The performance of the human nose in odour measurement

J.C. Walker
Sensory Research Institute, Florida State University, 1800 E. Paul Dirac Drive, Tallahassee, FL 32306-2741, USA (E-mail: jwalker@psy.fsu.edu)

Abstract Over the last 20 years or so, there has been steadily increasing activity in the area of applied human odour measurement. This has been especially true outside of the United States. Yet, for about 40 years, there has also been decreasing interest and activity, on the part of academic smell researchers, in rigorous quantitative measurement of the functional properties of the human olfactory system. There are some optimistic signs, however, that this situation may be improving. Applied meetings such as this one are reaching out to learn more about basic research in human olfaction and some research groups are venturing out to indoor air quality, environmental health, water quality and other applied areas. In this paper I hope to support and accelerate the increasingly fruitful interactions that are beginning. The paper aims to make four main points. First, some of the most important ways in which the laboratory differs from everyday life will be noted. Keeping these differences in mind lessens the risk that laboratory data will be used uncritically to make predictions of real-world responses to chemical stimuli. Next, the specific benefits that would accrue from more fruitful interactions between basic and applied researchers will be highlighted; this is perhaps best seen by noting problem areas resulting from too little cross-fertilisation. Third, the CEN standard for the measurement of odour thresholds will be discussed in light of what is known concerning both the functional aspects of the human olfactory system and the current state of knowledge concerning best methods for investigating this system. Finally, some recent work we have done that was designed to help characterise human odour responses and demonstrate improved methodology, will be briefly mentioned. The paper concludes with suggestions as to how the scientific basis of applied odour measurement may best be enhanced.

Keywords CEN standard; measurement; odour; olfactometry; threshold

Comparing laboratory and everyday life situations
The two pictures in Figure 1 illustrate some important contrasts. As depicted in the left panel, we most often experience whole-body exposures, for periods of minutes or hours, to temporally unstable chemical mixtures. These everyday exposures normally take place not when our main focus is on an evaluation of the air around us but when we are engaged in work, social interactions or other distractions. To obtain data that are simple functions of stimulus parameters, we collect data in the tightly controlled environment depicted on the right. Multiple, very brief (a few breaths or seconds) presentations of a single compound to only the nose are made and the participant normally renders only a verbal response after each stimulus trial. We must remember, however, that we are dealing with a whole organism and great care must be taken in understanding the events that take place between our presenting a stimulus and the experimental participant producing a (usually “digital” Yes or No) response. All of these differences illustrate that, in spite of the success we may speak of in terms of laboratory methods and the findings obtained using those methods, those of us in the research effort have much more work to do before we fully understand the impact of odour stimuli encountered in actual environments. This will initially involve elucidation of the odour intensity dimension, primarily using one chemical at a time presented to just the olfactory system. Then we may use this information to help develop an understanding
of the roles of mixture interactions, duration and the integration of multiple inputs in the brain as determinants of various responses.

Central integration of all of the chemosensory inputs stimulated by a given environmental exposure means that, despite the title of this paper, we cannot assume that responses we record in a laboratory or real-world setting are outcomes of processes taking place just in the nose or even just in the olfactory part of the brain. This point, illustrated in Figure 2, is essential in this group’s understanding of “odour impact”. There are many perceptual and physiological effects of short-term chemical stimulation of various chemosensory inputs. In addition to odour, these may include headache, irritation of nose, eyes or throat, exacerbation of asthma, slightly diminished cognitive function or mood alterations. In some cases the connections between the inputs stimulated and the effects are clear. For example, we know that irritating sensations and breathing changes may result from stimulation of parts of the trigeminal (5th cranial) nerve that carry responses to chemical stimuli from the eye to the brain (Walker et al., 1990). Even if we set aside nasal irritation, we can be sure only that odour perception and breathing (Walker et al., in preparation) changes may result from stimuli of the olfactory nerve. The biological basis of such effects as odour impact, nausea or headache is much less clear. It was once thought, and is still widely believed, that asthmatic episodes may be triggered by olfactory stimulation but the few careful attempts to demonstrate this have failed (e.g., Shim and Williams, 1986; Millqvist and Löwhagen, 1996). Thus we need to be extremely careful to distinguish between the impact of an odour per se and an impact that merely accompanies an odour perception, but might well be independent of olfactory activation.

Need for greater basic-applied collaborations

The considerable knowledge gaps noted above are consequential; regulations and remedial efforts do not benefit properly from the best science available nor is there an adequate pace of research to improve this situation. This situation is not unique to the efforts of those in IWA but can be seen also in the indoor air quality, environmental health and ventilation standard-setting areas as well. In all of these spheres, little use is made of the limited knowledge of the mechanisms that underlie various perceptual and physiological responses to olfactory and non-olfactory activation by airborne chemicals. Responsibility for more effective interaction rests, of course, not only with those developing standards or trying to solve real-world odour problems. Researchers who are one step removed from the “front lines” of applied odour measurement work will, ideally, become much more proactive in learning what the most pressing research needs are in various applied areas, working with
colleagues to design and conduct the most relevant studies and then actively disseminating these findings in ways most likely to be “digested” by applied research colleagues.

Comments on the CEN standard
From the perspective of basic odour research, the current CEN standard, titled “Air quality-determination of odour concentration by dynamic olfactometry”, provides good evidence of the need for the greater collaboration called for above. Many of the core elements of this standard do not rest on a base of scientific findings or methodology, the key concept of the odour unit (OU) is not useful or meaningful as it is defined in the standard and a number of approaches to odour measurement are sanctioned without the benefit of prior testing. For example, this standard prescribes how many rounds of testing are sufficient to obtain valid threshold measurements but this appears not to have a basis in empirical data. The concept of the odour unit (OU) is stretched beyond its legitimate value as a multiple by which an air sample has to be diluted to reach odour threshold. It is not meaningful to express this simple ratio in terms such as OU/m³ since it is not a unit of mass and is not properly considered in terms of any other physical or chemical scale. The concept of taking samples from a site and then testing at a remote location has understandable appeal because of logistical advantages. However, this approach is endorsed without testing to show that the results are equivalent to what would be obtained by comparing, with on-site tests, clean air to the odour source of interest. Numerous other difficulties may be cited.

Throughout the document, there is the appearance of rigour and this is particularly true in the use of many complex statistical formulae. Yet a closer reading illustrates that the guidelines are sufficiently ambiguous and open-ended that they accomplish little in terms of quality control. While it is certainly true that it is important to promulgate a “rules of the game” set of procedures, the maximum possible weight should be given to the admittedly limited collective wisdom of scientific research on the subject, with minimum weight given to the understood need for compromise. Such an approach would have made it clear where the research support for various aspects of odour measurement was robust and would also have exposed those areas most in need of additional careful laboratory study.
Some recent and useful methodology and findings

Laboratory work of concrete value in addressing some points of weakness in the CEN standard has been scarce. In fairness to the authors of this standard, it must be acknowledged that, although those from the more academic area of smell research may well have made better use of available literature on olfactory research, they would still have had to contend with quite inadequate knowledge in two key areas. First, until quite recently there was very little quantification of within- and between-individual variation in odour perception. Ignorance in this area was only one of the reasons for the second area of need; namely, the lack of agreement on what constitutes a valid measurement of the detection threshold for odour.

Two comprehensive studies were designed and conducted to address these deficiencies. The net effect of these projects is to provide partial support for some aspects of the CEN methodology. In addition, this recent work provides guidance as to how the CEN standard might be edited to make it vastly more simple to understand and follow. The first study (Kendal-Reed et al., 1998) was designed to explore responding throughout a range of concentrations of propionic acid that ranged from peri-threshold to intense in terms of odour magnitude. An unusual feature of this study, in which 31 individuals were tested with propionic acid, was that each person was tested for a total of ~10 hours (four 50-trial sessions). This was done to explore intra-individual stability. An earlier report by Stevens et al. (1988) indicated that thresholds vary by a factor of 10,000 over time for a given combination of individual and chemical. If this finding were supported by replication attempts, such massive within-individual variation would make it extremely problematic to make valid comparisons across chemicals or individuals.

We developed an iso-response method for determining the concentrations, for each combination of participant and session, corresponding to various percentages (10, 25, 50 and 75) of the way from the average odour strength rating on clean air trials up to the highest rating recorded for that session. Once odour responding was specified this way in units of chemical concentration, we could use various statistical techniques (Walker et al., 1999) to estimate the uncertainty surrounding each iso-response concentration (for a given session), the true variance due to changes over time for a given individual and inter-individual variation. Even with the lower two perceptual intensities, iso-response levels (10%, 25%), 19 of the participants showed only negligible variation from session to session. Of the remaining individuals, the standard deviations describing variation over time ranged from ~0.4 to 0.9 log unit. In sum, with sufficient sampling of the odour responses from each individual to carefully controlled chemical concentrations, stability over time for a given individual was at least 1000-fold better than had been reported earlier. Inter-individual variation was modest as well; over 90% of the individuals were contained within a concentration span of one log unit for the two lowest perceptual intensities.

These findings with propionic acid provided reasonably strong, though somewhat indirect, evidence that variation in threshold was much smaller than had been previously believed. Since this observation, if it could be substantiated by further work, would simplify and clarify the task of the applied odour measurement professional, we pursued the issue of odour threshold much more intensively in a study designed to address some lingering questions.

1. Propionic acid is a moderately effective trigeminal stimulus and therefore it is conceivable that the work with this stimulus might not have provided a “clean” measure of perception due only to olfactory nerve activation. To remove this possible concern, a compound with minimal nasal trigeminal potency (n-amyl acetate) was selected for a follow-up study.

2. The propionic acid study had covered the full dynamic range, and we approximated threshold as being the 10% iso-response concentration. Therefore, the plan was made to
focus only on the peri-threshold region in a second study, so that a great deal of data was collected each session in response to undetected, barely detected and easily detected stimuli.

3. One could have suggested that the variation over time for a given individual was not adequately estimated by our prior procedure, which included only four sessions. Therefore the decision was made to test for 12 or more sessions into a second study.

4. The need for a probability-based definition of threshold concentration was recognised and the study was therefore designed in such a way that simple binomial statistics might be used to explicitly define an exact threshold concentration.

The study that resulted from these considerations is summarised in a presentation at this meeting. Briefly seven individuals were tested, for at least 12 sessions, with ranges of concentrations of n-amyl acetate. Air-dilution offactometry (Prah et al., 1995) was used to generate different concentrations. Delivery of stimuli to a snug-fitting face mask was synchronised to the onset of exhalations. At the end of each presentation, participants responded “Yes” or “No” as to whether odour perception occurred, using a computer and mouse. Binomial statistics were used to evaluate responding on each concentration. The likelihood was calculated that the level of responding (or a higher rate of response Yes) observed with each odour stimulus would be seen on clean air control trials. Adjustments were made over the course of sessions so that, for virtually every session, the lowest concentration was not detected and the highest was easily detected. Binomial probability was plotted against concentration and a logistic regression equation was developed for use in determining the exact concentration associated, with a binomial probability of 0.05 being used as the criterion value for odour threshold. A given participant varied little in odour sensitivity over the course of 12 or more sessions. Threshold concentrations, calculated on the basis of all sessions where the participant was having some difficulty in detecting the stimulus, ranged from 0.0085 to 0.12 ppm (v/v) for the 7 participants.

This recent work can be quite useful in strengthening our understanding of the processes and parameters to be considered in any applied odour measurement effort. In addition, this work may allow us to devise guidelines that are much better supported by careful laboratory work and far easier to understand and execute than is the case with the present guidelines in the CEN standard. Thus some detailed consideration of the n-amyl acetate threshold results, in light of the CEN standard, is of value and raises several important issues.

Fortunately, and in marked contrast to the literature available at the time the CEN standard was written, people are surprisingly stable over time in terms of odour potency in response to precisely controlled chemical concentrations. In a sense, this very recent work provides a reasonably clear retroactive validation for one assumption underlying the CEN standard approach to threshold measurement; namely, the reliance on a single session to measure threshold. It is important to note, however, that the CEN approach likely overestimates the threshold concentration. This conclusion is based on sufficient random sampling (excluding the first session one in the n-amyl acetate study) of data from the n-amyl acetate study to yield the equivalent of, in CEN terms, three rounds. Then the rules stated in the standard were used to calculate a threshold for each individual. These are compared, in Figure 3, to the thresholds we calculated using the method described above based on all of the results for each participant. If one assumes that the data from the above-described n-amyl acetate study are more accurate, the comparisons in Figure 3 suggest an error, in the direction of under-estimation of sensitivity, of up to a factor of two due to some combination of the smaller sample size and calculation method proposed in the CEN standard. If the error were always, as seems reasonable to suppose, in the direction of underestimating the effect of the exposure, this would mean, of course, that the OU estimate would be in error as well. For example, if the threshold is calculated as twice as high as it actually is, a presumed
OU of 25 should actually be reported as one of 50. It should be noted that this comparison is perhaps an especially "gentle" one for the CEN standard. Since no data from the first session were allowed, the comparison minimised any adverse effects of a potentially inadequate time for the participant to acclimatise to the study. In addition, stimulus control and delivery were, of course, not allowed to factor into the comparison either, since sample response data were drawn from the above-described n-amyl acetate study. With relaxation of these aspects of the experimental protocol, while still staying within the guidelines in the standard, much greater error would be expected.

Conclusions
While the comparison above is only a cursory one, it does suggest that the laboratory method we have recently developed and validated could serve as a useful benchmark for quantitative evaluation of the CEN standard. It would be extremely valuable to learn how different are the outcomes of the two approaches when they are applied to yield thresholds to a given single chemical from the same set of individuals. Any differences observed could then be attributed entirely to differences in methodology and would be useful in identifying areas for improvement.

An ideal outcome of this and future examinations of the CEN standard, in light of recent laboratory work, is that the former is sharpened and continually improved so that it can be confidently used as a tool to generate data that withstand scientific and various other types of challenges. Apart from the enhanced measurement of odour thresholds, however, the enhanced standard could be used as a valuable research tool to tackle such critically important issues as the rules that govern odour responses to complex mixtures and those that might be discovered to translate odour intensity into odour impact.

Acknowledgements
Many of the ideas articulated in this keynote address have been developed in conversations with colleagues such as Martin Kendal-Reed, Bill Cain, Pawel Wargocki, Brigitta Danuser, Bob Bottcher, Xu-Feng Niu, Sandy Hall and Lloyd Beidler.
References


