Gestalt Principle in Auditory Perception

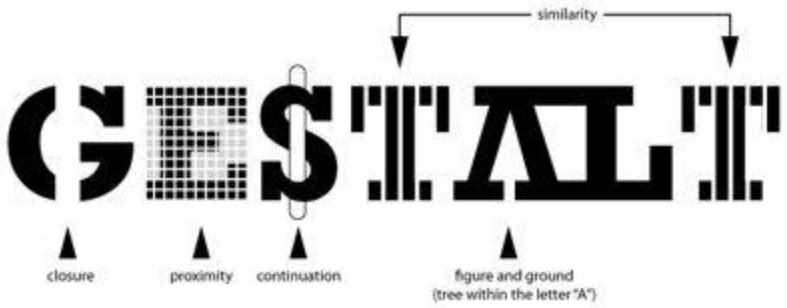
Yuhe Qin 2024.11.14

Introduction

Gestalt theory, widely employed in cognitive psychology, explains fundamental principles of visual perception.

In recent years, these principles have also been extensively applied to auditory perception.

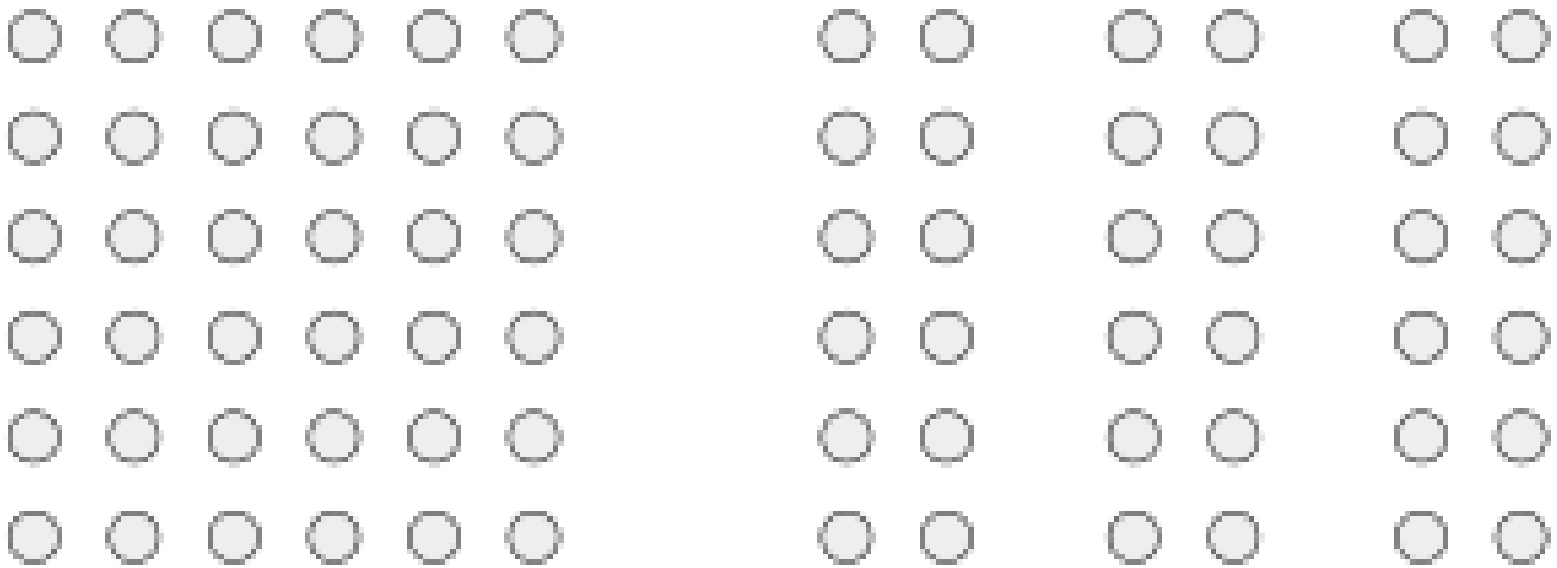
Inspired by a Gestalt computational model in visual perception*, we aim to *identify regular patterns in auditory perception* and *develop a corresponding computational model*.



**Jiayu Zhou et al., The gestalt computational model, arXiv preprint arXiv:2405.20583v1 (2024).*

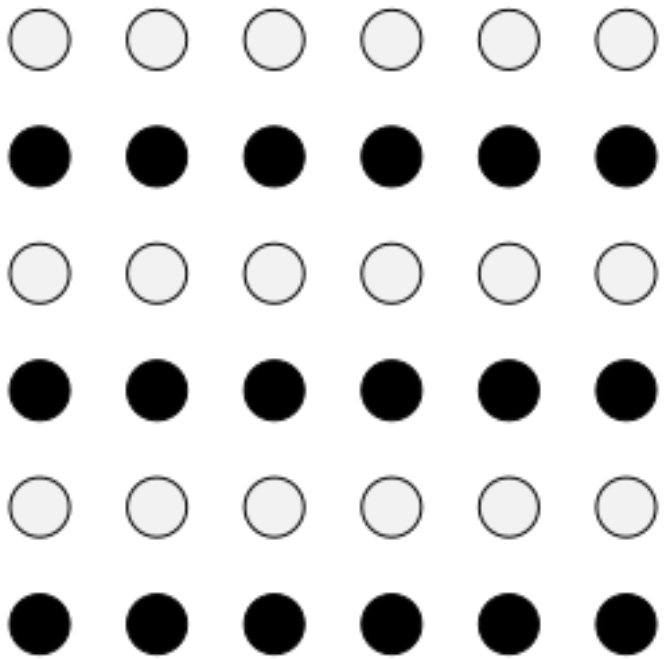
Gestalt Principle in Visual Perception

Proximity



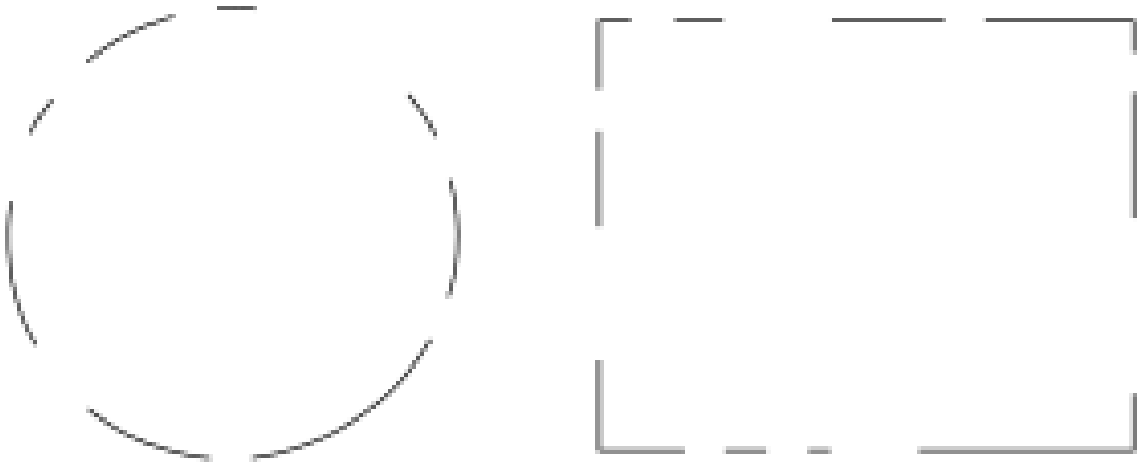
Gestalt Principle in Visual Perception

Similarity



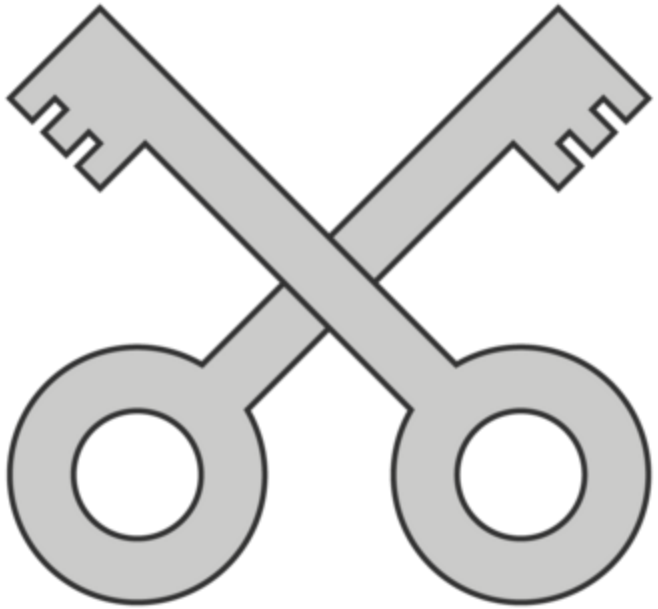
Gestalt Principle in Visual Perception

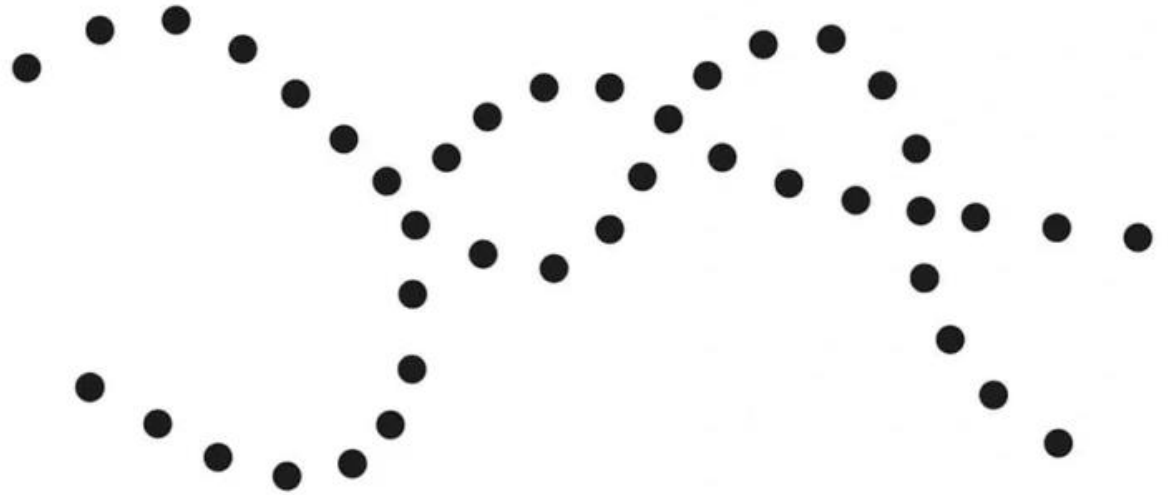
Closure



Gestalt Principle in Visual Perception

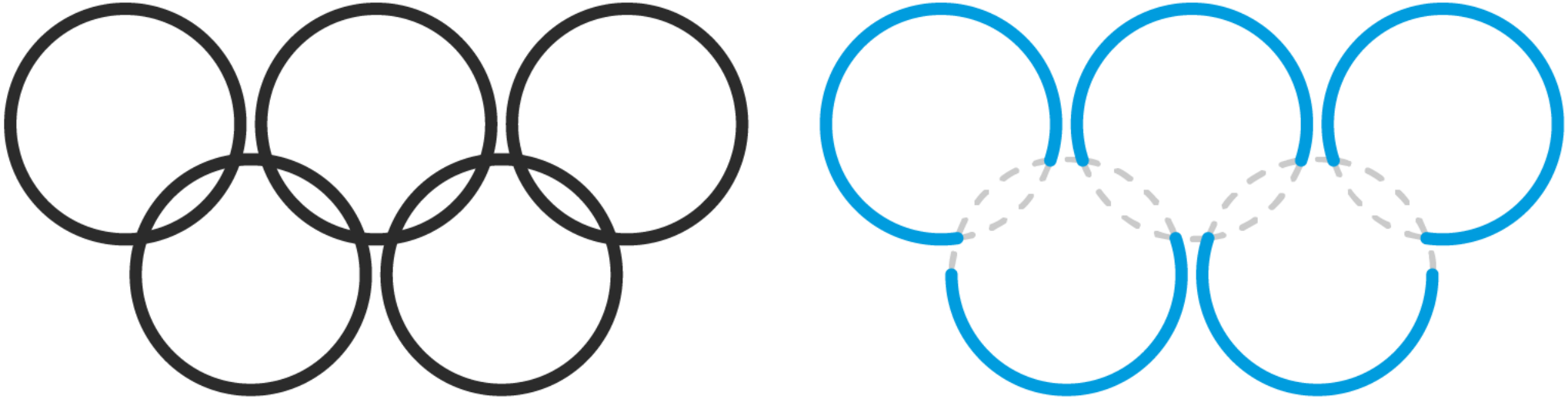
Good Continuation





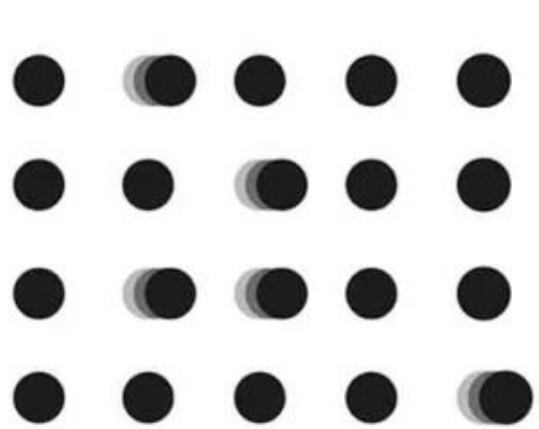
Gestalt Principle in Visual Perception

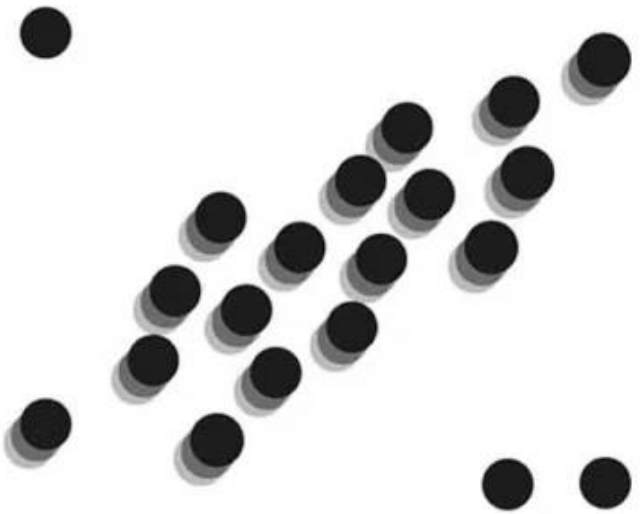
Simplicity



Gestalt Principle in Visual Perception

Common Fate





Gestalt Principle in Visual Perception

Rhythmic Law

Gestalt Principle in Auditory Perception

Factors Affecting Auditory Streaming

- ***Frequency*** (also pitch)
- ***Timbre:*** Analogy to color in vision
- ***Amplitude***

➢***Onsets and offsets***: Sudden changes in the intensity of the time signal

- ***Spatial cues***:

➢Interaural level differences (ILDs)

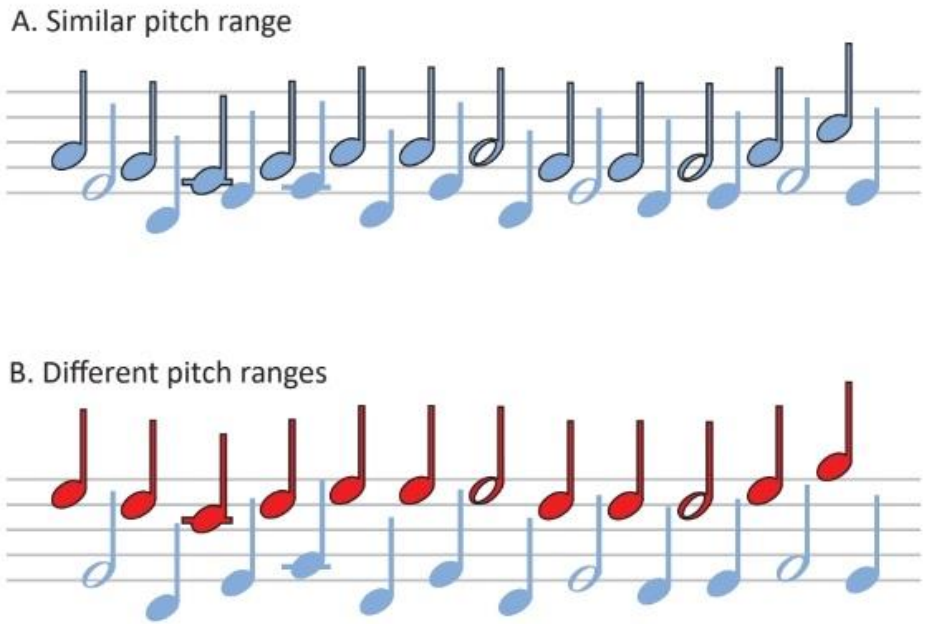
➢Interaural time differences (ITDs)

◆***Temporal modulation***: Similar frequencies or rhythmic patterns

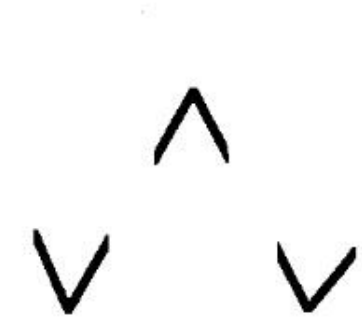
Gestalt Principle in Auditory Perception

Examples

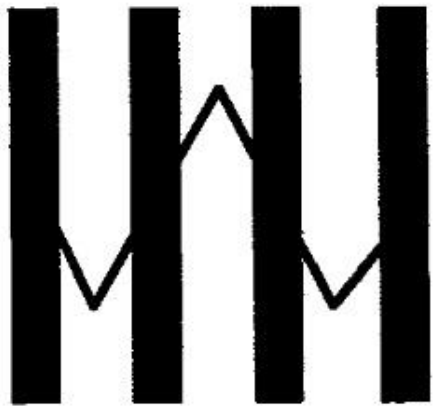
- ***Similarity in Pitch Range***



- ***Good Continuation***



NO BURSTS

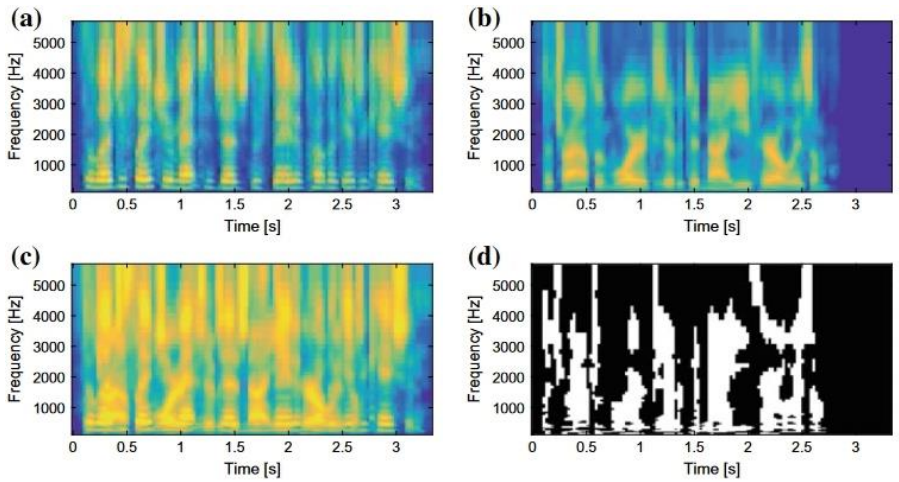


BURSTS

The Gestalt Computational Model in Auditory Perception

Research Plan

- ***Step1.*** Extract n features related to the auditory stream formation and perform time-delay embedding on each, embedding them into an m-dimensional space. These embeddings will then be combined to form a nm-dimension space.
- ***Step2.*** Compute the persistence diagram (PD) in this nm-dimensional space.



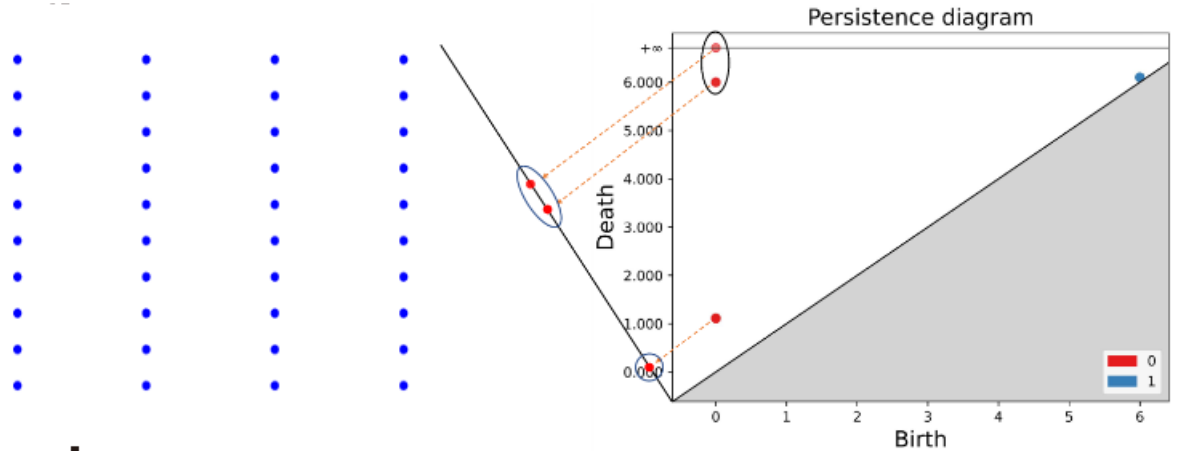
- Maybe by MFCC?
- Consider more ways to solve longer sequences…(maybe not by TDE)

➢ An idea: Maybe we can cut the long sequences into small sequences with gaps (like 50%)? Then apply TDE on the small sequence.

The Gestalt Computational Model in Auditory Perception

Research Plan

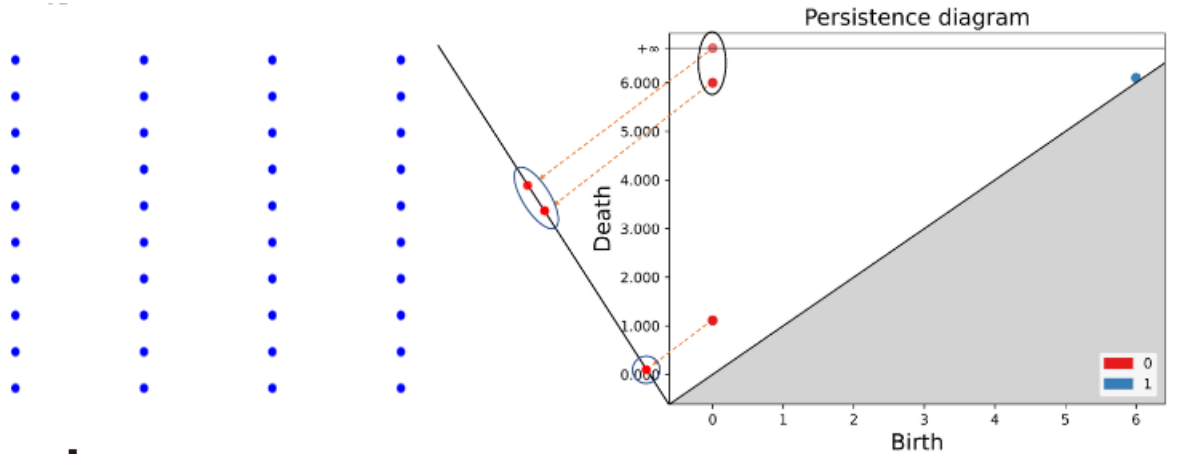
- ***Step3.*** The points in the PD are projected onto the line y = −x, and the projected points are clustered into two groups using a clustering method, such as the k-means algorithm (with k = 2).
- The group closer to the origin (0, 0) represents noise points, referred to as the "**noise class**", while the group farther from the origin represents significant points with greater persistence, referred to as the "**significant class**".

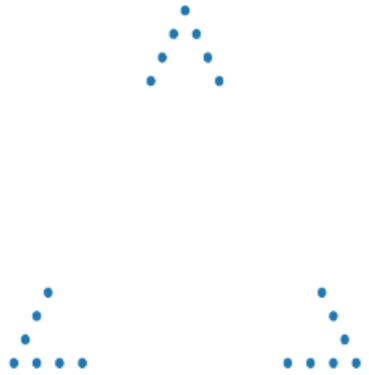


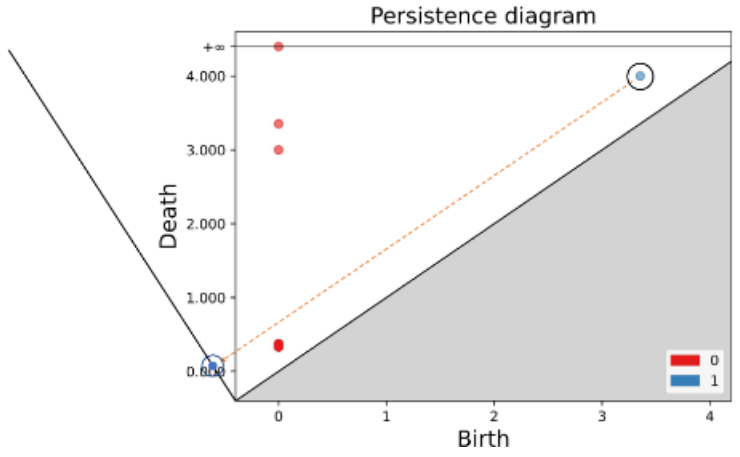
The Gestalt Computational Model in Auditory Perception

Research Plan

- ***Step4.*** A suitable threshold ε_g is determined.
- For 0-PD, ε_g should be larger than the greatest death time t_d of the noise points and smaller than the smallest death time of the significant points.
- For 1-PD, ε_g should be greater than the largest birth time t_b of the significant points and smaller than the smallest death time of the significant points. At the Vietoris-Rips (VR) complex with parameter ε_g , i.e., VR(ε_g), all significant topological features such as significant connected components or loops exist. By default, we set $\varepsilon = t_d$ for 0-PD or $\varepsilon = t_b$ for 1-PD.





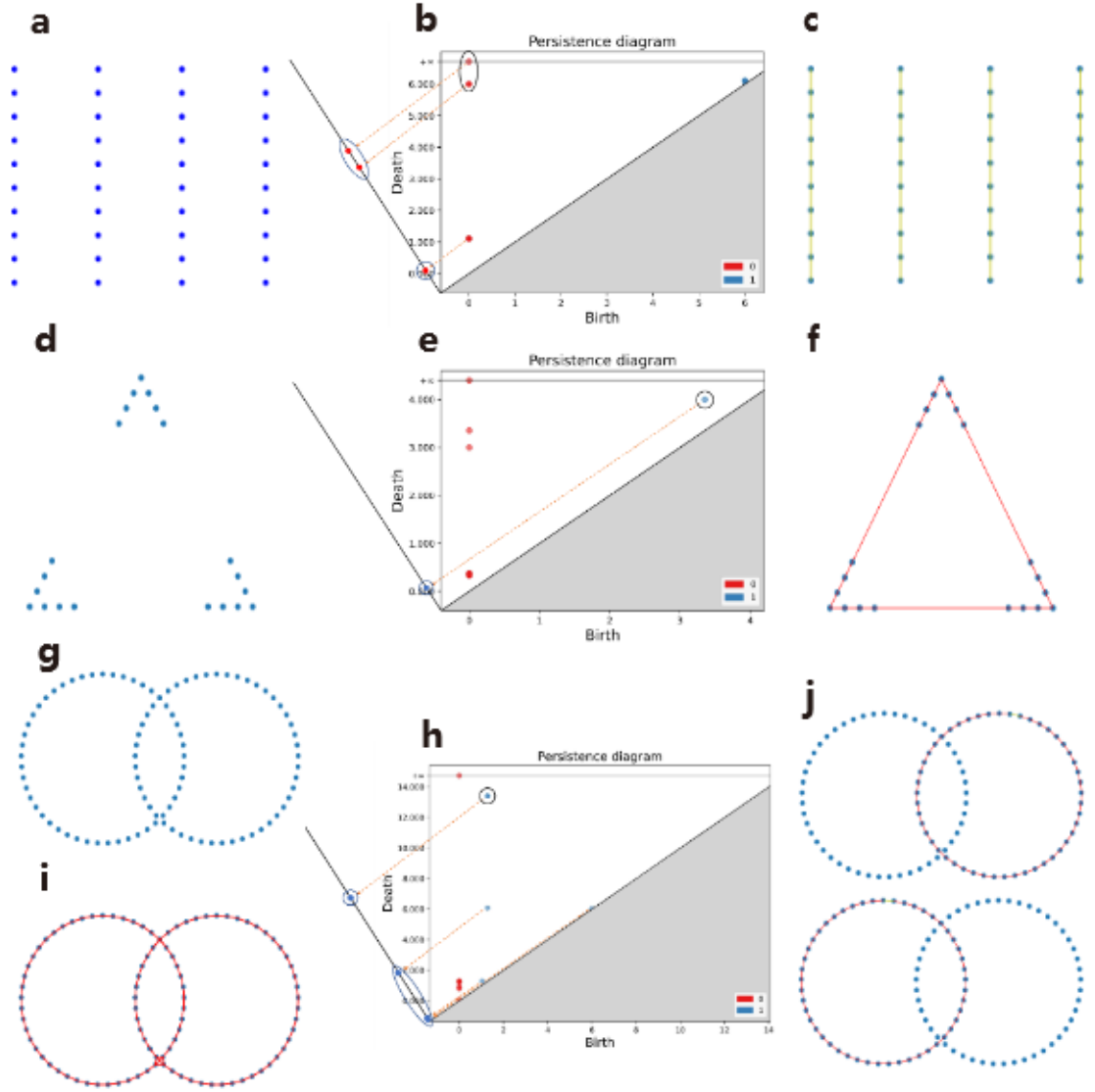


The Gestalt Computational Model in Auditory Perception

Research Plan

- ***Step5.*** The perceptual results are reconstructed from the VR complex with parameter ε_g , i.e., VR(ε_g).

We use Persistent Homology to process the point cloud formed by the characteristics of the sound, and then group the point cloud through the VR complex obtained to determine the formation of the auditory stream.
(Maybe use some other knowledge for processing?



New Page!

Restruction of the Gestalt Principle by VR Complex

Point Cluster & Threshold determination

Project the points in a _____ onto the line y = −x, and cluster the projected points into **noise class** and **significant class.**

A. 0-PD

Similarity, Proximity

ε_g: greater than the *largest death time* t_d *of the noise points* and less than the smallest death time of the significant points

B. 1-PD

Closure, Good Continuation, Simplicity

ε_g: greater than the *largest birth time* t_b *of the significant points* and less than the smallest death time of the significant points

Future Works

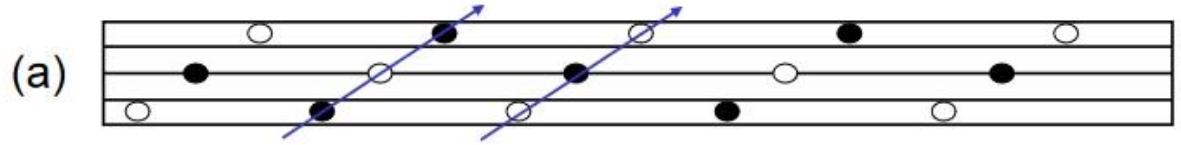
- Search for the mathematical expression of factors affecting auditory streaming and categorize them into those suitable for topological analysis and those that are not.
- Attempt to apply time-delay embedding and form persistence diagram.
- Explore potential synergies with the research conducted by Prof. Ximena Fernández* and Prof. Hongwei Lin to uncover additional possibilities.
- Attempt to study the emergence from factors to principles of auditory perception.

*Reise, Wojciech, Ximena Fernández, Maria Dominguez, Heather A. Harrington, and Mariano Beguerisse-Díaz. 2023. 'Topological Fingerprints for Audio Identification'.

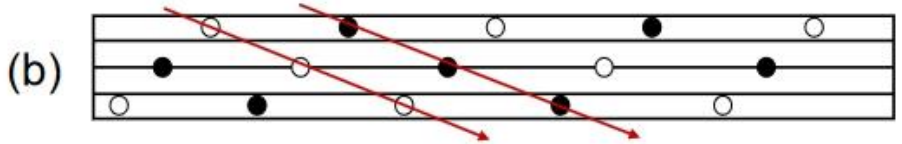
Problems

When constructing the model, I think it can be done by first building out the individual auditory streams and then assembling them together. Finally test whether our model can restore the separate auditory streams at the very beginning.

- In real-world scenarios, people's application of these principles is likely to be multifaceted, and further research is needed to understand how specific Gestalt rules are applied in practice.



If played slowly, listeners report an ascending sequence by different (alternating) instruments.



If played rapidly, listeners report two separate descending sequences, played by different instruments. This is an example of auditory stream segregation where grouping has produced two separate sequences.

Thank you!