Bringing in the outliers: A sparse subspace clustering approach to learn a dictionary of mouse ultrasonic vocalizations

ICASSP 2020

Jiaxi Wang[1], **Karel Mundnich**[1], Allison T. Knoll[2], Pat Levitt[2], Shrikanth Narayanan[1]

May 5th, 2020

[1]Signal Analysis and Interpretation Lab (SAIL), University of Southern California

[2]Department of Pediatrics and Program in Developmental Neuroscience, The Saban Institute, Children's Hospital Los Angeles, Keck School of Medicine, University of Southern California

- Mice vocalize during social interactions.

- Mice vocalize during social interactions.
- These vocalizations are in the ultrasonic range (between 30 and 150kHz).

- Mice vocalize during social interactions.
- These vocalizations are in the ultrasonic range (between 30 and 150kHz).
- Their properties vary across different social contexts.

- Mice vocalize during social interactions.
- These vocalizations are in the ultrasonic range (between 30 and 150kHz).
- Their properties vary across different social contexts.
- Therefore, vocal communications are used as a proxy to study different behavioral patterns and states.

- Mice vocalize during social interactions.
- These vocalizations are in the ultrasonic range (between 30 and 150kHz).
- Their properties vary across different social contexts.
- Therefore, vocal communications are used as a proxy to study different behavioral patterns and states.

We require a dictionary of vocalizations.

However, there is no consensus on the amount of vocalizations that different mice strains produce.

Traditional Methodology

Traditionally, the dictionaries of vocalizations (USVs) have been generated by hand, where researchers decide on arbitrary rules to differentiate between vocalizations.

Traditionally, the dictionaries of vocalizations (USVs) have been generated by hand, where researchers decide on arbitrary rules to differentiate between vocalizations.

This is a difficult task!

Example

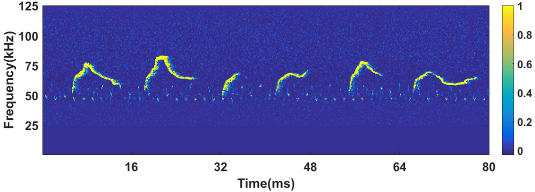


Figure 1: A spectrogram depicting a sequence of mouse ultrasonic vocalizations (USVs). Brighter colors represent higher energy in a given frequency band.

Some tools have been proposed to automate this process:

- MSA[1]
- VoICE[2]
- MUPET[3]
- DeepSqueak[4]

However, these methods do not consider in the clustering stage the space in which the USVs lie on!

[1] J. Chabout et al. "Male mice song syntax depends on social contexts and influences female preferences". In: *Frontiers in behavioral neuroscience* 9 (2015), p. 76.

[2] Z. D. Burkett et al. "VoICE: A semi-automated pipeline for standardizing vocal analysis across models". In: *Scientific Reports* 5 (2015), p. 10237.

[3] M. V. Segbroeck et al. "MUPET-Mouse Ultrasonic Profile ExTraction: A Signal Processing Tool for Rapid and Unsupervised Analysis of Ultrasonic Vocalizations". In: *Neuron* 94.3 (2017), pp. 465–485.

[4] K. R. Coffey, R. G. Marx, and J. F. Neumaier. "DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations". In: *Neuropsychopharmacology* 44.5 (2019), p. 859.

In this work

- We propose using a sparse subspace clustering (SSC) approach to find a dictionary of mouse ultrasonic vocalizations (USVs).

In this work

- We propose using a sparse subspace clustering (SSC) approach to find a dictionary of mouse ultrasonic vocalizations (USVs).
- We evaluate our approach against two baselines using:

In this work

- We propose using a sparse subspace clustering (SSC) approach to find a dictionary of mouse ultrasonic vocalizations (USVs).
- We evaluate our approach against two baselines using:
    - k-means

In this work

- We propose using a sparse subspace clustering (SSC) approach to find a dictionary of mouse ultrasonic vocalizations (USVs).
- We evaluate our approach against two baselines using:
    - k-means
    - Spectral clustering

In this work

- We propose using a sparse subspace clustering (SSC) approach to find a dictionary of mouse ultrasonic vocalizations (USVs).
- We evaluate our approach against two baselines using:
  - k-means
  - Spectral clustering

We show the advantages of using SSC over other methods.

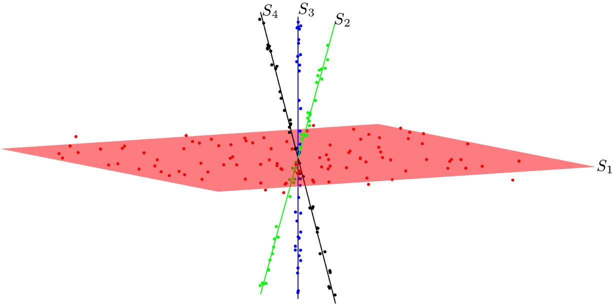


Figure 2: Clusters lying in subspaces of $\mathbb{R}^3$. Image retrieved from.[5]

We assume that vocalizations lie in subspaces of $\mathbb{R}^n$.

[5] https://viterbi-web.usc.edu/~soltanol/RSC.html

[6] E. Elhamifar and R. Vidal. "Sparse subspace clustering". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 2790–2797.

Let $\mathbf{s}$ be a vector containing a single USV, and $\mathbf{S}$ be the matrix containing all the vocalizations. SSC can be posed as LASSO:

$$\min_{\mathbf{Y}} \frac{1}{2}\|\mathbf{S} - \mathbf{S}\mathbf{Y}\|_2^2 + \lambda\|\mathbf{Y}\|_1 \quad s.t. \quad \mathbf{diag}(\mathbf{Y}) = 0, \tag{1}$$

where $\mathbf{diag}(\mathbf{Y})$ is a vector with the diagonal elements of $\mathbf{Y}$. Each column of $\mathbf{S}$ contains a vocalization and each column of $\mathbf{Y}$ contains the coefficients for the different USVs.

From $\boldsymbol{Y}$, a similarity matrix $\boldsymbol{A}$ is computed by:

$$\boldsymbol{A} = |\boldsymbol{Y}| + |\boldsymbol{Y}|^T. \tag{2}$$

The elements in $\boldsymbol{A}$ are the subspace similarities. To find the clusters, we apply spectral clustering[7] (with random walk Laplacian[8,9]) to $\boldsymbol{A}$.

[7] A. Y. Ng, M. I. Jordan, and Y. Weiss. "On spectral clustering: Analysis and an algorithm". In: *Advances in neural information processing systems*. 2002, pp. 849–856.

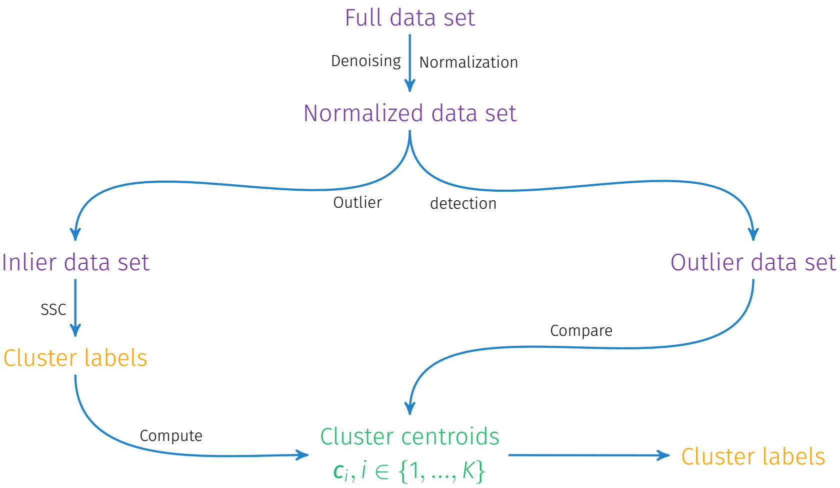
[8] U. Von Luxburg. "A tutorial on spectral clustering". In: *Statistics and computing* 17.4 (2007), pp. 395–416.

[9] J. Shi and J. Malik. "Normalized cuts and image segmentation". In: *Departmental Papers (CIS)* (2000), p. 107.

According to[10,11] the success of SSC depends mainly on three properties of the data:

1. Low affinity between subspaces
2. Enough samples for a certain subspace
3. The data not having too many outliers (where an outlier is defined as a data point that does not lie in any of the subspaces).

[10] M. Soltanolkotabi, E. Elhamifar, and E. J. Candés. "Robust subspace clustering". In: *Annals of Statistics* 42.2 (2013), pp. 669–699.
[11] M. Soltanolkotabi and E. J. Candés. "A geometric analysis of subspace clustering with outliers". In: *Annals of Statistics* 40.40 (2011), p. 2012.


Full data set
Denoising
Normalization
Normalized data set
Outlier
detection
Inlier data set
Outlier data set
SSC
Cluster labels
Compare
Compute
Cluster centroids
$c_i, i \in \{1, \ldots, K\}$
Cluster labels

- We use MUPET for USV detection and initial denoising.
- We normalize the vocalizations to a same size $F \times T$ using bi-cubic interpolations.
- We denoise by keeping only the parts of the spectrogram that are above a certain energy threshold.

We use the cosine similarity between vectors:

$$\cos(s, s') = \frac{s \cdot s'}{\|s\|\|s'\|}. \tag{3}$$

Detection Let s be a vocalization. We define s as an outlier if:

$$\max_{s \neq s'} \cos(s, s') < \tau, \tag{4}$$

where τ is a parameter to be set.

- **Inliers** We perform sparse subspace clustering in the inlier data set.
- **Outliers** We assign clusters to each one of the USVs in the outlier data set.
  - We define cluster centroids as as the mean of a cluster.
  - Then, we assign an outlier $\mathbf{s}_{out}$ into the most similar cluster k using:

$$k = \operatorname*{argmax}_j \cos(\mathbf{s}_{out}, \mathbf{c}_j). \tag{5}$$

- 40 records sampled at 250kHz emitted by the following mouse *strains*:
  - DBA/J2 (DBA)
  - C57B1/6J (C57)
- We used MUPET[12] to detect USVs (Figure 1)
- We detected and used approximately 9000 vocalizations.

[12] Segbroeck et al., "MUPET-Mouse Ultrasonic Profile ExTraction: A Signal Processing Tool for Rapid and Unsupervised Analysis of Ultrasonic Vocalizations".

USV sizes

- We set $F = T = 64$, such that USVs are represented by square images.
- We vectorize the images to vectors of length 64×64.

Outlier parameters

- DBA $\tau = 0.8$
- C57 $\tau = 0.7$

Number of clusters

- We set $K = 20, 40, 60$.

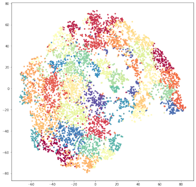
Qualitative

- We use t-SNE to check cluster separability
- We inspect the spectrograms to look at the qualities of the clustering

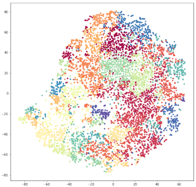
Quantitative

- We compute the harmonic mean of cosine distance between the centroids of the clusters:

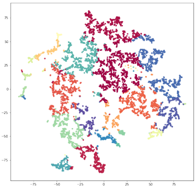
$$\bar{d}_{\cos}(C) = \left(\frac{1}{K(K-1)} \sum_{i \neq j} \frac{1}{1 - \cos(c_i, c_j)} \right)^{-1}.$$



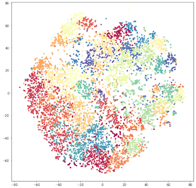
(a) DBA: k-means (baseline)



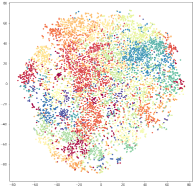
(b) DBA: CS + SC (baseline)



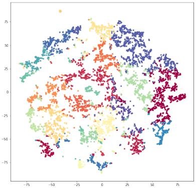
(c) DBA: LASSO-SSC



(d) C57: k-means (baseline)



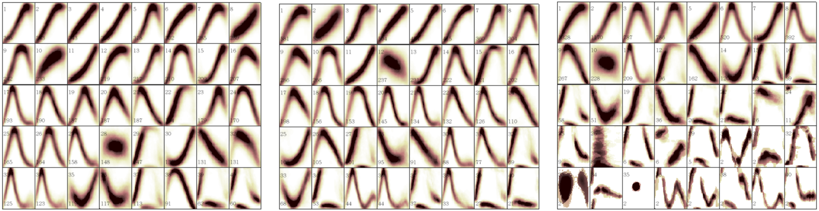
(e) C57: CS + SC (baseline)



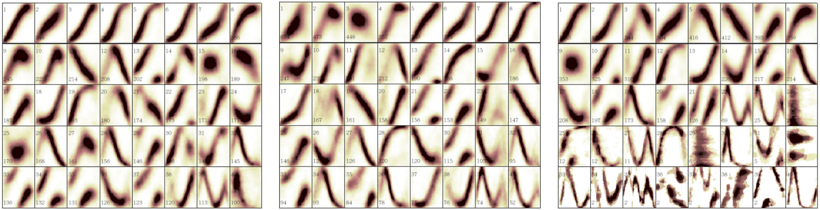
(f) C57: LASSO-SSC

Figure 3: t-SNE visualizations of the clusters in 2 dimensions. The embedding computed from the subspace similarity matrix ***A*** is able to better discriminate among different clusters compared to the cosine similarity between feature vectors.

Results: Cluster centroids



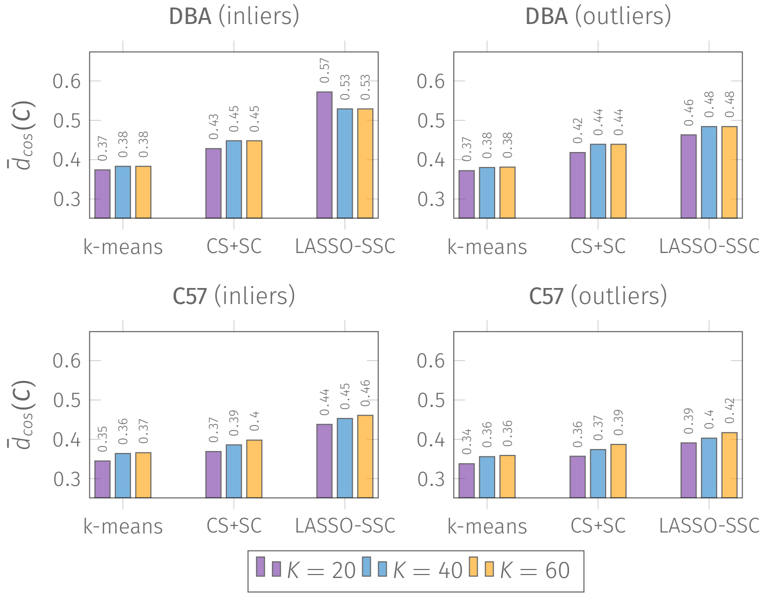
(a) **DBA**: k-means (baseline) (b) **DBA**: CS + SC (baseline) (c) **DBA**: LASSO- SSC (proposed)



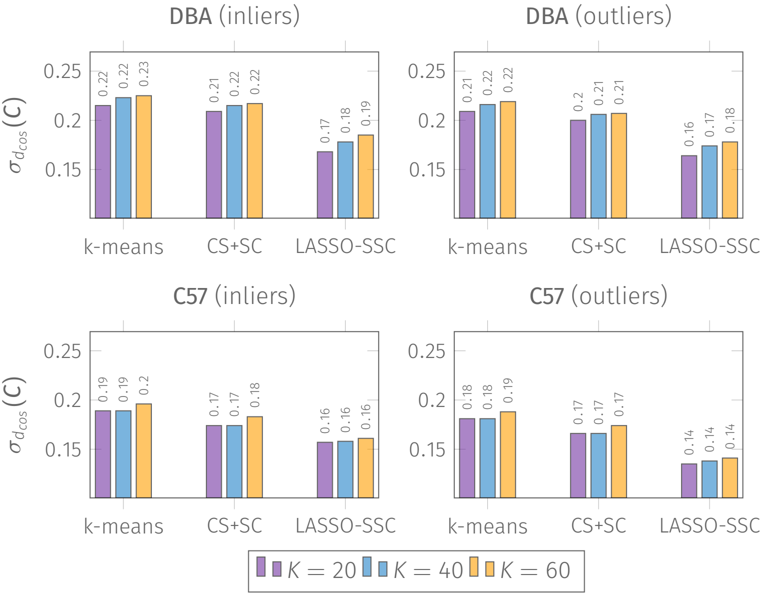
(d) **C57**: k-means (baseline) (e) **C57**: CS + SC (baseline) (f) **C57**: LASSO- SSC (proposed)

Figure 4: Centroids of clusters of inliers. In each figure, the top left window contains the cluster with most USVs, and decreases towards the right.

Results: Mean cosine similarities



Results: σ of cosine similarities



Conclusions

- We propose to learn a dictionary of mouse ultrasonic vocalizations using sparse subspace clustering.
- Our approach is able to find more diverse and cleaner cluster representations which translate to better dictionary representations.

We thank the USC Viterbi School of Engineering and the Feng Deng Foundation in Tsinghua for their support.

Code Availability & Questions

Code

https://github.com/usc-sail/mupet-subspace-clustering

Questions?

mundnich@usc.edu

References

J. Chabout, A. Sarkar, D. B. Dunson, and E. D. Jarvis. "Male mice song syntax depends on social contexts and influences female preferences". In: *Frontiers in behavioral neuroscience* 9 (2015), p. 76.

Z. D. Burkett, N. F. Day, O. Peñagarikano, D. H. Geschwind, et al. "VoICE: A semi-automated pipeline for standardizing vocal analysis across models". In: *Scientific Reports* 5 (2015), p. 10237.

M. V. Segbroeck, A. T. Knoll, P. Levitt, and S. Narayanan. "MUPET-Mouse Ultrasonic Profile ExTraction: A Signal Processing Tool for Rapid and Unsupervised Analysis of Ultrasonic Vocalizations". In: *Neuron* 94.3 (2017), pp. 465–485.

K. R. Coffey, R. G. Marx, and J. F. Neumaier. "DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations". In: *Neuropsychopharmacology* 44.5 (2019), p. 859.

E. Elhamifar and R. Vidal. "Sparse subspace clustering". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 2790–2797.

A. Y. Ng, M. I. Jordan, and Y. Weiss. "On spectral clustering: Analysis and an algorithm". In: *Advances in neural information processing systems*. 2002, pp. 849–856.

- U. Von Luxburg. "A tutorial on spectral clustering". In: *Statistics and computing* 17.4 (2007), pp. 395–416.
- J. Shi and J. Malik. "Normalized cuts and image segmentation". In: *Departmental Papers (CIS)* (2000), p. 107.
- M. Soltanolkotabi, E. Elhamifar, and E. J. Candés. "Robust subspace clustering". In: *Annals of Statistics* 42.2 (2013), pp. 669–699.
- M. Soltanolkotabi and E. J. Candés. "A geometric analysis of subspace clustering with outliers". In: *Annals of Statistics* 40.40 (2011), p. 2012.
- Freepik. *Hand drawn mice collection*. URL: *https://www.freepik.com/free-vector/hand-drawn-mice-collection_1588756.htm*.

This theme is a modified version of ***mtheme***, and can be found at:

https://github.com/usc-sail/presentation-template

Get the source of ***mtheme*** and the demo presentation from:

https://github.com/matze/mtheme