Chem. Senses 31: 379–391, 2006
doi:10.1093/chemse/bjj042
Advance Access publication March 9, 2006

Development of a Smell Identification Test Using a Novel Stick-Type Odor Presentation Kit

Sachiko Saito[1,2], Saho Ayabe-Kanamura[1,3], Yasuhiro Takashima[4], Naomi Gotow[1], Naomi Naito[5], Takashi Nozawa[1,6], Miyako Mise[5], Yuichi Deguchi[7] and Tatsu Kobayakawa[1]

[1]National Institute of Advanced Industrial Science and Technology, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan, [2]Saito Sachiko Taste and Smell Institute, [3]Health Science University, [4]International Society for Fragrance and Culture, [5]National Institute of Technology and Evaluation, [6]University of Tsukuba and [7]Takasago International Corporation

Correspondence to be sent to: Sachiko Saito, National Institute of Advanced Industrial Science and Technology, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan. e-mail: sachiko-saitou@aist.go.jp

Abstract

The odor identification is strongly influenced by the social and cultural factors; therefore, the odorants used in a smell identification test should be familiar to the test population. In addition, the device used in the test is desired to be simply handled and retain odor quality over time. We developed a novel stick-type odor presentation kit that consists of microcapsules of odorant incorporated into stable cream and the smell identification test using it. Thirteen odorants were selected to be familiar to the test population. In the test, we used two identification methods: one was a modified forced-choice paradigm with "detectable but not recognizable" and "no smell detected" added as choices and the other was a two-step identification paradigm where the participant first selected one of eight odor categories and then chose the specific odor name from the selected category. We verified the performance of the odor stick and the test by stability, using a test–retest paradigm, comparing this test with another smell test, and testing Japanese people from a range of age groups. We conclude that this kit is a useful odor presentation device, and the test using it works effectively as a smell identification test.

Key words: aged people, device, odor identification, odor stick, olfaction, smell test

Introduction

In everyday life, we use the olfactory system to detect danger such as leaking gas, fires, and rotten foods, underscoring the importance of our ability to discriminate between and identify odors. Impaired olfaction, which has been extensively reported about in aging individuals, significantly increases the risk of serious accident or disease. Moreover, in addition to assessing olfactory dysfunction, an effective means of measuring olfactory performance is expected to be used for diagnosis of Alzheimer's disease.

To present an odor in smell identification tests, natural or chemical odorants in plastic or glass jars have been used (Cain *et al.*, 1983; Hendriks, 1988; Nordin *et al.*, 1998). Additionally, several sophisticated odor presentation kits have been also developed to allow for more simple and practical tests. For example, Doty *et al.* (1984b) developed, as a simple and practical kit, the Pennsylvania smell identification test (UPSIT), in which the odorant was encapsulated and printed on paper that participants sniff after scratching. This test, however, was problematic when used in Japan because several unfamiliar odors were included (Zusho *et al.*, 1983); individual performance of odor identification is significantly influenced by societal and cultural background (Ayabe-Kanamura *et al.*, 1998). Later, Doty *et al.* (1996) developed the cross-cultural smell identification test (CC-SIT), which comprised 12 cross-cultural odorants. Another simple odor presentation device was developed by Kobal *et al.* (1996) and called the Sniffin' Stick. In this test, the odorant was presented at the tip of a device that resembled a felt-tip pen.

We have developed a new odor presentation kit that employs "odor stick," which is expected to preserve the odorant better than the cards because microcapsules of odorant were incorporated into a stable cream. This kit was devised to consist of various odors that are familiar to the intended test population while being both stable and easy to use. In this paper, we describe the odor stick manufacturing process and the method for odor presentation.

We also devised an odor identification test using these odor sticks called the odor stick identification test (OSIT). We proposed two identification methods that use different strategies for the identification process. The first is called the "four-plus alternative method," which is a modified forced-choice paradigm that assumes a top–down strategy in the identification process. The second is the "two-step identification method," which assumes a bottom–up strategy when the subject encounters an odor without advanced cues of specific odors. In this method, we identify a vague odor category (odor cluster)—for example, a "sweet smell" or a "rotten smell"—and look for the odor source around and then identify specific odor such as "rose" or "sweaty smelling clothes." We also used picture and word alternatives (henceforth, "picture–word alternatives") as well as "word alternatives" to identify the odor or odor cluster.

Next, we verified the efficacy of the odor stick by examining the quality of odor representation and the stability of the odor stick. We also evaluated the OSIT using four-plus alternative method by 1) evaluating the OSIT scores applied to healthy young and middle-aged people, 2) using a test–retest paradigm, 3) comparing the OSIT with CC-SIT, and 4) applying OSIT to the Japanese people from different age groups. Lastly, we examined two-step identification method and the effects of picture–word alternatives in the odor identification test.

Materials and methods

Development of materials

Odor stick

The odor stick manufacturing process. Odors used in identification tests should be familiar to the test subjects. For our study of odor identification in Japanese individuals, we selected 20 odors familiar to this cohort by surveying 178 individuals (age range: 20–89 years) regarding their familiarity and experience with 119 odors and by deriving from the Japanese cognitive odor space determined by Saito *et al.* (1999). We also selected dangerous odors encountered in daily life including those associated with gas leaks, rotting food, and burning. Japanese cypress was also selected as a typical Japanese odor in traditional Japanese houses or hot spring spas. We then identified specific odorants that corresponded to each odor. Some odors were excluded from the study due to the difficulty of finding appropriate odorant substances. To manufacture the odor sticks, we made powdery microcapsules of the odorants (5- to 15-μm diameter), using melamine resin as a base, and uniformly mixed the microcapsules with a melted base material composed primarily of Vaseline and glycerin. The concentration of microcapsules in the final suspension was set so that healthy people without olfactory impairment could easily identify the odor. We then poured the mixture into a mold shaped like a lipstick container. After cooling, the odorous cream formed a semi-solid odor stick that was encased like a lipstick and could be moved in and out by turning the case's base (Figure 1). Some odor substances were difficult to incorporate into powdery

A B

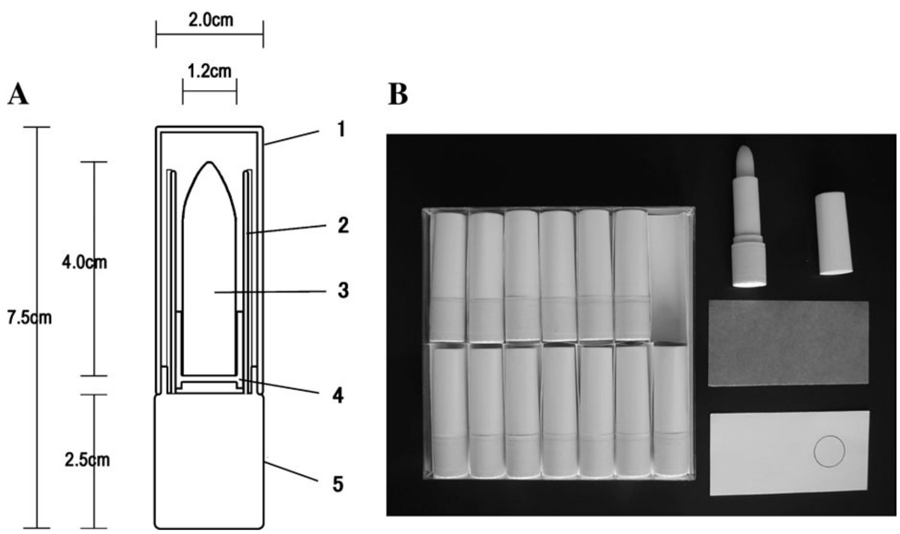

Figure 1 Odor sticks: a novel stick-type odor presentation kit. **(A)** Structure of odor stick (1, a cap; 2, a cylinder for the preservation of an odorous solid cream; 3, an odorous solid cream; 4, a lifting plate; 5, a base). An odorous solid cream shaped like a lipstick that could be moved in and out by turning the case's base. **(B)** Left side, odor sticks arranged in a stick container; upper right side, an example of an odor stick; middle right side, paraffin paper; bottom right side, a mount for paraffin paper.

microcapsules, and some odor sticks required an extremely high microcapsule concentration for detection, resulting in a cream that did not sufficiently solidify to permit delivery. In the end, we produced 13 different odor sticks for experimentation. The selected odors, their odor category based on Japanese odor classification system (Saito *et al.*, 1999), and the characteristics of the odor substances are shown in Table 1.

Odor presentation. The experimenter applied the odorous semisolid cream from an odor stick to a 2-cm circle on a 5.25×10.5–cm rectangular piece of thin paraffin paper, folded this paper in half, rubbed it to grind the microcapsules, and passed it to a participant. The participant then opened and sniffed the paper (Figure 2).

OSIT

Four-plus alternative method. We adopted two identification methods for the identification process in the OSIT. The four-plus alternative method was a modified forced-choice paradigm in which the choices "detectable but not recognized" and "no smell detected" were added as alternatives to an ordinal forced-choice method. Participants were instructed to try to choose the correct odor name from four alternatives; they were, however, allowed to choose "detectable but not recognized" or "no smell detected" if they could not identify the odor. This method was chosen to precisely measure an olfactory deficit, while reducing chance correct answers and minimizing the participants' possible anxiety associated with strict forced choice when they could not detect or identify any odor. The additional two alternatives were distinguished from the four odor choices by a broken line or parentheses (Table 2). We referred to this identification method as the four-plus alternative method to distinguish it from the standard four-alternative forced-choice paradigm. The three incorrect odor alternatives for each odorant were prepared based on the established eight clusters of Japanese odors (Saito *et al.*, 1999); one alternative each was, as a rule, selected from the same cluster, a neighboring one, and a remote cluster. Table 2 shows the four odor alternatives offered for each odorant. One odor had two correct answers because Japanese individuals commonly associate isovaleric acid with one of two everyday odors: in a preliminary study, this compound was identified by half of the participants as the popular Japanese food "Natto," which is made of fermented soybeans, whereas the rest of the participants described it as smelling like sweaty socks and clothes.

Odor identification performance was assessed based on the number of the odorants that were correctly identified.

Table 1 Odor cluster and odorant for selected 13 odors

Odor cluster	Odor	Odorant
Sweet odor (fruit, flower, confectionery)	Perfume	Compound
	Rose	Compound
	Japanese orange	Artificial flavor
	Milk	Artificial flavor
Spices	Curry	Natural flavor
	Roasted garlic	Natural flavor
Rotten, excreta	Putrid smell	Compound
	Sweaty smelling clothes/fermented soybeans[a]	Isovaleric acid
Gas, smoke	Cooking gas	Tetrahydrothiophene
Wood, grass, herb	Indian ink	Borneol
	Wood	Essential oil
	Japanese cypress	Essential oil
	Menthol	Menthol

[a]Popular Japanese food known as "Natto."

Two-step identification method. The two-step identification method was developed as an analogy of a strategy for odor identification situation in which the participants were presented with an odor without definite top–down cues of specific odor name in advance. Participants should search the odor souse from all possible odors. Participants were instructed to broadly categorize the odor and to select an appropriate odor cluster from eight alternatives of the established Japanese odor clusters (Table 3, left side). It was assumed that participants would catch the broad outline of the odor based on the odor cluster map, which had been formed with the odor experiences of test population. Next, the participant was asked to select a specific odor name from several alternatives in the selected cluster (Table 3, right side). If the participant, however, could not find an appropriate odor name in the selected cluster, the experimenter allowed the participant to make a selection from all the possible odors. The two-step identification method was designed to re-create a bottom–up identification strategy without advanced cues compared with the top–down strategy of the four-plus alternative method. It was assumed in this strategy that we would catch the broad outline of the odor to search the odor souse from all possible odors. The choices "detectable but not recognized" and "no smell detected" were also added in the two-step identification method. Participants received a score of 1.0 for the selection of the correct odor, whereas they received a score of 0.5 when the correct cluster was selected but the incorrect odor was chosen.

Picture–word alternatives in the identification test. We also used picture–word alternatives in addition to word alternatives in the identification test. In the four-plus alternative method, the four odor names were presented with an associated picture and the word, whereas the two alternatives of

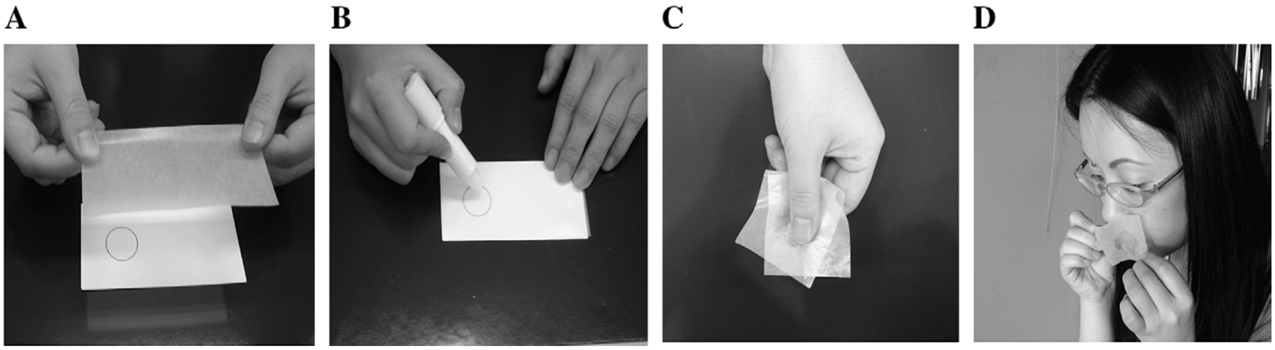

Figure 2 Method to present odor by an odor stick. **(A)** A paraffin paper is put on a mount. **(B)** An odor stick is applied to a circle with a 2-cm diameter on a 5.25 × 10.5–cm strip of paraffin paper. **(C)** A paraffin paper is folded in two and rubbed together to release the microencapsulated odorant. **(D)** A participant unfolds and sniffs the paraffin paper.

"detectable but not recognizable" and "no smell detected" were represented only by words (Figure 3). In the two-step identification method, the names of eight odor clusters and several odor names from each cluster were presented with an associated picture and the word, whereas the two extra alternatives were represented only by words.

Methods of verification

This study was conducted in accordance with the revised version of Helsinki Declaration and was approved by the National Institute of Advanced Science and Technology (Japan).

Odor stick (Exp. 1)

The quality and intensity of the odor sticks (Exp. 1-1). We examined the odor quality and perceived intensity of the 13 odor sticks.

Participants. Twenty healthy women (age range: 29–48 years) participated in this experiment. For this study, participants were deemed healthy if they reported no olfactory deficits.

Stimuli. Thirteen odor sticks that were manufactured 1 month prior to the study.

Procedure. A participant opened a piece of paraffin paper folded in half and sniffed it. She then expressed her perception of the odor quality using free descriptors and estimated the perceived intensity on a numerical visual analog scale (VAS) from 0 (no smell detected) to 5 (a very strong odor). The descriptive words "faint," "weak," "moderate," and "strong" were placed on the VAS next to the numerical ratings of 1, 2, 3, and 4, respectively. As a reference, we prepared a filter paper dipped in 0.01% (w/w) methyl cyclopentenolone solution and instructed the participants that its intensity was 3 on the VAS. The participants were given this reference odor before the experiment and subsequently whenever they felt that they needed to reestablish the reference point. The reference odor served to decrease the within-individual differences for 31 months because participants repeated the same experiments nine more times (every 3 or 4 months) for Exp. 1-2. Methyl cyclopentenolone is one of 10 odorants of the Japanese Standardized Olfactory Test (Zusho, 1983; Takagi, 1989). It has a neutral odor quality in hedonics and is described as smelling like something burnt or like caramel. Each evaluation of the 13 odorants took about 30 min to complete, with an interstimulus interval of 2 min.

Odor stick stability (Exp. 1-2). To verify the stability of the odor sticks, the odor quality and the perceived intensity were measured for 31 months after the sticks were manufactured.

Participants. The same 20 women from Exp. 1-1 participated until the sixth session, and 15 of the 20 women continued from the 7th through the 10th session. The first session refers to Exp. 1-1, whereas the subsequent nine sessions were conducted in Exp. 1-2.

Stimuli. The 13 odor sticks used in Exp. 1-1 were used in the 2nd though the 10th sessions at 4, 8, 11, 15, 18, 22, 25, 28, and 31 months after the odor sticks were manufactured. Odor sticks were kept in a refrigerator for the 31 months and taken out 1 h before the experiment.

Procedure. The odor quality was evaluated by participant's free descriptors, and the perceived intensity was estimated nine times after Exp. 1-1, every 3 or 4 months up to 31 months after the day the stick was manufactured. We called these nine sessions of Exp. 1-2 the 2nd session to the 10th session, while Exp. 1-1 was the first session. The procedure in each session was the same as in Exp. 1-1.

Evaluating the OSIT using the four-plus alternative method (Exp. 2)

The OSIT score of healthy young and middle-aged people (Exp. 2-1). We examined the results of OSIT given to healthy young and middle-aged people.

Table 2 Alternatives of each odor item in four-plus alternative method

Alternatives	Alternatives	Alternatives
1 Incense stick 2 Honey 3 Coffee 4 Perfume (5 Detectable but not recognizable) (6 No smell detected)	1 Wood 2 Coffee 3 Japanese horseradish 4 Roasted garlic (5 Detectable but not recognizable) (6 No smell detected)	1 Sulfur 2 Indian ink 3 Vanish 4 "Tatami" mat (5 Detectable but not recognizable) (6 No smell detected)
1 Grassy plants 2 Rose 3 Apple 4 Raisins (5 Detectable but not recognizable) (6 No smell detected)	1 Gasoline 2 Cooking gas 3 Caramel 4 Putrid smell (5 Detectable but not recognizable) (6 No smell detected)	1 Spoiled food 2 Leather 3 Wood 4 "Tatami" mat (5 Detectable but not recognizable) (6 No smell detected)
1 Banana 2 Apple 3 Peanut 4 Japanese orange (5 Detectable but not recognizable) (6 No smell detected)	1 Fermented soybeans 2 Leather 3 Sweaty smelling clothes 4 Soybean paste[a] (5 Detectable but not recognizable) (6 No smell detected)	1 Sulfur 2 Incense stick 3 "Tatami" mat 4 Japanese cypress (5 Detectable but not recognizable) (6 No smell detected)
1 Condensed milk 2 Chocolate 3 Cinnamon 4 Peanut (5 Detectable but not recognizable) (6 No smell detected)	1 Cresol disinfectant 2 Cooking gas 3 Sulfur 4 Sweaty smelling clothes (5 Detectable but not recognizable) (6 No smell detected)	1 Incense stick 2 Menthol 3 Mold 4 "Tatami" mat (5 Detectable but not recognizable) (6 No smell detected)
1 Coffee 2 Pineapple 3 Curry 4 Butter (5 Detectable but not recognizable) (6 No smell detected)		

Correct alternative is enclosed by oval.
[a]Popular Japanese food known as "Miso."

Participants. Participants were 104 healthy volunteers (43 males) with an age range of 24–49 years (mean age 33.3 years). This age range has been found to not, on average, exhibit a decline in odor identification performance as measured by the UPSIT (Doty *et al.*, 1984a) and the OSIT (the results of Exp. 2-4 in this paper).

Stimuli. Thirteen odor sticks were manufactured within 12 months.

Procedure. Participants received the odor items from an experimenter, sniffed them, and identified them in a four-plus alternative forced-choice paradigm.

Table 3 Word alternatives of each odor item in two-step identification method

Odor cluster	Odor name
1. Flower, fruit, confectionery, sweet	1-1. Perfume
	1-2. Japanese orange
	1-3. Milk
	1-4. Chocolate
2. Spice, savory herb	2-1. Curry
	2-2. Garlic
	2-3. Vinegar
	2-4. Coffee
3. Soy sauce, soybean paste	3-1. Soy sauce, soybean paste
4. Fishy, seashore	4-1. Tangle flakes, seashore
5. Feces, putrid smell	5-1. Feces putrid, smell
	5-2. Sweaty smelling clothes
	5-3. Fermented soybeans
6. Cooking gas, smoke, charred	6-1. Cooking gas, gasoline
	6-2. Charred
	6-3. Smoke
7. Plant, herb	7-1. Japanese cypress
	7-2. Menthol
8. Thinner, leather	8-1. Naphthalene
	8-2. Thinner
	8-3. Leather, gum
(9. Detectable but not recognizable)	
(10. No smell detected)	

The internal consistency (test–retest) of the OSIT (Exp. 2-2). We evaluated the test–retest reliability to verify the internal consistency and stability of the OSIT. Forty-seven volunteers (15 males; age range: 20–83 years; mean age: 41.0 years) participated in two sessions that were separated by 122 days ± 26.8 (mean ± SE, n = 47). The procedures of odor presentation and odor identification were the same as described above. Feedback was not given to a participant after the completion of the first session.

Comparison of the OSIT with the CC-SIT (Exp. 2-3). Both the OSIT and the CC-SIT were applied to 108 healthy volunteers (58 males; age range: 20–81 years; mean age: 49.4 years, SD = 21.5). The two tests were given to an individual on different days, and the order of the tests was randomized among participants. The familiarity of each odor item was judged using a VAS from 0 (completely unknown) to 10 (well known).

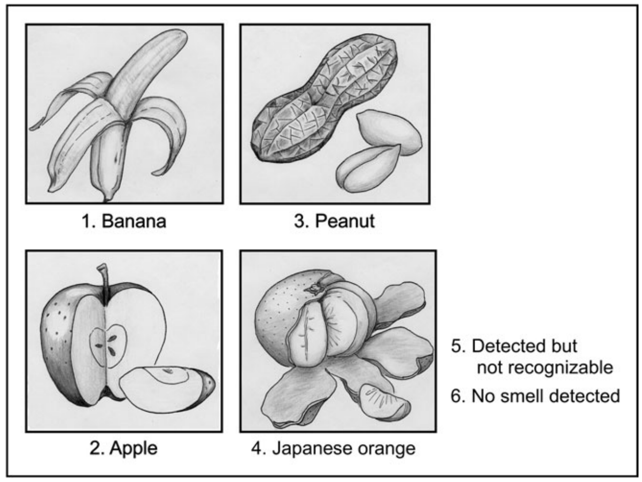

Figure 3 An example of identification tool "picture–word alternatives" in four-plus alternative method. This shows four alternatives and two added alternatives, "detectable but not recognizable" and "no smell detected," for an odor item, "Japanese orange."

Testing Japanese people from different generations with the OSIT (Exp. 2-4). The OSIT was given to 448 healthy Japanese volunteers (205 males). The breakdown of the age groups was 153 participants (67 males) between 20 and 29 years, 77 participants (46 males) between 30 and 39 years, 73 participants (23 males) between 40 and 49 years, 48 participants (18 males) between 50 and 59 years, 52 participants (29 males) between 60 and 69 years, 39 participants (18 males) between 70 and 79 years, and 6 participants (4 males) between 80 and 89 years. These tests were carried out in a university, several national or private institutes, and a community center.

Examination of the two-step identification method (Exp. 3)

We compared the identification rate measured using the two-step identification method with that measured using the four-plus alternative method. Of the 80 healthy participants who volunteered for this study, 42 volunteers (21 males; age range: 21–79 years; mean age: 53.5 years) were tested with the two-step identification method, whereas 38 volunteers (19 males; age range: 23–78 years; mean age: 51.3 years) were tested with the four-plus alternative method. Participants identified the 13 odor items according to the description in the Development of Materials section. In analysis, we added the results of 107 participants from 20s to 70s obtained in Exp. 2-3 to the data of four-plus alternative method and used the data of 145 participants (76 males; age range: 21–77 years; mean age: 49.7 years) in four-plus alternative method.

Examination of using picture–word alternatives (Exp. 4)

We compared the results from OSIT in which picture–word alternatives were used to identify the odors with those obtained from OSIT in which the odorants were identified by only word alternatives.

Participants. One hundred thirteen healthy volunteers (age range: 19–30 years) participated. For the four-plus alternative method, 40 people (20 females; mean age: 22.9 years) were tested using word alternatives to identify the odors, whereas 40 people (20 females; mean age: 23.9 years) were tested using picture–word alternatives. For two-step identification method, 15 people (five females; mean age: 21.7 years) were tested using word alternatives, whereas 18 persons (six females; mean age: 22.8 years) were tested with picture–word alternatives.

Stimuli. Nine of the original 13 odor sticks were used (perfume, excreta/putrid smell, Japanese cypress, and ''menthol'' were removed because they were either similar to other odors or found to mainly stimulate trigeminal nerve).

Procedure. The testing method was the same as has been described in the Development of materials section.

Results

Odor stick (Exp. 1)

The quality and intensity of the odor sticks (Exp. 1-1)

To assess odor quality, we analyzed the odor descriptors reported by 20 participants for each odorant. Nineteen participants' data were analyzed for odor item ''Indian ink'' because of missing datum. We counted the number of participants who reported descriptors included in the same odor category because it was reported that the odor descriptions were various and discord among individuals in a free-choice paradigm (Saito *et al.*, 1997). We used the eight categorical clusters of Japanese odors established by cluster analysis (Saito *et al.*, 1999). Table 4 shows, for each odor item, the number of participants who reported expected odor descriptors, who reported odor descriptors classified in the same odor cluster as the expected odor, and who reported descriptors classified in the other clusters. For example, 19 participants (95%) reported menthol for menthol item and one participant reported descriptor of the same cluster. For rose item, only six participants (30%) identified rose; however, all other participants identified it with other odor descriptors in the same odor cluster such as perfume, flower, or cosmetics. Therefore, all participants identified rose item with either the expected descriptor or the similar descriptors in the same odor cluster. For cooking gas and roasted garlic items, the smallest 14 participants (70%) reported either the expected odor descriptor or similar ones in same cluster. However, the number of participants was significantly larger ($P = 0.037$), by binominal test, than that who reported descriptors in other clusters.

The average perceived intensity for each item ranged from 2.8 to 4.3, and the overall average perceived intensity was 3.58.

Table 4 Representation of odor by odor stick item

Odor stick item	Expected odor descriptors (a)	Descriptors in same cluster (b)	a+b	Descriptors in other clusters
Japanese cypress	0.65	0.35	**1.00**	0.00
Perfume	0.40	0.60	**1.00**	0.00
Menthol	0.95	0.05	**1.00**	0.00
Wood	0.70	0.30	**1.00**	0.00
Rose	0.30	0.70	**1.00**	0.00
Sweaty smelling clothes/fermented soybeans	0.50	0.45	**0.95**	0.05
Curry	0.90	0.05	**0.95**	0.05
Putrid smell	0.65	0.30	**0.95**	0.05
Indian ink	0.53	0.37	**0.89**	0.11
Milk	0.40	0.50	**0.90**	0.10
Japanese orange	0.45	0.35	**0.80**	0.20
Cooking gas	0.60	0.10	**0.70**	0.30
Roasted garlic	0.45	0.25	**0.70**	0.30

Numerals show the proportion of participants to total participants.

Odor stick stability (Exp. 1-2)

For each odor item, the consistency of the identified odor across the sessions was evaluated for each participant. The consistency was examined by a criterion where odor descriptors across sessions were from one category in the eight categories. We first tried to establish a baseline odor category for each participant during the first three sessions. We were able to establish the baseline odor category for all the participants and each odor with the exceptions of one individual for ''Japanese orange,'' two for ''cooking gas,'' three for ''Indian ink,'' and one for ''wood.'' We then counted the number of participants that identified odors, in the subsequent 4th to 10th sessions, that were inconsistent with their established baseline. The number was significantly smaller, by binominal test, than that who reported the consistent ones in all odor items except for wood. For wood, the number was not significantly smaller after the 8th session (or the 25th month). These results showed that the odor quality was consistently recognized until at least the 7th session or 22 months after the odor sticks were manufactured.

For analysis of perceived intensity over time, we used the data from the 14 participants who attended all the sessions. For each odor item, differences in perceived intensity across the 10 sessions were examined by one-way analysis of variance (ANOVA). Multiple comparisons (Dunnett's method) were applied to identify sessions in which significant changes were reported with respect to the first session. Figure 4 shows the average perceived intensity for each of the 10 sessions spanning 31 months after the manufacture of the odor sticks.

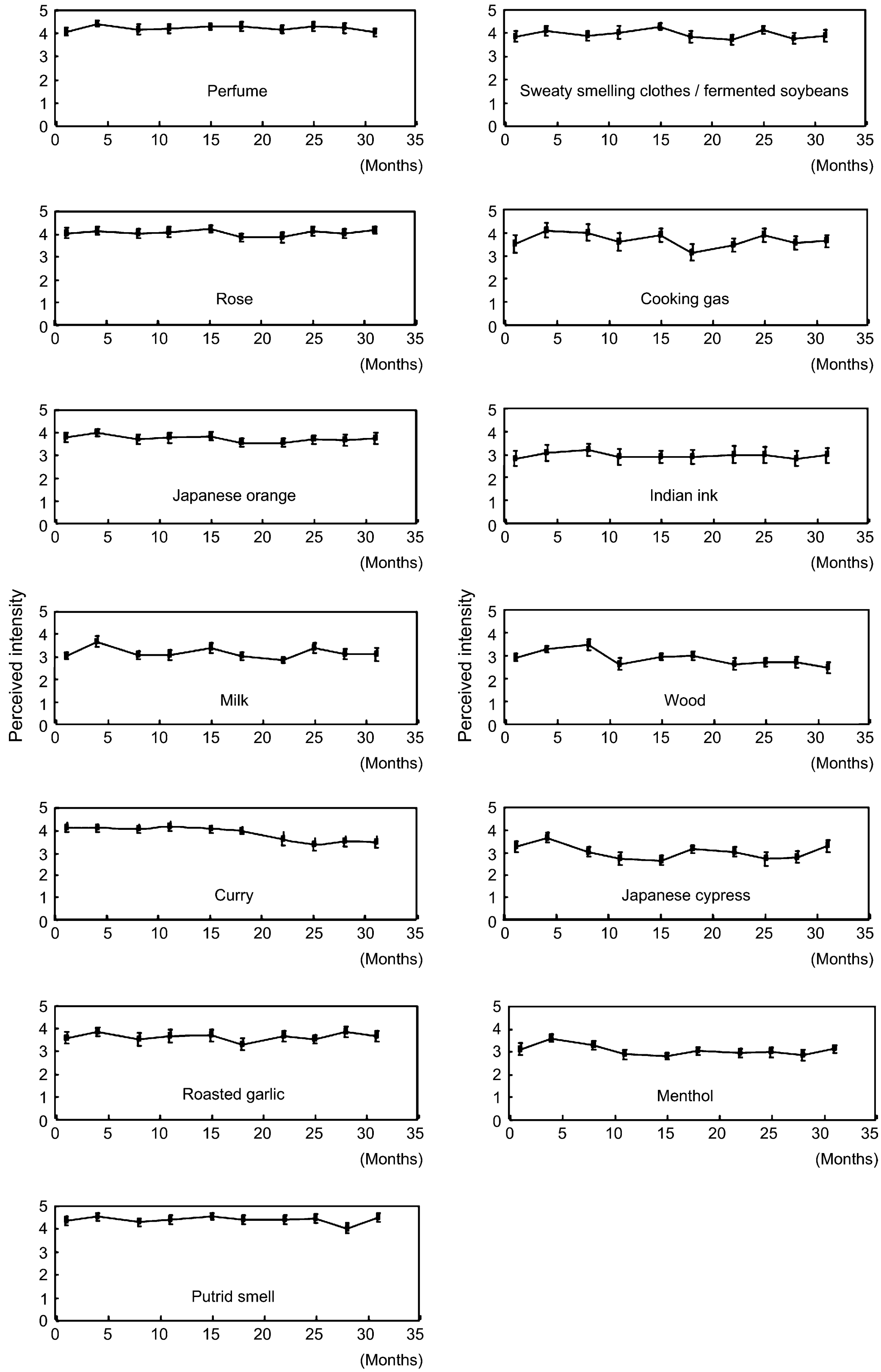

Figure 4 Stability of 13 odor sticks in perceived intensity spanning 31 months. Each graph shows the change of average perceived intensity of each odor item at 1, 4, 8, 11, 15, 18, 22, 25, 28, and 31 months after the odor sticks were manufactured. Horizontal axis shows months after the manufacture, and vertical axis shows perceived intensity in each graph (bars show SE, n = 14).

Significant differences in intensity were found among the 10 sessions by ANOVA for the following odors: curry [F (9, 117) = 3.49, $P < 0.01$], cooking gas [F (9, 117) = 2.53, $P < 0.05$], wood [F (9, 117) = 2.91, $P < 0.01$], Japanese cypress [F (9, 117) = 2.72, $P < 0.01$], and menthol [F (9, 117) = 2.07, $P < 0.05$]. We could not, however, find significant differences between the perceived intensity in the first session and that in the other sessions using multiple comparisons.

Evaluating the OSIT using the four-plus alternative method (Exp. 2)

OSIT score in applying healthy young and middle-aged people (Exp. 2-1)

Every odor item was identified correctly by at least 80% of participants (mean, 90.1%; range, 80.8–98.7%).

Internal consistency (Test–retest) of OSIT (Exp. 2-2)

We calculated Pearson's correlation coefficient between test and retest to compare this coefficient in previous studies (Doty *et al.*, 1985; Kobal *et al.*, 1996; Nordin *et al.*, 1998; Hummel *et al.*, 2001; Frank *et al.*, 2003, etc.). The correlation coefficient between two tests was r_{47} = 0.772 ($P < 0.001$) with respect to the identification rate for each participant (Figure 5). There was no significant difference [t (46) = −0.182, NS] between the average identification rate (±SE) on the initial test (0.827 ± 0.027, n = 47) and that for the retest (0.828 ± 0.028, n = 47).

Comparison of the OSIT with the CC-SIT (Exp. 2-3)

Pearson's correlation coefficient between OSIT score and CC-SIT score was 0.701 ($P < 0.01$). Average identification rate (±SE) using the OSIT (0.825 ± 0.018, n = 108) was significantly higher than that for CC-SIT (0.738 ± 0.016, n = 108) [t (107) = 6.57, $P < 0.001$]. In particular, "turpentine," "paint thinner," "rose," "lemon," and "smoke" showed lower identification rate than 0.7(0.509, 0.509, 0.519, 0.528, and 0.648, respectively) in CC-SIT, whereas Japanese orange showed lower identification rate than 0.7 (0.509) in OSIT. The average estimated familiarity score (±SE) for the odorants was significantly higher for OSIT (7.22 ± 0.160, n = 108) than that for CC-SIT (6.67 ± 0.163, n = 108) [t (107) = 3.87, $P < 0.001$].

Testing Japanese people from different generations with the OSIT (Exp. 2-4)

The averaged identification rate from each 10-year age group was calculated. Figure 6 shows the change of odor identification rate in each 10-year age group. The data from the 80- to 89-year-old age group were excluded from the statistical analysis because of the small number of participants. ANOVA showed significant difference among 10-year age groups [F (5, 436) = 19.0, $P < 0.001$]. Multiple comparisons resulted in 60- to 69-year-old age group and 70- to 79-year-old age group showing lower identification rate than 20- to 29-year-old, 30- to 39-year-old, and 40- to 49-year-old age groups ($P < 0.001$ in all cases). Seventy- to 79-year-old age group showed significantly lower identification rate than 50- to 59-year-old age group and 60- to 69-year-old age group ($P < 0.001$, $P < 0.05$, respectively).

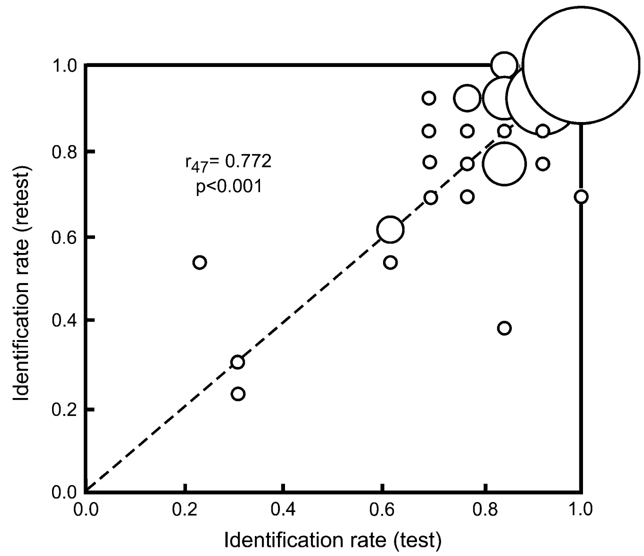

Figure 5 Correlation between identification rates (scores) obtained with test and retest (n = 47). Horizontal axis shows OSIT scores of test, and vertical axis shows that of retest. The larger the circle, the more data points converge on that coordinate. The largest circle shows eight data points, the next largest one four, the smallest one shows one datum, the next smallest one two data points, and the middle one three data points. The coefficient of correlation was r_{47} = 0.772 ($P < 0.001$).

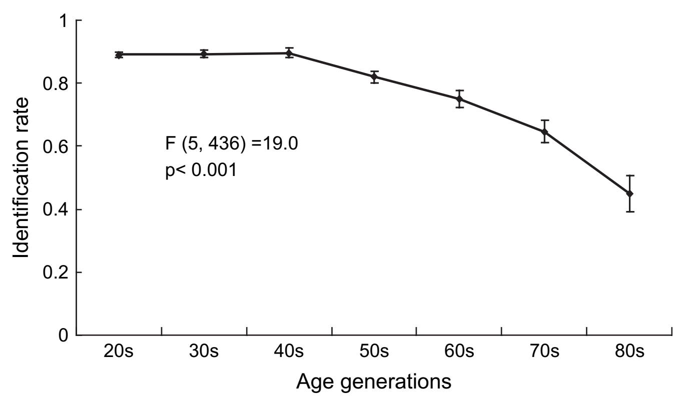

Figure 6 Change of average odor identification rate (+SE) in each 10-year age group (n = 448). ANOVA showed significant difference among 10-year age groups from 20s to 70s [F (5, 436) = 19.0, $P < 0.001$] without the data from 80s because of small number of subjects. Multiple comparisons showed that 60s and 70s showed lower identification rate than 20s, 30s, and 40s ($P < 0.001$ in all cases). The 70s showed significantly lower identification rate than 50s and 60s ($P < 0.001$, $P < 0.05$, respectively).

Examination of two-step identification method (Exp. 3)

To compare the two-step identification method to the four-plus alternative method, we classified the participants into 12 groups by sex and their 10-year age group ($2 \times 6 = 12$) and calculated the Pearson's correlation coefficient of the averaged identification rate of each group between two methods. The correlation coefficient was significantly high ($r_{12} = 0.817$, $P < 0.01$, Figure 7). We also calculated the correlation coefficient of identification rate of each odorant measured with the two methods and found that it was also significantly high ($r_{13} = 0.735$, $P < 0.01$).

Figure 8 shows frequency distribution of identification rate in both methods. The variance of identification rate of two-step identification method (0.053) was larger than that of four-plus alternative method (0.033) [F (41, 144) = 1.47, $P < 0.05$], and average identification rate (±SE) of two-step identification method (0.55 ± 0.035, n = 42) was lower than that of four-plus alternative method (0.81 ± 0.015, n = 145) [t (57)= −6.60, $P < 0.001$].

Examination of using picture–word alternatives (Exp. 4)

The results of average identification rates in each method were shown in Figure 9. For the four-plus alternative method, the average identification rate (±SE) using word alternatives was 0.90 (±0.016, n = 40), whereas that using picture–word alternatives was 0.94 (±0.013, n = 40). For two-step identification method, the average identification rate (±SE) using word alternatives was 0.64 (±0.061, n = 15), whereas that using picture–word alternatives was 0.65 (±0.053, n = 18). There was no significant difference between the results using word alternatives and that using picture–word alternatives in both four-plus alternative method [t (78) = −1.90, n.s.] and two-step identification method [t (31) = −0.14, n.s.].

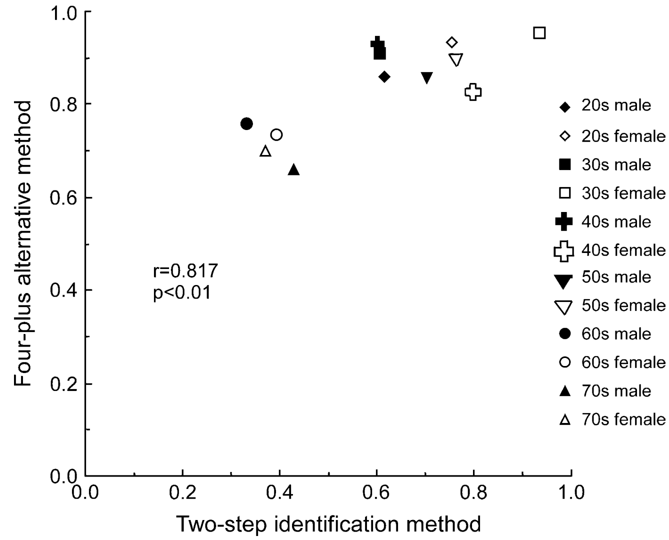

Figure 7 Relation between two-step identification method and four-plus alternative method. The average identification rates were calculated in 12 groups by sex and 10-year age group from 20s to 70s ($2 \times 6 = 12$) for both methods, and the Pearson's correlation coefficient between the average identification rates obtained by two-step identification method and those obtained by four-plus alternative method was calculated. The correlation coefficient was $r_{12} = 0.817$ ($P < 0.01$).

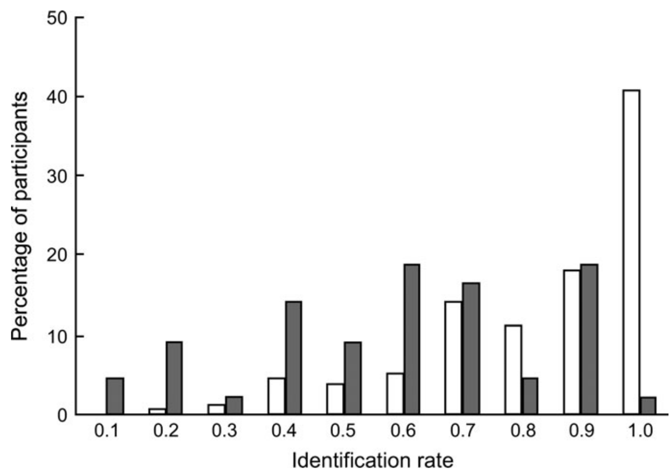

Figure 8 Frequency distribution of identification rates obtained by four-plus alternative method (white bars) and by two-step identification method (gray bars). The x-axis shows the identification rate, and the y-axis shows the percentage of participants. The participants (%) who showed the identification rate from 0.0 to 0.1 was shown at 0.1 of the horizontal line. The variance of identification rate of two-step identification method was larger than that of four-plus alternative method [F (41, 144) = 1.47, $P < 0.05$].

Discussion

Performance of odor stick

We applied a free-choice paradigm in this study to assess the representation of odor quality, and more than 70% participants reported descriptor of the expected odor or cluster in all odor items. It suggests that these odor sticks successfully represented the expected odor, considering the variability and breadth of possible descriptors in a free-choice paradigm. In addition, the odors were perceived to be intermediate in intensity. We assume that the odor quality and intensity of these odor sticks would be appropriate for use in smell tests.

In Exp. 1-2, we examined odor quality and perceived intensity up to 31 months after the sticks were manufactured. These results suggest that all the odor sticks retained their qualitative smell for at least 22 months after their manufacture. In addition, the perceived intensity in each of the 10 sessions that were conducted in the 31 months after the manufacture of the odor sticks was examined. Significant changes in intensity were found for some of the odorants. We, however, could not find significant differences between the perceived intensity in the first session and that in the other sessions for any of the odor items. These results suggest that the differences found in the perceived intensity among

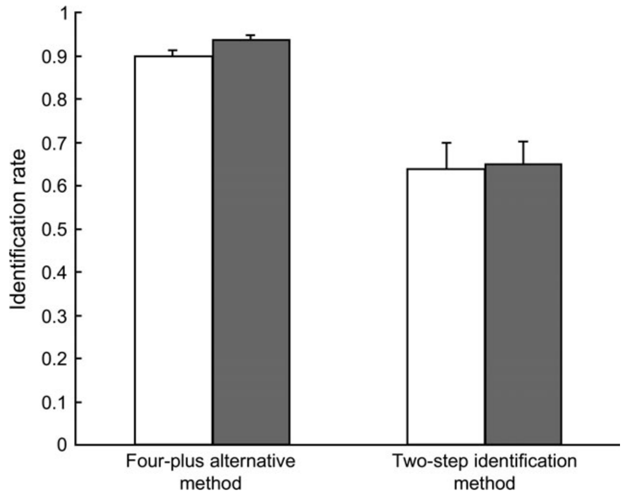

Figure 9 Average identification rate (±SE) obtained by using "word alternatives" (white bars) and that obtained by using "picture–word alternatives" (gray bars). Nine odor items were applied to 113 young healthy volunteers (age range: 19–30 years). There was no significant difference between average identification rate obtained by "word alternatives" and that obtained by "picture–word alternatives" in both the four-plus alternative method (left side) and the two-step identification method (right side).

the 10 sessions were due to cognitive factors rather than the degradation of the odor sticks. Indeed, it has been reported that prior experience or knowledge regarding an odor greatly affects the perception of that odor (Distel *et al.*, 1999). Our results suggest that both the quality and perceived intensity of the odor sticks were stable for at least 22 months.

Performance of the OSIT

We tested healthy participants with the newly developed OSIT. Each odor item was identified correctly by at least 80% of participants. We therefore believe that Japanese people can easily identify these odor items. Wood and Japanese orange, however, showed the lowest identification rate of 80.8%, which may be due to manufacturing difficulties or mismatching of the alternatives in the forced-choice paradigm. In particular, the odor of Japanese orange tended to be misidentified as apple, an alternative from the same odor category. The identification rate could be improved by using different alternatives in the same category. We currently are considering replacing the Japanese orange odorant with more stable simple odorant.

Japanese individuals commonly associate isovaleric acid with two everyday odors: fermented soybeans or sweaty smelling clothes. We prepared two correct answers for this compound because it would be unfair for a participant who identified this odor as fermented soybeans if only sweaty smelling clothes was prepared as correct answer. As we anticipated, about half (45.6%) of the participants in Exp. 2-1 identified this odor as fermented soybeans, and another half (49.5%) identified it as sweaty smelling clothes. While, isovaleric acid was selected to be an odor component by Japanese Standardized Olfactory Test (Zusho, 1983; Takagi, 1989), and as it is a stable, simple odor that is easily identified. Therefore, we would like to include this odorant in OSIT. Therefore, we chose this response format for isovarelic acid, though it is irregular. We, however, must be careful to compare the identification rate with those of other odorants because the probability of correct answer is 0.5, which is twice of other items. We are now considering solving this irregularity by using one correct answer format for this odorant.

Exp. 2-2 revealed that the statistically significant test–retest reliability was 0.772. Factors that might affect the reproducibility of the performance of the test–retest paradigm include changes in sensitivity to the odorant or learning effect induced during the interval between the test and retest, the number of test items, and the experimental environment in cases like seasonal factors or differences in the person administering the test (Liu *et al.*, 1995). According to the previous studies using UPSIT to investigate the effect of the interval between the tests, no statistical significance was found among correlation coefficients in these studies (r_{69} = 0.95 for 2-week interval, Doty *et al.*, 1985; r_{53} = 0.918 for 6-week interval, Doty *et al.*, 1984b; r_{100} = 0.91 for an average of 9 days interval, Frank *et al.*, 2003). In the present study, the interval between sessions was widely ranged; however, there were no significant differences in the performance of participants tested with longer and shorter intervals.

It has been also reported that using fewer odorants results in lower correlation coefficient between the identification rates from the test and the retest. For example, the correlation coefficients were 0.95 for UPSIT (40 items) and 0.96 for Geur Identification Test Utrecht (GITU, 36 items) (Hendriks, 1992). On the other hand, the coefficients were 0.71 for CC-SIT and 0.77 or 0.68 for GITU, in which the numbers of odor items were reduced to 12 (Doty *et al.*, 1995) and 18 (Hendriks, 1988), respectively. In other tests that used a small number of odorants, the correlation coefficient between the two sessions was approximately 0.7 [Hummel *et al.*, 2001 (r_{113} = 0.78, 12 items); Nordin *et al.*, 1998 (r_{19} = 0.68, 16 items); Hummel *et al.*, 1997 (r_{104} = 0.73, 12 items); Kobal *et al.*, 1996 (r_{52} = 0.75, 7 items)]. The correlation coefficient calculated in the present study was comparable to studies that employed a similar number of odor items, thereby verifying the internal consistency of OSIT.

Comparing the OSIT with the CC-SIT demonstrated the significant Pearson's correlation coefficient (r_{90} = 0.754) between the scores obtained when the two tests were given to Japanese patients with olfactory deficits (Hashimoto *et al.*, 2004). In this study, the correlation coefficient between two tests (r_{108} = 0.701) was also significantly high for a wide range of age groups, suggesting the utility of OSIT as a smell identification test.

The UPSIT has been used to show differences between different age groups in odor identification performance (Doty *et al.*, 1984a). Our result also demonstrated the usefulness of OSIT in measuring the decline in olfactory performance that is observed in older individuals.

We compared the OSIT scores obtained by two-step identification method with that obtained by four-plus alternative method to consider the characteristics of two-step identification method. For each group categorized based on the sex and age of the individuals, the average identification rate measured by two-step identification method was highly correlated with that measured by four-plus alternative method. In addition, the correlation between the identification rates for each of the odorants measured with the two methods was also high. These suggest that the two-step identification method measures similar aspects of olfactory performance correlated to four-plus alternative method. On the other hand, the two-step identification method resulted in lower identification rates than the four-plus alternative method. This reflects the difficulty of odor identification without limited concrete clues, such as the actual concrete odor name, because a participant was instructed to select one from eight odor categories in two-step identification method, while a participant was instructed to select one from four concrete odor names in four-plus alternative method. The variance of identification rate was larger using the two-step identification method than the four-plus alternative method. Hendriks (1992) reported that the variance of identification rate became larger when the number of alternatives became larger. Future study should focus on whether this large variance is a result of the simple difficulty of judgment or the individual differences in identification performance.

Concerning using picture–word alternatives to identify the odors, there were no significant differences in the averaged identification rate of young adult participants, compared to only word alternatives. The participants, however, reported that the identification test using the picture–word alternatives was interesting and made the test easier. We believe that the picture–word alternatives will be useful in the test for children, the elderly, and people with dementia.

In this study, we applied a modified forced-choice paradigm, in which the choices ''detectable but not recognized'' and ''no smell detected'' were added to the four odor alternatives. We thought that adding the two alternatives worked effectively to measure the precise degree of the abnormal performance while also minimizing the patient's possible anxiety associated with strict forced choice. Studies using the four-plus alternative method on the patients with olfactory dysfunction have indicated strong correlations (r_{120} = −0.802 in Kobayashi *et al.*, 2004; r_{110} = −0.766 in Hashimoto *et al.*, 2004; r_{85} = −0.84 in Miwa *et al.*, 2004) with the Japanese Standardized Olfactory Test. These studies show that this test method is effective to measure the olfactory abnormal performance in clinical use. The OSIT scores applied to wide age-ranged normal people also showed a significant correlation with the Japanese Standardized Olfactory Test (r_{108} = −0.69 in Saito *et al.*, 2001), suggesting the usefulness in measurement of impaired olfaction in the aged people.

We also called this test OSIT-J because odor items were selected for Japanese people. Kobayashi *et al.* (2005) tried to apply the OSIT-J to the participants in the United States and found that several odor items were not familiar to them. They suggested that the OSIT-J was effective in identifying the US participants with normal smell function and that the identification of test odorants having a cultural bias was critical when evaluating olfactory function tests for use in different populations. Therefore, the test odorants should be considered in application of OSIT (or OSIT-J) system to other countries.

Acknowledgements

The authors would like to express their sincere thanks to Ms Masuda Y. and Yasuoka N. for their help during the experiments.

References

Ayabe-Kanamura, S., Schicker, I., Laska, M., Hudson, R., Distel, H., Kobayakawa, T. and **Saito, S.** (1998) *Differences in perception of everyday odors: a Japanese-German cross-cultural study*. Chem. Senses, 23, 31–38.

Cain, W.S., Gent, J., Catalanotto, F.A. and **Goodspeed, R.B.** (1983) *Clinical evaluation of olfaction*. Am. J. Otolaryngol., 4, 252–256.

Distel, H., Ayabe-Kanamura, S., Martinez-Gomez, M., Schicker, I., Kobayakawa, T., Saito, S. and **Hudson, R.** (1999) *Perception of everyday odors—correlation between intensity, familiarity and strength of hedonic judgment*. Chem. Senses, 24, 191–199.

Doty, R.L., Marcus, A. and **Lee, W.W.** (1996) *Development of the 12-items cross-cultural smell identification test (CC-SIT)*. Laryngoscope, 106, 353–356.

Doty, R.L., McKeown, D.A., Lee, W.W. and **Shaman, P.** (1995) *A study of the test-retest reliability of ten olfactory tests*. Chem. Senses, 20, 646–656.

Doty, R.L., Newhouse, M.G. and **Azzalina, J.D.** (1985) *Internal consistency and short-term test-retest reliability of the University of Pennsylvania smell identification test*. Chem. Senses, 10, 297–300.

Doty, R.L., Shaman, P., Applebaum, S.L., Giberson, R., Siksorski, L. and **Rosenberg, L.** (1984a) *Smell identification ability: changes with age*. Science, 226, 1441–1443.

Doty, R.L., Shaman, P. and **Dann, M.** (1984b) *Development of the University of Pennsylvania smell identification test: a standardized microencapsulated test of olfactory function*. Physiol. Behav., 32, 489–502.

Frank, R.A., Dulay, M.F. and **Gesteland, R.C.** (2003) *Assessment of the sniff magnitude test as a clinical test of olfactory function*. Physiol. Behav., 78, 195–204.

Hashimoto, Y., Fukazawa, K., Fujii, M., Takayasu, S., Muto, T., Saito, S., Takashima, Y. and **Sakagami, M.** (2004) *Usefulness of the odor stick identification test for Japanese patients with olfactory dysfunction*. Chem. Senses, 29, 565–571.

Hendriks, A.P.J. (1988) *Olfactory dysfunction*. Rhinology, 26, 229–251.

Hendriks, A.P.J. (1992) *Gitu versus UPSIT. A comparison between two clinical tests of odour identification*. Chem. Senses, 17, 838–839.

Hummel, T., Konnerth, C.G., Rosenheim, K. and **Kobal, G.** (2001) *Screening of olfactory function with a four-minute odor identification test: reliability, normative data, and investigation in patients with olfactory loss*. Ann. Otol. Rhinol. Laryngol., 110, 976–981.

Hummel, T., Sekinger, B., Wolf, S.R., Pauli, E. and **Kobal, G.** (1997) *'Sniffin' Sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold*. Chem. Senses, 22, 39–52.

Kobal, G., Hummel, Th., Sekinger, B., Barz, S., Roscher, S. and **Wolf, S.** (1996) *"Sniffin' Sticks": screening of olfactory performance*. Rhinology, 34, 222–226.

Kobayashi, M., Nishida, K., Nakamura, S., Oishi, M., Majima, Y., Maeda, T., Furuta, S., Takashima, Y. and **Saito, S.** (2004) *Suitability of the odor stick identification test of Japanese in patients suffering from olfactory dysfunction*. Acta Otolaryngol. Suppl., 124, 74–79.

Kobayashi, M., Saito, S., Kobayakawa, T., Deguchi, Y. and **Costanzo, R.M.** (2005) *Japanese odor stick identification test (OSIT-J): comparison of data from U.S. and Japanese subjects*. In AchemS XXVII.

Liu, H.C., Wang, S.J., Lin, K.P., Lin, K.N., Fuh, J.L. and **Teng, E.L.** (1995) *Performance on a smell screening test (the MODSIT): a study of 510 predominantly illiterate Chinese subjects*. Physiol. Behav., 58, 1251–1255.

Miwa, T., Furukawa, M., Tsukatani, T., Ikeno, S., Yata, Y., Horikawa, I., Deguchi, Y. and **Saito, S.** (2004) *Clinical usefulness of odor stick identification test for patients with olfactory disturbance*. J. Otolaryngol. Jpn., 107, 956–965 (in Japanese).

Nordin, S., Bramerson, A., Liden, E. and **Bende, M.** (1998) *The Scandinavian odor-identification test: development, reliability, validity and normative data*. Acta Otolaryngol., 118, 226–234.

Saito, S., Iida, T., Sakaguchi, H. and **Kodama, H.** (1997) *Description of quality of pollution odor.* J. Odor Res. Eng., 28, 32–43 (in Japanese).

Saito, S., Kobayakawa, T., Kuchinomachi, Y., Ayabe-Kanamura, S. and **Takashima, Y.** (1999) *A smell test based on odor recognition by Japanese people, and its application*. In Bell, G.A. and Watson, A.J. (eds), Tastes & Aromas: The Chemical Senses in Science and Industry. UNSW Press, Sydney, Australia, pp. 75–82.

Saito, S., Masuda, Y., Kobayakawa, T., Gotou, N., Mizoguchi, C. and **Takashima, Y.** (2001) *Relation between the Japanese standardized olfactory test and the odor stick identification test for Japanese in 108 Japanese participants aged 20 to 81 years*. Jpn. J. Taste Smell Res., 8, 143–149 (in Japanese).

Takagi, S.F. (1989) *Standardized olfactometries in Japan—a review over ten years*. Chem. Senses, 14, 25–46.

Zusho, H. (1983) *Olfactometry in Japan*. Rhinology, 21, 281–285.

Zusho, H., Yamamoto, K., Negishi, M., Kobayashi, K., Doty, R.L. and **Applebeaum, S.** (1983) *Comparison of Japanese and American subjects on the University of Pennsylvania smell identification test*. In Proceedings of the 17th Japanese Symposium on Taste and Smell; 145–148 (in Japanese).

Accepted February 1, 2006