Cluster Analysis of Resonances

Clustering is an unsupervised machine learning technique that involves grouping similar data points based on a specific parameter, such as density or similarity. There are various models known in the literature, with K-means being a conventional one that generates a fixed number of clusters associated with a central point. The Markov Cluster Algorithm, introduced by Stijn Van Dongen (2008), is more appropriate for graphs/networks. Hierarchical clustering on the other hand is often used for the analysis of social network data and biological data analysis (Hexmoor, 2023; Yeturu, 2020). An important drawback of both K-means and hierarchical clustering for our application is that they do not automatically determine the number of clusters. Density-based algorithms, however, such as Mean-Shift, DBSCAN, and HDBSCAN, are more appropriate for this particular problem: resonances require a density-based approach (sudden changes of dense regions imply new musical objects), and they are also capable of automatically determining the amount of clusters based on the input data. The preview provided in Figure 8.1 highlights the benefits of using density-based algorithms for denoised resonance data. The resonances of the figures are extracted from the sound of a flute playing the note A4. The four lines at different frequencies successively represent the fundamental, first, second, and third harmonic overtone, and the cluster algorithms are solely performed on the onset and frequency features.

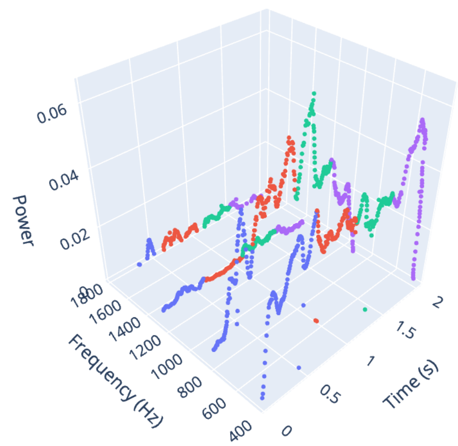

Figure 8.1: Resonances clustered by the K-means algorithm ($K = 4$) are represented with different colors.

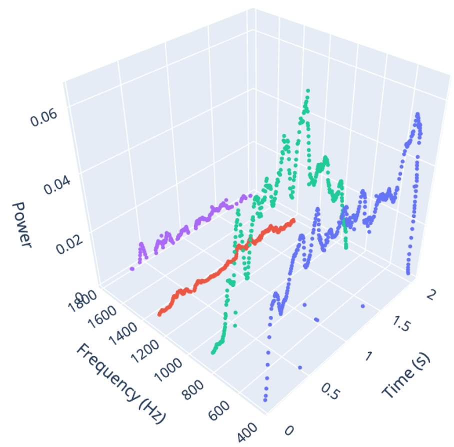

Figure 8.2: Resonances clustered by the DBSCAN algorithm ($\epsilon = 0.4$, $minPts = 4$) are represented with different colors. The labeling mimics how a human would draw circles around resonance groups to extract specific features, which exactly aligns with our desired outcome.

8.1 Density-based Cluster Algorithms

First, let us provide a concise overview of the three density-based cluster algorithms mentioned above. The iterative Mean-Shift algorithm moves each data point towards the mean of its respective region to form clusters. This is a centroid-based algorithm and works best for blob-shaped data. DBSCAN is capable of identifying outliers as noise, unlike the Mean-Shift method, and performs effectively on densely populated data with irregular shapes. HDBSCAN is a variation of DBSCAN introduced by Campello, Moulavi, and Sander (2013). In this algorithm, DBSCAN's principle of border points (see further) is abandoned, and only core points are considered as part of the cluster. Even though this method may be beneficial for handling noisy data, DBSCAN is nonetheless deemed to be the most appropriate clustering model for this problem, as it is better to implement custom filtering methods for noise reduction before performing clustering.