Test–Retest Reliability and Validity of the Sniffin’ TOM Odor Memory Test

Ilona Croy, Cora Zehner, Maria Larsson, Gesualdo M. Zucco and Thomas Hummel

1Smell & Taste Clinic, Department of Otorhinolaryngology, TU Dresden, Fetschertstrasse 74, 01307 Dresden, Germany, 2Gösta Ekman Laboratory, Department of Psychology, Stockholm University, Frescati Hagväg 9A, 106 91 Stockholm, Sweden and 3Department of General Psychology, University of Padua, Via Venezia 8, 35100 Padova, Italy

Correspondence to be sent to: Thomas Hummel, MD, Smell & Taste Clinic, Department of Otorhinolaryngology, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany. e-mail: thummel@mail.zih.tu-dresden.de

Accepted October 20, 2014

Abstract

Few attempts have been made to develop an olfactory test that captures episodic retention of olfactory information. Assessment of episodic odor memory is of particular interest in aging and in the cognitively impaired as both episodic memory deficits and olfactory loss have been targeted as reliable hallmarks of cognitive decline and impending dementia. Here, 96 healthy participants (18–92 years) and an additional 19 older people with mild cognitive impairment were tested (73–82 years). Participants were presented with 8 common odors with intentional encoding instructions that were followed by a yes–no recognition test. After recognition completion, participants were asked to identify all odors by means of free or cued identification. A retest of the odor memory test (Sniffin’ TOM = test of odor memory) took place 17 days later. The results revealed satisfactory test–retest reliability (0.70) of odor recognition memory. Both recognition and identification performance were negatively affected by age and more pronounced among the cognitively impaired. In conclusion, the present work presents a reliable, valid, and simple test of episodic odor recognition memory that may be used in clinical groups where both episodic memory deficits and olfactory loss are prevalent preclinically such as Alzheimer’s disease.

Key words: assessment, human, memory, olfaction nose, psychophysics, smell

Introduction

Declarative memory is well documented for information pertaining to our primary senses; sight and hearing although less is known regarding our ability to represent and retrieve conscious olfactory information (Eichenbaum 1997; Larsson 2002). The 2 main tasks used in the study of declarative odor memory are odor recognition and odor identification. The former task is related to episodic memory and the latter draws on semantic memory functions (Schab 1991).

In contrast to nondeclarative memory, that does not require conscious awareness about any previous experience, declarative memory is dependent on conscious processing of information and categorization (Schacter 1992). It has been proposed that the preconditions for declarative olfactory memory may not be optimal. For example, olfactory information often goes unnoticed and barely evokes attention in humans (Köster et al. 2002; Sela and Sobel 2010) and semantic activations that are a prerequisite for optimal episodic memory functioning are typically restricted. A verbal abstraction of incoming information enables a conscious categorization that favors a distinct representation and subsequent successful retrieval. Although humans in general are sensitive in detecting odors and can theoretically discriminate among billions, our ability to identify an odor verbally is extremely limited (Bushdid et al. 2014). Reviews indicate that unaided free identification of odors by young laypersons varies between 22% and 57% with set sizes ranging from 7–80 items (Cuevas et al. 2009). However, although spontaneous odor naming (i.e., free identification) is difficult, identification improves dramatically with the provision of response alternatives (Hummel et al. 2007).

Episodic recognition of odors received very little attention until the 1970s. In pioneering work, Engen and Ross (1973) presented target odors to participants in a laboratory setting. At shorter and longer retention intervals, subjects were exposed to the target odors intermixed with new distractor odors. Their task was to determine whether each odor was “old” or “new.” The results indicated a relatively robust ability to accurately recognize odors with little long-term loss.
Subsequent work on episodic odor recognition memory has shown that performance is susceptible to a range of different experimental conditions. Recognition performance varies with test set size and the degree of perceptual similarity between odors such that larger set sizes and qualitatively similar odors are related to poorer recognition (Lawless 1978; Schab 1991). Also, odor hedonics influence recognition such that memory for unpleasant odors is better than for pleasant ones (Larsson et al. 2009) and more resistant to the negative repercussions of the aging process (Konstantinidis et al. 2006). However, the bulk of evidence indicates that a key factor for successful odor recognition relates to the degree of semantic activation the olfactory information evokes (Jehl et al. 1995, 1997; Frank et al. 2011). Strong and positive relationships between odor familiarity and odor naming on recognition are well documented (Larsson 1997) although there is also evidence at variance (Zucco 2003; Stevenson 2010). Hence, odor recognition performance may be conceived of as a joint result of perceptual odor information and semantic associations to the odors (Schab 1991).

Currently, there are a number of tests available for the assessment of olfactory function (see Hummel and Welge-Lüüssen 2006, for a review). These tests are developed to measure olfactory sensitivity, discrimination, and identification (e.g., Hedner et al. 2010). However, few attempts have been made to develop an olfactory test that captures episodic retention of olfactory information (see, e.g., Zucco 2011). The possibility of measuring episodic odor memory is of particular interest in aging and in the cognitively impaired as both episodic memory deficits and olfactory loss have been targeted as reliable hallmarks of cognitive decline and impending dementia (Bäckman et al. 2005; Olofsson et al. 2010). Hence, a combination of 2 significant behavioral markers for neuronal change may further increase the possibility to detect subtle cognitive change among individuals.

Here, we present Sniffin’ TOM (test of odor memory), a short and an easily distributed episodic odor recognition test that is composed of 16 everyday familiar odors. Subjects ranging between 18 and 92 years smelled 8 odors in a learning phase with intentional encoding instructions. Recognition memory was assessed directly after encoding and participants were instructed to decide whether each odor presented was “old” or “new.” To prevent visual identification of the odors participants were blindfolded. For the analyses of recognition memory, hits and false alarms were calculated.

After completion of the Sniffin’ TOM, participants were again presented with the 16 odors for identification. Here, participants were asked to identify each odor by providing its name. A liberal criterion was used for scoring (e.g., citrus fruit for orange, or Christmas/dentist for clove). After this, 4 different written response alternatives were provided, with one of the names being equivalent to the target odor (cued identification). Also, verbal ability was assessed by the German “Wortschatztest” (Schmidt and Metzler 2000). This test consists of 42 rows, each encompassing 5 nonsense and 1 actual word that participants were instructed to identify. The difficulty of the task increased throughout the test.

A retest of the Sniffin’ TOM (but not of the identification test) took place 17 days later (first to second administration ranging from 7 to 64 days). The healthy participants underwent the tasks described above although the cognitively impaired participants only underwent one odor recognition test.

Materials and methods

Participants

A total of 96 healthy subjects participated (58 women, 38 men; mean age 45 years, age range 18–92 years). All participants provided written informed consent. All subjects had a score of 30 in the Mini–Mental State Examination (MMSE) (Folstein et al. 1975). The sample was divided in 3 age groups; the group of young adults were in the age range between 18 and 30 years (mean age 24.5 years; 17 women, 16 men), middle-aged adults in the range between 31 and 60 years (mean age 48.6 years; 27 women, 14 men), and old adults were 61+ years (mean age 69.8 years; 14 women, 8 men). Subjects with severe diseases that might affect olfactory function (e.g., chronic renal failure, diabetes) were excluded from the study. Additionally, 19 older participants with slight cognitive impairment (mean MMSE score = 24, mean MMSE score range 20–27) were included (5 men, 14 women, range 73–82 years, mean age 81.8 years, standard deviation [SD] = 6.4). The study was conducted according to the guidelines of the Declaration of Helsinki on Biomedical Research Involving Human Subjects; approved by the Ethics Committee of the TU Dresden.

Procedure

The set of odor items in the episodic memory task mainly comprised the 16 odor stimuli included in the “Sniffin Sticks” odor identification test, with 2 exceptions: anise was substituted with gasoline and orange with mushroom (Reden et al. 2006). At encoding, subjects were presented with 8 odors that were presented in felt-tip pens (“Sniffin’ Sticks”). The presentation followed 3 different pseudorandomized schemes (see Table 1), where each participant received one of the schemes and the same one was used for retesting. The presentation interval between odors was 15–20 s. Participants were instructed to memorize as many odors as possible. The presentation of the 8 target odors was followed by an immediate yes–no recognition test. Here, participants were presented with the 8 target odors randomly intermixed with 8 distractor odors. Participants were instructed to decide whether each odor presented was “old” or “new.” To prevent visual identification of the odors participants were blindfolded. For the analyses of recognition memory, hits and false alarms were calculated.

Materials and methods

Participants

A total of 96 healthy subjects participated (58 women, 38 men; mean age 45 years, age range 18–92 years). All participants provided written informed consent. All subjects had a score of 30 in the Mini–Mental State Examination (MMSE) (Folstein et al. 1975). The sample was divided in 3 age groups; the group of young adults were in the age range between 18 and 30 years (mean age 24.5 years; 17 women, 16 men), middle-aged adults in the range between 31 and 60 years (mean age 48.6 years; 27 women, 14 men), and old adults were 61+ years (mean age 69.8 years; 14 women, 8 men). Subjects with severe diseases that might affect olfactory function (e.g., chronic renal failure, diabetes) were excluded from the study. Additionally, 19 older participants with slight cognitive impairment (mean MMSE score = 24, mean MMSE score range 20–27) were included (5 men, 14 women, range 73–82 years, mean age 81.8 years, standard deviation [SD] = 6.4). The study was conducted according to the guidelines of the Declaration of Helsinki on Biomedical Research Involving Human Subjects; approved by the Ethics Committee of the TU Dresden.

Procedure

The set of odor items in the episodic memory task mainly comprised the 16 odor stimuli included in the “Sniffin Sticks” odor identification test, with 2 exceptions: anise was substituted with gasoline and orange with mushroom (Reden et al. 2006). At encoding, subjects were presented with 8 odors that were presented in felt-tip pens (“Sniffin’ Sticks”). The presentation followed 3 different pseudorandomized schemes (see Table 1), where each participant received one of the schemes and the same one was used for retesting. The presentation interval between odors was 15–20 s. Participants were instructed to memorize as many odors as possible. The presentation of the 8 target odors was followed by an immediate yes–no recognition test. Here, participants were presented with the 8 target odors randomly intermixed with 8 distractor odors. Participants were instructed to decide whether each odor presented was “old” or “new.” To prevent visual identification of the odors participants were blindfolded. For the analyses of recognition memory, hits and false alarms were calculated.

After completion of the Sniffin’ TOM, participants were again presented with the 16 odors for identification. Here, participants were asked to identify each odor by providing its name. A liberal criterion was used for scoring (e.g., citrus fruit for orange, or Christmas/dentist for clove). After this, 4 different written response alternatives were provided, with one of the names being equivalent to the target odor (cued identification). Also, verbal ability was assessed by the German “Wortschatztest” (Schmidt and Metzler 2000). This test consists of 42 rows, each encompassing 5 nonsense and 1 actual word that participants were instructed to identify. The difficulty of the task increased throughout the test.

A retest of the Sniffin’ TOM (but not of the identification test) took place 17 days later (first to second administration ranging from 7 to 64 days). The healthy participants underwent the tasks described above although the cognitively impaired participants only underwent one odor recognition test.
Olfactory testing took place in quiet, well-ventilated rooms. The test procedure lasted about 1 h. Subjects were asked not to eat or drink anything except water, not to smoke, or not to brush teeth, and so forth 1 h prior to testing.

Statistical analysis

The SPSS 21.0 (SPSS Inc.) was employed for statistical evaluation. Age-related differences were investigated for the 3 age groups (18–30, 31–60, and 61+ years). The number of correctly recognized items (hits + correct rejections) was analyzed as well as hits and false alarms. Hit rates, false alarm rates, and correctly recognized odors as a function of age and sex are provided in Table 2.

ANOVAs were used to investigate possible effects of the factors “age group” (3 levels) and “sex” (2 levels) on performance. Bonferroni tests were used for post-hoc comparisons. Correlational analyses were performed using Pearson statistics. The alpha level was set at 0.05.

Results

Sniffin’ TOM distribution

Overall, participants recognized 13.5 out of the 16 items correctly with a SD of 1.7. To define a decreased performance, the 10th percentile of the distribution of the number of correctly recognized items in the young age group (18 and 30 years) was used to differentiate out normal from impaired performance (see Doty et al. 1984b; Kobal et al. 2000) (Table 2). Hence, a score of <12 reflects impairment.

Test–retest reliability

Test–retest reliability of Sniffin’ TOM was satisfactory, as indicated by the correlation coefficient for performance in session 1 with session 2 ($r_{96} = 0.70, P < 0.001$). Accordingly, the Bland–Altman plot reflects a good congruence between the results from the 2 sessions (see Figure 1). Notably, the variation in time intervals between the 2 test sessions did not influence recognition performance in the first and second test sessions ($r_{96} = 0.01, P = 0.91$) indicating a relatively stable memory performance over time.

Influences of age and sex on episodic odor memory

Overall, Sniffin’ TOM performance decreased significantly with increasing age, ($F(2,90) = 8.09, P < 0.001$) although there was no significant main effect of sex on the number of correctly recognized odors ($F < 1$). Post-hoc Tukey tests indicated that the older age group showed a poorer

---

**Table 1** Stimuli used in the Sniffin’ TOM and odor identification

<table>
<thead>
<tr>
<th>Sniffin’ TOM Encoding version</th>
<th>Odor identification test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Licorice, Cherry, Menthol, Soap</td>
</tr>
<tr>
<td>2</td>
<td>Vanilla, Chocolate, Onion, Pineapple</td>
</tr>
<tr>
<td>3</td>
<td>Petrol, Juniper, Strawberry, Pineapple</td>
</tr>
<tr>
<td>4</td>
<td>Coconut, Banana, Walnut, Cherry</td>
</tr>
<tr>
<td>5</td>
<td>Cucumber, Raspberry, Rose, Cherry</td>
</tr>
<tr>
<td>6</td>
<td>Chocolate, Onion, Menthol, Apple</td>
</tr>
<tr>
<td>7</td>
<td>Petrol, Orange, Ham, Cinnamon</td>
</tr>
<tr>
<td>8</td>
<td>Mushroom, Rhum, Honey, Pine</td>
</tr>
<tr>
<td>9</td>
<td>Bread, Fish, Cheese, Ham</td>
</tr>
<tr>
<td>10</td>
<td>Tobacco, Coffee, Wine, Smoke</td>
</tr>
<tr>
<td>11</td>
<td>Perfume, Glue, Leather, Grass</td>
</tr>
<tr>
<td>12</td>
<td>Cloves, Banana, Cinnamon, Mustard</td>
</tr>
<tr>
<td>13</td>
<td>Beer, Peppermint, Pine, Onion</td>
</tr>
<tr>
<td>14</td>
<td>Banana, Wine, Lemon, Leather</td>
</tr>
<tr>
<td>15</td>
<td>Apple, Mint, Garlic, Carrot</td>
</tr>
<tr>
<td>16</td>
<td>Mustard, Pear, Cheese, Turpentine</td>
</tr>
</tbody>
</table>

The test was presented randomly in 3 versions. Each participant was randomly assigned to one version of 8 target odors and to another one for the retest. One-third of the participants were therefore presented the odors from encoding version one. After encoding, they were presented with all 16 odors and had to indicate for each odor, if they smelled it before. For the identification test, the participants were presented with the same 16 odors again and were asked to first provide the name of the odor (free identification) and second to choose the right name out of the list of 4 cues, presented in the last column.
recognition performance than the young and the middle aged \((P < 0.05)\), although the latter 2 groups did not differ significantly. The interaction between age and sex was significant \((F(2,90) = 4.53, P = 0.013)\). The source of this interaction was that the number of correctly recognized odors decreased steadily in women with increasing age, whereas men exhibited a stable performance between the ages of 31 and 60 years (see Table 2). Separate analyses on the hit and false alarm rates indicated significant age-related impairments in hit rates \((F(2,93) = 7.08, P = 0.03)\) although age did not influence false alarm rates \((F < 1)\). Tukey tests indicated that the hit rates were significantly lower among the old age group as compared with the 2 younger age groups \((P < 0.01)\), although performance did not differ between these 2 groups. No other effects were reliable.

### Relationships among verbal ability, odor identification, and Sniffin’ TOM

Age affected both free \((F(2,93) = 10.7, P = 0.001)\) and cued \((F(2,93) = 6.30, P = 0.03)\) identification negatively, although the main effect of sex and the age by sex interactions were unreliable for both types of identification \((Fs < 1)\). For both types of identification tests, young adults outperformed the middle-aged and old age groups and these 2 groups identified an equal number of odors. To determine relationships
between semantic factors and Sniffin’ TOM, a set of correlations with age partialled out was performed. The results showed that the number of correctly recognized odors was unrelated to performance in the vocabulary test ($r_{90} = 0.12$, $P = 0.24$) and there was a significant correlation between odor recognition memory and odor identification ($r_{11} = 0.44$, $P < 0.001$). This outcome suggests that odor identification is related to better retention of odors. Interestingly, free odor identification was higher for correctly recognized odors than for faulty recognition ($t(83) = 4.3$, $P < 0.001$; see Figure 2) although cued identification performance did not differentiate between correctly recognized odors and the errors ($t(83) = 0.7$, $P = 0.5$).

Influence of cognitive impairment on Sniffin’ TOM

Fifteen of the 19 subjects with slight cognitive impairment (79%) recognized only 12 or less odor items correctly, indicating a decreased performance in this test. Consequently, odor recognition was significantly worse compared with healthy age-matched controls ($t(39.1) = 3.1$, $P = 0.004$).

Discussion

The Sniffin’ TOM proved to be a valid odor memory test that has the advantage of being short enough to also enable testing of people with cognitive impairment. As noted above, available olfactory tests typically are designed to measure absolute sensitivity or identification to assess the integrity of the olfactory sensory system. To the best of our knowledge this is the first study that evaluates the reliability of an episodic odor recognition test. Notably, the test performance proved relatively stable over a test interval averaging 17 days with a retest coefficient of 0.70. The retest reliability is not as good as expected. However, one has to keep in mind, that the test only uses 8 target items. Similar reliability coefficients have been obtained for short versions of the well-established odor identification tests such as the Sniffin Sticks and the Smell Identification Test (Doty et al. 1996; Hummel et al. 2001). Also, people with mild cognitive impairment showed impaired episodic odor recognition suggesting that performance varies with the cognitive status of the subject. A recent meta-analysis found Alzheimer’s disease to be related to reduced odor identification and recognition (Rahayel et al. 2012). One can speculate that reduced odor memory scores might have a predictive value for such disease in later life. However, long-term studies are needed here.

Participants were able to recognize about 80% of the presented odors that is in congruence with previous work (Olsson et al. 2009). Also, replicating other studies, chronological age exerted a negative impact on episodic odor recognition overall (Murphy et al. 1991; Larsson and Bäckman 1993; Zucco 2011). The analyses on hit rates and false alarm rates indicated that the age-related impairment primarily was related to a decrease in hit rates rather than an increment in the generation of false alarms. In accordance with a wealth of findings, aging affected odor identification ability negatively. The results clearly indicated that both free and cued identification were impaired with increasing age (Doty et al. 1984a; Larsson et al. 2004; Hummel et al. 2007). Although provided with retrieval support at testing (odor label cues) older adults showed an impaired identification performance. This outcome suggests that sensory/discriminatory difficulties rather than cognitive factors underlie the age deterioration in odor identification (Larsson et al. 1999).

Empirical evidence suggests a female superiority in episodic memory for verbal and visual information (Herlitz and Rehnman 2008). These documented findings have also been generalized to olfactory information in most cases (Oberg et al. 2002; Larsson et al. 2003). In contrast to these findings but in accordance with a previous study on odor recognition memory (Zucco 2011), the present results of the Sniffin’ TOM revealed no influence of sex on hit rates, false alarm rates, or overall recognition performance.

![Figure 2](https://academic.oup.com/chemse/article-abstract/40/3/173/614200/attachment/725817/141200) Sniffin’ TOM performance in relation to age and cognitive impairment. A significant decrease of odor memory can be observed with age ($P < 0.001$). People with mild cognitive impairment perform significantly worse than age-matched healthy participants ($P < 0.01$).
It is of theoretical interest to determine the role of verbal mediation in episodic odor recognition. If olfactory memory would be independent of verbal or linguistic factors, then olfactory processing would be fundamentally different from the cognitive processing carried out in other modalities. Although there is still some controversy regarding the role of semantic factors in episodic odor memory (see Herz and Engen 1996, for a review Zucco 2003; Stevenson 2010), a large number of studies underscore the positive influence of familiarity and identification on olfactory retention (e.g., Larsson and Bäckman 1993; Cain and Potts 1996; Olsson et al. 2009; Frank et al. 2011). In congruence with these observations, the present findings indicate a reliable relationship between identification and recognition memory performance such that a higher ability to identify odors was related to better memory (Jehl et al. 1995, 1997; Yeshurun et al. 2008; Olsson et al. 2009). Importantly, this relationship was independent of age. Still, it is noteworthy that although participants were able to identify about only one-third of the odors, the correct recognition rate was about 80%. Hence, although identifiability may drive proficiency in episodic odor recognition it cannot fully explain the variability in memory performance. Other candidates that may contribute to recognition proficiency are perceptual odor information (e.g., hedonics), imagery, or personal associations. Thus, episodic odor recognition performance may be conceived of as a joint and integrated result of perceptual, semantic, and episodic information.

In conclusion, the Sniffin’ TOM is a reliable, valid, and simple test of episodic odor recognition memory. In particular, the test may be used in clinical groups where both episodic memory deficits and olfactory loss are prevalent in preclinical stages such as Parkinson’s disease (Haehner et al. 2009), schizophrenia (Moberg et al. 2014), depression (Zucco and Bollini 2011; Croy et al. 2014), and Alzheimer’s disease (Masurkar and Devanand 2014).

Acknowledgements

In a previous study on an extended version of the Sniffin’ Sticks test T.H. received help from the Burghart company (Wedel, Germany) in that the company generously provided portions of the test kit.

References


