Chem. Senses 33: 17–21, 2008
doi:10.1093/chemse/bjm058
Advance Access publication August 30, 2007

Different Odor Tests Contribute Differently to the Evaluation of Olfactory Loss

Jörn Lötsch[1], Heinz Reichmann[2] and Thomas Hummel[3]

[1]Pharmazentrum Frankfurt/ZAFES, Institute of Clinical Pharmacology, Johann Wolfgang Goethe-University, Theodor Stern Kai 7, D-60590 Frankfurt am Main, Germany, [2]Department of Neurology, University of Dresden Medical School, Fetscherstrasse 74, D-01307 Dresden, Germany and [3]Smell and Taste Clinic, Department of Otorhinolaryngology, University of Dresden Medical School, Fetscherstrasse 74, D-01307 Dresden, Germany

Correspondence to be sent to: Thomas Hummel, Smell and Taste Clinic, Department of Otorhinolaryngology, University of Dresden Medical School ("Technische Universität Dresden"), Fetscherstrasse 74, 01307 Dresden, Germany. e-mail: thummel@mail.zih.tu-dresden.de

Abstract

In a clinical context, the importance of the sense of smell has increasingly been recognized, for example, in terms of the evaluation of neurodegenerative disorders. In this study, 2 strategies of olfactory testing, a simple one and a more complex one, were compared with respect to their suitability to assess olfactory dysfunction. Odor threshold (T), discrimination (D), and identification (I) were assessed in a control sample of 916 males and 1160 females, aged 6–90 years, and in 81 men and 21 women, aged 38–80 years, suffering from idiopathic Parkinson's disease (IPD). Sums of the 3 subtest results T, D, and I yielded threshold discrimination identification (TDI) scores reflecting olfactory function. Sensitivity of any of the 3 subtests to confirm the diagnosis established by the composite TDI score was assessed separately for each test. Principal component analyses were applied to determine any common source of variance among the 3 specific subtests. Sensitivities of the subtests to provide the diagnosis established by the composite TDI score were 64% (T), 56% (D), and 47% (I), respectively. In IPD patients, each of the subtests provided the correct diagnosis (sensitivity >90%), as olfaction was impaired in 99% of the patient group. Two principal components emerged in both controls and IPD patients, with eigenvalues >0.5. The first component received high loadings from all factors. The second component received high loadings from odor threshold, whereas loadings from odor discrimination and identification were much smaller. In conclusion, combined testing of several components of olfaction, especially including assessment of thresholds, provides the most significant approach to the diagnosis of smell loss.

Key words: olfaction, psychophysics, smell

Introduction

The clinical significance of the sense of smell, for example, with respect to the diagnosis of neurodegenerative disorders (Hawkes 2006) such as early diagnosis of Parkinson's disease (Berendse and Ponsen 2006; Haehner et al. 2007) has been increasingly appreciated.

Dissecting the sense of smell leads to at least 3 different components, namely, 1) the perception of odors at low concentrations (odor threshold), 2) the nonverbal distinction of different smells (odor discrimination), and 3) the ability to name or associate an odor (odor identification). Some psychophysical tests assessing olfactory performance include separate subtests for the assessment of each of these components (Hummel et al. 1997) or some of them (Cain and Rabin 1989; Thomas-Danguin et al. 2003; Lam et al. 2006), whereas others rely on just one single component (Doty et al. 1984). Whether testing of two or more components of olfaction assesses different or redundant aspects is controversial. The suggestion that tests of single components of olfactory function measure a common source of variance (Doty et al. 1994) is questioned by the observation of impaired odor identification but unchanged odor thresholds after focal brain excision (Jones-Gotman and Zatorre 1988). To provide guidance for the adoption of a simple or a more complex olfactory testing strategy, we analyzed the differential contribution of olfactory tests to the diagnosis of overall olfactory performance in 2076 controls and 102 patients with idiopathic Parkinson's disease (IPD).

Materials and methods

Subjects and testing of olfactory function

Controls were (916 men, 1160 women, aged 6–90 years, mean age 35.2 ± 16.2 years) either presented at the ENT clinic with complaints related to taste or smell or recruited as healthy participants in various studies concerned with olfactory function. Results from IPD patients (81 men, 21 women, aged 38–80 years, mean age 60.8 ± 10.1 years, unified Parkinson's disease rating scale staging 24.7, range 4–58 [Ramaker et al. 2002]) were pooled from previous studies (including, e.g., Muller, Mungersdorf, et al. 2002; Muller, Reichmann, et al. 2002; Mueller et al. 2005; Hummel et al. 2007).

Olfactory function was assessed birhinally with the "Sniffin' Sticks" test (Burghart GmbH, Wedel, Germany) (Hummel et al. 1997). In this, validated test odors are presented in felt-tip pens. For odor presentation, one pen at a time—cap removed—is placed in front of the nostrils at a distance of approximately 1–2 cm (for a detailed description of the test procedures, please see Hummel et al. 1997).

Odor thresholds were obtained for the rose-like odor phenylethyl alcohol presented in sixteen 1:2 dilution steps starting from a 4% solution. Using a 3-alternative forced-choice task and a staircase paradigm starting at low phenylethyl alcohol concentrations, one pen with the odorant and 2 blanks were presented at each dilution step. Two successive correct identifications or one incorrect identification, respectively, triggered a reversal of the staircase. Odor threshold was represented by the mean of the last 4 out of 7 staircase reversals (normal values: >6 for men, >6.5 for women, respectively [Hummel et al. 2007]). Odor discrimination was determined with 16 triplets of pens, 2 containing the same odorant and the third a different one (for detailed listing of the used odorants, see Hummel et al. 1997), employing a 3-alternative forced-choice task (normal value: ≥11 correct discriminations). Odor identification was determined with 16 odors (i.e., orange, peppermint, turpentine, cloves, leather, banana, garlic, rose, fish, lemon, coffee, anise, cinnamon, liquorice, apple, and pineapple) using a 4-alternative forced-choice task with presentation of a list of 4 descriptors for each pen (normal value: ≥12 correct identifications). The clinical evaluation of olfactory performance was based upon the composite "threshold discrimination identification score" (TDI) represented by the sum of the scores from the 3 subtests (Wolfensberger et al. 2000). Pathologic olfactory function was indicated by TDI ≤30.5, with the separation of hyposmia from functional anosmia at TDI = 15.5 (Hummel et al. 2007).

Statistics

The following approach was chosen in order to investigate whether the 3 single olfactory tests provide redundant information or whether each test provides clinically relevant specific information: First, sensitivity and specificity, with respect to the distinction between normal and pathologic olfactory function according to TDI scores, were assessed separately for each test, and for all 3 pairs of 2 subtests each (sensitivity [%] = 100 × correctly positive/[correctly positive + false negative]; specificity [%] = 100 × correctly negative/[correctly negative + false positive]) (Altman and Bland 1994). Second, correlations between the 3 olfactory tests were computed (Spearman's r). By employing 16 items in each test (i.e., 16 triples of odors in the discrimination test and 16 different odors in the identification test) and using 3- or 4-alternative forced-choice tasks, odor discrimination or identification scores up to 5.33 or 4, respectively, are equal to chance, whereas the staircase paradigm of threshold measurement almost excludes chance. Therefore, scores within the range of chance were set to 0, and the chance level was subtracted from all scores above chance. Thus, the chance-corrected values then ranged between 1 and 16 for odor thresholds (i.e., no correction needed) but between 0 and 10.67 or 12 for discrimination and identification, respectively. To obtain similar scaling for all 3 subtests, the values were linearly rescaled to a range of 0–16 as rescaled value = (actual value − 1)/(16 − 1) × 16 for threshold, as rescaled value = actual value/10.67 × 16 for discrimination, and as rescaled value = actual value/12 × 16 for discrimination. Third, the intercorrelation matrix of these rescaled values was subjected to principal component analyses (Pearson 1901) retaining factors with eigenvalues higher than 0.5. This limit was chosen rather than the conventional limit at 1 (Kaiser 1974; Doty et al. 1994) because the latter may miss important factors (Ivanenko et al. 2004). Forth, the results of olfactory subtests were submitted to hierarchical cluster analysis, clustering variables rather than cases in order to assess similarity between the 3 subtests (Systat 11, Systat Software, Inc., San Jose, CA).

Results

Controls

TDI values ranged from 9 to 47.5 (average 34.5 ± 5.5). The fraction of 20.1% subjects showing pathologic olfactory function (15 subjects were anosmic, TDI ≤ 15.5, and 402 hyposmic, TDI≤30.5; Figure 1) was greater than in an average population (Hummel et al. 2007) due to subjects presenting with complaints related to taste or smell being included.

The sensitivity for identifying pathologic olfactory function (as determined by TDI scores) was 64% for odor threshold, 56% for the discrimination test, and 47% for the identification task (test specificity 92%). For example, identifications of pathologic TDI were correctly positive in 268 cases, correctly negative in 1529 cases, false positive in 130 cases, and false negative in 149 cases. With each pair of subtests, test sensitivity remained below 90% (86% for threshold and discrimination, 84% for threshold and identification,

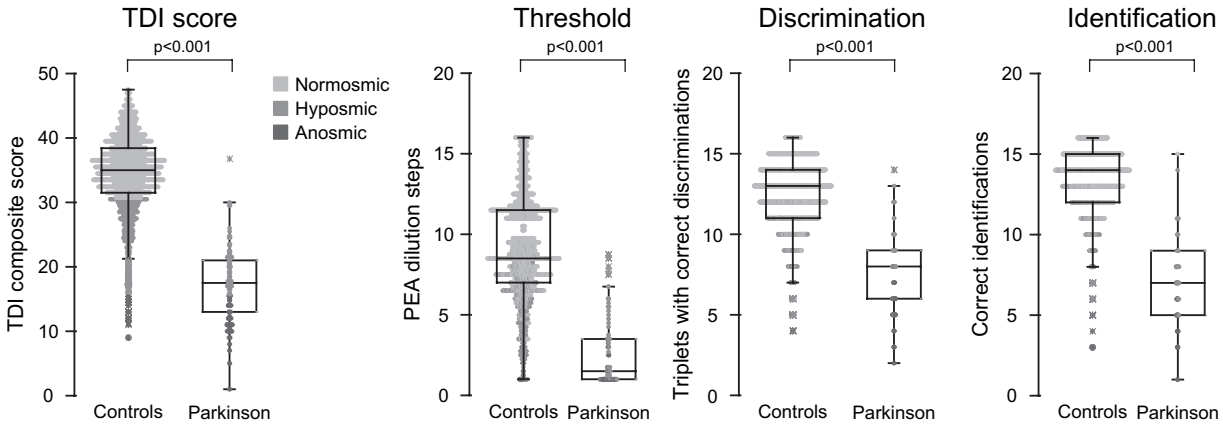

Figure 1 Observed results (single values as dots, statistical summaries in overlaid box plots) of the composite TDI score (left), which is obtained as the sum of the results of testing of odor threshold, discrimination, and identification (right), displayed separately for controls (n = 2076) and IPD patients (n = 102) and color coded for the TDI-based olfactory diagnoses. The boxes span the 25th to 75th percentiles, with the median crossing the box as a horizontal line and the whiskers spanning values within 1.5 times the 25th to 75th percentiles. The statistical differences are the results from Wilcoxon test group comparisons.

and 78% for discrimination and identification) and test specificity decreased to 84%.

Odor threshold, odor discrimination, and identification (rescaled to correct for chance) were weakly but significantly correlated among each other (r^2 = 0.07 for threshold vs. discrimination, 0.08 for threshold vs. identification, and 0.07 for discrimination vs. identification; $P < 0.001$). Two principal components emerged with eigenvalues >0.5 (Figure 2). The first component, explaining 55% of the total variance, bore high loadings from all factors (0.71, 0.74, and 0.77 for odor threshold, discrimination, and identification, respectively). The second component, explaining 24% of the total variance, received the main loading from odor threshold (−0.67), whereas loadings from odor discrimination and identification were smaller (0.49 and 0.15, respectively). Hierarchical cluster analysis showed that discrimination and identification were joined first, whereas the resulting cluster was joined by thresholds at a greater distance (Figure 3).

IPD patients

Olfactory function was pathologic in 99% of the IPD patients (37 anosmic, 64 hyposmic; average TDI score 17.2 ± 5.7; Figure 1). The significant differences in TDI scores and single test results ($P < 0.001$, Wilcoxon tests) persisted when comparing the patients with 102 age- and sex-matched controls. Neither the patients' age nor the duration of confirmed IPD nor the severity of IPD were correlated with olfactory test results.

With most patients having pathologic results in each subtest, diagnosis of pathologic olfactory function was obtained at a test sensitivity of >0.9 when solely based on any one of the 3 subtests. When paired subtests were tried for establishing the olfactory status, test sensitivity increased to 99% and test specificity remained high at 100%.

Results from correlational and principal component analyses in IPD patients were similar to those obtained in controls (Figure 2, numerical details not given). Clustering of the subtests was tighter (Figure 3), with threshold and discrimination or threshold and identification as closely linked in IPD patients as discrimination and identification in controls.

Discussion

In the present study, the composite TDI score was used as the standard for the diagnosis of normal or pathologic olfactory function. This score has been repeatedly shown to provide a clinically relevant estimate of olfactory function (Kobal et al. 2000; Muller, Reichmann, et al. 2002). With test sensitivities below 66%, the 3 individual tests for odor threshold, discrimination, and identification performed poorly in terms of providing the correct olfactory diagnosis of the non-IPD subjects. Test sensitivity remained <90% with paired subtests. The dissimilarity of the subtests was further supported by their weak correlations, with values of r^2 <0.1, which probably became significant merely due to the large number of cases. Thus, using just one of the subtests bears the risk to miss the diagnosis of olfactory loss. The information obtained with the complete test set, combining differential aspects of olfactory function, may be especially important in the diagnostics performed at early stages of diseases such as IPD.

Odor threshold exhibited more distinct properties than the discrimination and identification subtests, which are both based upon presentation of suprathreshold odors. This difference was indicated by the tighter relation of the latter 2 as compared with threshold in the cluster analysis (Figure 3) and, in addition, by results of the principal component analysis that revealed 2 main components, one being related to odor threshold and the other one to odor identification and

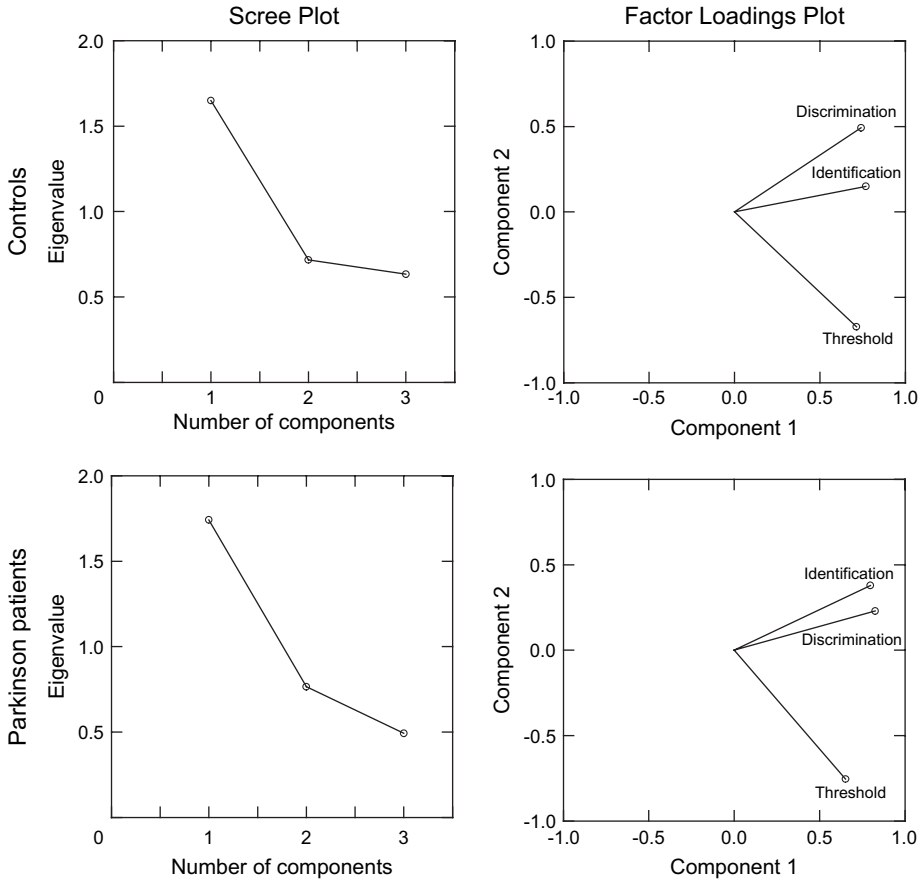

Figure 2 Principal components of olfactory subtests. Left: scree plot of the eigenvalues against their associated component identifying large values that separate well from smaller eigenvalues and thus identifying a useful number of factors to retain. Right: component plot of the principal components of variance among odor threshold, odor discrimination, and identification resulting when retaining factors with eigenvalues >0.5.

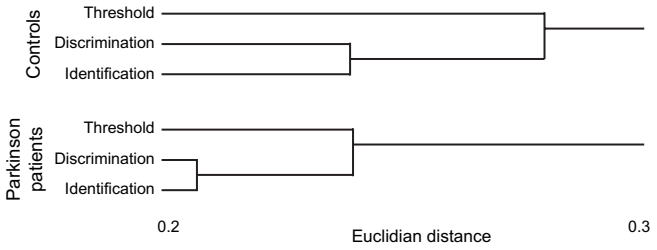

Figure 3 Hierarchical clustering of performance in the tests of odor threshold, odor discrimination, and odor identification, indicating that identification and discrimination belonged to the same cluster, whereas threshold belonged to a separate cluster. Clustering was tighter in the IPD patients than in the controls.

discrimination, with thresholds on the one hand and discrimination on the other hand almost orthogonally to each other, which supports their independency (Figure 2). This consistency among statistical test results indicates that the less conservative statistical criterion applied in the principal component analysis better reflected olfactory dimensions than a conservative analysis (Doty et al. 1994), which yielded only one component.

A differential role of memory for threshold testing as opposed to discrimination and identification testing may contribute to the distinct role of threshold compared with the latter 2 subtests. Although formally a 3-alternative paradigm is used, the pen containing the odor may be identified immediately without necessary reference to the nonsmelling pens. In contrast, during odor discrimination testing, the subject has to memorize the smell of the other pens before completing the task and memorizing odors is also required, at least to some degree, for odor identification. This hypothesized difference in the role of odor memory for threshold and discrimination/identification testing is in line with the report that odor identification, but not thresholds, was associated with AIDS-related dementia (Hornung et al. 1998). However, similar correlations have been reported between results from odor memory tests and results from tests of odor identification or threshold (Tourbier and Doty 2007), an issue that should be addressed in greater detail in future studies.

In control subjects, the combined testing of all 3 components of olfactory function consistently appeared to be superior for detecting olfactory loss compared with the use of any subtest applied alone. This situation was different in the IPD

group where olfactory function was generally severely compromised. Under this condition, the 3 subtests shared a much greater common variance than in the control group. This was indicated by the tighter clustering and greater test sensitivity and specificity toward olfactory impairment found for single as well as paired subtests. Thus, once IPD has become clinically manifest, olfaction appears to be already severely perturbed, which is in line with the absence of longitudinal changes of olfactory function during progression of IPD (Doty et al. 1988; Herting et al. 2007). It suggests that in later stages of IPD, associated with severe olfactory loss, testing of olfactory function could be reduced to one test. However, this conclusion may be premature, especially when considering the difficulty of IPD patients to identify or recognize odors while being able to detect odors (Masaoka et al. 2007).

We conclude that a composite analysis of several components of olfaction, especially including assessment of odor thresholds, provides the most meaningful approach to human olfactory function. Assessment of a single olfactory function cannot replace a more complex approach for diagnosis of early stages of olfactory dysfunction.

References

Altman DG, Bland JM. 1994. Diagnostic tests. 1: sensitivity and specificity. Br Med J. 308:1552.

Berendse HW, Ponsen MM. 2006. Detection of preclinical Parkinson's disease along the olfactory trac(t). J Neural Transm Suppl. 321–325.

Cain WS, Rabin MD. 1989. Comparability of two tests of olfactory functioning. Chem Senses. 14:479–485.

Doty RL, Deems D, Steller S. 1988. Olfactory dysfunction in Parkinson's disease: a general deficit unrelated to neurologic signs, disease stage, or disease duration. Neurology. 38:1237–1244.

Doty RL, Shaman P, Kimmelman CP, Dann MS. 1984. University of Pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. Laryngoscope. 94:176–178.

Doty RL, Smith R, McKeown DA, Raj J. 1994. Tests of human olfactory function: principle component analysis suggests that most measure a common source of variance. Percept Psychophys. 56:701–707.

Haehner A, Hummel T, Hummel C, Sommer U, Junghanns S, Reichmann H. 2007. Olfactory loss may be a first sign of idiopathic Parkinson's disease. Mov Disord. 22:839–842.

Hawkes C. 2006. Olfaction in neurodegenerative disorder. Adv Otorhinolaryngol. 63:133–151.

Herting B, Schulze S, Reichmann H, Haehner A, Hummel T. Forthcoming 2007. A longitudinal study of olfactory function in patients with idiopathic Parkinson's disease. Mov Disord.

Hornung DE, Kurtz DB, Bradshaw CB, Seipel DM, Kent PF, Blair DC, Emko P. 1998. The olfactory loss that accompanies an HIV infection. Physiol Behav. 15:549–556.

Hummel T, Kobal G, Gudziol H, Mackay-Sim A. 2007. Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. Eur Arch Otorhinolaryngol. 264:237–243.

Hummel T, Sekinger B, Wolf S, Pauli E, 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.

Ivanenko YP, Poppele RE, Lacquaniti F. 2004. Five basic muscle activation patterns account for muscle activity during human locomotion. J Physiol. 556:267–282.

Jones-Gotman M, Zatorre RJ. 1988. Olfactory identification deficits in patients with focal cerebral excision. Neuropsychologia. 26:387–400.

Kaiser HF. 1974. Analysis of factorial simplicity. Psychometrika. 39:31–36.

Kobal G, Klimek L, Wolfensberger M, Gudziol H, Temmel A, Owen CM, Seeber H, Pauli E, Hummel T. 2000. Multicenter investigation of 1,036 subjects using a standardized method for the assessment of olfactory function combining tests of odor identification, odor discrimination, and olfactory thresholds. Eur Arch Otorhinolaryngol. 257:205–211.

Lam HC, Sung JK, Abdullah VJ, van Hasselt CA. 2006. The combined olfactory test in a Chinese population. J Laryngol Otol. 120:113–116.

Masaoka Y, Yoshimura N, Inoue M, Kawamura M, Homma I. 2007. Impairment of odor recognition in Parkinson's disease caused by weak activations of the orbitofrontal cortex. Neurosci Lett. 412: 45–50.

Mueller A, Abolmaali ND, Hakimi AR, Gloeckler T, Herting B, Reichmann H, Hummel T. 2005. Olfactory bulb volumes in patients with idiopathic Parkinson's disease—a pilot study. J Neural Transm. 112:1363–1370.

Muller A, Mungersdorf M, Reichmann H, Strehle G, Hummel T. 2002. Olfactory function in Parkinsonian syndromes. J Clin Neurosci. 9: 521–524.

Muller A, Reichmann H, Livermore A, Hummel T. 2002. Olfactory function in idiopathic Parkinson's disease (IPD): results from cross-sectional studies in IPD patients and long-term follow-up of de-novo IPD patients. J Neural Transm. 109:805–811.

Pearson K. 1901. On lines and planes of closest fit to a system of points in space. Lond Edinburgh Dublin Philos Mag J Sci. 6:559–772.

Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. 2002. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord. 17:867–876.

Thomas-Danguin T, Rouby C, Sicard G, Vigouroux M, Farget V, Johanson A, Bengtzon A, Hall G, Ormel W, De Graaf C, et al. 2003. Development of the ETOC: a European test of olfactory capabilities. Rhinology. 41: 142–151.

Tourbier IA, Doty RL. 2007. Sniff magnitude test: relationship to odor identification, detection, and memory tests in a clinic population. Chem Senses. 32:515–523.

Wolfensberger M, Schnieper I, Welge-Lussen A. 2000. "Sniffin' Sticks": a new olfactory test battery. Acta Otolaryngol. 120:303–306.

Accepted August 3, 2007