of a mobile interface that allows users to report unusual and/or unpleasant odors via a combination of visual and textual cues. We expect that this kind of ‘social environmental sensing’ will considerably improve early detection of environmental incidents in the near future.

In future work, we plan to evaluate our mobile prototype in terms of usability, but also in terms of classification accuracy of provided responses. The main difficulties for user testing are potentially big differences in terms of perceived stress and urgency between actual crisis events and scenarios in which to test the system. Incident situations are difficult to setup and it is not known whether something that works in a ‘faked’ experiment setting will also work in real life.

In theory though, the implemented system should allow to gather more data in a relatively short time to further boost the performance of the underlying Bayesian network-based gas detection system.

Further future work will investigate the effect of adding elements of ‘playfulness’ to the application, for example through badges for ‘saving’ a region as rewards for useful responses. Other ideas include high scores to compete against friends, neighbors, or people in user’s social network by way of social comparison.

Finally, we plan to integrate and fuse the information provided by users of the mobile application with existing information from a distributed gas sensor network, meteorological data, and information about the location of factories and refineries to take into account as many information sources as possible for the best informed localization and estimation of environmental incidents.

3. Observation Models for Complaint Types

The modeling parameters are a critical part of each BN, i.e. the conditional probabilities defining the CPTs. In this chapter we focus on the estimation of the CPT parameters that relate variables corresponding to complaints and some observed phenomena. In other words, such a CPT captures a “sensor model” for humans. In this work we could extract such relations from the database compiled by the DCMR milieudienst in Rijnmond. The database captures complaints/reports of citizens collected during incidents with known causes.

We used this database for the extraction of the CPTs that support classification of gases. While the database was too small to be able to extract reliable models for all relevant substances, we managed to find a few useful observation models that support identification of a relevant class of gases.

The CPT parameters were estimated by using the maximum likelihood approach. In principle we counted how often complainers responded with yes/no to a question typical for a certain type of a gas family. For example "Does the smell remind you of rotten eggs?". If the answer were YES, then this would increase the likelihood of having pollution with sulfuric components otherwise the likelihood would be reduced. Such a question could be asked via a web-page or by an automated response system when complainers call a special number provided by the DCMR.

3.1 Database analysis

The DCMR database contains records of all complaints including the location of the reporter, the date and the type, subtype, and sub-subtype of smell that was reported for the period 2002 to 2010. It also contains records of incidents for the same period. If the DCMR experts were able to find the cause of the incident, the incident record contains also the location of the source and the nature of the gas being leaked.

In addition, if the DCMR experts were able to associate a complaint with the incident causing the complaint, a reference to the incident is added in the complaint record. A complaint is always associated with an incident.

However the data are recorded at the DCMR under observational and operational constraints, which can distort the statistics contained in the database in different ways:

- For some incidents, the DCMR experts were not able to assess the nature of the leaked gas, which means the ground truth is not available. These cases are discarded.
- Complaints may be associated with a case without any inspection. This can occur in case of complaint waves, a large inflow of complaints in a big incident. For efficiency reasons at the operational level, the telephone operators at DCMR record all complaints after the first ten as the same type and belonging to the big incident without further investigation. As part of the filtering, thus all complaints after the first ten associated with the same case are discarded.
- In order to be able to obtain statistically significant results, we have to use cases which have enough complaints. Based on the previous experiments using anomaly reports we estimated that the minimal number of complaints associated to a case to perform source localization with a good performance is 6. Thus cases with less than 6 complaints associated to them are discarded.

The results of the filtering described above produce the dataset (incidents and related complaints) that we used for the estimation of the CPTs.

3.2 Selection of the studied gas types

The variety of industrial activities in the area of the port of Rotterdam implies that incidents involving a large variety of gases and smells are observed. The size of the dataset is too small to allow reliable estimation of the CPTs for all types of gases. To identify the types for which computed probability may be robust enough, we need enough cases, for which reports are consistent, that is a certain ratio of the reports signal the same type of a gas or a closely related smell.

The DCMR uses a list of possible subtypes of smells. Based on these subtypes, we extracted the category that was most frequently represented among the reports associated with the case and computed the percentage of complaints falling in that subtype for this case, that is, we computed a report consistency ratio for this incident. Then for each subtype we computed the number of cases available in the dataset and determined the consistency ratio.

Subtype consistency ratio	Incidents	Complaints
>50%	70	3046
>60%	58	1650
>70%	38	1227
>80%	25	1054
>90%	16	842

Table 7: Number of incidents and associated complaints for various consistency ratios.

Subtype consistency ratio	Sulfur	Mercaptan	Solvents	Chlorine	Chemical	Oil	…
>50%	5	4	1	5	31	13	
>60%	3	2	1	5	26	12	
>70%	2	1	1	5	18	9	
>80%	2	1	0	5	11	6	
>90%	1	1	0	3	9	2	

Table 8: Number of cases for different consistency ratios for the most represented subtypes.

Subtype consistency ratio	Sulfur	Mercaptan	Solvents	Chlorine	Chemical	Oil	…
>50%	164	803	16	131	589	649	
>60%	135	155	14	123	452	564	
>70%	105	136	9	114	287	526	
>80%	102	134	2	107	197	495	
>90%	27	130	1	66	171	439	

Table 9: Number of complaints associated with different consistency ratios, for the most represented subtypes

As can be seen from these statistics, there are only a limited number of subtypes for which there are sufficiently high numbers of complaints from which CPT values can be reliably extracted. The Mercaptan category is shown as an example of a subtype for which there are many complaints, but only a small number of incidents. This indicates that the associated incidents are large. Using the filtering as indicated in section 3.1, only the first 10 complaints associated to a given incident should be taken into account. Therefore we put a lower bound of 5 on the number of incidents a certain subtype must have. In this way we consider only the subtypes for which sufficient numbers of complaints exist. Based on these numbers, we decided to use the following subtypes:

- Chemical
- Oil
- Chlorine
- Sulfur
- Other

3.3 Mapping of sub- subtype into subtype

After discussion with the experts from the DCMR it appears that, if given the choice between the limited subset of subtypes mentioned in section 3.2, human reporters will recognize some of the sub- subtypes belonging to other subtypes as belonging to the one we use.

For instance, if the subtype aromatic components are not proposed, people would associate the sub-subcategory benzene to the subtype chemical. We decided to slightly change the mapping of sub- subtypes into subtype to fit with the subset of subtypes that we use.

Appendix A shows the original hierarchy of subtypes, sub- subtypes, used by the DCMR for the type of complaint “Smell”. Table 10 shows the new mapping used between sub- subtypes and subtypes.

New subtype	Contains the sub-subtype
Chemicals	All sub-subtypes from DCMR subtype Chemical. All sub-subtypes from DCMR subtype Aromatic except “General”. All sub-subtypes from DCMR subtype Corrosive Acrid Irritant. All sub-subtypes from DCMR subtype Disinfectant. All sub-subtypes from DCMR subtype Solvent sub-subtypes from Burnt Latex and Burnt Plastic from the DCMR subtype Fire smell.
Oil	All sub-subtypes from DCMR subtypes oil.
Chlorine	All sub-subtypes from DCMR subtypes chlorine.
Sulfuric	All sub-subtypes from DCMR subtypes sulfuric.
Others	All other sub- subtypes

Table 10: mapping of existing sub-subtypes of smell to the new subtypes used for CPTs computation

Incident records provide the name of the leaked product if the DCMR experts were able to determine it. This product is used to categorized incidents based in the same subtypes as for the complaints.

3.4 CPT Computation

The CPTs can be computed by applying the following steps:

- Filter the dataset to obtain the usable dataset as described in section 3.1
- We will compute CPTs for observation models for each incident category. For each of the categories, we split up the incidents into two groups, one with incidents where the ground truth (i.e. the indicated leaked product) corresponds to the studied category and one where the ground truth does not correspond to the category being studied.
- For each incident in each group, we count the number of complaints of each subtype.

We calculate the values for P(E_i^{Smell} | *GasX*) and P(E_i^{Smell} | *Not(GasX)*) where E_i ranges over all possible subtypes. The resulting CPTs are shown in the following sections.

Chlorine Subtype

The perception model P(E_i^{Smell} |*Cl*) shown in Table 11 is based on 8 incidents for which the cause of complaints was Chlorine and 52 incidents for which Chlorine was not the cause. These 8 incidents have a total of 59 complaints attached to them of which 39 were of the subtype chlorine. The non-Chlorine incidents had 421 complaints in total.

	Chlorine related Incident = true	Chlorine related Incident = false
$E_{chlorine}^{Smell}$	0,661±0.1208	0,002±0.0046
E_{other}^{Smell}	0,339±0.1208	0,998±0.0046

Table 11: Conditional probability tables capturing relations P(E_i^{Smell} |Chlorine Gas).

Chemical Subtype

The perception model P(E_i^{Smell} |*Chemical*) in Table 12 is based on 19 incidents for which cause of the complaints was a Chemical and 41 cases that could not be attributed to a chemical. The 19 Chemical incidents have a total of 168 complaints of which 134 were of the subtype chemical.

	Chemical related incident= true	Chemical related incident= false
$E_{chemical}^{Smell}$	0,798±0.0607	0,255±0.0478
E_{other}^{Smell}	0,202±0.0607	0,245±0.0478

Table 12: Conditional probability tables capturing relations P(E_i^{Smell} |Chemical Gas).

Oil Subtype

The perception model P(E_i^{smell} |*Oil*) in Table 13 is based on 8 incidents for which cause of the complaints was Oil and 52 incidents where the oil was not the cause. These incidents corresponding to Oil have a total of 62 complaints of which 46 were of the subtype Oil.

	Oil related incident= true	Oil related incident= false
E_{oil}^{Smell}	0,742±0.1089	0,083±0.0265
E_{other}^{Smell}	0,258±0.1089	0,917±0.0265

Table 13: Conditional probability tables capturing relations P(E_i^{Smell} |Oil).

Sulfuric Subtype

We extracted the perception model P(E_i^{Smell} |*S*) shown in Table 14. The estimation of the parameters in Table 14 was based on 13 incidents for which the cause of incident (the leaked product) was of a sulfuric type and 47 incidents where the cause was different. The 13 incidents caused by sulfuric compounds were associated with a total of 106 complaints of which 29 were of the subtype sulfur.

	Sulfur related Incident= true	**Sulfur related Incident = false**
E_{sulfur}^{Smell}	0,273±0.0848	0,029±0.0169
E_{other}^{Smell}	0,727±0.0848	0,971±0.0169

Table 14: Conditional probability tables capturing relations P(E_i^{Smell} |Sulfuric Gas).

3.5 Discussion of the resulting observation models

The CPTs were extracted for a subset of relevant chemical substances which are (i) relevant for the environmental monitoring and (ii) for which observation models can be extracted by using statistically sound methods.

In the presented work, we decided to use BNs with binary variables. Namely, it turned out that many conditional probabilities required for the definition of the multinomial CPTs cannot be reliably estimated due to the lack of the data. By using binary variables and clustering of certain categories, the number of the parameters is reduced.

We should note that the extracted CPTs are based on the data that corresponds to the data collection that was based on the intermediation by a DCMR operator. Such an expert can interpret various complaints and describe them in the database by using the classification terms/codes for subtypes and sub-subtypes used by the DCMR. In other words, the CPTs are observation models which are valid in case an operator is mediating. Consequently, if an automated system were handling the complaints, we are likely to obtain slightly different CPT parameters. However, the evaluation of the DCMR database provided the following important implications:

- The complainers are often able to provide the right clues associated with a specific chemical; i.e. they provide the operator with the information that allows a mapping to correct subtypes. This is a critical element of any automated or manual approach to the gas detection and classification. Especially reliable detection is possible for the subtypes Chlorine, Chemical and Oil. The subtype Sulphur is obviously more difficult to describe by untrained people.
- The CPTs describing the observations for a system with automated querying are likely to retain the same greater than/smaller than relations between the parameters as the CPTs corresponding to the approach with the mediating DCMR experts. Note that these simple relations are critical for good performance, despite of the fact that the true parameters might deviate from the estimated parameters significantly. This assumption is especially plausible, if the system provides the complainers with a limited set of descriptions for each type of the chemical using a suitable interface.

The observation models for such an approach could be obtained by asking the operators to use predefined queries that would also be used by the automated system. In this way, we would be gathering the data that reflects the performance of the complainers under very similar conditions as the automated approach.

4. Conclusions.

The results of our studies show that people will more accurately identify smells with the help of an interface that offers relevant stimuli. Participants that were presented with additional stimuli generally provided more accurate odor descriptions than those that had to provide a description without a cue. The findings also show that participants found the combination of word and picture stimuli most enjoyable and useful. We have used the results of this study to develop a first prototype of an intelligent interface that automatically generates such cues to assist human smell identification (see chapter 2). Computer-assisted human perception is an interesting field of research and the work described in this paper shows the potentially powerful ways in which human olfactory perception can be enhanced. We believe this merits further research and future work will continue to evaluate and develop interactive tools to assist human smell.

Furthermore, we developed improved Bayesian models that support classification of a relevant subset of gaseous pollutants. In particular, we have proposed model structures and extracted CPT parameters from the complaint database of the DCMR. A significant effort was invested in the development of a set of perception models which are key to effective automated classification of gases and the localization of sources. The results provide a statistical evidence that citizens can provide useful information which supports classification of an important class of chemicals.

5. References

[1] Cai, G. (2005) Extending Distributed GIS to SupportGeo-Collaborative Crisis Management. Geographic Information Science, 11, 1, 4-14.

[2] Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. Journal of Applied Social Psychology 22:14, 1111-1132.

[3] DiMicco, J.M., Hollenbach, K.J., Pandolfo, A. andBender, W. (2007). The Impact of Increased Awareness while Face-to-Face. Special Issue on Awareness SystemsDesign, Human-Computer Interaction. Volume 22 (2007),Number 1.

[4] Ellis, C., Gibbs, S.J. and Rein, G.L. (1991). Groupware: Some issues and experiences. Communications of the ACM 34(1): 38-58. [5] Feibush, E., Gagvani and N., Williams, D. (2000). Visualization for Situational Awareness. IEEE Graphics, September/October, pages 38 - 45, 2000.

[6] ISO 9241:11. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. ISO/IEC.

[7] MacEachren, A. M. and Brewer, I. (2004). Developing a Conceptual Framework for Visually-enabled Geocollaboration. International Journal of Geographical Information Science, 18, 1-34.

[8] Sharp, H., Rogers, Y. and Preece, J. (2007). Interaction design: Beyond human-computer interaction. Hoboken, N.J.: Wiley.

[9] Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion Into the Technology Acceptance Model. Information Systems Research (11:4), pp. 342–365.

[10] Winterboer, A., Martens, M.A., Pavlin, G., Groen, F.C.A., Evers, V., (2011). DIADEM: A System for Collaborative Environmental Monitoring, ACM conferenceon Computer Supported Collaborative Work (CSCW '11), Video abstract, Hangzhou, China

[11] Wu, A., Zhang, X., Convertino, G. and Carroll, J.M. (2009). CIVIL: Support Geo-collaboration with Information Visualization. Proceedings of the ACM 2009 international conference on Supporting group work (pp. 273-276). Sanibel Island, FL: ACM.

[12] Zerger, A. and Smith, D. (2003) Impediments to using GIS for real-time disaster decision support. Computers, Environment and Urban Systems 27(2): 123-41.

[13] http://www.youtube.com/watch?v=I_u5trC_IfA(Visited 9th of February 2011)

[14] http://www.dcmr.nl/binaries/publicatie/2010/dcmrmilieumeldingenverslag-2009.pdf (Visited 9th of February 2011)

[15] Ademoye, O.A., and Ghinea, G., *Synchronization of Olfaction-Enhanced Multimedia*. IEEE Transactions on Multimedia, Vol. 11, No. 3, 2009.

[16] Berlin, B. and Kay, P. (1969) Basic *Color Terms: Their Universality and Evolution*. Berkeley: University of California Press

[17] Bodnar, A., Corbett, R., Nekrasovski, D. (2004). AROMA: ambient awareness through olfaction in a messaging application. Proceedings of the 6th internatational Conference on Multimedia Interfaces, 2004, State College, PA.

[18] Brewster, S.A., McGookin, D.K. and Miller, C.A. Olfoto: Designing a smell-based interaction. In Proceedings of CHI'06, Montreal, Canada , pp 653-662.

[19] Cain, W. S. (1979). To know with the nose: Keys to odor identification. Science, 203, 467-470.

[20] Chastrette, M. Classification of odors and structure-odor relationships, in Olfaction, Taste, and Cognition, C. Rouby, B. Schaal, D. Dubois, R. Gervais, and A. Holley, Eds., Cambridge Unicersity Press, 2002, pp. 100–116.

[21] Chu, S. and Downes, J.J. (2000) Odour-evoked Autobiographical Memories: Psychological Investigations of Proustian Phenomena. Chemical Senses. 25(1): 111-116

[22] Davis, F.D., Bagozzi, R.P., Warshaw, P.R. (1992) Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. Journal of Applied Soc. Psych. 22:14, 1111-1132

[23] De Wijk, R., & Cain, W.S. (1994). Odor quality: Description versus free and cued identification. Perceptions & Psychophysics, 56, 1, p. 12-18.

[24] Demattè ML, Sanabria D, Spence C. Cross-modal associations between odors and colors. Chem. Senses (2006),31:531–538.

[25] Doty, R., Shaman, P. and Dann, P. (1984) Development of the U of Penn Smell Identification Test: a standardized microencapsulated test of olfactory function. Physiology & Behavior. 32 (1984), pp. 489-5O2.

[26] Ghinea, G. and Ademoye, O.A. 2009. Olfaction-enhanced multimedia: bad for information recall?. Proc. of the IEEE Conf on Multim. and Expo'09, pp. 970-973.

[27] Gilbert, A.N., Martin, R., Kemp, S.E. (1996) Cross-modal correspondence between vision and olfaction: The color of smells. American Journal of Psych, 109(3):335-351.

[28] Godinot N. and Sicard G. (1995). Odor categorization by human-subjects: an experimental approach. Chem. Senses 20:101

[29] Gottfried, J. and R. Dolan (2003). The Nose Smells What the Eye Sees: Cross-modal Visual Facilitation of Human Olfactory Perception. Neuron 39(2): 375-386.

[30] Grabbe, J.W., McCarthy, A.L., Brown, C.M., Sabo, A.A. (2010), Olfaction and Emotion Content Effects for Memory of Vignettes, C. Res. in Psych. 6 (1): 53-60.

[31] Gutierrez-Osuna, R. (2004) Olfactory Interactions, Encycl. of Human.-Comp. Inter., W. S. Bainbridge (Ed.), Berkshire Publishing, pp. 507-511, 2004.

[32] Harel, D., Carmel, L., & Lancet, D. (2003). Towards an odor communication system. Comp. Bio. & Chem., 27, pp. 121-133.

[33] Herz, R.S. (1997). Emotion experienced during encoding enhances odor retrieval cue effectiveness. Am. J. of Psychology, 110, 489-505.(1997)

[34] Kaye, J. Symbolic Olfactory Display. MIT, Media, Arts and Sci., School of Archit. and Pl., Masters Thesis, 2001.

[35] Kaye, J., Aromatic Output for HCI. Interactions, 11 (1) (2004), pp. 49-61.

[36] Keller, A. and Vosshall, L.B. (2004) Human olfactory psychophysics, Current Biol., vol. 14:20, pp. 875–878.

[37] Khan, R.M., Luk, C.H., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., Sobel, R. (2007). Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. J. Neurosci. 27:10015–23

[38] Köster, E.P. 2002, The Specific Characteristics of the Sense of Smell" In Olfaction, Taste, and Cognition, eds. C. Rouby, B. Sehaal, D. Dubois, R. Gervais & A. Holley, Re-issue edn, Cam U Press, U.K., pp. 27-44.

[39] Lyman, B. and McDaniel, M. (1990). Memory for odors and odor names: Modalities of elaboration and imagery. J. of Ex. Psych.: Learn, Mem, and Cog, 16(4):656-664.

[40] Morrot, G., Brochet, F., Dubourdieu, D. The Color of Odors. Brain and Lang., Vol. 79:2, 2001, p.309-320.

[41] Murphy, C., Cain, W.S., Gilmore, M.M., Blair Skinner, R. (1991). Sensory and semantic factors in recognition memory for odors and graphic stimuli: Elderly versus young persons. The A. J. of Psycho. 104(2):161-192.

[42] Pavlin, Gregor, Wijngaards, Niek and Nieuwenhuis, Kees, *Towards a Single Information Space for Environmental Management through Self- Configuration of Distributed Information Processing Systems*, in: Proceedings of the TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe, pages 94-103, Masaryk University, 2009.

[43] Pearce, T. C., (1997). Computational parallels between the biological olfactory pathway and its analogue `The Electr. Nose': Bio. Olfact. BioSystems, 41(2), 43-67.

[44] Sakai, N, Imada, S., Saito, S., Kobayakawa, T., and Deguchi, Y. (2005). The effect of visual images on perception of odors. Chem. Senses 30:i244.

[45] Schaal, B., Soussignan, R., and Marlier, L., Olfactory Cognition at the Start of Life: The Perinatal Shaping of Selective Odor Responsiveness, in Olfaction, Taste, and Cognition, C. Rouby, B. Schaal, D. Dubois, R. Gervais, and A. Holley, Eds., Re-issue ed. NY: Cam U. Press, 2002, pp. 421–440.

[46] Schiffman, S. S. Physicochemical correlates of olfactory quality. Science, 1974, 185:112–17.

[47] Schiffman S. S., Robinson DE, Erickson RP (1977) Multidimensional-scaling of odorants: examination of psychological and physiochemical dimensions. Chem. Senses, 2:375–390.

[48] Schiffman, S. S., & Pearce, T. C. (2003) Introduction to olfaction: perception, anatomy, physiology, and molecular biology. In T. C. Pearce, S. S. Schiffman, H. T. Nagle and J. W. Gardner (Eds.), Handb. of Machine Olfaction: Electronic Nose Technology, Wiley-VCH.

[49] Shepard, R. (1967). Recognition memory for words, sentences, and pictures. J. of Verbal Learning and Verbal Behavior 6(1): 156-163.

[50] St. Croix Sensory Inc. (2003), A Detailed Assessment of The Science and Technology of Odor Measurement, Figure 3.6, pp 23, June, 2003.

[51] Tortell, R., Luigi, D., Dozois, A., Bouchard, S., Morie, J. & Ilan, D., 2007, The effects of scent and game play experience on memory of a virtual environment, Virtual Reality, Volume 11, Number 1, 61-68.

[52] Venkatesh, V. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 11 (4): 342–365.

[53] Wilson, D. A., & Stevenson, R. J. (2003). Olfactory perceptual learning: the critical role of memory in odor discrimination. Neurosc. and Biobehav. Reviews. Vol 27, Issue 4, June 2003, Pages 307-328.

[54] Wilson, D.A., Stevenson, R.J. (2006) Learning to smell: Olfactory Perception from Neurobiology to Behavior. Johns Hopkins University Press.

[55] Winterboer, A., Cramer, H.S.M., Groen, F.C.A., Pavlin, G., Evers, V. (2009). 'Do you smell rotten eggs?' Evaluating interactions with mobile agents in crisis response situations, Extended abstract at MobileHCI'09, Bonn Germany.

[56] Yeshurun, Y. and N. Sobel (2010). An Odor is Not Worth a Thousand Words: From Multidimensional Odors to Unidimensional Odor Objects. Annual Review of Psychology 61: 219-241.

Appendix A: Subtypes and sub-subtypes of smell used by DCMR

The DCMR uses a list of possible subtypes of smells. Based on these subtypes, we computed for each case, we extracting the category that was the most represented among the reports associated to the case and computed the percentage of complaints falling in that subtype for this The table below presents the subtypes and sub-subtypes of smell used by DCMR to categorize complaints.

General	10109 General 10101 Suffocating/Heavy 10103 Industrial Smell 10111 dirty/foul 10113 Sweet/sickly 11429 Acid 11719 Perfume/Soap
Aromatic	10201 General 10203 Benzene 10207 Cresol/Phenol 10211 Naphthalene/Camphor 10213 Styrene/Polyester
Cooking smells	10301 general 10303 Barbeque/Grill smell 10305 Deep fried smell 10309 Pizza smell 10311 Shoarma smell 10315 Fish smell
Corrosive Acrid Irritant	10401 General 10405 irritating for the eye/Causing tears 10407 Irritating in the throat/asphyxiating
Fire smell	10501 General 10503 Coal/Locomotive smell 10507 Open Chimney/All burner smell 10508 sooty smell/smoke smell 10511 Searing smell 10513 Burnt oil

	10515 burnt plastic 10517 burnt latex **10510 Tobacco Cigarette smell**
Chemical	10601 General 10603 Acetate/Ester/Sour candy 10605 Acrylic 10607 Amine/Fish smell 10609 Ammonia 10611 Vinegar 10621 Gassy 10623 Insecticide/Herbicide 10627 Plastic 10629 latex
Chlorine	10703 General 10701 bleach-/Swimming pool smell
Grain/Flour	10803 general 10805 Grass/Hay 10813 Fodder 10815 Fish Flour smell
Natural Fertilizer/Manure	10901 General 10903 Chicken Manure
Mercaptan	11009 Algemeen 11001 Natural Gas 11003 Cauliflower/Brussels sprouts 11005 Cat urine 11007 Garlic/Onions
Oil	11121 General 11103 Asphalt/Bitumen/Tar 11105 Gasoline 11109 Crude 11113 gasoil/Diesel 11115 Petroleum/Kerosene

	11119 Naphtha 11125 Pyrolysis gasoline/Pygas 11127 Refinery smell 11129 Fuel oil
Disinfectant	11213 General 11203 Ether 11205 Iodine 11207 Lysol/Hospital smell
Solvent	11301 General 11303 Acetone 11309 Glue 11317 Paint solvent 11327 Paint/Turpentine/Thinner
Putrefaction	11401 General 11405 Cesspool/feces 11407 Biodegradable Waste/Compost smell 11413 vomit smell/acid 11417 sewage smell 11421 Rotten fish 11423 Offal/cadaver smell 11425 Urine smell 11427 Garbage smell
Food	11523 General 11501 Butter/rancid 11503 Food smell 11507 Seasoned/Spicy smell 11516 Raw Fish
Sulfuric	11611 general 11607 SO2/SO3 11613 H2S/Rotten eggs
Traffic smell	11717 General 11715 Engine running

Others	11701 Metal smell/welding fumes 20203 Smog