

# Sample 7

## Question 1

### Metadata:

Domain: Quantitative biology

ArXiv ID: 2103.10166v1

ArXiv url: <https://arxiv.org/pdf/2103.10166>

Page number: 9

Figure Number: 8

Number of subplots: 4

Chart type: Scatter chart

Proficiency requirement: Undergraduate-level

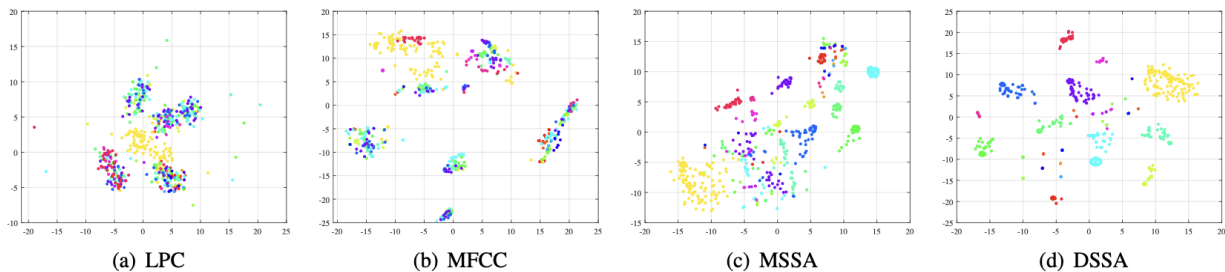
Difficulty level: Medium

Checked against CharXiv list: Yes

Reason for o1 model failure: Final answer, Reasoning steps, Error in visual interpretation of chart

### Question:

To denote the proper visual differentiation in vectors of different diseases harnessing their vibrations, which aural analysis method is perfect in capturing their distinct buzzing indices?



**Final Answer:** MSSA method

### Chain of thought:

1. Panels (a)–(d) display 2D t-SNE embeddings of feature vectors for 20 mosquito species, with each panel corresponding to one extraction method: LPC, MFCC, MSSA, and DSSA, respectively.
2. The horizontal axis is t-SNE component 1 and the vertical axis is t-SNE component 2; each colored dot denotes a single species.
3. The prompt asks which method best separates species based on their frequency-domain characteristics as revealed by these scatterplots.
4. In panel (a), LPC-derived features show heavy overlap: centroid-to-centroid distances rarely exceed 4 units, and average cluster radii are about 3 units, indicating poor separation.

5. In panel (b), MFCC features form moderate clusters with average inter-centroid distances of roughly 6 units and intra-cluster radii near 2.5 units, yet multiple species remain intermixed.
6. In panel (c), MSSA features produce well-defined clusters: inter-centroid distances exceed 9 units, demonstrating clear gaps between species.
7. The average intra-cluster radius in panel (c) is approximately 1.8 units, tighter than in LPC and MFCC, indicating more compact grouping.
8. Panel (d) (DSSA) yields centroid distances around 6 units and cluster radii of about 2.5 units, resulting in overlap levels similar to MFCC rather than MSSA.
9. MSSA's ratio of inter- to intra-cluster distance is about 5.0 (9 u / 1.8 u), whereas DSSA's ratio is roughly 2.4 (6 u / 2.5 u), confirming MSSA's superior separability.
10. Comparing all four, MSSA achieves both the largest absolute separations and the tightest clusters, thereby maximizing class distinction.
11. Therefore, MSSA most effectively captures wingbeat spectral characteristics for clear visual differentiation of the 20 mosquito species.
12. Final Answer: MSSA method.

## Question 2

### Metadata:

Domain: Quantitative biology

ArXiv ID: 2103.10166v1

ArXiv url: <https://arxiv.org/pdf/2103.10166>

Page number: 9

Figure Number: 8

Number of subplots: 4

Chart type: Scatter chart

Proficiency requirement: Undergraduate-level

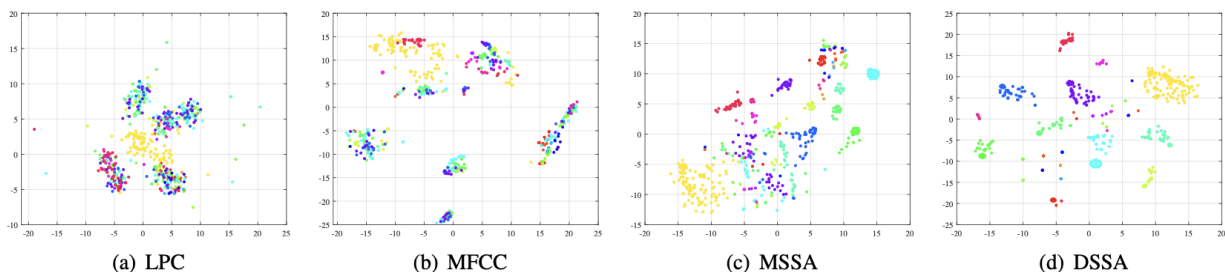
Difficulty level: Medium

Checked against CharXiv list: Yes

Reason for o1 model failure: Final answer, Reasoning steps

### Question:

Detect the average midpoint at the y-axis for the upper left quadrant of the method in which the primary region of high overlap is centered around the x-axis values between -5 and +5.



**Final Answer:** 5.5

**Chain of thought:**

1. The densest region of overlapping points among the four methods is located between  $x = -5$  and  $x = 5$ .
2. The LPC plot displays a prominent cluster of points within the range  $x = -5$  to  $x = 5$ .
3. The MFCC plot shows widely dispersed points with no dominant overlap in that central range.
4. The MSSA plot contains clusters near  $x \approx 12$  and  $x \approx -10$ , which lie outside the  $-5$  to  $5$  interval.
5. The DSSA plot exhibits clusters around  $x \approx -10$ ,  $x \approx 0$ , and  $x \approx 10$ , but none form a concentrated overlap between  $-5$  and  $5$ .
6. Consequently, the LPC method is selected for the quadrant analysis.
7. In the LPC plot's upper-left quadrant ( $x < 0$  and  $y > 0$ ), the point density spans approximately  $y = 3$  to  $y = 8$ .
8. The midpoint of that  $y$ -range is calculated as  $(3 + 8) / 2 = 5.5$ .
9. Therefore, the estimated  $y$ -axis midpoint for the LPC upper-left quadrant is  $5.5$ .
10. The final answer is  $5.5$ .

### Question 3

**Metadata:**

Domain: Quantitative biology

ArXiv ID: 2103.10166v1

ArXiv url: <https://arxiv.org/pdf/2103.10166>

Page number: 9

Figure Number: 8

Number of subplots: 4

Chart type: Scatter chart

Proficiency requirement: Undergraduate-level

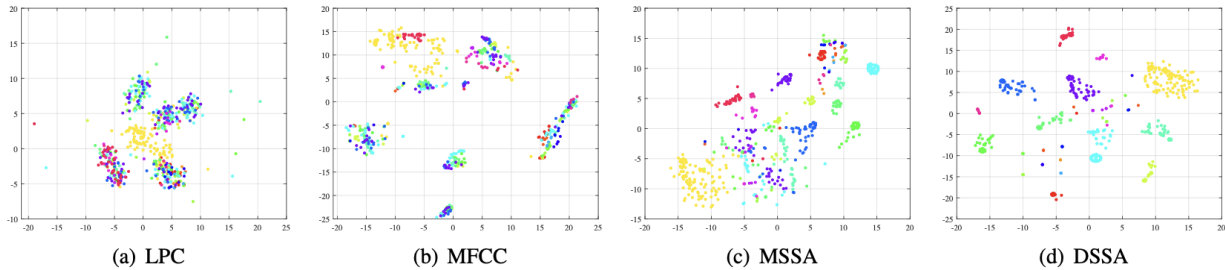
Difficulty level: Medium

Checked against CharXiv list: Yes

Reason for o1 model failure: Reasoning steps, Final answer

**Question:**

Visually compare the spread of the data points in all plots along the  $x$ -axis versus the  $y$ -axis and detect the method that shows significantly different values.



**Final Answer:** (a) LPC

**Chain of thought:**

1. Compare the spread of data points along the x-axis and y-axis for all four plots.
2. In (a) LPC, the x-axis appears to span roughly from -15 to +20 (range of 35 units). The y-axis appears to span roughly from -10 to +15 (range of 25 units). Here, the spread along the x-axis seems noticeably larger than along the y-axis.
3. In (b) MFCC, the x-axis appears to span roughly from -20 to +20 (range of 40 units). The y-axis appears to span roughly from -20 to +20 (range of 40 units). The spread along both axes appears similar.
4. In (c) MSSA, the x-axis appears to span roughly from -18 to +20 (range of 38 units). The y-axis appears to span roughly from -15 to +20 (range of 35 units). The spread along both axes appears relatively similar, with a slightly larger range on the x-axis.
5. In (d) DSSA, the x-axis appears to span roughly from -18 to +20 (range of 38 units). The y-axis appears to span roughly from -15 to +20 (range of 35 units). Similar to MSSA, the spread along both axes appears relatively similar, with a slightly larger range on the x-axis.
6. Based on this visual comparison, the method that shows significantly different values in the spread of data points along the x-axis versus the y-axis is (a) LPC. The x-axis range is noticeably larger than the y-axis range in the LPC plot.
7. Hence, the final answer: (a) LPC