Available online at www.sciencedirect.com
SCIENCE DIRECT®

NEUROCOMPUTING

ELSEVIER Neurocomputing 58–60 (2004) 1019–1025
www.elsevier.com/locate/neucom

On the dimensions of the olfactory perception space

Amir Madany Mamlouk*, Thomas Martinetz

Institute for Neuro- and Bioinformatics, University of Lübeck, Ratzeburger Allee 160, Lübeck 23538, Germany

Abstract

In recent works, large databases of stimuli and their corresponding olfactory perceptions have been analyzed to gain an insight into the organization of olfactory perception. Maps of these perceptions have provided evidence that the olfactory perception space is high dimensional. Based on these results, the question of the dimensionality of olfactory perception space can be asked using a new perspective. In this paper the problem of dimensionality is approached more rigorously and upper bounds on the dimensionality of the olfactory perception space are estimated.

Keywords: Olfactory perception; Multidimensional scaling; Principal component analysis; Dimension reduction

1. Introduction

In color vision, it has been found that three dimensions are sufficient to span the color perception space [6]. We are able to compose all colors that humans can perceive using a linear combination of three primary colors, giving every color a three dimensional vector representation in a Euclidean space. In this space, distances between the perception quality of colors can be measured easily. For olfaction, knowledge about the structure of such a perception space would reveal essentially new and much more rigorous ways to quantify relationships between individual odor sensations (like, e.g. between *cherry* and *apple*).

* Corresponding author. Tel.: +49-451-3909-508; fax: +49-451-3909-545.
E-mail address: madany@inb.uni-luebeck.de (A. Madany Mamlouk).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.01.161

Recently, Chee-Ruiter [2] proposed a map for olfactory perception based on a set of odor quality descriptions. She intended to reveal structural information of the olfactory perception space (also called odor space). This map was formulated as a directed graph without a particular dimensional embedding. In an extension of this framework [7] the proposed observations were projected into the nearest Euclidean space by using multidimensional scaling (MDS) [5]. Unfortunately, this nearest space was still quite high-dimensional. Therefore, self-organizing maps [9] were then applied to derive a two-dimensional map which preserves the high-dimensional topology as well as possible.

The Euclidean embedding via MDS provided strong quantitative support for the long-held belief that olfactory perception space is high dimensional [8]. However, we do not know the features that are characteristic for odor space. We do not even know the number of features we are looking for. Therefore, it seems to be essential for the further understanding of olfactory perception to find at least constraints for the minimal number of features that can explain the variety of sensations of smell.

Historically, there have been only a few approaches that explicitly try to estimate the dimensions of olfactory perception space. In 1915 Henning [3] proposed his prism of odors. He used six primary odors to span a three-dimensional prism. In 1974, Schiffman [10] argued that the odor space can be embedded in only two dimensions. These results were based on two different sets of experimental data. One consisted of 50 odorants compared to nine standard odorants, which were chosen to cover a wide range of olfactory quality. Consequently, a successful embedding into at most nine dimensions has to be expected. The second set of data was acquired by Woskow [11] consisting of 25 arbitrarily chosen stimuli that were compared qualitatively by subjects. Some of these stimuli were mentioned to be so similar that the boiling points had to be checked to ensure that the bottles were labeled correctly. In this case, it is reasonable to expect the according dimensionality to be much smaller than 24. For both setups, an initial factor analysis revealed at least eight dimensions.

Assuming that the olfactory perception space might be more complex than these initially low-dimensional data sets, we will present a more realistic embedding by using a significantly larger source of information. In the following, we will again focus on an estimation of the dimensionality of the odor space. We will explain the techniques which were employed and present our results on data that has been used so far to produce maps of olfactory perception [8].

2. Data

Aldrich Chemical Company [1] has published a catalog in which several chemicals are profiled by a set of several hundred descriptors. These descriptors (like *fruity* or *cherry*) are the terms we commonly use to describe smell sensations. Hence, each descriptor can be regarded as a point in the olfactory perception space.

Chee-Ruiter analyzed these data and obtained a collection of 851 stimuli, each described by 278 descriptors. This set was then reduced onto 171 descriptors by removing descriptors that are evoked by only a single chemical. The dissimilarities of these 171 descriptors were estimated using a symmetric distance measure which can be interpreted as a weighted version of a cross-entropy measure [2,7].

This dissimilarity matrix is used to analyze the dimensionality of the odor space. It should be mentioned again that these values are not directly evaluated by test subjects. Instead, we took verbal odor descriptions and measured the similarities between them using a mathematical measure.

3. Methods

We are looking for a metric representation of the data that is as low dimensional as possible and still conserves all neighborhood relations. Similar odors should become close neighbors in this representation, whereas very dissimilar odors should be clearly seperated.

Generally, given a set of points in a high-dimensional metric space, the first step to evaluate the intrinsic dimensionality of this set would be to calculate its principal components and the corresponding eigenvalues. The eigenvalues give the variance of the distribution of these points. For unused dimensions this variance is zero. In other words, the number of eigenvalues greater than zero equals the smallest number of dimensions needed [4].

However, the odor space data are not given by a set of points but by a dissimilarity matrix. If we interpret this matrix as a distance matrix of points in a certain Euclidean space, we can employ MDS [5], which finds the set of points that corresponds to the dissimilarity matrix as much as possible. If the data are metric, i.e. if the dissimilarity matrix is a distance matrix, for a $(n \times n)$ matrix, $(n-1)$ dimensions are always sufficient to obtain an exact mapping. However, since there is no known metric measure that estimates the dissimilarities between different odors accurately, we have to expect the data to be non-metric to a certain degree. In this case, by definition an error-free representation of the data would not be possible for any metric embedding. MDS tries to minimize the error between the distances of the embedded points and the dissimilarities of the corresponding odors. Therefore, if $d_{\mathrm{sim}}(i,j)$ is the dissimilarity between two odor descriptors O_i and O_j, and if $d_{\mathrm{eucl}}^{(n-1)}(i,j)$ denotes the Euclidean distance between the two descriptors after embedding n points into an $(n-1)$ dimensional Euclidean space, the residual error

$$\Delta(i,j) = (d_{\mathrm{sim}}(i,j) - d_{\mathrm{eucl}}^{(n-1)}(i,j))$$

is the non-metric defect of the single dissimilarity $d_{\mathrm{sim}}(i,j)$. This non-metric defect remains constant for embedding dimensions $D > (n-1)$. However, it might be that the individual embedding error $d_{\mathrm{sim}}(i,j) - d_{\mathrm{eucl}}^{D}(i,j)$ remains on this minimal level even for embedding dimensions $D < (n-1)$. Then we assume that the data can be embedded into an intrinsic subspace. This intrinsic dimension is given by the smallest embedding dimension for which the embedding error does not increase significantly. This allows us to define a metric embedding error

$$d_{\mathrm{metric}}^{D}(i,j) = d_{\mathrm{eucl}}^{D}(i,j) - \Delta(i,j).$$

Thus, the overall metric embedding error can then be expressed as

$$\varepsilon^D_{\text{metric}} = \sum_{i,j} |d_{\text{sim}}(i,j) - d^D_{\text{metric}}(i,j)|.$$

Interestingly, the quality of the embedding into a metric space directly depends on the *metric* embedding quality while the non-metric defects can be expected to be constant.

4. Results

Initially, we applied MDS on the (171×171) dissimilarity matrix to get 171 points in $\mathbb{R}^{170}$ with a minimal metric embedding error. On these points we performed principal component analysis to analyze the eigenvalues of this embedding. In Fig. 1a the sorted eigenvalues are plotted. For dimensional reduction tasks, usually the largest eigenvalues that form more than 90% of the energy are discarded. In this case, 68 dimensions are a reasonable size for an upper dimensional bound. This bound is marked in Fig. 1a and b. Incidentally, there are zero eigenvalues. This confirms our assumption of having an intrinsic dimensionality lower than $(n-1)$.

Fig. 2 shows scatter plots for different dimensions. In such a plot the dissimilarities are plotted against their corresponding distances. As expected, in Fig. 2a it can be seen that the embedding even in $\mathbb{R}^{170}$ is far from perfect, due to the non-metricity of the data. As described, we used the difference to a perfect metric embedding as an estimate of the constant non-metric defect. The metric embedding quality for 32 dimensions as well as for 16 and for 64 dimensions—after removing the non-metric defect—can be

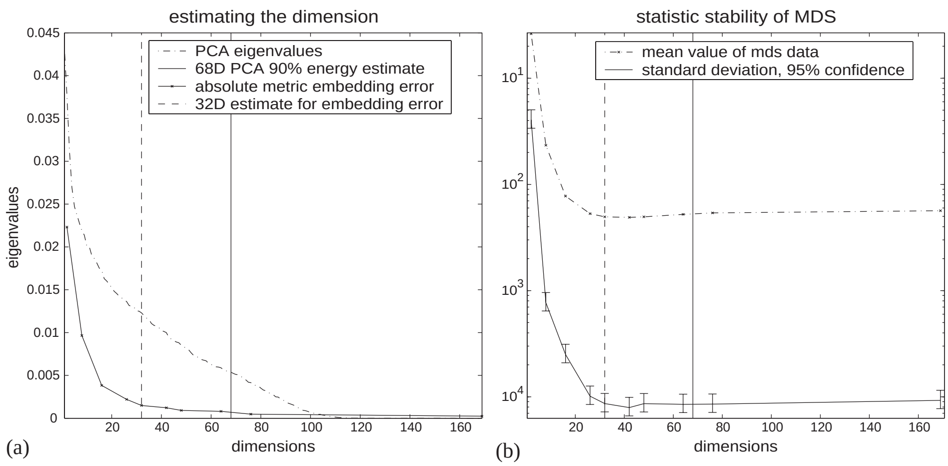

Fig. 1. (a) Dashdotted line shows the sorted eigenvalues for 170D, vertical line marks 90% energy cut-off (68D). The metric embedding error (solid curve) increases significantly for dimensions smaller than 32 (see dashed vertical line). The error was scaled to fit in the energy graph. (b) Standard deviation and mean values for simulation runs of MDS. Values become smaller the more the scaled distances correspond to the given dissimilarities.

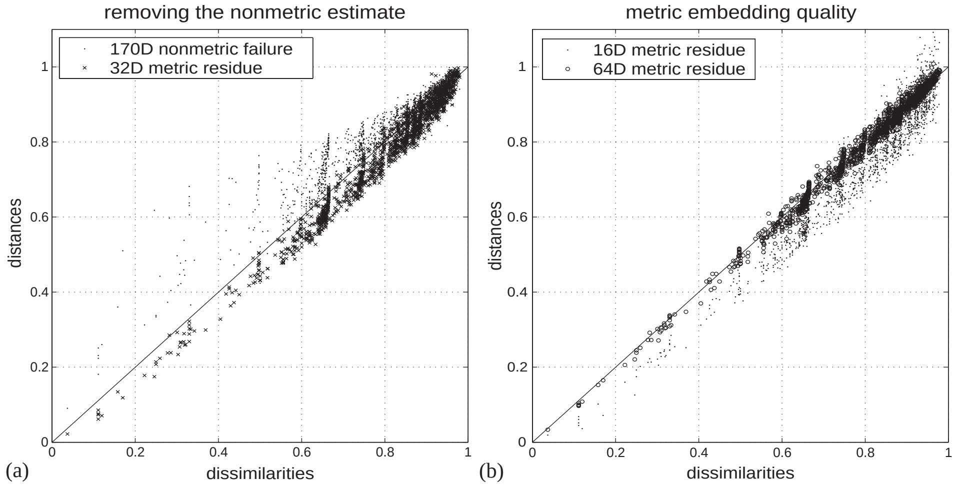

Fig. 2. Scatter plots of odor data after MDS. (a) Small dots illustrate the non-metric defect, i.e. embedding of 171 points into 170D. Crosses show the metric embedding quality for 32D. (b) For 16D (small dots) the overall metric embedding quality is not as good as for 32D, for 64D (circles) it is similar to 32D.

seen in the scatter plots in Fig. 2. Interestingly, while there is a significant increase of embedding quality between 16 and 32 dimensions, the enormous expansion to 64 dimensions does not notably increase the quality.

This becomes even clearer if we take a closer look at the development of the overall metric embedding error $\varepsilon^D_{\mathrm{metric}}$ for different dimensions. In Fig. 1a, $\varepsilon^D_{\mathrm{metric}}$ is shown, for several dimensions between 2 and 170. Please note that from 170 down to 32 dimensions this error increases only marginally. This would be a typical behavior for a 32 dimensional structure, because then and only then the dimensional reduction from 170 to 32 would not affect the embedding quality.

To avoid local minima, typically several runs of MDS with randomized start values are performed. In Fig. 1b, the overall results for different dimensions are shown. For dimensions lower than 32, the standard deviation of the resulting distances $d^D_{\mathrm{eucl}}(i,j)$ as well as the mean of the corresponding stress values [5] increase significantly. This can be taken as further evidence for our bounds. The more the metric embedding forces the distances to differ from the given dissimilarities, the more local minima occur and the more ambiguous are the results. As a consequence, the standard deviation of the $d^D_{\mathrm{eucl}}(i,j)$ as well as the mean of the stress values increase.

5. Discussion

For the olfactory system, it remains an open issue to define quantitatively the complexity of the underlying nervous system. On the other hand, it seems to be intuitively clear that the thousands of olfactory receptors combine to a somewhat more simple representation on the perceptual level, as smelling is a very natural task.

Woskow [11] and Schiffman [10] suspected olfactory perception to be on a complexity level as low as eight or even two dimensions, respectively. Not only have 30 years not been enough to understand these few dimensions of odor space, but both experimental setups have turned out to be inadequate to describe a high-dimensional structure in general.

In this work, results are presented that imply a high-dimensional olfactory perception space. Chee-Ruiter's data [2] combined with our methods reveal an upper bound of 68 dimensions and a lower bound of at least 32 Euclidean dimensions for the olfactory perception space.

With the results presented here, new questions might be posed especially about the neural organization underlying the olfactory perception.

References

[1] Aldrich (Ed.), Flavor and Fragrances Catalog, Sigma Aldrich Chemicals Company, Milwaukee, WI, 1996.

[2] C.W.J. Chee-Ruiter, The biological sense of smell: olfactory search behavior and a metabolic view for olfactory perception, Ph.D. Thesis, California Institute of Technology, Pasadena, CA, 2000.

[3] H. Henning, Der Geruch, Barth, Leipzig, 1916.

[4] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 1986.

[5] J.B. Kruskal, M. Wish, Multidimensional Scaling, Sage Publications, Beverly Hills, CA, 1978.

[6] M.W. Levine, Fundamentals of Sensation and Perception, 3rd Edition, Oxford University Press, New York, 2001.

[7] A. Madany Mamlouk, Quantifying olfactory perception, Diploma Thesis, University of Luebeck, Germany, 2002.

[8] A. Madany Mamlouk, C. Chee-Ruiter, U.G. Hofmann, J.M. Bower, Quantifying olfactory perception: mapping olfactory perception space by using multidimensional scaling and self-organizing maps, Neurocomputing 52–54 (2003) 591–597.

[9] H. Ritter, T. Martinetz, K. Schulten, Neural Computation and Self-Organizing Maps: An Introduction, Addison-Wesley, Reading, MA, 1992.

[10] S.S. Schiffman, Contributions to the physicochemical dimensions of odor: a psychophysical approach, Ann N Y Acad Sci. 237 (1974) 164–183.

[11] M.H. Woskow, Multidimensional scaling of odors, Ph.D. Thesis, University of California Los Angeles, Los Angeles, CA, 1964.

Amir Madany Mamlouk graduated in computer science with a minor in biomathematics and bioinformatics at the University of Lübeck, Germany. He did research for his diploma thesis at the California Institute of Technology. Currently, he is a research associate with the Institute for Neuro- and Bioinformatics at the University of Lübeck. His main research interests are in methods for dimension reduction, artificial neural networks, and self-organizing structures. He focuses on their application on biological systems, especially on the understanding of the sense of smell.

Thomas Martinetz studied Physics and Mathematics in Munich and Cologne. From 1988 to 1991 he did his Ph.D. as a member of the Theoretical Biophysics Group at the Beckman Institute of the University of Illinois at Urbana-Champaign, focussing on research on self-organising neural networks. In 1991 he joined the Neuroinformatics Research Center of Siemens AG in Munich. In 1996 he became head of the Complex Systems Group at the Institute for Neuroinformatics of the University of Bochum, and CEO of the Zentrum für Neuroinformatik GmbH, a company for the application of Neuroinformatics in various commercial areas. Since 1999 he is Director of the Institute for Neuro- and Bioinformatics at the University of Lübeck. The main lines of research at his institute are in biological information processing, in particular in neural vision, pattern recognition, and learning.