Castro et al. Flavour 2014, 3(Suppl 1):P3
http://www.flavourjournal.com/content/3/S1/P3

Flavour

POSTER PRESENTATION Open Access

Categorical, low-dimensional decomposition of human odor space with non-negative matrix factorization

JB Castro[1], SP Quinn[2], A Ramanathan[3], CS Chennubhotla[4*]

From 1st International Workshop on Odor Spaces
Hannover, Germany. 4-7 September 2013

Recent studies using Principal Components Analysis (PCA) support low-dimensional models of odor space, in which one or two dimensions - with hedonic valence featuring prominently - explain most odor variability. Here we use non-negative matrix factorization (NMF) - a non-linear optimization method - to discover an alternative, reduced-dimensional representation of the Dravnieks odor database (144 odors x 146 descriptors). NMF is theoretically well-suited for these types of analyses, as odor profiling data is inherently non-negative (e.g. descriptors either apply, or do not). We divided the dataset into training and testing halves, and found that RMSD testing error attained a minimum for subspace choice of 25, motivating this as an upper bound for odor perceptual space dimensionality. More parsimonious representations were found by comparing reconstruction errors (fraction of unexplained variance) of NMF with reconstruction errors of PCA on scrambled data (PCAsd). For subspace sizes > 10,

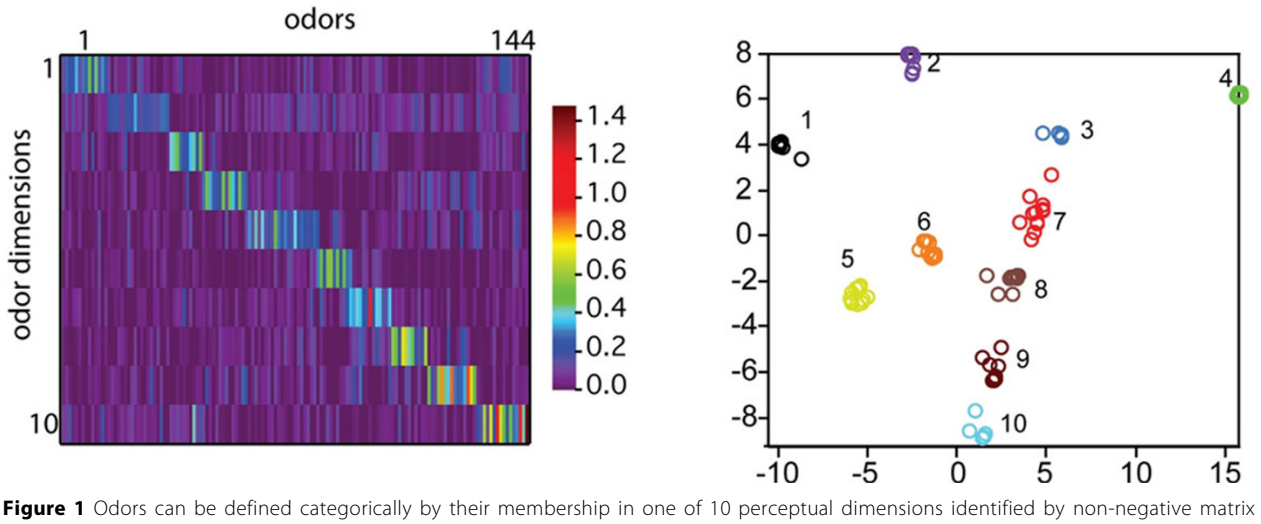

Figure 1 Odors can be defined categorically by their membership in one of 10 perceptual dimensions identified by non-negative matrix factorization (NMF). **Left:** Matrix of odor weights (H). Each column corresponds to a single monomolecular odorant. Rows are the 10 perceptual dimensions discovered by NMF. Prominent block—diagonal structure suggests that a given odorant is best defined by its membership in one perceptual dimension. **Right:** two-dimensional embedding of the matrix H, using stochastic neighbor embedding. Each point corresponds to a column of H. Colors indicate clusters identified by eye.

[4]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA
Full list of author information is available at the end of the article

NMF error was indistinguishable from PCAsd error, indicating no gain in retaining more than 10 perceptual dimensions. As is typical of NMF basis sets, the 10 odor dimensions we obtain are sparse (only a small subset of the 146 descriptors apply), and categorical (represent a positive valued quality). Moreover, these 10 dimensions were near-orthogonal, with a mean angle of 73 degrees between all pairs of basis vectors. Investigating the distribution of odors in this 10-dimensional space, we find marked clustering (Figure 1), with each odor being well-defined by its membership in a single dimension, and to the exclusion of others. In ongoing work, we are using graph-kernel methods to define a rudimentary mapping between physicochemical features of odorants and the 10 descriptor dimensions.

Authors' details
[1]Department of Psychology, Bates College, Lewiston, ME, USA. [2]Joint CMU-Pitt Ph.D. Program in Computational, Biology, University of Pittsburgh, Pittsburgh, PA, USA. [3]Computational Data Analytics Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [4]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA.

Published: 16 April 2014

doi:10.1186/2044-7248-3-S1-P3
Cite this article as: Castro *et al.*: Categorical, low-dimensional decomposition of human odor space with non-negative matrix factorization. *Flavour* 2014 **3**(Suppl 1):P3.