# Preliminaries 2024

## 26 May

Created consensus files for KC and ORN data, with and without scaling (for each), and verified that psg\_align\_knit\_demo and psg\_consensus\_demo give the same results. KC data need not be scaled (z-scored), but ORN data perhaps should be scaled (delta-F/F).

Plotted consensus data at dim 3 with each dataset in a separate color.

To do

document pipeline saved by psg\_align\_knit\_demo

Make a dataset by lumping together all the Kenyon cell data rather than trying to find a consensus.

Look for affine alignments, cutpoint alignments (as function of number of dimensions) with psg\_geomodels\_test and psg\_procrustes\_regr\_demo. With psg\_geomodels\_test and KC as reference, prelim looks good with 5 dims for ORN’s and 3 dims for KC;

Run for 6 dims for ORN’s, 6 dims for KC’s, only one cutpoint.

## After 31 May

Made a pooled dataset from the 7 orn terminal datasets with no missing stimuli – they are very much like the consensus datasets – see orn7sets\_pooledVconsensus\_psg\_consensus\_demo\_02Jun24.fig

Analyze the pooled to see how the PC’s project on individual flies->done, see [kc|orn7sets]\_pooled[-sm]\_pcacontrib.fig

Geomodel analysis with the pooled data and consensus data – affine model is best, projective or affine with a cut is not better than affine as nested model. psg\_geomodels\_run\_orn\*\_kc\*.[mat,txt] and plots from 07Jun24

**The combined datasets:**

**KC-soma**:

hlid\_odor17\_coords\_kc\_soma\_nls\*: pooled, consensus, pooled with mean subtracted (pooled-sm), and consensus without scaling (consensus\_noscale)

**ORN terminals:**

hlid\_odor17\_coords\_orn\_terminals7sets\*: pooled, consensus, pooled-sm, consensus\_noscale (the 7 datasets with no nans)

hlid\_odor17\_coords\_orn\_terminals: consensus, consensus\_noscale (9 datasets, some have nan’s)

## After 09 Aug

Do statistics of nested models of lower dimensions – set to run with psg\_geomodels\_run, esp for affine, for pooled and for consensus data

Look at which odorants are good fits, which are bad in a modeled transformation from ORN to KC (and compare with consistency within ORN and within KC)

Look at which odorant pairs go to smaller distances, which to larger in map from ORN to KC (model-free)

In affine fit: find principal directions and degree of stretching

Model KC data (random connectivity) from Remy: to analyze in same way

Later:

Add less-good KC sets “silver” to consensus, they will have missing stimuli so can’t add to pooled.

Add other odorants

## After 16 Aug

In psg\_majaxes:

Align the eiv’s in adj and ref plots

Allow for an arbitrary ordering; Preferred ordering (this is in hlid\_opts.display\_orders)

kc\_ord = ['2h', 'IaA', 'pa', '2-but', 'eb', 'ep', 'aa', 'va', 'B-cit', 'Lin', '6al', 't2h', '1-8ol', '1-5ol', '1-6ol', 'benz', 'ms']

Optionally also plot the PC’s from MDS analysis for adj and ref

Allow for vertical size of heatmaps to be identical

Note: coordinates for consensus datasets need not be in descending order of rms, as this may only hold for MDS data. For the single datasets, each dimension is cumulative, but for consensus, this need not be the case. In fact, for kc data, the first coordinate of the consensus does not account for as much of the variance as the first two coords.

> clear

>> load('...\data\kc\_soma\_nls\hlid\_odor17\_coords\_**kc\_soma\_nls\_consensus**\_scale.mat')

>> sum(dim4.^2)

ans =

1.0e+03 \*

1.3949 5.5530 4.7714 4.4327

>> sum(dim8.^2)

ans =

1.0e+03 \*

1.2316 5.5720 4.7769 4.4101 1.9527 2.2851 1.8897 1.5157

>> load('...\data\kc\_soma\_nls\hlid\_odor17\_coords\_**2022-11-10-1-megamat0-sm**.mat')

>> sum(dim4.^2)

ans =

1.0e+03 \*

7.8803 6.3985 4.3655 3.3402

>> sum(dim8.^2)

ans =

1.0e+03 \*

7.8803 6.3985 4.3655 3.3402 2.2785 2.1822 1.9912 1.4315

>> load('...\data\kc\_soma\_nls\hlid\_odor17\_coords**\_2022-11-10-1-megamat0**.mat')

>> sum(dim4.^2)

ans =

1.0e+04 \*

1.4145 0.6692 0.6393 0.4329

>> sum(dim8.^2)

ans =

1.0e+04 \*

1.4145 0.6692 0.6393 0.4329 0.2760 0.2261 0.2147 0.1758

>>

>> load('...\data\orn\_terminals\hlid\_odor17\_coords\_**orn\_terminals7sets\_consensus**\_scale.mat')

>> sum(dim4.^2)

ans =

5.0720 2.6578 3.0839 2.4933

>> sum(dim8.^2)

ans =

5.8641 2.9537 3.1949 2.6913 1.2219 1.3141 0.4937 0.7318

>> load('...\data\orn\_terminals\**hlid\_odor17\_coords\_2023-06-22-01**.mat')

>> sum(dim4.^2)

ans =

39.5501 2.7433 2.6053 1.6668

>> sum(dim8.^2)

ans =

39.5501 2.7433 2.6053 1.6668 1.0099 0.7548 0.5345 0.4364

>>

## 27 Aug

Model data files have stimuli in a different order than data. hlid\_csv2coords\_demo revised to reflect this.

See psg\_geomodels\_run\_orn7setsConsensusScale\_kcHemibrainModel\_6d.[txt,mat] for mapping of orn’s onto model.

See hlid\_majaxes\_27aug24.txt,pdf for comparison of major axes – hemibrain model keeps the same major axes in model KC’s as in orn’s, and does not have the same degree of stretching.

## 30 Aug

Interesting to analyze “uniform” model, which lacks the connectivity data. If connectivity matters, it should be an even poorer fit.

See hlid\_majaxes\_30aug24-unifor.txt,pdf: similar to hemibrain model.

Also files in

/plots/[uniform-001|uniform-100|hemibrain|kc4recs…]\_orn\_consensus\_procrustes\_vs\_affine.fig, etc, to show that the models have similar behavior to the data in that affine model is best, and, in nested-by-dimension analysis, 3d in orn space and 4,5,6 d in kc/model space is best

Also interesting to analyze a model in which the hemibrain connectivity is used, but not at the level of correlations – this is “hemidraw” model, see files of 04Sep24 and 05Sep24

Eventually see if the principal directions identified in psg\_majaxes, when mapped back to ORN (glomerular) space, has any relationship to the correlation patterns in the connectivity data.

## 06 Sep

Agreed that all models are quite different from real KC data.

Could look at “directions” in the correlation matrix of ORN projections, or, in the (either 2 or 3) 50 x 2000 matrices of ORN to PN/KC projections – to see what is consistent.

Could also reframe analysis in ORN space instead of odor space, going back to the original data files – and forming a consensus across recordings. But note that not all ORNs are recorded in each prep,and, in some cases, not all 17 odors are present. So we coujld only use the flies with all odors present. But there are still some options to how to combine across flies]. We could do a missing-data consensus alignment (procrustes\_consensus) as a first step. Or a missing-data PCA, e.g., we are missing data for every odorant in one ORN x prep “line” in [ORN, odorant, prep], so, for each odorant, there is an [ORN x prep] sheet with isolated missing data, and we fill it in by afalwt.m or similar – and then do a weighted average across preps? [Also, since some orn data are missing – perhaps one should redo the orn coordinates used above as a missing-data consensus]?

Could we make use of the 2-odor mixture data on a 4x4 grid?

## 13 Sep

Grant ideas:

### Prelim figures:

representation in ORN space, representation in KC space, transformation between them

### Things to propose:

Choice of stimulus sets to distinguish chemical similarity from chemical pathway similarity; overlap of stimuli between sets to “knit together” representations

Hyperbolic models

Analysis of connectivity matrix and its principal directions ->see hlid\_majaxes\_compare. Suggestive of correspondence of some principal axes of transformation with connectivity matrix (“diff” but not “mean”)

Modulation of inhibition

Activity patterns in APL and its dimensionality, and its relationship to the principal directions

Analyze compartments of ORNs – e.g., dorsal only?

# Oct 18 2024

To do: Tests for robustness of correlations of principal axes with connectivity matrices, possibly also compare connectivity matrices with each other. Bootstrap: (barely) not significant at 5% lower CL,

Yang and Remy to put together datasets of KC recordings with normal APL and with APL inhibition turned off.

These will have a slightly non-overlapping set of stimuli.

25Oct24 These are invoked, for example, by kc\_soma\_nls-process\_26May24.txt

# Tetanus toxin datasets from Yang Zheng, designed to look at APL influence

These are kc recordings; file names will have convention: yyyy-mm-dd\_\_fly[0x]\_\_[identifier]

Main control is TNTin\_label. [inactive tetanus toxin, labeled – 6 datasets]

Main experimental condition is TNT\_label [active tetanus toxin, labeled – 8 datasets – APL should be turned off]

Auxiliary controls are TNTin\_nolabel and TNT\_nolabel [ 2 or 3 datasets each]

These have 17 odorants, including PFO (paraffin oil), which is the solvent for several of the odors and elicits very little response. Five of the odorants are not in Remy’s dataset; they were chosen because they seemed to lead to distinctive patterns of activity in the APL. First step is to compare label with in\_label. But we also want to compare the set of 12 odors that overlap with Remy’s. Eventually, we should be able to build an ORN library by merging a third dataset, which has the 5 unique odors in the TNT dataset, and also “diagnostic” odors; Remy’s dataset has the diagnostic odors as well as her 17 standard odors.

Basic pipeline (both with and without subtracting the mean):

To put Hong Lab data files into psg format, use hlid\_rastim2coords\_demo.m, invoked, for example, by hlid\_rastim2coords\_[TNT|TNTin]\_[label|nolabel].txt, creating files without subtracting the mean and with subtracting the mean (files -sm). All remaining analyses done without subtracting the mean.

To do statistics on alignment, use psg\_align\_stats\_demo.m, invoked, for example, by

kc\_TNT-label\_align\_stats\_29Oct24.[txt|fig|mat], which is modeled after kc\_4recs\_psg\_align\_stats\_demo\_08May24.txt

To create the consensus, can use psg\_align\_knit\_demo.m. To plot the individual values and the consensus, can use psg\_consensus\_demo. These are invoked, for example, by kc\_TNT-label\_process\_29Oct24.[txt|fig|mat], which is modeled after kc\_soma\_nls-process\_26May24.txt, and create data files such as

../kc-tnt/hlid\_odor17\_coords\_TNT-[no]label\_consensus\_scale.mat

Make sure to keep data files in same order when selecting from user interface (use creation time to order).

Analyses done 29Oct24 and 30Oct24, but all files labeled 29Oct24.

# Nov 01 2024

Look at eigenvalues of connectivity matrix “diff” randomized with one or more PC’s subtracted, missing data used to fill in. See if this was already done: <https://elifesciences.org/articles/62576>

TNT dataset: proceed with consensus, all four conditions, and look at transformations between. Mean not subtracted, no rescaling

Ran psg\_geomodels\_run on Procrustes and affine, for TNT-label (adj) and TNTin-label (ref), followed by hlid\_majaxes.

Need to run psg\_geomodels\_run on expanded library of models, and then look at psg\_geomodels\_run on other pairs of conditions.

# Nov 12 2024

Connectivity: Look at the correlations between the first few eigenvectors of the fafb and other dataset.

Principal directions, etc: Looks like 3,4, 5, and 6d models are worthwhile, affine better than projective for non-inhibited (TNT-label) to inhibited (TNTin-label),(except for isolated dimension pairing) and affine also better than piecewise affine.

To create consensus across all control conditions (TNTin-label, TNT-nolabel, TNTin-nolabel), and rerun (“TNT-3c”).

Shows 3,3,4,6d models worthwhile, some benefit of going to projective from affine. Files:

HongLab/psg\_geomodels\_run\_TNT\_label\_TNT3c\_ConsensusNoScale.[\*txt,mat],

psg\_geomodels\_summ-> HongLab/plots/tnt\_label\_tnt3c\_consensus\_noscale\_\*.fig,

then hlic\_majaxes->HongLab/plots/hlid\_majaxes\_TNT\_label\_TNT3c\_ConsensusNoScale[\_just AffineOffset].pdf

File without \_justAffineOffset also includes affine offset and piecewise affine

# Nov 22 2024

Looks like 4d model is expanded a bit w.r.t. 3d model (with inhibition turned off), and maybe that is why its first eigenvector is compressive.. See 4d model in …/plots/kcTNT\_label\_consensus\_4d\_22Nov24.fig

anova after consensus – see psg\_align\_vara\*\_29Nov24.[txt.mat,fig]: shows signif difference TNTlabel vs all controls, and no sig dif among the controls

# Dec 06 2024

Main thing (Betty points out) is that we need to think about response variability, the “dimensions” it occupies – may not be spherical. Especially important since inhibition leads to an overall contraction of the responses, which are quantified by z-scores.

Maybe do PCA on individual trials, and compare dimensions with PCA on mean-across replicates? But this could be swamped by instrumental noise.

# Dec 13 2024

To do: PCA on individual trials with stimulus mean subtracted; can do this with Remy’s ORN and kc data, and Yang will reformat her data for single trials.

(Spherical? Same axes as stimulus space? Low-dim but different axes?)

As follow-up, if this is consistent across flies, can think about ways to pool data across flies

# Dec 20 2024

Next up: TNT datasets.

Need a way to characterize the non-sphericity. Put single trials into a response space, keeping origin at the zero response, and look for anisotropy, both in the space and out of it.

Analyses 28Dec24 of kc-soma-nls datasets with hlid\_rastim\_trial\_[pca|pcs2|vis].m: Response variability is not isotropic; residuals (single-trial responses minus trial-average) are mostly aligned with responses, but they also tend to lie in a lower-dimensional space (not just one-diensional), and or orthos (single-trial responses with trial-average projected out) also show that responses to similar odorants have similar residuals

Analyses 31Dec24 of kc-soma-nls with hlid\_rastim\_trial\_vis.m: to demonstrate that rotation of resids to consensus is critical, without the transform (files \*noxform), the residuals appear to be pointing in random directions.

Analyses of 01Jan25 of kc-soma-nls with hlid\_rastim\_trial\_vis ->vis2.m: shows that first eigenvector does not explain. There’s also excess covariace in the second eigenvector. Also see this with the ORN datasets, but possibly not as strongly. See hlid\_rastim\_trial\_vis\_01Jan25\_[kc-soma-nls|orn]\_d6.[mat,pdf]

Also holds up when mean is subtracted (if\_recenter=1) prior to covariance calculation (analyses of 02Jan25: hlid\_rastim\_trial\_vis\_02Jan25\_[kc-soma-nls|orn]\_d6.pdf

And holds when the first two eigenvectors are fixed, with or without recentering prior to covariance calculation (analyses of 08Jan25: hlid\_rastim\_trial\_vis2\_08Jan25\_[kc-soma-nls|orn]\_d6.pdf

# Jan 03 2025

Agree that it looks like variance is mostly confined to a plane, plane depends on stmiuls. Less so in ORN data.

BH: maybe because of cells with two rates of adaptation?  
to look at TNT datasets!

# Jan 15 2025

Initial analysis of TNT datasets (TNTlabel, 8 files, vs TNT3c, all controls pooled, 11 files), single-trial:

hlid\_rastim\_trial\_vis, up to 4 dims, doesn’t show a major difference in noise pattern, but overall response sizes are very different. hlid\_rastim\_trial\_vis\_15Jan25\_[TNTlabel|TNT3c]\_consensus\_[raw|msub\_norm]\_stim\_[d123|d234].fig

Also, both TNTlabel and control show the squarhing of variance, hlid\_rastim\_trial\_vis2\_15Jan25\_[TNTlabel|TNT3c]\_d4.pdf

Heatmaps of distances in hlid\_rastim\_trial\_pca\_15Jan25\_[TNTlabel|TNT3c].pdf

# Jan 17 2025 meeting

We are puzzled why the TNTlabel datasets seem to show better separation than the TNT three controls.

One possibility (deferred for now) is to have some kind of within-fly measure of signal to noise.

Another possibility is that the quality of the consensus is different in the three control groups.

To do: consensus (without scaling) for the 3 controls, and then, for the megamats, and then an omnibus consensus between them.

# Jan 21 2025

After some mods in psg\_align\_knit\_demo, ran on previously-made files:

./data/kc\_soma\_nls/hlid\_odor17\_coords\_kc\_soma\_nls\_consensus\_noscale.mat (4 datasets, consensus without scaling)

./data/kc\_tnt/hlid\_odor17\_coords\_TNT3c\_consensus\_noscale.mat (11 datasets, consensus without scaling)

This produces

hlid\_odor17\_coords\_kc\_soma\_tnt3c\_knitted\_noscale[\_meta].mat (a consensus alignment, without scaling)

hlid\_odor17\_coords\_kc\_[soma|tnt3c]\_[aligned|component]\_noscale.mat : stimuli lined up, with NaN’s where missing; aligned: stimuli lined up without consensus transformion, component: stimuli lined up and transformed into consensus

So aligned and component files differ only by a rotation, and the component files will match the knitted if the odorant only occurs in one set of experiments. If odorants are present in both datasets and there is no scaling, the consensus will be the mean. However, the knitted datasets and the components may have different centroids, so \*do not\* move centroid to the origin.

psg\_consensus\_demo properly plots these datasets, and one can also use it to form a consensus with scaling between the two components.: psg\_consensus\_demo\_21Jan25.txt

psg\_consensus\_demo\_21Jan25\_kc\_soma\_tnt3c\_d1234[\_scaling].fig

# Jan 24 2025 meeting

Agree that alignment between kc\_soma and tnt3c datasets look very good

To set up single-trial decoding analysis, across flies

* Allow for downsampling from total number of animals, to enable comparison between datasets of different sizes
* Ways to leave out trials:
  + a few trials of a few odorants, randomly across flies (? Within one trial number ? randomly)
  + All stimuli from all flies but from one trial number (1,2,3)
* Ways to treat trials numbers: all in one bag, or, each trial number is separately kept and aligned
* How to do consensus across flies and trials: with or without scaling allowed
* Kind of distance to base decision on
  + Euclidean
  + Cosine
  + Least RMS distance? Least distance to any in cluster? Harmonic mean?
* Dimensionality of space
* Normalize responses first?

To create: confusion matrix

# Jan 31 2025 meeting

Folds with fewer omitted have better “pancakes”, but this is pretty minor. A strategy with one stimulus randomly chosen from each repeat of each dataset seems like a good compromise –this will yield 17 folds since there are 17 stimuli.

Pilot studies with leaving out a full repeat (from one animal) and aligning this as a whole indicate that this is not a good way to go; as dimension increases, fraction correct increases to ceiling. It’s using all of the stimuli to align each of the stimuli, so it’s not really leaving out a trial.

To proceed with decoding.

Analyses (hlid\_rastim\_trial\_decode\_confmtx\_06Feb25\_TNT3c.[txt,mat.pdf] and

hlid\_rastim\_trial\_decode\_confmtx\_07Feb25\_TNTlabel.[txt,mat,pdf],

replotted with hlid\_rastim\_trial\_decode\_confmtx ->file names \*details[1|2|3|4] with confusion matrices (raw, mean sub, and with and without normalization),

\*summ with line plots of fraction correct

with

show that decoding is better with label than without (for Euclidean, cosine, and Mahalanobis).

Decoding does not saturate, even with 7 dimensions.

Go up to dim 12, but (for speed) only one downsample and only one cross-validation configuration:

Analyses (hlid\_rastim\_trial\_decode\_confmtx\_10Feb25\_[TNT3c|TNTlabel].[mat,pdf,fig]

These saturate around dim 7

# Feb 14 2025 meeting

We agree that turning off inhibition seems to improve decoding. But maybe what (normal) APL inhibition does is to increase animal-to-animal variability? So try a single-animal decode.

This can be done with simple modifications o fhlid\_rastim\_trial\_decode: downsample to one dataset at a time. But leave out only one trial (from just one repeat). So there are 17x3 folds, and only one cross-validation configuration.

However – see hlid\_rastim\_trial\_decode\_confmtx\_15Feb25\_[TNT3c|TNTlabel].[mat,pdf,fig], the superiority of the TNTlabel (APL inhibition removed) persists even with within-prep decoding.

Also: Another difference between my analysis and Yang’s is that she,working within single flies, does not do dimension reduction first; just looks at cosine correlation. So maybe she can do her analysis but only after a dimension-reduction within a fly?

Reanalyzing up to 16 dimensions hlid\_rastim\_trial\_decode\_confmtx\_17Feb25\_[TNT3c|TNTlabel].[mat,fig]

shows that the advantage of the TNTlabel goes away – suggesting that it is due to the denoising.

Need to remove the denoising completely – e.g., analyze in a space of dimension equal to number of KCs, or, equivalentluy, number of stims x number of trials.

# Feb 21 2025 meeting and later thoughts

## Papers to look at (from Betty)

Babadi & Sompolinsky 2014

https://www.sciencedirect.com/science/article/pii/S0896627314006461

Litwin-Kumar et al

<https://www.cell.com/neuron/fulltext/S0896-6273(17)30054-5>

also, Task-dependent optimal representations for cerebellar learning, e-life, 2023

Angus Silver + colleagues:

https://www.nature.com/articles/s41467-017-01109-y

There’s also a (long) review article from Angus Silver on mechanisms of pattern separation that tries to argue for the utility of dimensionality over metrics like correlation. It’s a useful read:

https://www.cell.com/neuron/fulltext/S0896-6273(19)30071-6

## Ideas

APL may sparsify, and this could have the gleamed advantages for coding random activity patterns but — for these odors, they only cover a low dim space when we look agt what is needed to decode (3 to 7),

APL may also be a gain control or make differing intensities look more similar,

The Babadi paper uses a task in which the N odors have to be segregated into two categories, and any combination may be in one category or another. This is very different from representing proximity relationships.

## Effective dimensionality (participation ratio)

This is sum(eigs)^2/sum(eigs^2)

in hlid\_participation\_ratio\_24Feb25.txt, for single-trial and avgs across repeats, for mean-sub and not mean-sub, with and without normalization

Could apply Chanwoo’s correction

## To do

Unembedded decoding -> done

**Determine transformation** between labeled and unlabeled space, for all 3 controls pooled

Already done for comparison with a single control, found it was affine, HongLab\plots\tnt\_label\_tntin\_label\_consensus\_noscale\_procrustes\_vs\_affine.fig -> this was done, ca. 18Nov24 to 19Nov24, files in HongLab or HongLab\plots:

psg\_geomodels\_run\_TNT\_label\_TNT3c\_ConsensusNoScale.[txt,mat]

tnt\_label\_tnt3c\_consensus\_noscale\_[Procrustes|affile|projective]\*.fig

hlid\_majaxes\_TNT\_label\_TNT3c\_ConsensusNoScale.txt

hlid\_majaxes\_TNT\_label\_TNT3c\_ConsensusNoScale\*.pdf

Affine is much better than Procrustes; projective is slightly better than affine (dim range 3-6).

Major axis analysis: one eigenvector stays the same length; others are reduced (factor of 0.6-0.8)

**Calculate information** from confusion matrix, with T-P bias correction, or, via nearest-neighbor distance, within-fly or across-fly; can use xv\_confmtx\_make with out\_sample=in\_sample and aux output to retrieve the disparities; can do this with ORN data, and each KC dataset.

# Feb 28 2025 meeting

Yang: analyses of cosine and Euclidean distances between pairs, raw and after PCA, in TNT and controls: looks like when APL is active, there is greater separation of distances between same-odor responses, and distances between different-odor responses. But this needs to be looked at with ROC analysis.

*We actually have multiple tasks:*

* two odors presented, is it same or different, with each stim drawn randomly from N=17 odors
* two odors presented, is it same or different, each stim drawn randomly from only a pair of odors (Yang can re-analyze)
* which odor (“decoding”)
* learn an arbitrary correspondence (Babadi/Sompolinsky) – here, dimension expansion is useful
* any of the above tasks with or without confounds by changes in concentration
* any of the above tasks with mixtures, and with mixtures, determining whether an odor is a component of the mixture

The “arbitrary correspondence” task seems non-physiological, since it ignores correlations that must be there from nature. But it still might be a good idea to have some capacity to learn unexpected correspondences.

The APL might be useful for any or all of these, or be useful for some and detrimental for others, and it might just be for housekeeping

# Mar 02 2025

Information in confusion matrices

hlid\_rastim\_trial\_decode\_confmtx\_02Mar25\_info.txt: re-analyzed results from TNT3c, TNTlabel in hlid\_rastim\_decode\_confmtx with info analysis added:

Decoding via embedding, pooling flies

hlid\_rastim\_trial\_decode\_confmtx\_[06Feb25\_TNT3c|07Feb25\_TNTlabel] (d1-d7)

hlid\_rastim\_trial\_decode\_confmtx\_[10Feb25\_TNT3c|10Feb25\_TNTlabel] (d1-d12)

Decoding via single flies

hlid\_rastim\_trial\_decode\_confmtx\_[05Feb25\_TNT3c|15Feb25\_TNTlabel] (d1-d7)

hlid\_rastim\_trial\_decode\_confmtx\_[17Feb25\_TNT3c|17Feb25\_TNTlabel] (d1-d16), also saved some confusion matrix heatmaps

In general, info and fraction correct are parallel, *except* for high-dimensional embeddings and single-fly decoding (analyses of 17Feb25), even when the fraction correct declines at high dimension, the information stays high, i.e., errors are systematic

# Mar 05 2025

Analysis without embedding: controls have higher fraction correct and higher info than label. Note that no embedding is equiv to 51 dimensions (17 stimuli x 3 trials), so the 16-dim embedding still does some de-noising. Also there is a decline in fraction correct with embeddings 10-16, but not in information: i.e., the errors are systematic.

hlid\_rastim\_trial\_decode\_confmtx\_05Mar25\_[TNT3c|TNTlabel]\*.(mat,fig,txt)

# Mar 07 2025 meeting

Further discussion on task types

Why is there such an improvement between 16-dim embedding and no embedding?

* PCA: for each repeat considered separately, about 90% of variance explained with 16 dims -- Look at intermediate numbers of dimensions?
* Maybe try to build a model that has this behavior?

Also to do: fix labeling problem in hlid\_rastim\_trial\_decode\_confmtx; try decoding with cosine but nearest neighbor (like Yang)

# Mar 14 2025

hlid\_rastim\_trial\_decode now has several options:

if\_singleprep: all decoding done within a single fly. Allows for if\_noembed (which decodes based on ROI data, without trying to find a common representational space across preps)

if\_embedbyprep: 0: embedding is done on each repeat separately, and then the several repeats are aligned within and across preps by Procrustes. 1: embedding is done on all repeats within the same prep. Repeats are aligned across preps. This allows for embedding in a dimension that is greater than the number of stimuli, but less than nstims x nrepts. Alignment across preps matches responses only if they come from the same stimulus and the same repeat. So this requires decoding across flies.

Not implemented is a way of embedding all nstims x nrepts together (i.e,. allowing for high-dimensional rembeddings), but then only aligning based on stimulus identity. That would allow decoding within flies, but a high-dimensional embedding.

Also, not implemented is alignment across flies, but scrambling the repeats.

# Mar 14 2025 meeting

Should do nstims x nrepts together, decode across flies, by drop one, dims [1 2 3 4 8 12 16 24 32 48], TNT3c and TNTlabel, aligning (as is current) by stimulus and repeat

Then do this, scrambling repeat number (to see if capturing fine structure of repeat-dependence is crucial), need to add scrambling repeat number to software

Then do this within fly, but this needs software logic mods to allow for within-fly decoding AND embed-by-fly, so that there is alignment within flies by ignoring the repeat number

# Mar 17 2025

Analysis with embedding of the three repeats together, so one can look at dims [1 2 3 4 6 8 12 16 20 24 32 40 48]

Either decode within prep, dropping just one response, or decode across preps, dropping one response, and aligning the trials with respect to [stim x rept] (i.e,. all 51 trials)

With hlid\_rastim\_trial\_decode\_17Mar25.txt:

TNT3c does better than TNTlabel for within-prep:

hlid\_rastim\_trial\_decode\_inprep\_TNT[3c|label]\_17Mar25\*.[mat,fig]

Approx. equal, sometimes slight advantage for TNTlabel, in across-prep:

hlid\_rastim\_trial\_decode\_x-prep\_TNT[3c|label]\_17Mar25\*.[mat,fig]

With hlid\_rastim\_trial\_decode\_17Mar25a.txt:

hlid\_rastim\_trial\_decode\_o-prep\_TNT[3c|label]\_17Mar25\*.[mat,fig]: across-fly decoding but leaving out all trials from one fly. Similar to x-prep.

Across-prep is slighty better than within-prep, but perhaps this is because in the across-prep, there are more “standards” in addition to the other-repeat responses from the same fly (even though the centroid is used for decoding). Same-fly confound is removed in the o-prep runs.

Need to scramble repeats

# Mar 21 2025 meeting

* Backburner scrambling repeats – won’t be a sharp control as, even though it will make the alignment not as good, the centroids of the three repeats will still be aligned.
* Let’s return to the transformation of the space between APL-off and APL-on, with the idea of identifying a direction with a consistent change. Do a shuffle on the label vs 3c, forming consensus after each shuffle, and then aligning, finding the axes with greatest and least distortion. Do this parametric in number of dimensions (same for TNTlabel and TNT3c). Check that same convention (adjusted and reference as was used in the past. Then, if the actual data is different than the shuffle, do a bootstrap or jackknife to put error bars on these directions. These directions could then be used for a future behavioral study, and possibly, the weightings of the odorants on them could be interpretable.

Maybe do this with aligning the mean responses, vs aligning the individual trials?

* Remy may be able to create ORN datasets with more odorants, overlapping partially with existing ORN datasets, and covering the odorants used in the TNT experiments. These could be consolidated by a “missing data” algorithm, and a space constructed – both for intrinsic interest and comparison with the TNT-KC datasets

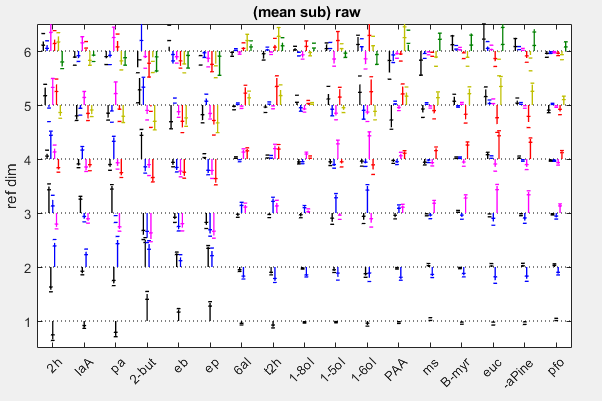
# Mar 28 2025 meeting

* Re statistical analysis of transformation from APL-off to APL-on: looks promising. Add “censoring” of shuffles to include only shuffles with substantial relabels. Also add absolute size of largest stretch, although this may be confounded by technical aspects of how the APL-off and AL-on responses are recorded
* Would be worth looking at the transformation from ORN to APL-off, and from ORN to APL-on, but this needs ORN datasets with the APL odor panel
* Would also be worth trying to knit together KC megamat and KC APL-on datasets, and Remy is working on formatting ORN datasets with the APL odor panel

# Apr 04 2025 meeting

* Yang finds that with cosine distance, nearest-neighbor, single-trial, within-fly decoding, then TNTlabel is worse than TNT3c whether done with PCA or not; need to compare this -> hlid\_rastim\_trial\_decode\_confmtx\_04Apr25\* shows that I also find that this is the case, over the entire range of dimensions.
* Yang will do 2-AFC same-different decoding, across experimental conditions
* Statistics of APL-off to APL-on: makes sense to keep only 50% of most-entropic trials
  + Compared adj and ref with heatmaps of hlid\_majaxes\_TNT\_label\_TNTin\_label\_ConsensusNoScale\_justAffineOffset.pdf (generated by hlid\_majaxes.m), and there is good agreement of the analysis of the raw data via hlid\_hlid\_geom\_transform\_stats.m with that previous analysis, for dimensions 2 and 3. Not so for dims 4 and 5, but top two eigenvalues are quite similar. Note also that stimulus order was native stimulus order prior up to 05Apr2025, and defaults to kcmerge (in hlid\_setup) beginning 06Apr25.
  + Need to do bootstrap confidence limits on projections, maybe also on cosines or dot-products as a more global measure

# Apr 18 2025 meeting

* We found “biological” directions in the odor space representation in KC’s: the eigendirections of the APL-off to APL-on transformation. In an N-dimensional model (e.g., N=3), this means that the way that each of M odors modulates the activity of each of K Kenyon cells, via the APL, can be reduced to an MxN matrix of the activation of one of the N APL “modes” by one of the M odors, and an N x K matrix of the modulation of each of the K Kenyon cells by one of these N modes. We can’t compare the NxK matrix across individuals (since the Kenyon cells cannot be identified between animals), but we have found that the MxN matrix is conserved. The N columns are orthogonal, since they are the principal directions of the transformation. This set of directions is arguably more invariant than the principal components of the representation, since the latter is heavily driven by which odors are used.
  + For dimension 3, one could ask whether the weightings of the odors onto the axes have any structure – are they sparse? Are they anti-sparse?
  + 
* We previously found the principal directions of the distortion from ORN space to KC (APL on) space, albeit with a non-overlapping set of odorants. This is also an M x N matrix (in odor space), along with an N x R matrix (of odorant activations of receptors). The latter is conserved across individuals (since ORN’s are identifiable and have conserved sensitivity profiles). The MxN matrix of principal directions need not have been conserved, but seemed to be – since it was based on the consensus ORN response.
  + We might do statistics on the ORN to KC space principal directions (bootstrap but can’t do shuffle)
  + It is natural to ask, how does the ORN to KC principal directions compare to the KC-APLoff to KC-APLon principal directions? This will require new ORN datasets with the 17 KC odorants.

# Apr 23 2025 meeting

* Reinforced above plan, and also, to look at 22-odor space by knitting together the 17-odor megamat set and the 17-odor APL set (with 12 overlaps, and 5 odors unique to each set)
* The neuron from APL to the glomeruli is incredibly ramified, may be tree-like.
* How can we quantify selective APL -> KC inputs? Work in the dimension-reduced space of each fly’s KC odor?
* Remember that in a fly dataset with an ORN ROI missing, it is probably because it is unresponsive.
* Can bring odor space into ORN space using a “universal” ORN to odor matrix, via PCA/missing data across flies

# May 2 2025 meeting

* Data to receive:
  + “orn\_megamat\_wd” datasets are the same as the original orn\_terminals, but have diagnostic odors added (single trial, original 17 panels); should be in single-trial format
  + “orn\_validation2\_wd” datasets have a set of 17 test odors that overlaps in 12 with orn\_megamat, also in single-trial format. Also has diagnostic odors
* Odors are designated by strings of chemical abbrev, then @, then concentration. In some cases, the same chemical is used at one concentration as a test odor, and at a second (typically lower) concentration as a diagnostic odor. In orn\_terminals, only the test odors are present; they are listed in trial\_info.stim in the order delivered. Taking the first 17 and putting them in alphabetical order is the standard read order,

{'1-5ol @ -3.0'}

{'1-6ol @ -3.0'}

{'1-8ol @ -3.0'}

{'2-but @ -3.0'}

{'2h @ -3.0' }

{'6al @ -3.0' }

{'B-cit @ -3.0'}

{'IaA @ -3.0' }

{'Lin @ -3.0' }

{'aa @ -3.0' }

{'benz @ -3.0' }

{'eb @ -3.0' }

{'ep @ -3.0' }

{'ms @ -3.0' }

{'pa @ -3.0' }

{'t2h @ -3.0' }

{'va @ -3.0' }

This is also in response\_amplitude\_stim.stim, and it matches the commented-out label\_list from hlid\_setup:

opts\_plot=filldefault(opts\_plot,'label\_list','typenames');

%opts\_plot=filldefault(opts\_plot,'label\_list',...

% {'1-5ol','1-6ol','1-8ol','2-but','2h','6al','B-cit','IaA','Lin','aa','benz','eb','ep','ms','pa','t2h','va'});

%

if ~exist('opts\_multm\_def') opts\_multm\_def=struct; end

opts\_multm\_def=filldefault(opts\_multm\_def,'color\_norays\_list',{'k','b','c','m','r',[0.7 0.7 0],'g',[0 0.5 0],[0.5 0.5 0.5]}); %colors for individual datasets

disp('opts\_read, opts\_plot, opts\_multim\_def initialized for hlid data');

%

if ~exist('display\_orders')

display\_orders=struct;

display\_orders.kcclust={'2h','IaA','pa','2-but','eb','ep','aa','va','B-cit','Lin','6al','t2h','1-8ol','1-5ol','1-6ol','benz','ms'};

display\_orders.kctnt={'1-5ol','1-6ol','1-8ol','2-but','2h','6al','IaA','PAA','eb','ep','ms','pa','t2h','B-myr','euc','-aPine','pfo'};

display\_orders.kcmerge={'2h','IaA','pa','2-but','eb','ep','6al','t2h','1-8ol','1-5ol','1-6ol','PAA','ms','B-myr','euc','-aPine','pfo'};

end

Note that hlid\_rastim\_trial\_read.m verifies that the trial-averages computed locally are equal to the trial averages in the file.

# May 3 2025

JV: One thing we did not discuss yesterday: was, whether to do the alignment between flies (and between non-overlapping odor sets) based on the average of the three trials within flies, vs, individually aligning all three repeats (effectively treating the 17 odors x 3 trials as 51 separate odor-trial data points).

We did the latter (17 x 3) for the decoding, and I think it made sense – since we were decoding single trials, and there was systematic repeat-to-repeat variability.

But here, for knitting together sets to make a richer olfactory space, I don’t think it makes sense: we would be aligning noisier data, and in a sense fooling ourselves in thinking that we had 51 pieces of evidence, ignoring the obvious fact tht the three trials of the same stimulus are correlated. And also, if the stimuli are presented in a different order across flies, we would be adding in the variability of the sequence effects prior to the alignment, rather than first trying to average it out.

EJH: I agree that doing the alignment on the trial-averaged responses in the ORNs feels more principled. Our immediate goal isn’t necessarily to understand trial-to-trial reliability but to get the best estimate of each stimulus’ coordinates, so averaging away both random and systematic noise seems better?

# May 9 2025 meeting

More on merging ORN files: each panel is non-overlapping: megamat17 has 17 odorants, validation2 has 22 odorants. But there should be no “scale factor” between them; each has same genotype. So we just have to make sure that hlid\_orn\_merge does not apply a scale factor when it does the missing-data PCA.

This required a fix in hlid\_orn\_merge. Original version used the regression slopes (afalwt\_fit.x\_true), which only is a fit to the absolute size of the data when multiplied by afalwt\_fit.b\_norm. So, now, when if\_restore\_size=1, this multiplication is carried out.

The new files all have “wd” in the name, indicating “with diagnostic odors”

# May 27 2025 analysis

\* Created standard coordinate files from each single-trial dataset of megamat17\_wd and validation2\_wd, using hlid\_rastim2coords\_demo.mat: see hlid\_rastim2coords\_demo\_[megamat|validation2]\_wd\_27May25.[txt,pdf], files are in HongLab/data/[orn\_megamat17\_wd|orn\_validation2\_wd].

These files (trial-averaged coordinates) have names like

HongLab/data/orn\_megamat17\_wd/ hlid\_odor17\_coords\_2023-04-22-2.mat

and

HongLab/data/validation2 \_wd/hlid\_odor22\_coords\_2023-11-19-1.mat

\* Verified that the coordinates in data/megamat17\_wd (e.g., hlid\_odor17\_coords\_2023-04-22-2) and data/orn\_terminals (same file name) are identical.

\* checked consistency, via psg\_align\_stats\_demo, and created consensus files within megamat17 and validation2—see psg\_align\_stats\_demo\_27May25.txt. File names are

hlid\_odor17\_coords\_orn\_megamat\_wd\_consensus\_[no]scale\_renorm.mat in …data/orn\_megamat\_wd

hlid\_odor22\_coords\_orn\_validation2\_wd\_consensus\_[no]scale\_renorm.mat in ../orn\_validation2\_wd

and

hlid\_odor17\_coords\_orn\_terminals\_consensus\_[no]scale\_renorm.mat in ../orn\_terminals

which is equivalent to

hlid\_odor17\_coords\_orn\_terminals\_consensus\_[no]scale.mat made in May 2024

(made with plots/orn-terminals\_9recs\_psg\_align\_stats\_demo\_21May24.[fig,.mat] from the original megamat17 files)

\* Merged the two disjoint panels and compared, see hlid\_orn\_merge2\_28May25.txt, files put in ../data/orn\_merged and plots/hlid\_orn\_merge2\_odor39\*.fig

\* Comparing the odor spaces for each of the components, i.e., comparing the PC’s of the merged and component sets (the columns of v) in resps(stim,glomerulus)=u\*s\*v’) shows that the two odor spaces share the first PC but then differ (with some overlap), so that the first 5 pc’s of the combined space project strongly into each of the first three pc’s of the component space. See hlid\_orn\_merge2\_29May25.txt and ../plots/hlid\_orn\_merge2\_odor39\_[inter|union]\_resps\_dots.mat

\* The two panels seem to occupy distinct halves of space. This is supported by Fisher Discriminant analysis; for dim 2 and above, a linear classifier separates megamat17 from validation2. See hlid\_orn\_merge2\_fisherdisc.txt

# May 30 2025 meeting

\* Main discussion about the separation of the two subsets: cannot be due to a scaling issue between the subsets (same linear span), but perhaps a matter of differences in optical sectioning in the two sets of experiments, making the ventral glomeruli less-resolved in one case? But seems unlikely, discrimination is not based on ventral vs dorsal. Megamat17 has simpler molecules, validation2 is mostly monomolecular components from natural odorants

\* map the Fisher discriminant into ORN space, to confirm that it’s not dorsal vs ventral -> done, see .../data/orn\_merged/FisherDiscriminantInORNspace.[txt,fig]

\* future: add more odorant datasets

\* look at ORN’s in ORN space, KC’s in KC space (rather than odors in odor space)

# June 2 2025 analysis

proceed with looking at the subset of odors that are in the KC TNTlabel and TNT3c, and the mapping from glomeruli, using “union” merge -> done, see psg\_geomodels\_run\_orn039union\_TNT[Label|3c]Noscale\_7d.[mat,txt] and ../plots/psg\_geomodels\_run\_orn039union\_TNT[Label|3c]Noscale\_7d\*.fig

The label shows a larger difference between Procrustes and affine than the control.

Comparing principal axes: see hlid\_majaxes\_06Jun25.(txt,pdf). First 3 axes look similar in control and label, but axis 4 is more attenuated in control than in label (page 5 vs 15). (Initial files with names 02Jun25 were incorrect, had stimulus ordering and labelling problem, fixed with psg\_commonstims and edits in psg\_majaxes: the problem was that when stimuli were reordeered to find common stimuli in psg\_geomodels\_run, this reordering was not replicated in psg\_majaxes)

Looking at projective and piecewise affine: These are no better than affine.

See psg\_geomodels\_run\_orn039union\_TNT[Label|3c]Noscale\_6d.[mat,txt] and ../plots/psg\_geomodels\_run\_orn039union\_TNT[Label|3c]Noscale\_6d\*.fig

# June 6 2025 meeting

In the 4-d model, it does look like eiv 4 is more squeezed from ORN to KC in control vs APL-off – but the 4-d model is shaky, via the nesting test. Also, it’s not clear what it maps into in KC space, since the KC coords, which were built by consensus, are not PCs.

So – need to build the KC coords as PCs, and redo. See

plots/kc\_TNT-label\_align\_stats\_29Oct24.txt,

plots/kc-TNT3c\_process\_13Nov24.txt

The corresponding files in pc rotation, made with psg\_align\_knit\_demo, after modifications to do pc rotation:

data/kc\_tnt/kc\_TNT-label\_process-pc\_09Jun25.txt -> data/kc\_tnt/hlid\_odor17\_coords\_TNT-label\_consensus-pc\_noscale.mat

data/kc\_tnt/kc\_TNT3c\_process-pc\_09Jun25.txt ->data/kc\_tnt/hlid\_odor17\_coords\_TNT3c\_consensus-pc\_noscale.mat made

Then, psg\_geomodels\_run, psg\_geomodels\_run\_orn039union\_TNT[label|3c]-pc\_7d.[txt,mat], analyses with psg\_geomodels\_summ are identical (saved in ../plots/psg\_geomodels\_run\_orn039union\_TNT[Label|3c]-pc\_7d\*.fig), and hild\_majaxes\_10Jun25.[txt|pdf], eivenvalues are identical, and principal axes are identical, but (as expected) bases in the KC space are different, since now they are PC’s

# June 20 2025 meeting

Revisiting the mapping from ORN to KC space: looks like the finding of a selective change for dim 4 is fragile. psg\_geomodels\_run\_orn039union\_TNT3cNoscale-pc\_7d\_nestbydim.fig shows that only dim 3 can be trusted.

Comparing eigenvalues of the affine mapping (c=control, t=TNT) – note that dim 5 goes in the opposite direction

e3c = 20.2900 9.0000 4.3000

e3t = 26.6800 10.9200 5.3000

e3t./e3c = 1.3149 1.2133 1.2326

e4c = 23.5200 18.9300 10.1900 2.7800

e4t = 31.4000 25.3800 14.2400 5.6800

e4t./e4c = 1.3350 1.3407 1.3974 2.0432

e5c = 21.5700 19.7400 10.4300 7.6300 3.4500

e5t = 28.3900 26.1900 14.6200 10.9900 0.6700

e5t./e5c = 1.3162 1.3267 1.4017 1.4404 0.1942

OTOH we noticed that for the dim 3 mapping (control), coord 1 (pca1) corresponds to principal directin 3 of the transformatoin, coord 2 to direction 2, and coord 3 (pca3) to direction 1. This raises the possibility that the transformatoin is “sphering” the ORN coding volume – if each eigenvalue of the transformation is inversely proportional to the sqrt(variance) of each PC explained. For the tnt-active conditin (inhibition turned off), the proportinalities are similar, but the directions don’t match as well.

So the next thing to look at is

\* whether this sphering holds quantitatively (for this odor set): both in terms of magnitudes of the PC’s and the eigenvalues,

\* Does it hold for hte ORN to KC megamat set? Look to see if analysis of the transformation was done in PC coordinates

\* Pay attention to whether mean is subtracted or not – probably makes sense NOT to subtract the mean – and to what extent the first PC is similar to the mean

\* We also have a KC dataset with 22 stimuli (from knitting the KC megamat and KC-3c), can look at transformation from. ORN data with same 22 odorants – to do in PC space.

# June 21 2025 emails

(Preliminary to the above): : The merged ORN dataset (megamat17+validation2) had 39 odorants, 16 of which were in common with the KC-TNT dataset (all of the TNT’s except for PFO). So the transformations I analyzed were based on mapping the first K PC’s of the ORN space into the first K PC’s of the KC space, using those spaces but rstricing to the 16 odorants in common. BUT – those PC’s were constructed for all of the odorants (39 and 17), rather than the 16 in common. So they are not the same as the PC’s for the space of the 16 in-common odorants, in a sense, some of the PC’s may be misleading in that they capture variances of the stimuli in only one of the two stimulus sets. So I think that this analysis needs to be redone, again, creating the spaces just for the 16 odorants in common. Maybe it won’t be very different, but maybe it will shed some light on what we were puzzled by – for example, that the only model that was statistically significant was the 3D model. So let’s see – half a step back, before more steps forward.

Implementation: One possibility: modifying hlid\_coords\_svd so that it can take an input that lists the stimulus names (use that field, since it contains concentration designators) to use for SVD. This list needs to be kept in coord\_opts, which hs other options related to the svd, such as a mean subtraction flag.. Then hlid\_orn\_merge2 needs to be rerun to create new ORN datasets with SVD done just on the 16 odors in common with the TNT experiment. But then hlid\_odor17\_coords\_TNT[3c|-label]\_consensus-pc\_noscale.mat need to be modified to also just use the 16 odors. But this needs to be done as a pipe, OR, psg\_align\_knit\_demo.m needs to be modified so that output files are only made based on PCs for a subset of coords.

So – better to write a pipe in which, assuming a full set of PC’s, it is projected into the PC spce of a smaller number of stimuli – with stimuli listed in stim\_labels, since that is the only field guaranteed to be present. This can be added to psg\_coord\_pipe\_proc, which will warn if not all the dimensions are present. This can be applied both to the union files for the ORN’s and the kc files.

Remember

For consensus, models of successively higher dimension need not strictly contain a model of lower dimension, AND, the coordinates are not necessarily PC’s; this can be applied to ORN or KC data. But psg\_align\_knit\_demo has an option to rotate each model into its PCs, and a procedure in psg\_coord\_pipe\_proc can rotate each model into its PCs. Consensus necessarily removes the mean.

For knitting, ie., consensus with substantial non-overlap of stimuli, situation is same as consensus, both regarding nesting and removing the mean.

For alignment via offset-and-scale (only for ORN’s), each successive dimension contains the previous dimension, and it is in PC’s (since this is what hlid\_coords\_svd does.). The mean is not removed, and may project onto multiple PC’s.

Another procedure is added to psg\_coord\_pipe\_proc that projects each model into PC’s of a subset of stimuli – NOT a rotation. If the starting dataset has a full set of PC’s, and is nested from dim 1 up to max, then this should be the same as calculating PC’s with the restricted set of stimuli from the original data. For dimensions greater than the number of stimuli in the restricted set, the coordinates should be zero.

# June 24 2025

ORN and TNT-label and TNT-3c coord sets made with new psg\_coord\_pipe\_proc, using only the 16 stimuli in common.

“Full nesting” used, i.e,. coords reconstructed from all 37 (or 17) stimuli, and then projected into the first N PC’s as determined by the 16 stimuli in common.

In data/orn\_merged: hlid\_odor039\_coords\_merged\_union-ovlp16.mat, psg\_coord\_pipe\_proc\_orn-ovlp16.txt, and orn-ovlp16\*.fig

In data/kc-tnt: hlid\_odor17\_coords\_TNT[3c|-label]\_consensus-pc\_noscale-ovlp16.mat and psg\_coord\_pipe\_proc\_kctnt-ovlp16.txt, no figure files made, since this only omits one stimulus (pfo)

Looking at transformations: psg\_geomodels\_run\_orn039union\_TNT[3c|Label]-pc-ovlp16\_7d.[txt,.mat] and in plots/psg\_geomodels\_run\_orn039union\_TNT[3c|Label]Noscale-pc-ovlp16\_7d\_\*.fig: nested analysis supports affine transformation up to dim 3 or 4 – which was not the case with the reconstruction based on 39 stimuli.

# June 27 2025 meeting

Enhancements to hlid\_majaxes\_eval: Look at ratios of var explained in pc space, Look at kurtosis on each axis, pc and prj, arith and geometric mean

Rebuild the 16-odor space by concatenating and restricting to the 16 relevant stmuli before doing PCs, both as a check and to allow mapping back to ORN space in hlid\_majaxes

# June 30 2025

hlid\_orn\_kc\_check.m (and hlid\_orn\_kc\_check\*.docx, hlid\_majaxes\_eval\*.doc) properly cross-check, showing that the reason that sometimes the variance explained in hlid\_majaxes\_eval is non-monotonic is that those eigenvectors are calculated on the adj (orn) data, around zero, but then the coordinates in the results structure used by psg\_geomodels\_run are centered across the 16 stimuli in common. So those are no longer the principal axes.

Interesting: hlid\_majaxes\_eval2: also looks at the shape of the adj dataset, after the best-fitting affine transformation. **This is NOT in the direction of sphering ... the so the affine model is not good enough to capture that feature – the affine model over-shrinks the small directions.**