On the Dimensions of the Olfactory Perception Space

Amir Madany Mamlouk [1] and Thomas Martinetz

Institute for Neuro- and Bioinformatics
University of Lübeck, Germany

Abstract

In order to generate maps of olfactory perception space a huge database of stimuli and their corresponding perceptions has been compiled. Using this database as a source of information, the question of the dimensionality of olfactory perception space can be asked again. Maps of these data have provided strong evidence that the olfactory perception space is high dimensional. In this paper this problem of dimensionality is approached more rigorously and upper bounds on the dimensionality of olfactory perception space are estimated.

Key words: Olfactory Perception; Multidimensional Scaling; Principal Component Analysis

1 Introduction

In color vision, it has been found that three dimensions are sufficient to span the color space [6]. That is, we are able to compose all perceptible colors using a linear combination of three primary colors, what gives every color a three dimensional vector representation in an Euclidean space. In this space, we can easily measure distances between the perception quality of colors.

Regarding the fact that still very little is understood about the larger organization of the human sense of smell, knowledge about the structure of such a space for olfactory perception could reveal essentially new and much more rigorous ways to quantify relationships between single odor sensations (like *cherry* and *apple*) than this has been possible before.

To obtain a first insight into the structure of olfactory perception space (also called odor space), Chee-Ruiter [2] derived a map for olfactory perception from a set of odor quality descriptions. This map was formulated as a directed

[1] email: madany@inb.uni-luebeck.de

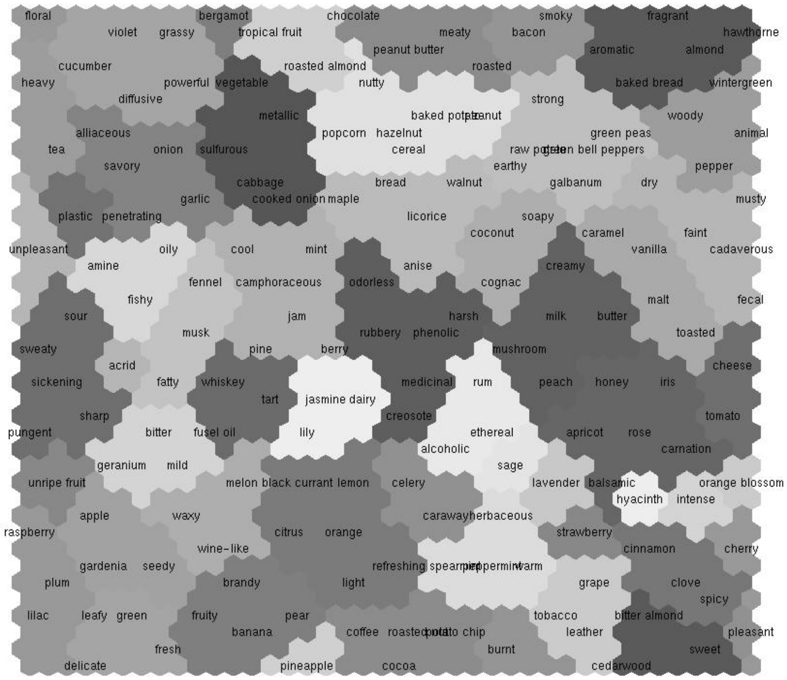


Fig. 1. Olfactory perception map for scaled data of the Aldrich database. The map is labeled with odor descriptors. Each descriptor was plotted onto its nearest neighbor on the SOM. The shaded areas correspond to k-means clusters.

graph. Madany Mamlouk et al. [8] extended this framework and projected the corresponding distance map into the nearest Euclidean space by using multidimensional scaling [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 similarities as much as possible. This map [8] is shown in Figure 1.

The approach via multidimensional scaling (MDS) provided strong quantitative support for the long held belief that olfactory perception space is high dimensional [7,8]. 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 constraints for the number of features that are necessary to explain the variety of sensations of smell. This would greatly simplify the task of understanding olfactory perception.

There have been former approaches that tried to estimate the dimensions of olfactory perception space. In 1915 Henning [3] proposed a 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. She examined two sets of data. One consisted of 50 odorants compared to nine standard odorants, which were chosen to cover a wide range of olfactory quality. The second one was a data set proposed by Woskow [12] consisting of 25 arbitrary chosen stimuli that were compared qualitatively by subjects.

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 the results on the same data which were used to produce the map in Figure 1.

2 Data

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

Chee-Ruiter [2] analyzed this 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 a single chemical only within the whole database. The dissimilarities of these 171 descriptors were estimated using a distance measure which can be interpreted as a weighted version of a cross-entropy measure [2,7].

In this paper we use this dissimilarity matrix to analyze the dimension of the data. It should be mentioned again, that we used no psychophysical measure by asking subjects to judge the similarity of two stimuli, but we used psychophysical profiles of chemicals to derive these similarities objectively.

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 nearby neighbors in this representation, whereas very dissimilar odors should have a large distance.

Given a set of points in a high dimensional metric space, the first step to evaluate the intrinsic dimension of this set would be to calculate its principal components and the corresponding eigenvalues [4]. 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 to the smallest number of dimensions needed.

Our input data is not a set of points but a dissimilarity matrix. If we interpret this matrix as a distance matrix of points in an Euclidean space, we can employ a method called multidimensional scaling [11,5]. Within an Euclidean space of our choice, MDS finds the set points which 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 not more than $(n-1)$ dimensions are needed to obtain an exact map. 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 embedding the data is not possible for any metric dimension.

Multidimensional scaling is a numerical method which tries to minimize the error between the distances of the embedded points and the dissimilarities of the corresponding odors. Therefore, we can embed non-metric data as well, but the embedding will never be perfect. If $d_{\text{sim}}(i,j)$ is the dissimilarity between two odor descriptors O_i and O_j, and if $d_{\text{eucl}}^{n-1}(i,j)$ denotes the Euclidean distance between the two descriptors after embedding into an $(n-1)$ dimensional Euclidean space, the residual error

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

is the non-metric defect of the single dissimilarity $d_{\text{sim}}(i,j)$. This non-metric defect remains constant for embedding dimensions $D > (n-1)$. However, it might be that the individual embedding error $d_{\text{sim}}(i,j) - d_{\text{eucl}}^{D}(i,j)$ remains on this minimal level even for embedding dimensions $D < (n-1)$. Then we assume that the data have a lower-dimensional intrinsic dimension. This intrinsic dimension is given by the embedding dimension for which the embedding error increases. This allows to define a metric embedding error

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

The embedding error is then given by the sum of the metric embedding error and the non-metric defect. The overall metric embedding error is then given by

$$\varepsilon_{\text{metric}}^{D} = \sum_{i,j} |d_{\text{sim}}(i,j) - d_{\text{metric}}^{D}(i,j)|.$$

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

4 Results

In a first step we applied multidimensional scaling on the (171×171) dissimilarity matrix to produce 170 dimensional points. On these points we did principal component analysis (PCA) to analyze the eigenvalues of this embedding. In Figure 3.a the sorted eigenvalues are plotted. For dimensional reduction tasks, usually the smallest eigenvalues that form less than 10% of the energy are cut off the data. Here, 68 dimensions are a reasonable number for an upper dimensional bound. This bound is marked in Figure 3.a and b. Incidently, there are zero eigenvalues, which fulfills our assumption of having a dimensionality lower than $(n-1)$.

Figure 2 shows results we obtained with multidimensional scaling. As expected, in Figure 2.a it can be seen that the embedding even in $\mathbb{R}^{170}$ is by far

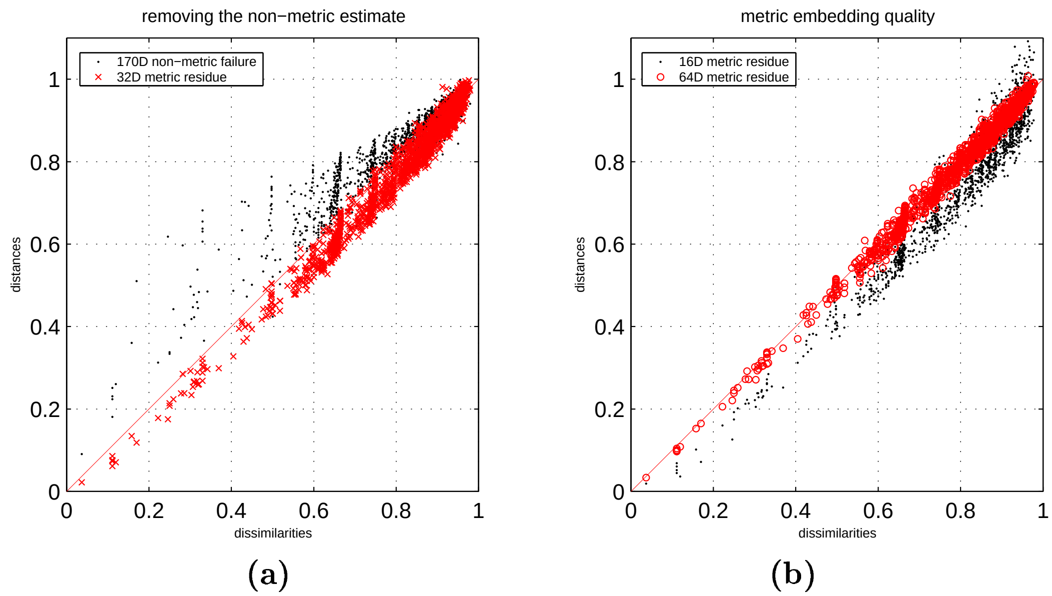
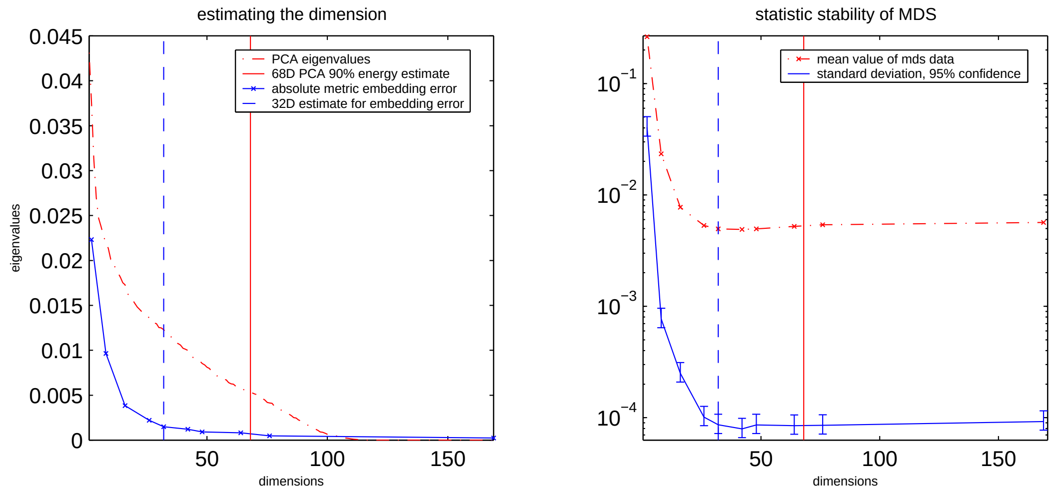


Fig. 2. Metric embedding quality of MDS on odor data. The scatter plot shows the dissimilarities against corresponding distances from MDS. **(a)** The small dots are illustrating the non-metric defect, because for 170D they should be on a straight line. The crosses show the metric embedding quality in 32D. **(b)** For 16D (small dots) the overall metric embedding quality is not as good as for 32D. The result for 64D (marked by crosses) is similar to the one obtained for 32D.

not perfect because of the non-metricity of the data. We can use this difference to the expected perfect embedding for metric data 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 seen in the scatter plots in Figure 2. Interestingly, while there is a significant increase of embedding quality between 16 to 32 dimensions, the enormous expansion to 64 dimensions does not notably increase the quality.

This becomes even clearer if we take a closer look onto the development of the absolute metric embedding error for different dimensions. In Figure 3.a the absolute metric embedding error can be seen for several dimensions between 2 and 170. Please note that from 170 down to 30 dimensions this embedding error increases only marginal. Such a behavior would be typical for a 30 dimensional Euclidean structure, because then and only then the dimensional reduction from 170 to 30 would not effect the quality of embedding.

To avoid local minima, typically several runs with randomized start values are performed with MDS. In Figure 3.b the results of these runs for the different dimensions can be seen. For dimensions lower than 30 suddenly the standard deviation of the resulting distances as well as the mean of the corresponding stress values are increasing. This is another evidence for our bounds, because the more the metric embedding stresses the distances compared to the dissimilarities, the more ambiguous are our results, and the higher is the standard deviation of the results as well as the mean of the stress values.



(a) **(b)**

Fig. 3. **(a)** The eigenvalues from 1 to 68 express 90% of the energy, the metric embedding error increases significant for dimensions smaller than 30. The embedding error has been scaled down. **(b)** Standard deviation and mean values of stress values for simulation runs of MDS. The stress values become smaller the more the scaled distances and the given dissimilarities correspond.

5 Conclusion

In our data we found strong evidence for a high dimensional olfactory perception space. This dissents with Schiffman's results that this space is rather low dimensional [10]. Schiffman's first set described 50 odorants by nine dimensional vectors, which consequently led to an embedding of lower or equal than nine dimensions. The second set by Woskow [12] consisted of only 25 chemicals in all. He mentioned that some of these stimuli were so similar that they had to check the boiling points to make sure that the bottles are labeled correctly [12]. But if some of the stimuli are that similar, it is reasonable to expect the according dimensionality to be much smaller than 24. In both cases, factor analysis revealed at least eight dimensions [12,10].
With our data it is possible now to explore potential dimensionalities of olfactory perception of up to 170. Principal Component Analysis gave a first upper bound of this dimensionality at about 60-70. Further analysis based on Multi Dimensional Scaling revealed an upper bound at about 30 to 40, at least for the data we have so far. Obviously, with the data available today we are able to gain a much better view onto the structure of the olfactory perception space than this was possible almost thirty years ago.

References

[1] Aldrich, editor. *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.* PhD thesis, California Institute of Technology, Pasadena, CA, 2000.

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

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

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

[6] M. W. Levine. *Fundamentals of Sensation and Perception.* Oxford University Press, New York, 3rd edition, 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, and J. M. Bower. Quantifying olfactory perception: Mapping olfactory perception space by using multidimensional scaling and self-organizing maps. *to appear with Neurocomputing (De Schutter E, ed)*, 2003.

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

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

[11] R. N. Shepard, A. K. Romney, and S. B. Nerlove. *Multidimensional Scaling: Theory and Applications in the Behavioural Sciences, Volume I – Theory.* Seminar Press, New York, 1972.

[12] M. H. Woskow. *Multidimensional Scaling of Odors.* PhD thesis, University of California Los Angeles, Los Angeles, CA, 1964.