University of Zurich[UZH]

Determining the Optimal Number of Vowel Clusters in a Wide Range of Fundamental Frequencies using Unsupervised Learning

MSc Thesis Defense

Simon Frischknecht

September 16, 2021

Table of Contents

University of Zurich[UZH]

Introduction

Introduction

- For humans, the higher the fundamental frequency (fo), vowels become less intelligible
- Recognition of corner vowels (a, i, u) up to around 1 kHz fo
- Goal of this thesis: Use machine learning algorithms to see how they solve this task
- In particular: apply clustering algorithms to groups vowels

University of Zurich[UZH]

Speech Data and MFCC

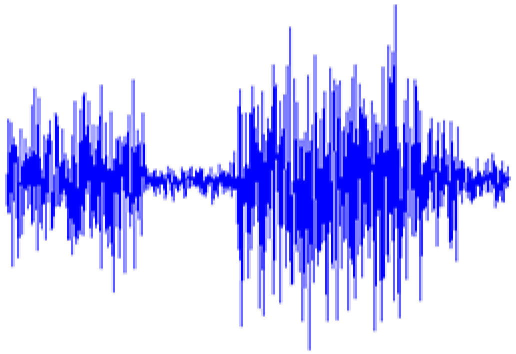
Vowel Speech Data

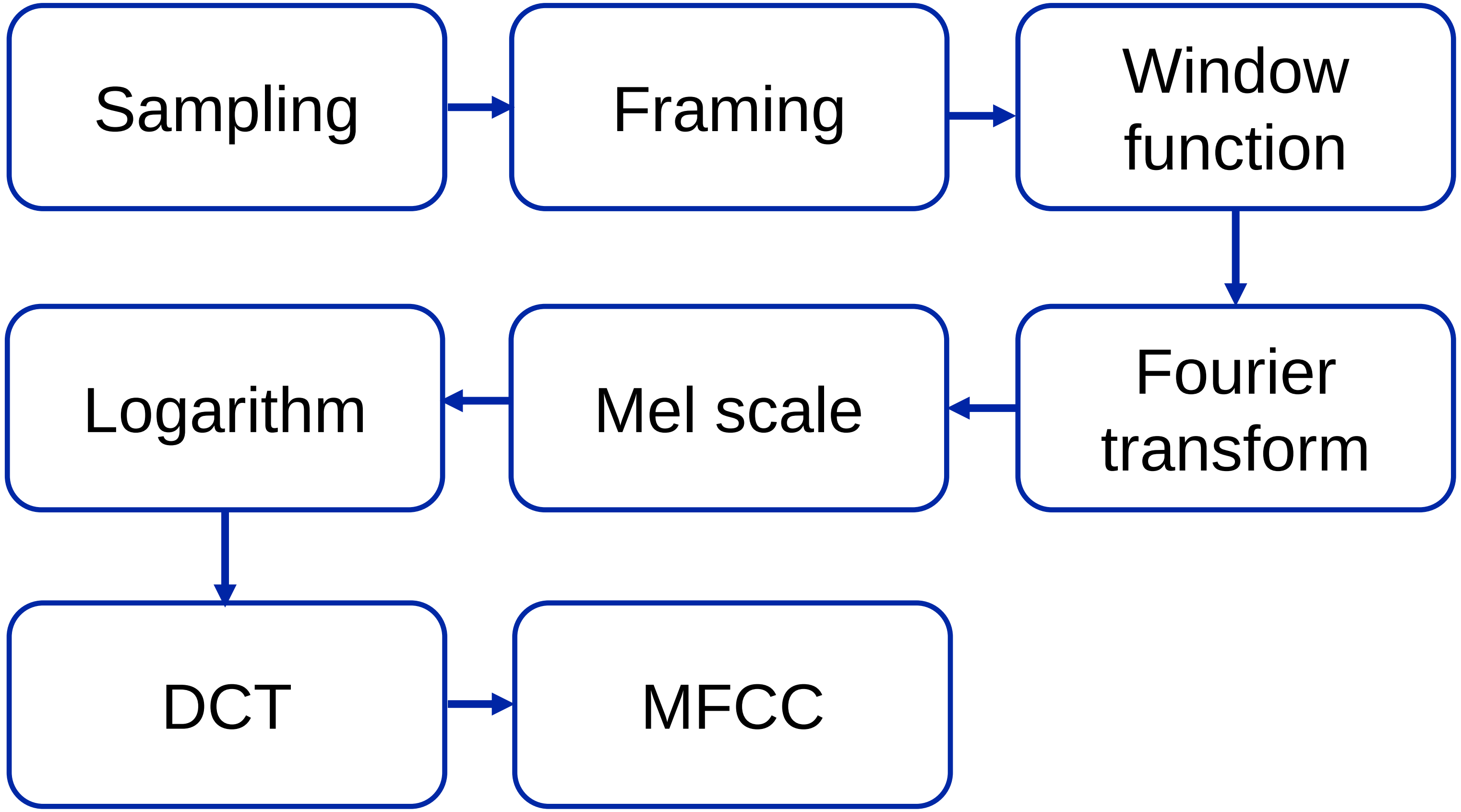
- Eight isolated Standard German vowels (a, e, i, o, u, ä, ö, ü)
- Recorded by four professional female actresses
- Data comes from larger corpus
- Wide range of fundamental frequencies (fo)
- Restrict to 10 fo levels 220, 330, 440, 523, 587, 659, 698, 784, 880, 988 Hz

Mel-Frequency Cepstral Coefficients (MFCCs)

- MFCCs are a numeric representation of speech signals
- It is a standard in many modern speech recognition systems
- Input for clustering algorithms
- It is based on several transformations:
  - Discrete Fourier Transform
  - Mel scale
  - Discrete Cosine Transform

MFCC Generation

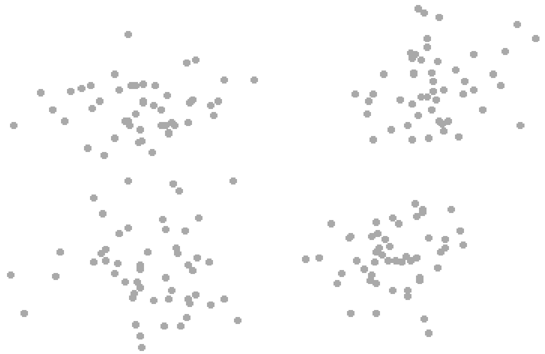




University of Zurich[UZH]

Clustering Methods

Clustering Methods



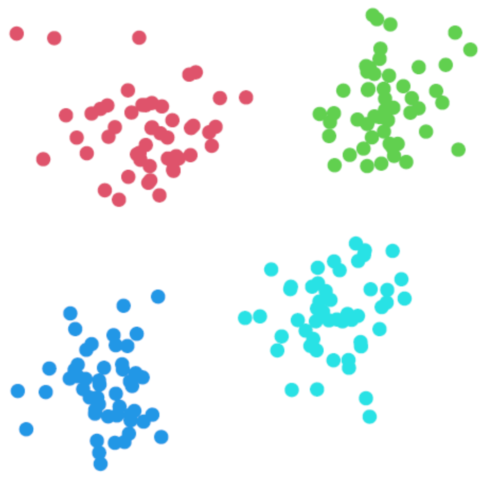
- Clustering methods are unsupervised learning algorithms
- No labels are used (in contrast to supervised learning)
- These algorithms have the goal of finding groups in a dataset:
  - “Observations in the same group should be as similar as possible, and observations in different groups should be as dissimilar from each other as possible.”
- Big challenge: how many clusters are there?

Clustering Methods

- List of used clustering algorithms:
  - k-Means
  - Gaussian Mixture Model
  - Hierarchical Clustering
  - Spectral Clustering
  - Mean Shift
  - DBSCAN
  - Affinity Propagation

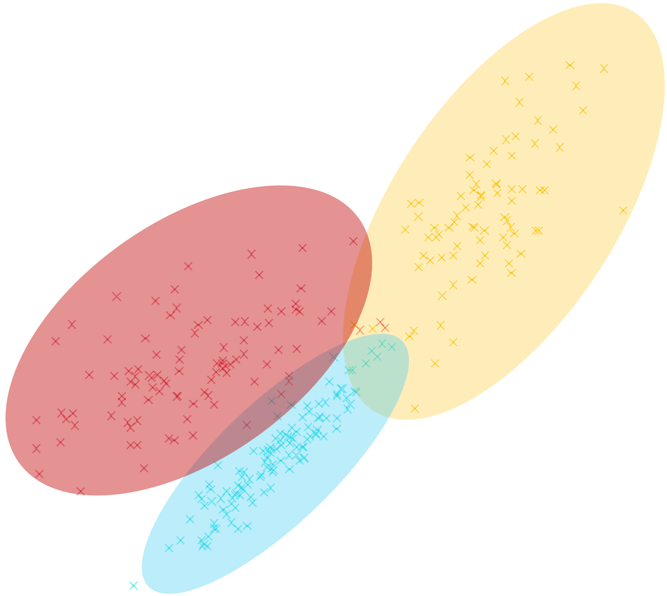
K-Means

- Simple iterative algorithm
- Computationally fast
- Solves non-convex optimization problem
- Finds different solutions based on initialization and number of iterations
- Needs to define the number of clusters k a priori



Gaussian Mixture Model (GMM)

- K-means can be seen as a special case of a GMM
- It is a probabilistic technique that assumes normally distributed clusters
- High flexibility, can easily overfit (Bias-variance tradeoff)
- Computationally expensive (estimate covariance matrices)
- Needs to define the number of clusters k a priori



Mean Shift

- Different to k-means and GMM
- Clustering method based on kernel density estimation
- Two parameters: kernel function and bandwidth
- Find local maxima of estimated density
- No need to define the number of clusters k

Cluster Validation Techniques

Clustering Validation Techniques

Often needed to find optimal number of clusters

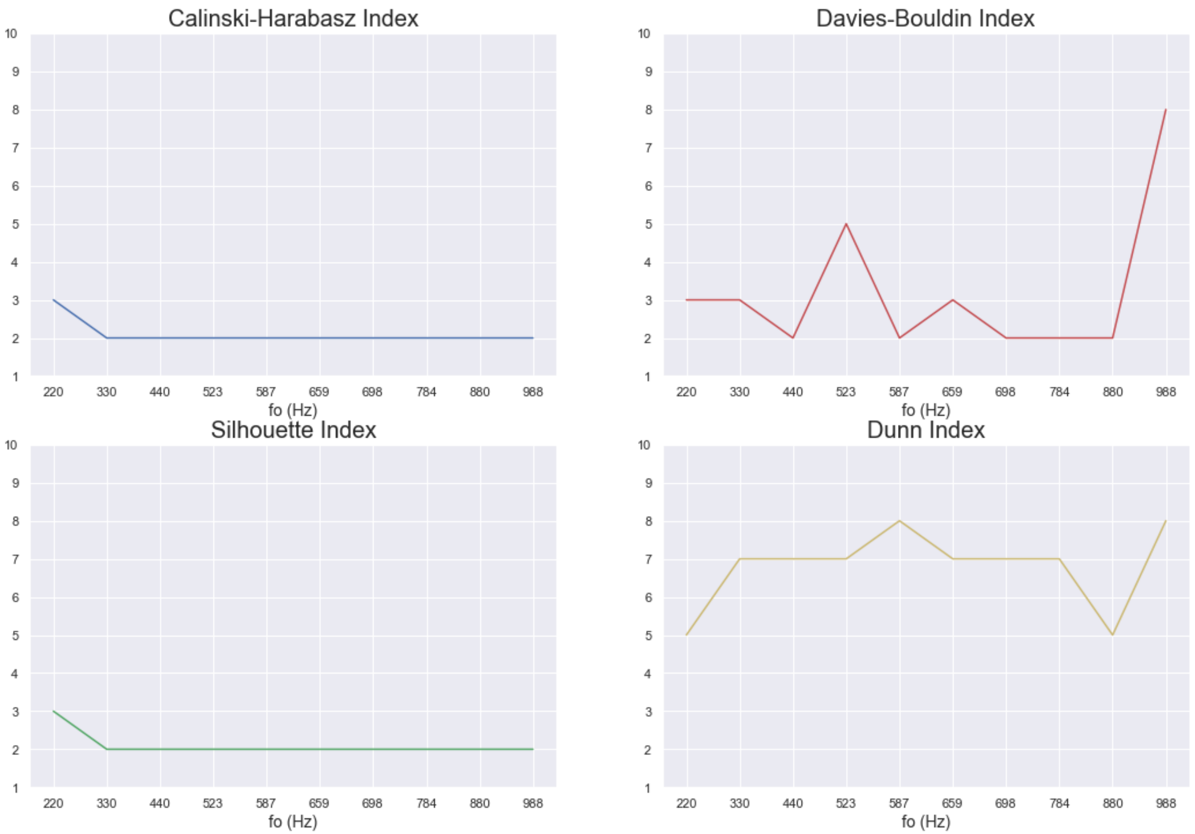
- **Internal Evaluation**
  - Do not make use of the true labels
  - More realistic for most real-world applications
  - Analyzed in thesis: Davies-Bouldin Index, Calinski-Harabasz Index, Dunn Index, Silhouette Index
- **External Evaluation**
  - Make use of the ground truth
  - Potential to find optimal results
  - Analyzed in thesis: Rand Index, Fowlkes-Mallows Index, Mutual Information, V-Measure

Some clustering algorithms estimate the optimal number of clusters by itself

Implementation and Evaluation

K-Means

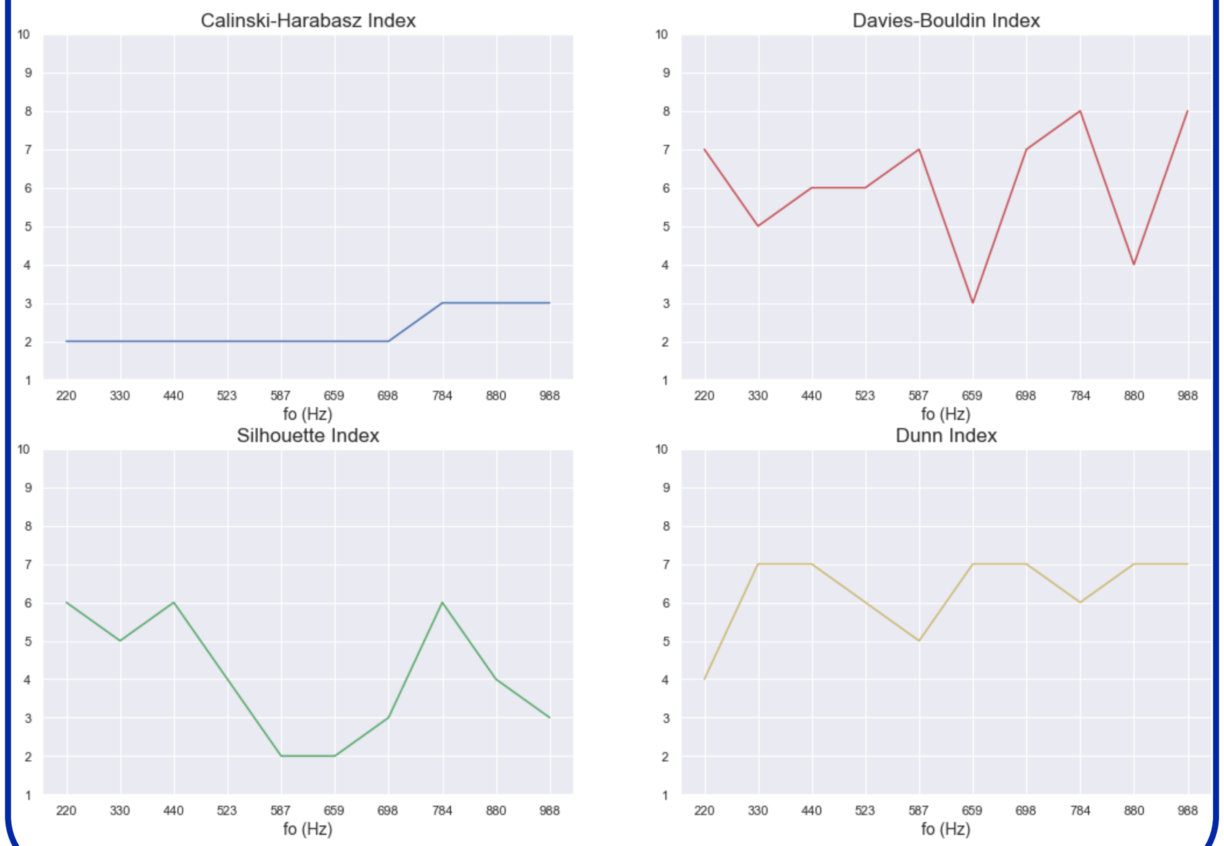
Internal Evaluation



External Evaluation

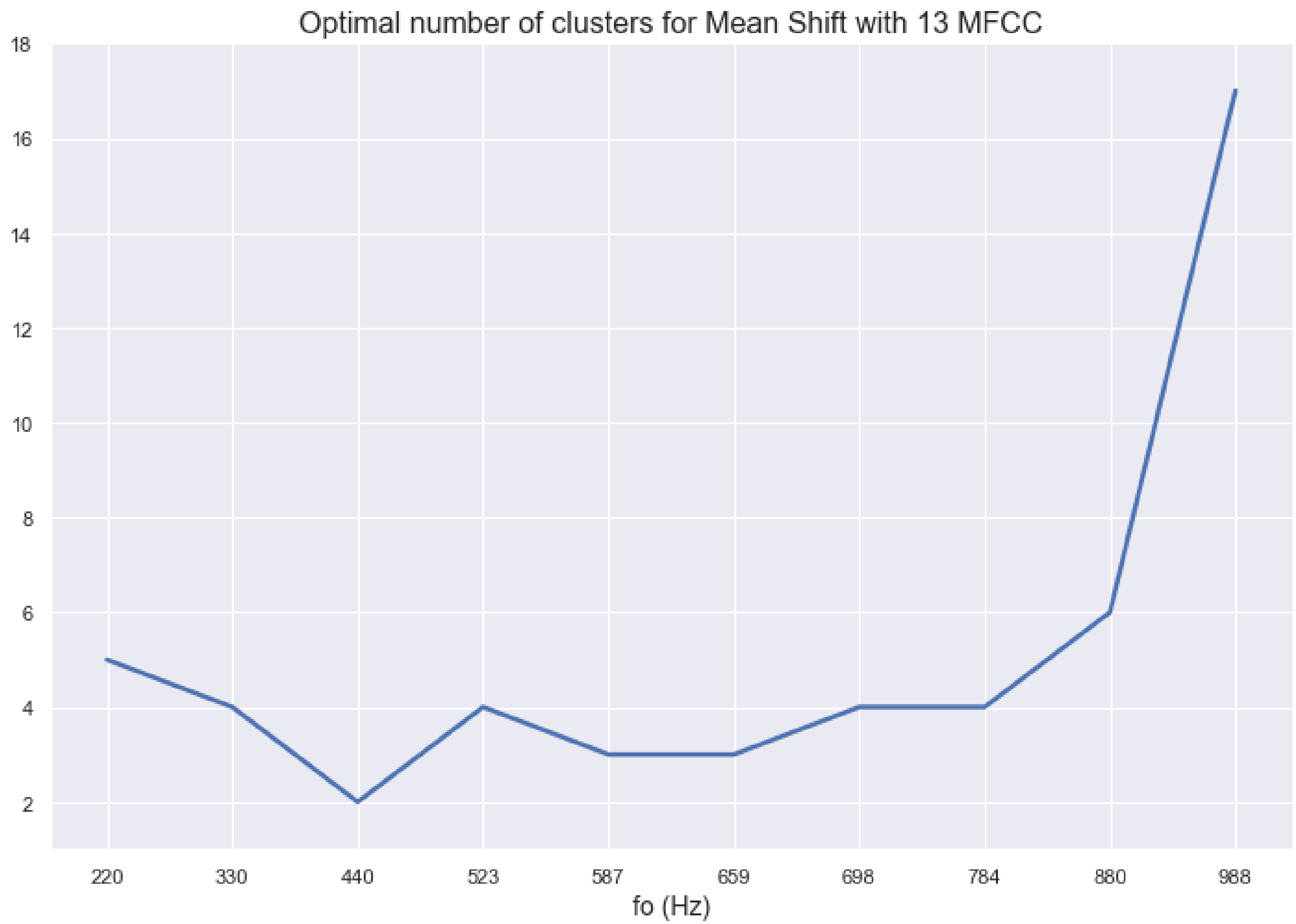
Gaussian Mixture Model

Internal Evaluation



External Evaluation

Mean Shift



Optimal number of clusters for Mean Shift with 13 MFCC
18
16
14
12
10
8
6
4
2
220
330
440
523
587
659
698
784
880
988
fo (Hz)

Conclusion

Conclusion

- Seven clustering methods analyzed
- No “best” clustering algorithm
- Clustering results depend on validation criterions
- Exception: mean shift, affinity propagation and DBSCAN estimate clusters on its own
- Internal evaluation criterions frequently underestimate the true number of clusters
- External validation methods partially suggest the correct number of groups
- fo has no significant influence on the results

It was fun to work on this topic!

Thank you for your attention!

University of Zurich[UZH]

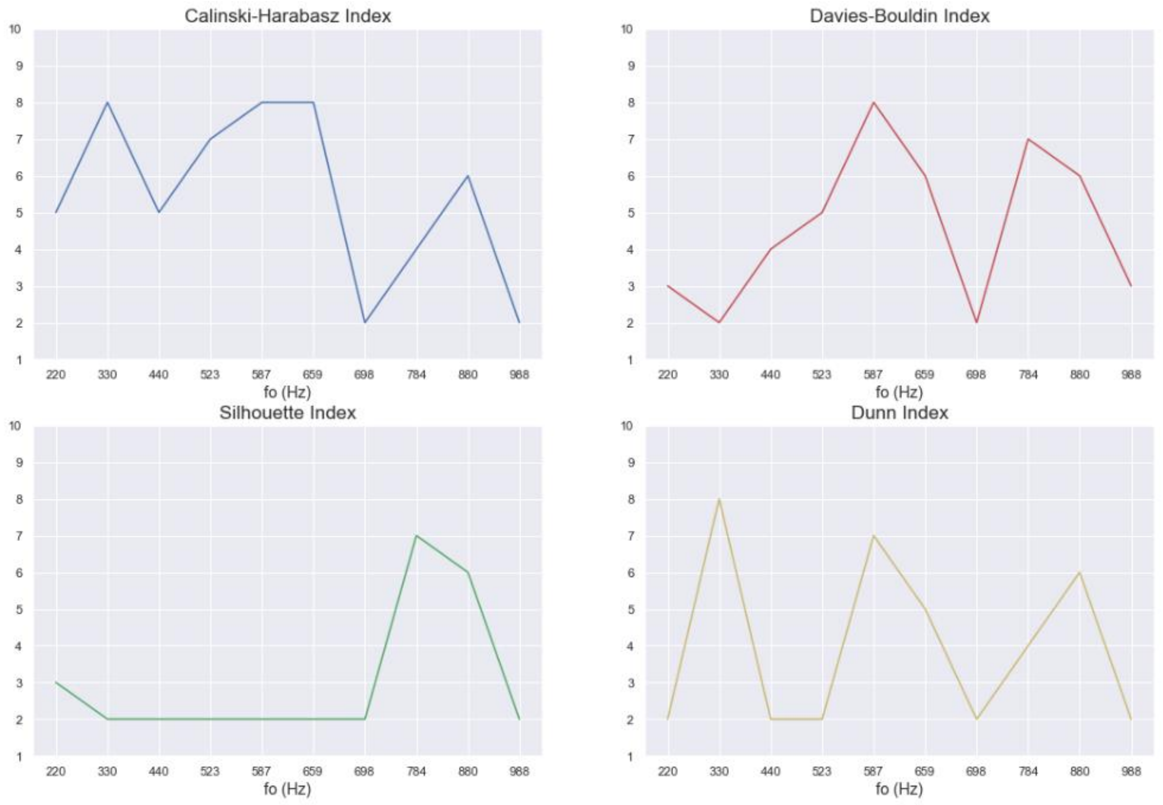
References

- Friedrichs, D., Maurer, D., & Dellwo, V. (2015). The phonological function of vowels is maintained at fundamental frequencies up to 880 Hz. The Journal of the Acoustical Society of America, 138(1), EL36-EL42.
- Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017). Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks. The Journal of the Acoustical Society of America, 142(2), 1025–1033.
- Kathiresan, T., Maurer, D., & Dellwo, V. (2019). Highly spectrally undersampled vowels can be classified by machines without supervision. The Journal of the Acoustical Society of America, 146(1), EL1-EL7.

University of Zurich[UZH]

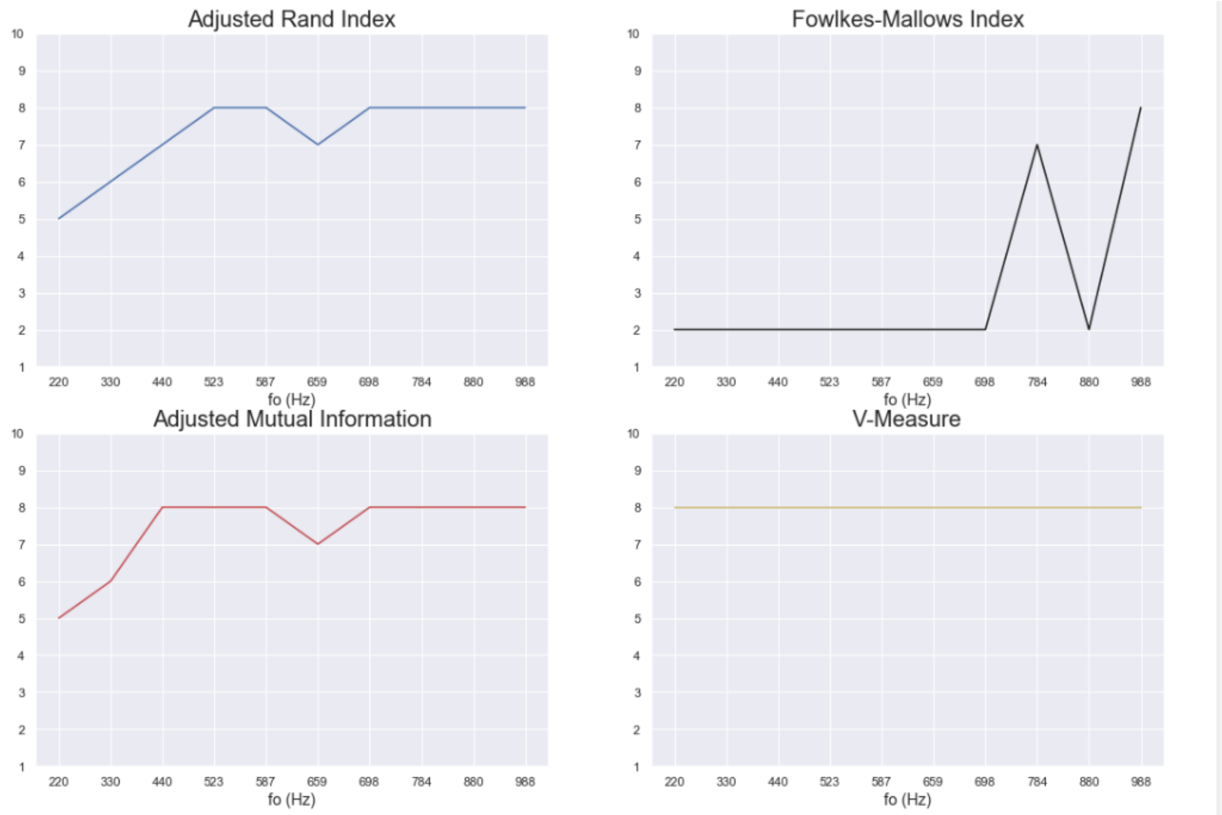
Appendix

Hierarchical Clustering (Internal)

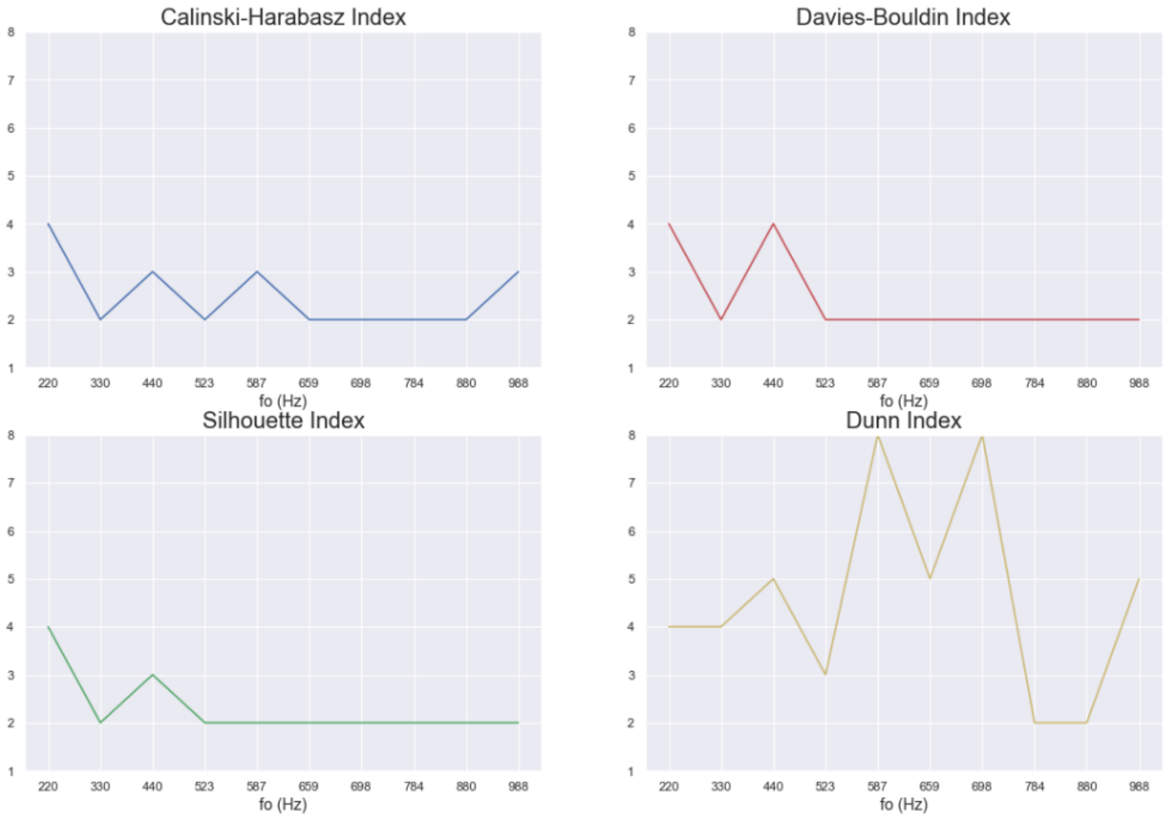


University of Zurich[UZH]

Hierarchical Clustering (Exernal)

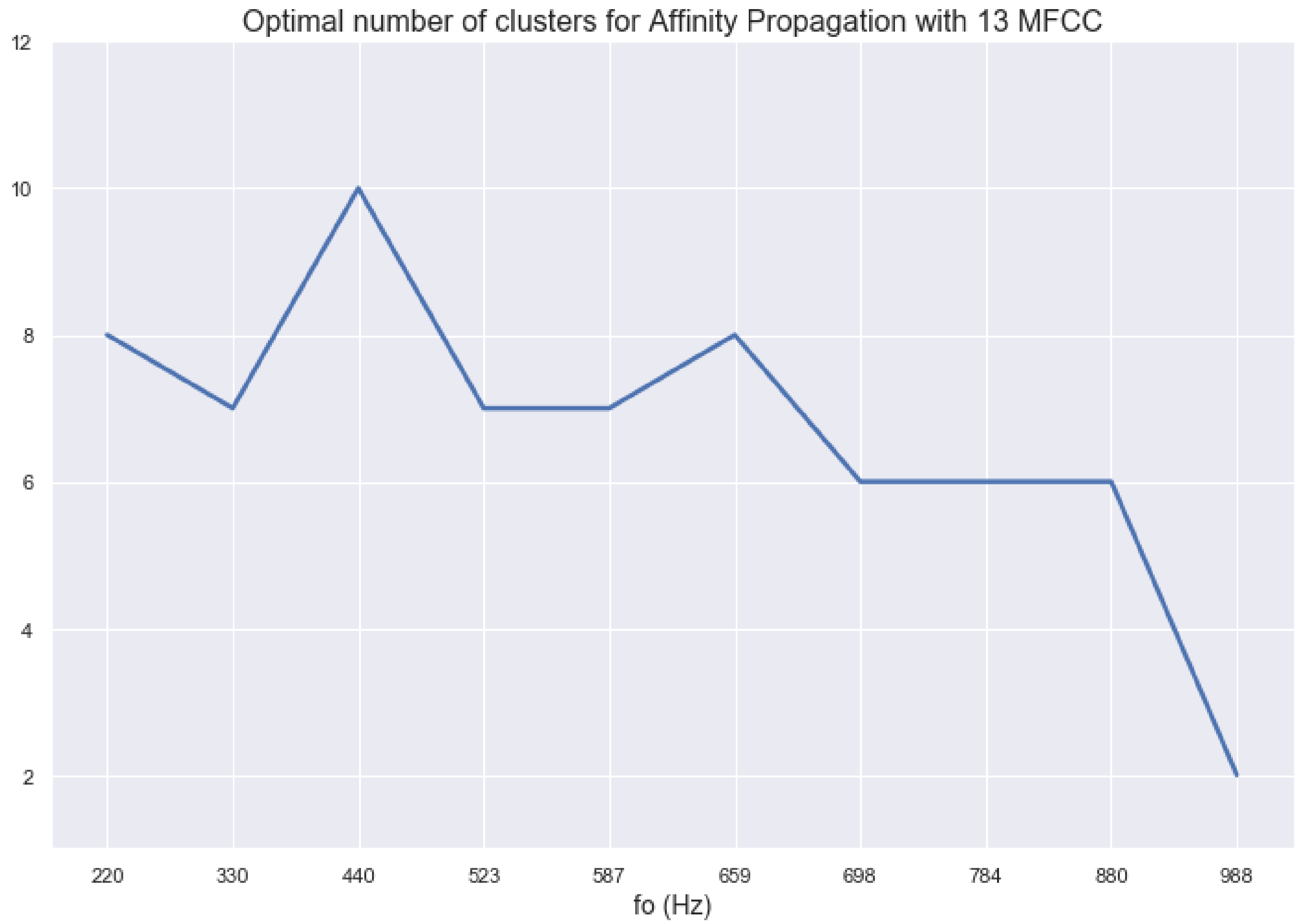


Spectral Clustering (Internal)

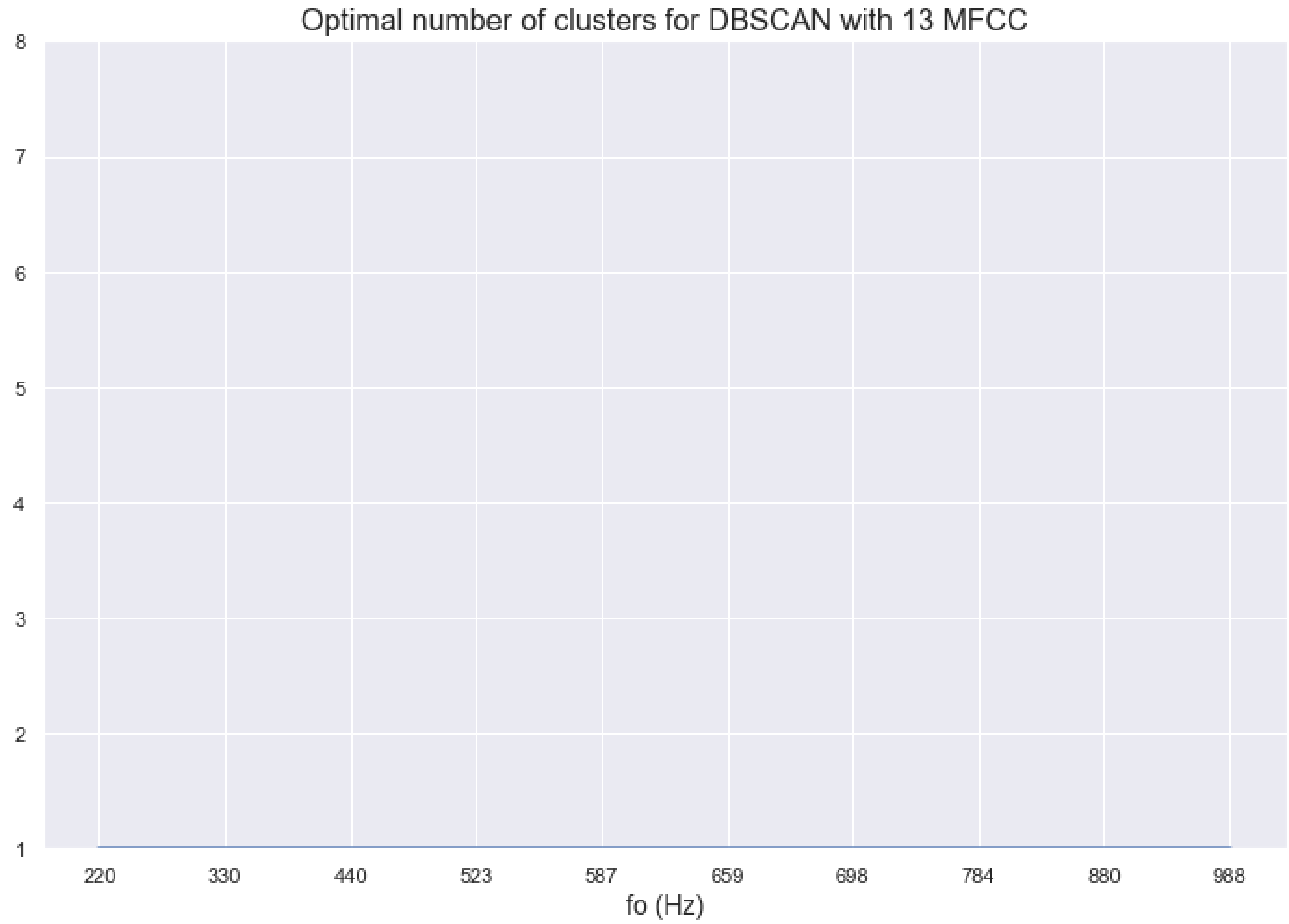


Spectral Clustering (External)

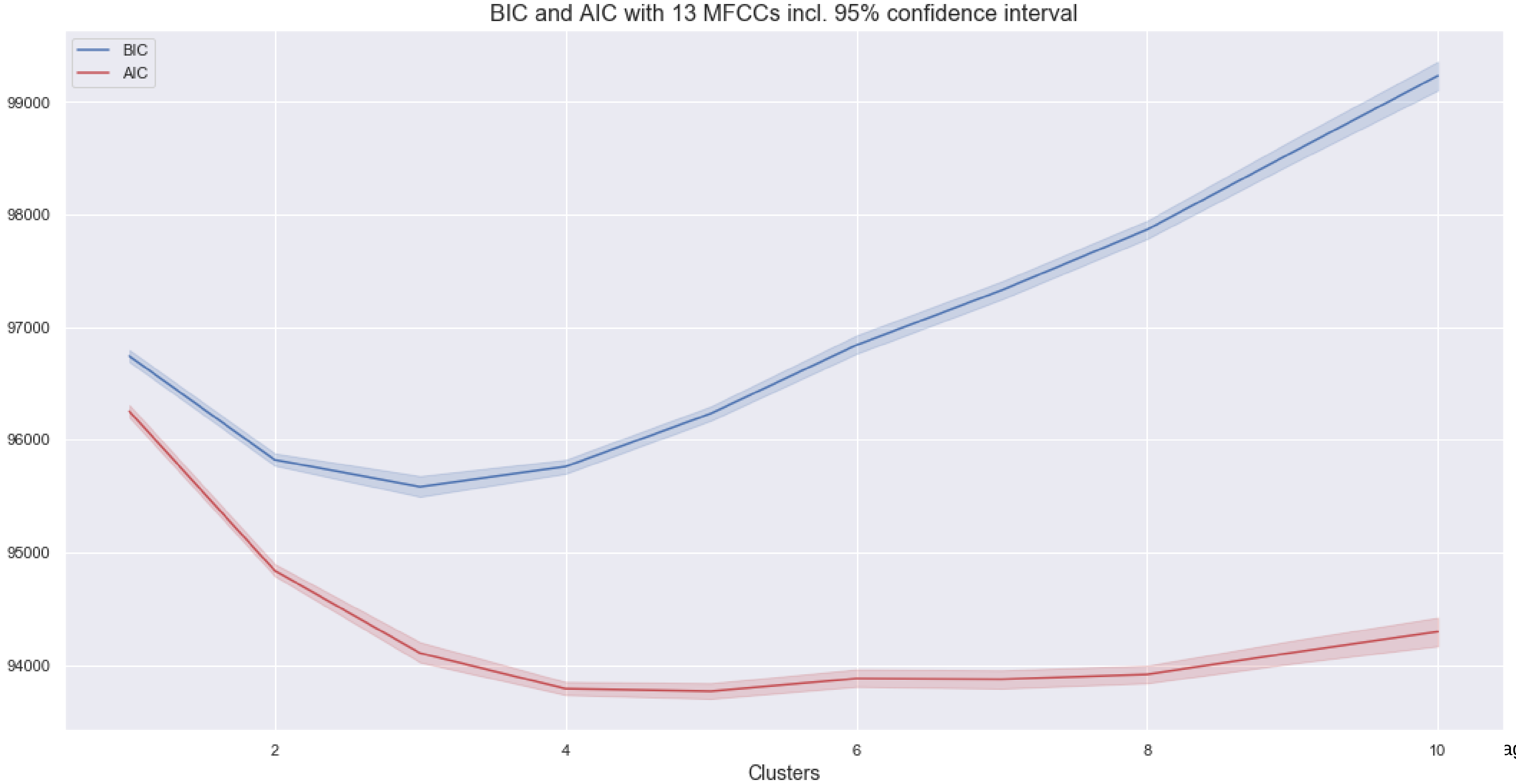
Affinity Propagation



DBSCAN (no validation criterions)



GMM (BIC, AIC)


BIC and AIC with 13 MFCCs incl. 95% confidence interval
BIC
AIC
99000
98000
97000
96000
95000
94000
2
4
6
8
10
Clusters