

XỬ LÝ TÍN HIỆU SỐ

Nhận dạng nguyên âm không phụ thuộc người nói

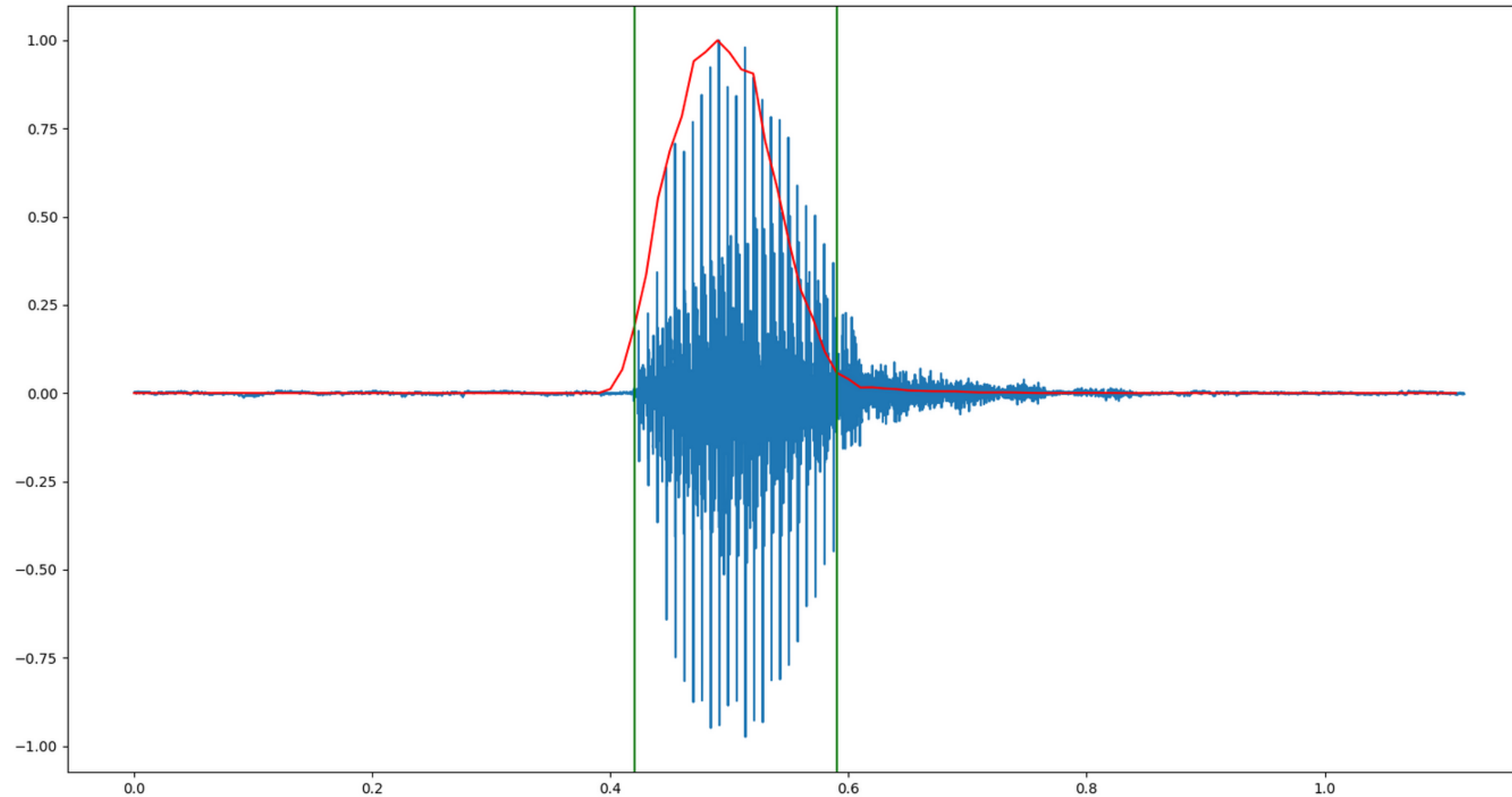
Giảng viên hướng dẫn: TS. Ninh Khánh Duy

Sinh viên thực hiện: Huỳnh Phú Quý
Hoàng Quang Hùng

Phân công:

- Hoàng Quang Hùng:
 - + Chọn thuật toán phân đoạn nguyên âm khoảng lặng
 - + So khớp và nhận dạng nguyên âm
 - + Khảo sát các cặp N và K để tìm ra cặp tham số cho độ chính xác cao nhất
- Huỳnh Phú Quý:
 - + Trích xuất vector đặc trưng của các nguyên âm (FFT, MFCC)
 - + Tạo CSDL các vector đặc trưng của các nguyên âm
 - + So sánh hiệu quả 2 loại vector đặc trưng FFT và MFCC

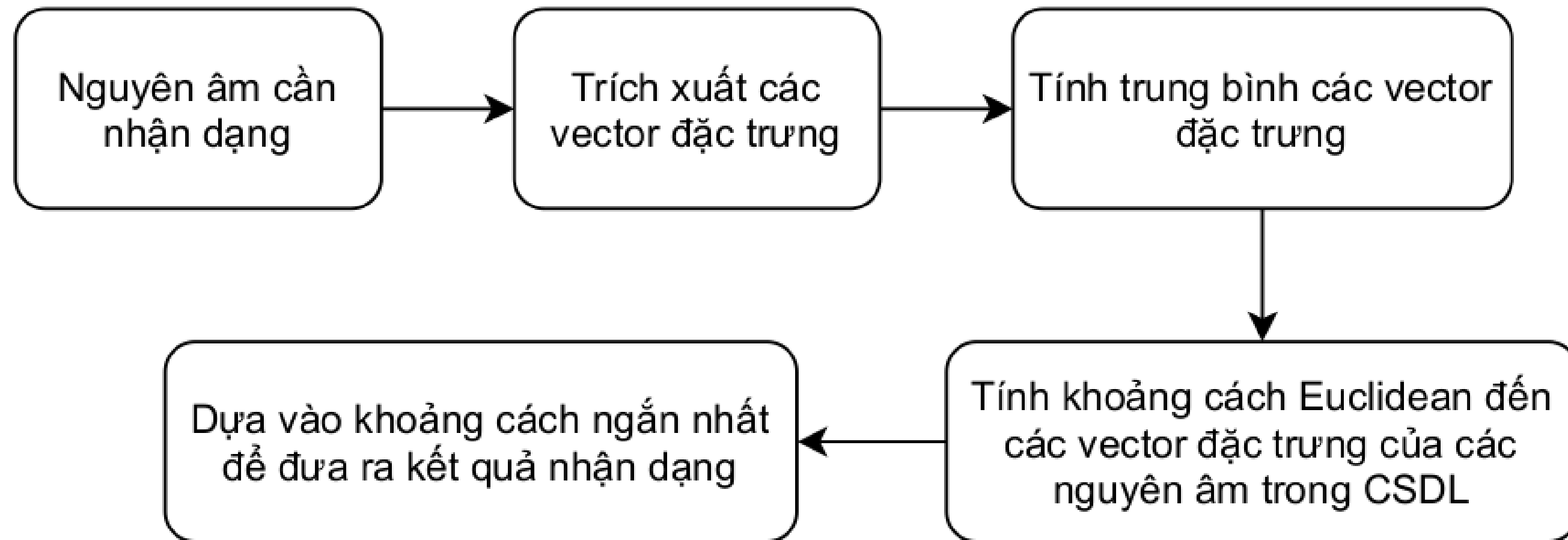
Chọn thuật toán phân đoạn nguyên âm khoảng lặng



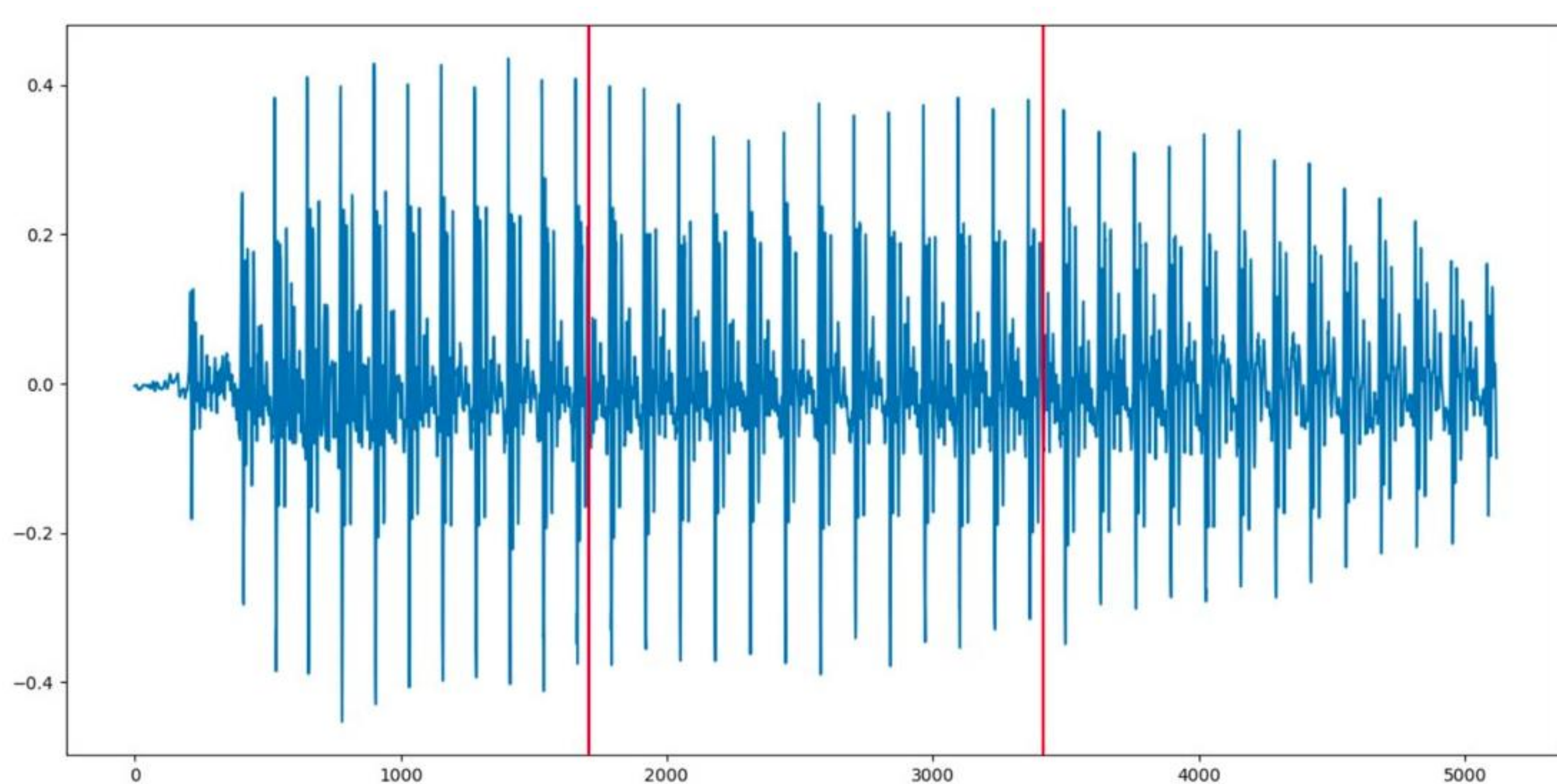
- Sử dụng thuật toán STE
- Sử dụng thuật toán Binary Search

threshold = 0.06827

So khớp và nhận dạng nguyên âm

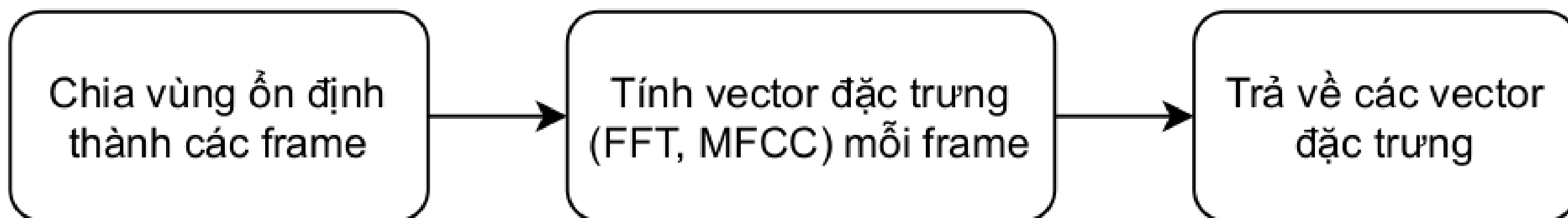


Trích xuất các vector đặc trưng của các nguyên âm

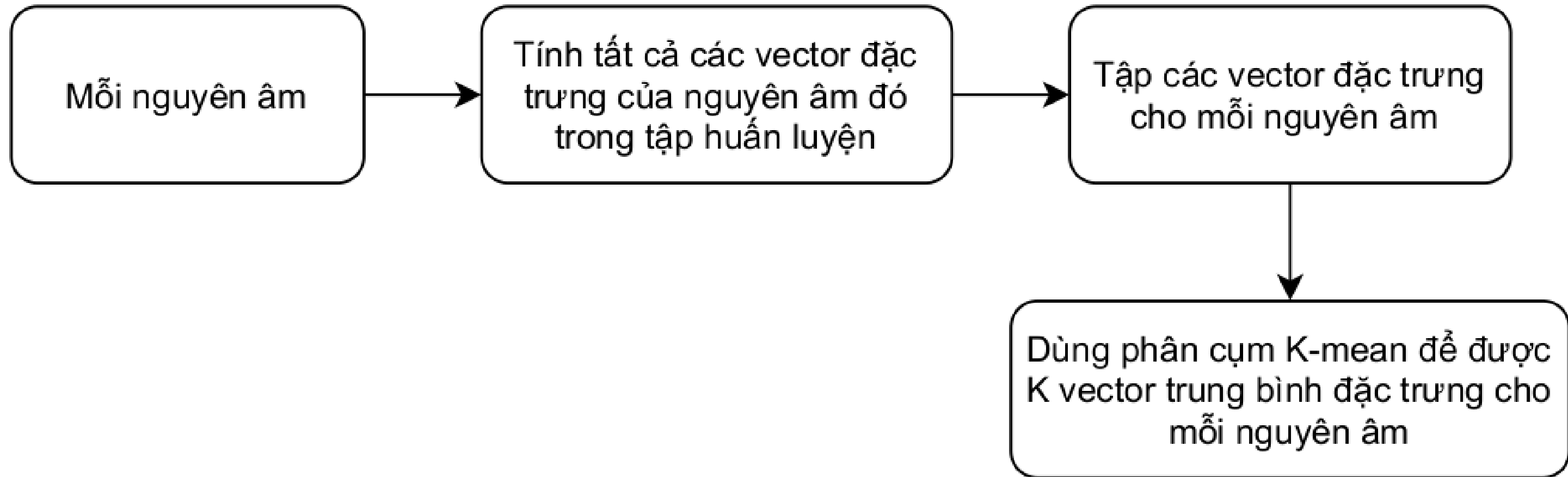


Đoạn nguyên âm a của 01MDA

Trích xuất các vector đặc trưng của các nguyên âm



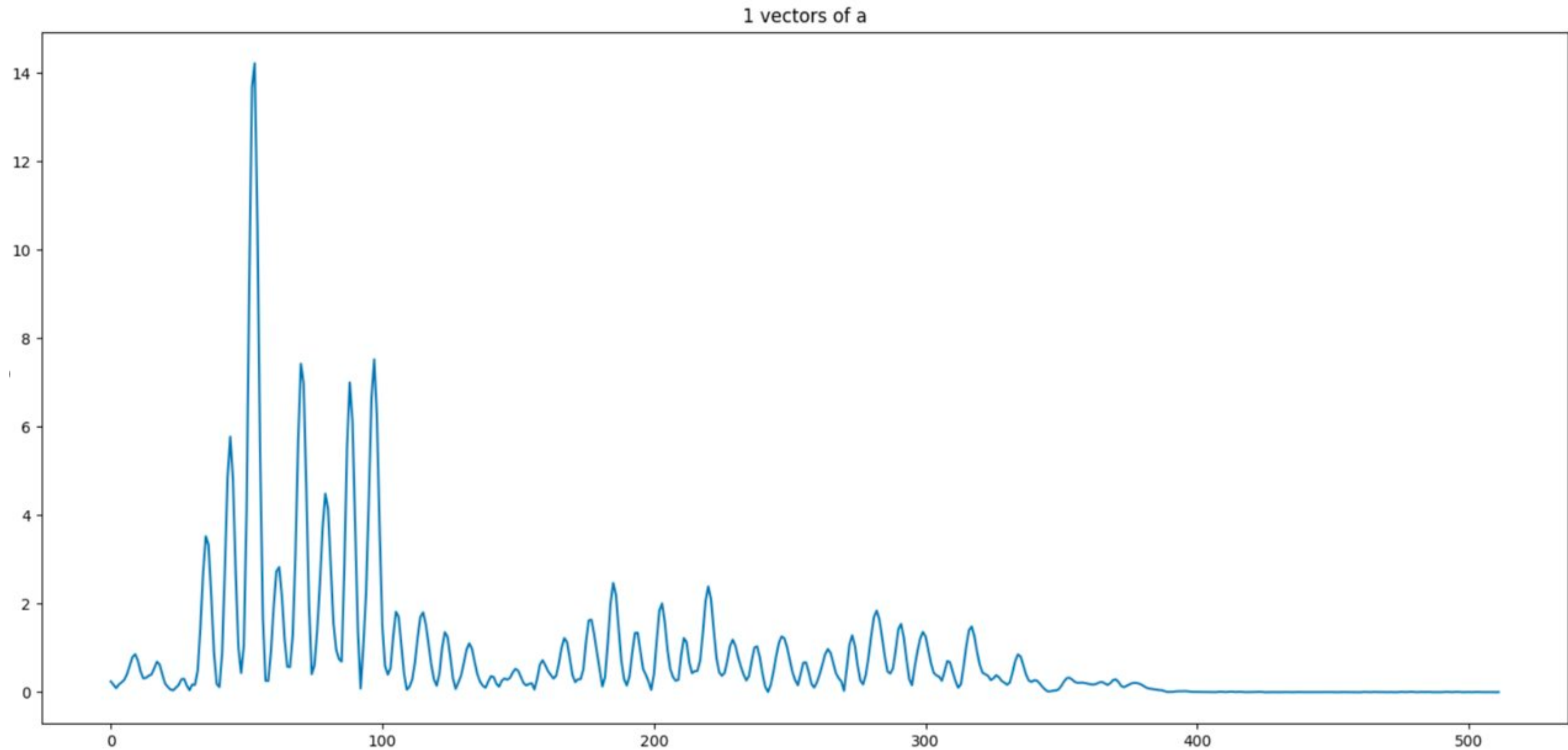
Tạo CSDL các vector đặc trưng của các nguyên âm



So sánh 2 loại vector đặc trưng FFT và MFCC

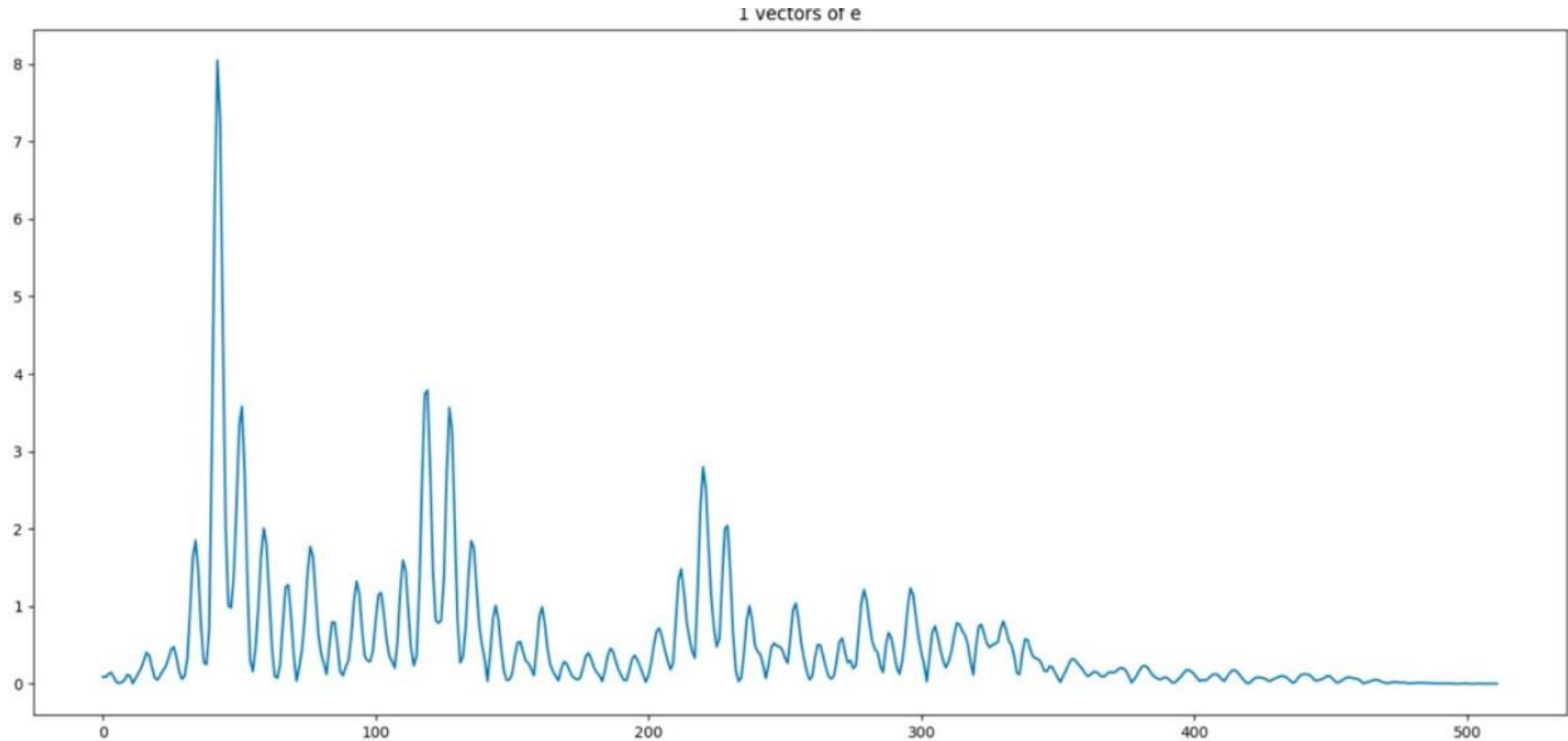
- Vector FFT có số điểm FFT là 1024, số chiều vector là 512
- Vector MFCC có số chiều là 13 / 26 / 39
- Cả 2 loại vector đều dùng chung kích thước frame 0.03s, frame shift 0.01s, $K = 1$ và 2

Vector đặc trưng FFT



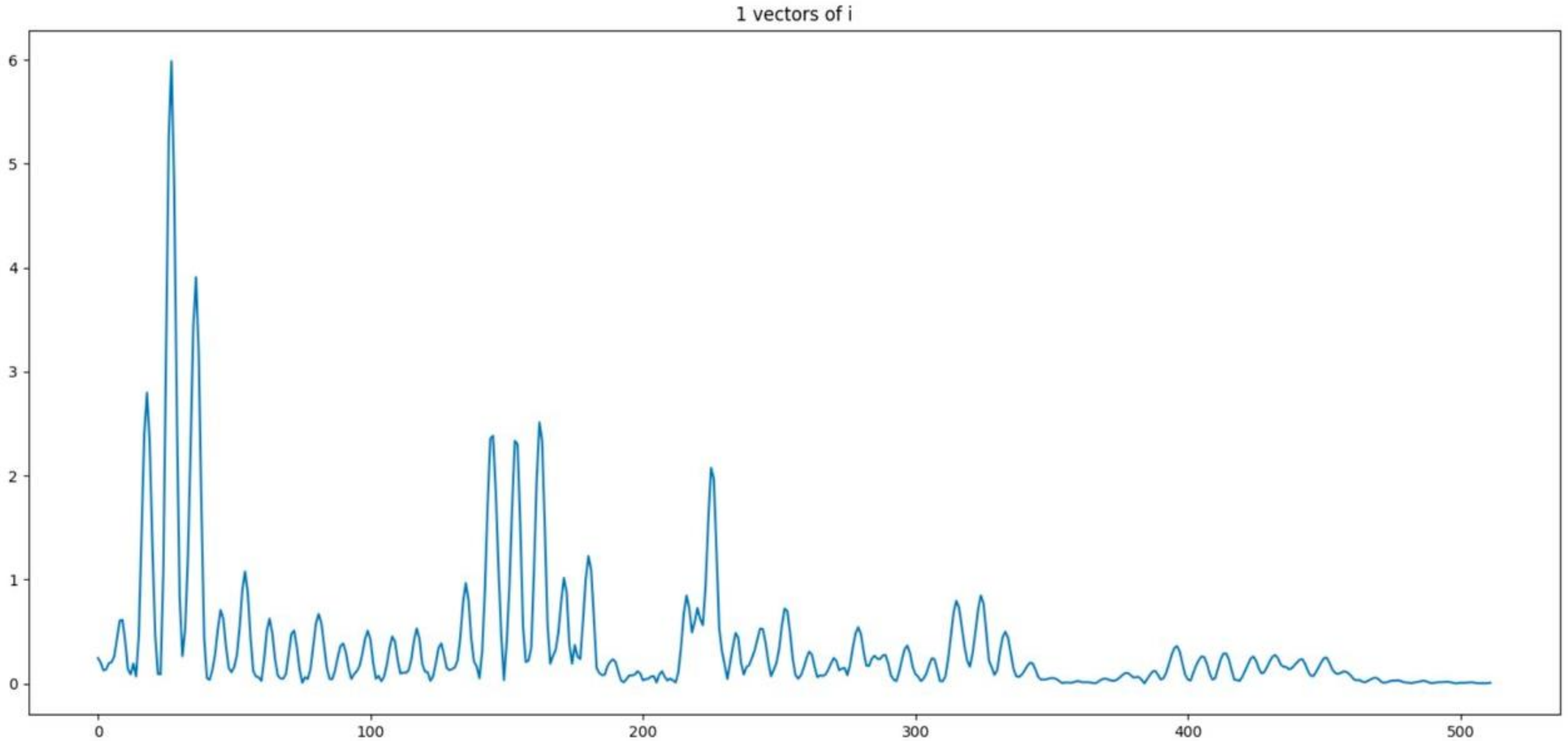
Vector đặc trưng FFT của nguyên âm a

Vector đặc trưng FFT



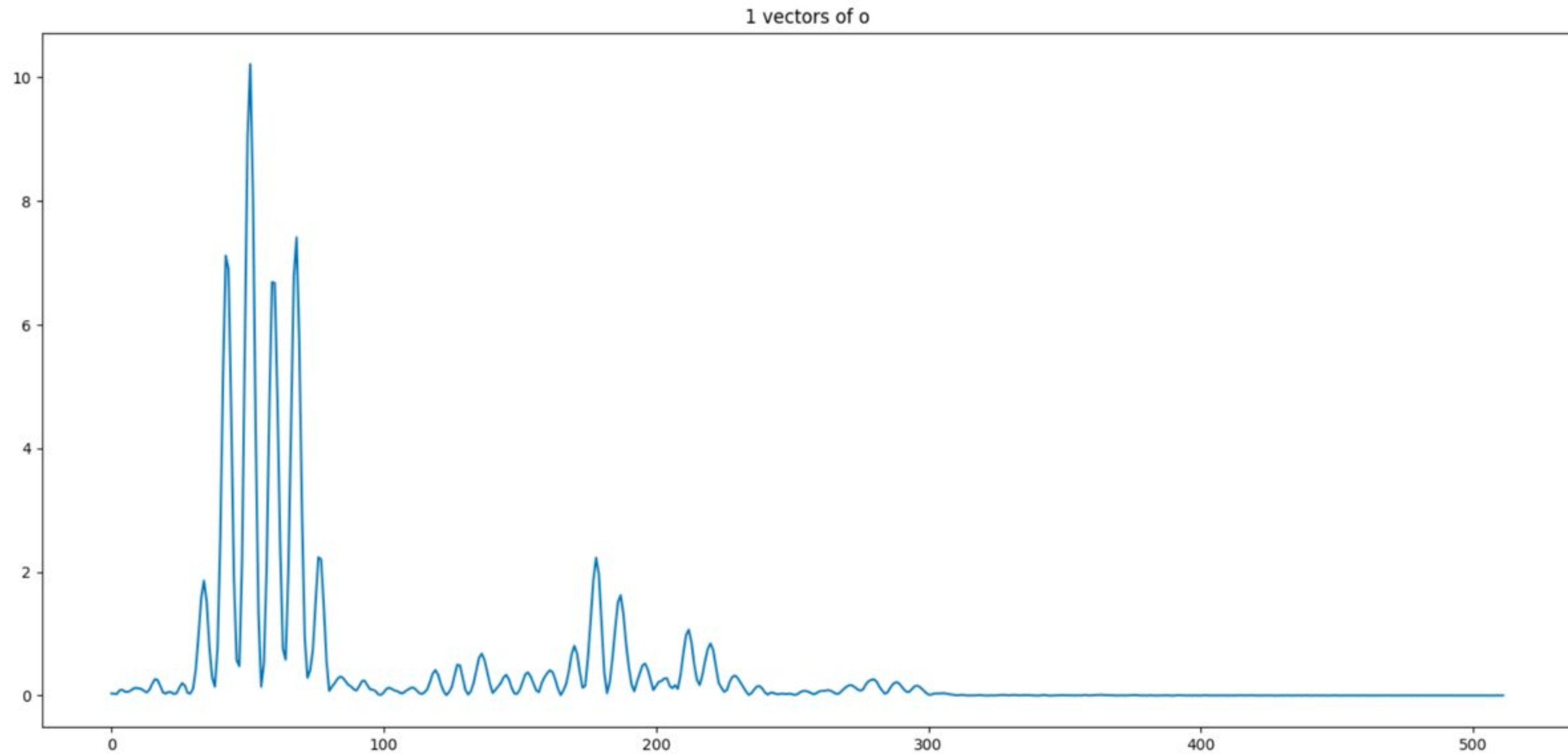
Vector đặc trưng FFT của nguyên âm e

Vector đặc trưng FFT



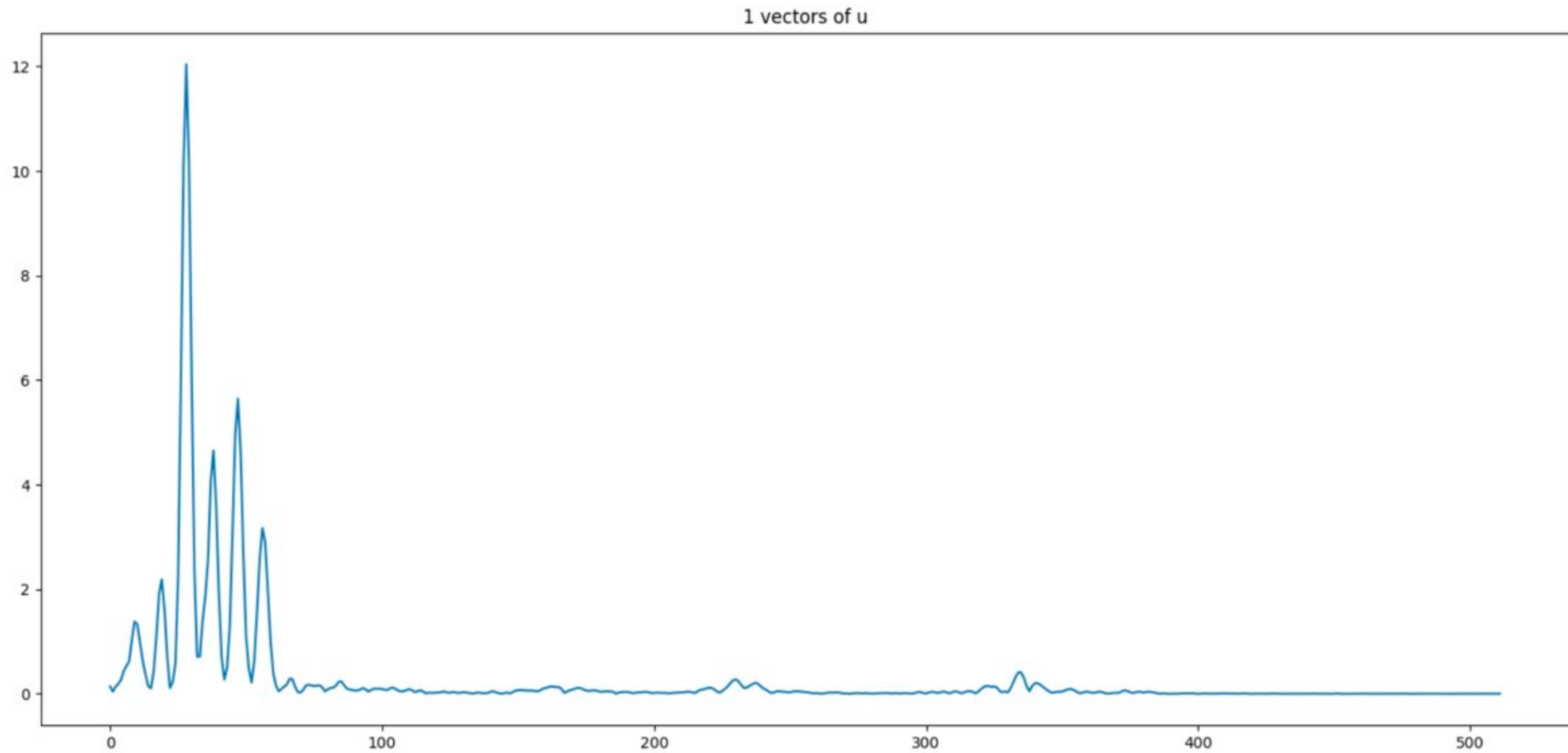
Vector đặc trưng FFT của nguyên âm i

Vector đặc trưng FFT



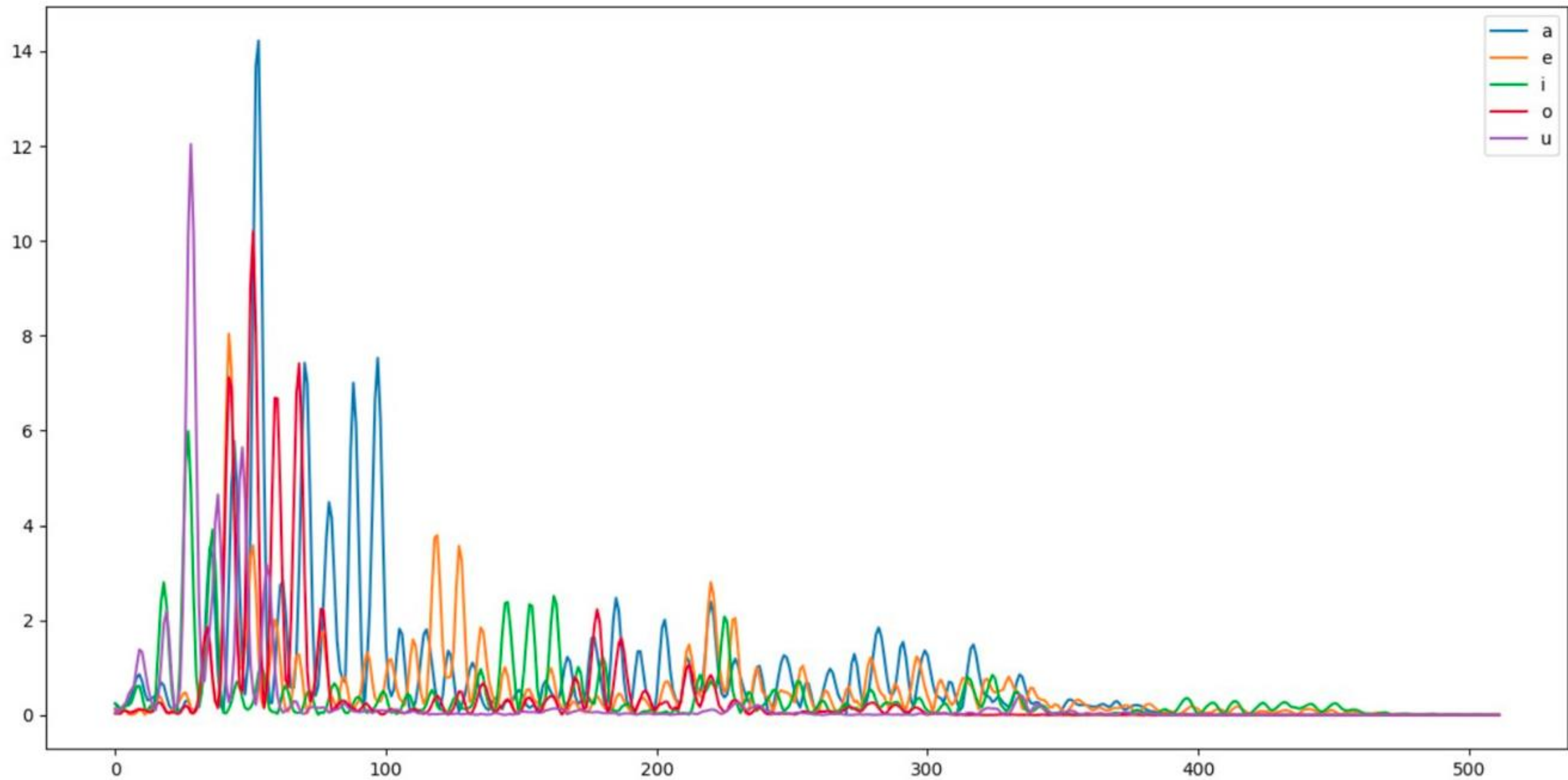
Vector đặc trưng FFT của nguyên âm o

Vector đặc trưng FFT



Vector đặc trưng FFT của nguyên âm u

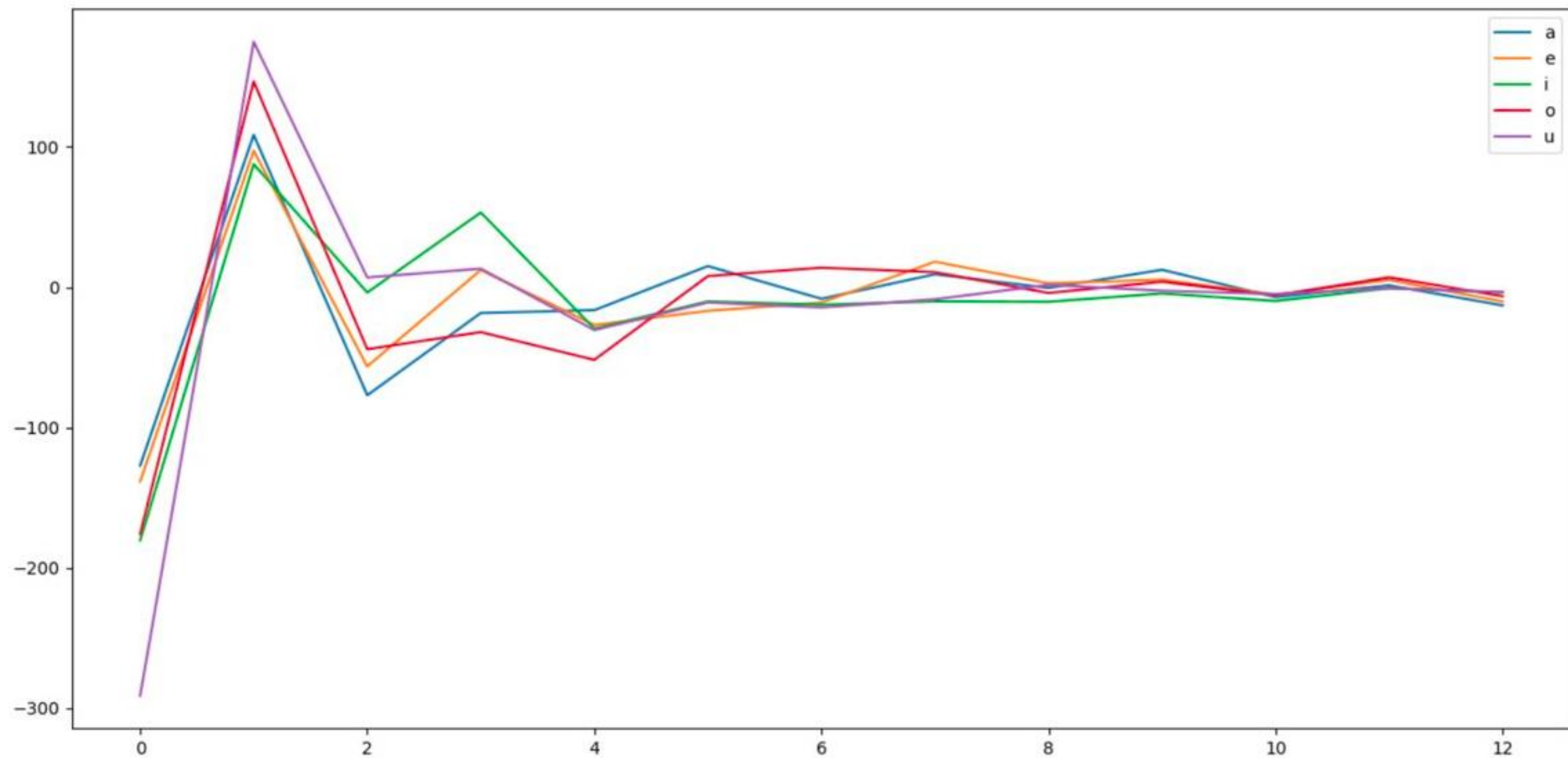
Vector đặc trưng FFT



Tổng hợp các vector đặc trưng FFT của 5 nguyên âm

→ Có sự khác biệt giữa các vector đặc trưng FFT giữa 5 nguyên âm nên có thể sử dụng chúng để nhận dạng nguyên âm

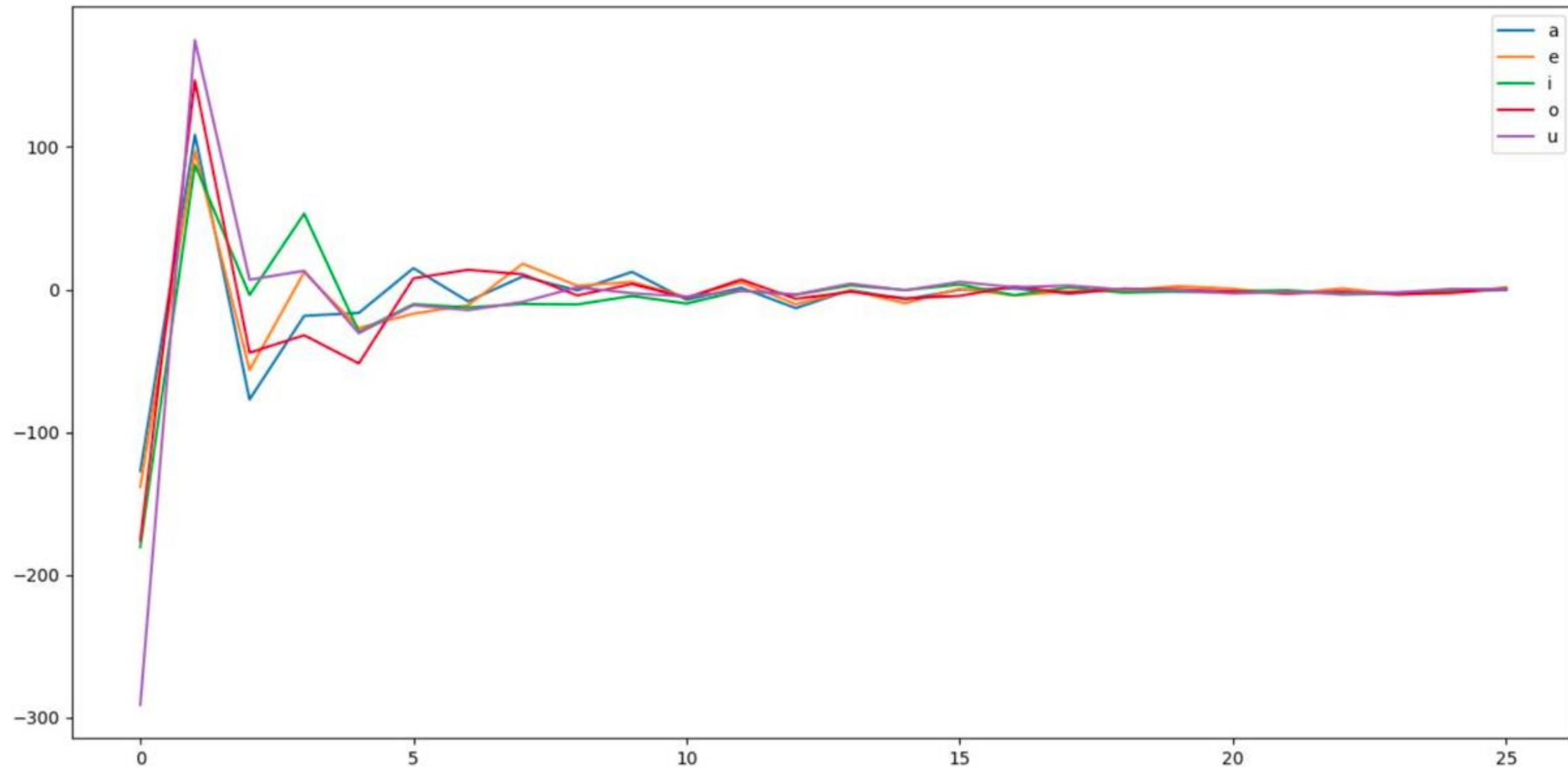
Vector đặc trưng MFCC



Vector đặc trưng MFCC của các nguyên âm với $N = 13$

→ Các vector MFCC giữa các nguyên âm phân biệt rõ ràng, giúp dễ dàng nhận dạng các nguyên âm dựa vào vector MFCC

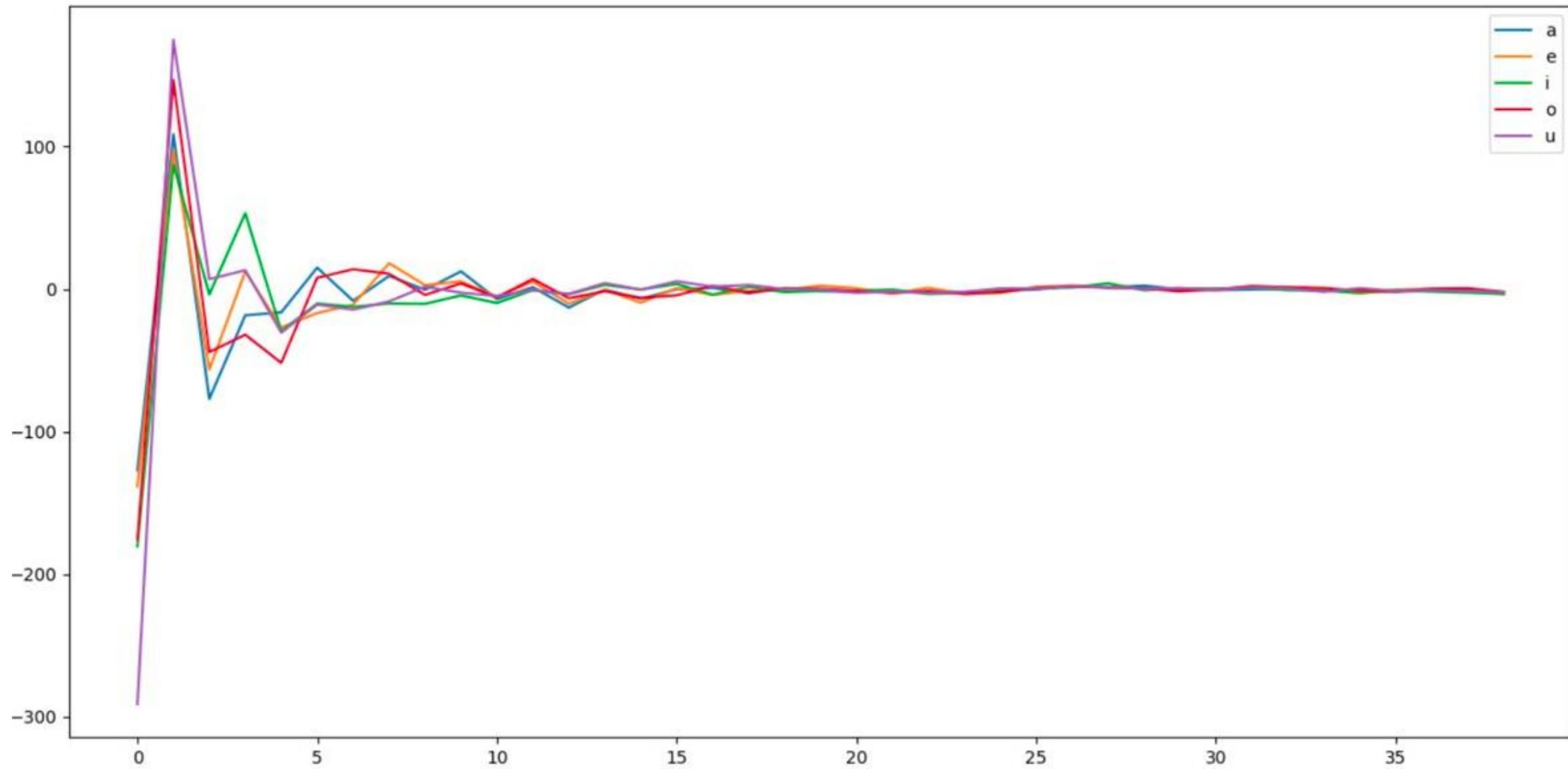
Vector đặc trưng MFCC



Vector đặc trưng MFCC của các nguyên âm với $N = 26$

- Từ hệ số thứ 14 – 17 vẫn có sự tách biệt giữa các vector các nguyên âm
- Từ hệ số 17 trở đi, các đường vector trùng nhau nhiều hơn và khó phân biệt

Vector đặc trưng MFCC



Vector đặc trưng MFCC của các nguyên âm với $N = 39$



SO SÁNH ĐỘ CHÍNH XÁC GIỮA 2 LOẠI VECTOR ĐẶC TRƯNG

FFT							MFCC					
N = 512	/a/	/e/	/i/	/o/	/u/		N = 13, 26, 39	/a/	/e/	/i/	/o/	/u/
/a/	19	0	0	2	0		/a/	19	1	0	1	0
/e/	0	18	2	1	0		/e/	0	19	2	0	0
/i/	0	0	18	0	3		/i/	0	0	20	0	1
/o/	2	2	0	17	0		/o/	5	1	0	15	0
/u/	0	0	3	0	18		/u/	0	0	1	1	19
ĐCX	85.71%						ĐCX	87.62%				

Độ chính xác 2 loại vector đặc trưng với $K = 1$
 \Rightarrow Với $K = 1$, vector đặc trưng MFCC có ĐCX cao hơn FFT 1.9%

FFT							MFCC					
N = 512	/a/	/e/	/i/	/o/	/u/		N = 13, 26, 39	/a/	/e/	/i/	/o/	/u/
/a/	18	0	0	3	0		/a/	18	1	0	2	0
/e/	1	18	2	1	0		/e/	1	19	1	0	0
/i/	0	0	18	0	3		/i/	0	1	18	0	2
/o/	5	1	0	15	0		/o/	3	0	0	18	0
/u/	0	0	6	0	15		/u/	0	0	1	1	21
ĐCX	80%						ĐCX	89.52%				

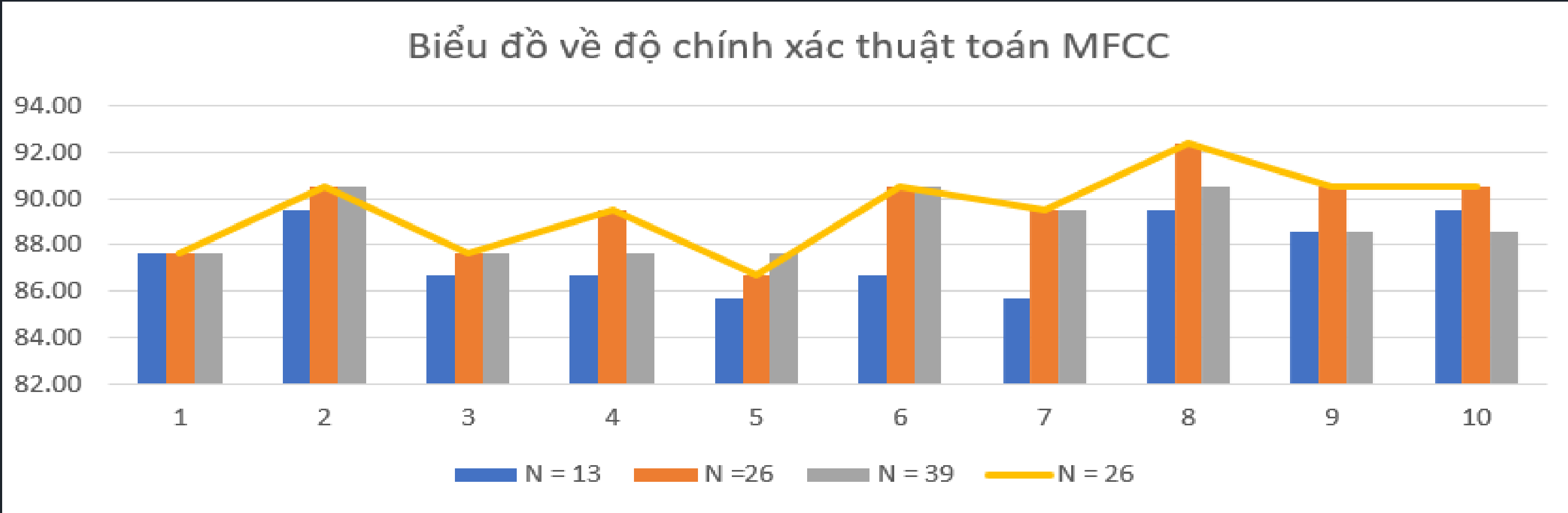
Độ chính xác 2 loại vector đặc trưng với $K = 2$
 \Rightarrow Với $K = 2$, vector đặc trưng MFCC có ĐCX cao hơn FFT 9.52%

- Vector MFCC cho độ chính xác cao hơn vector FFT và ĐCX cải thiện khi tăng số cụm K



KHẢO SÁT CÁC THAM SỐ,
ĐỘ CHÍNH XÁC CỦA THUẬT TOÁN

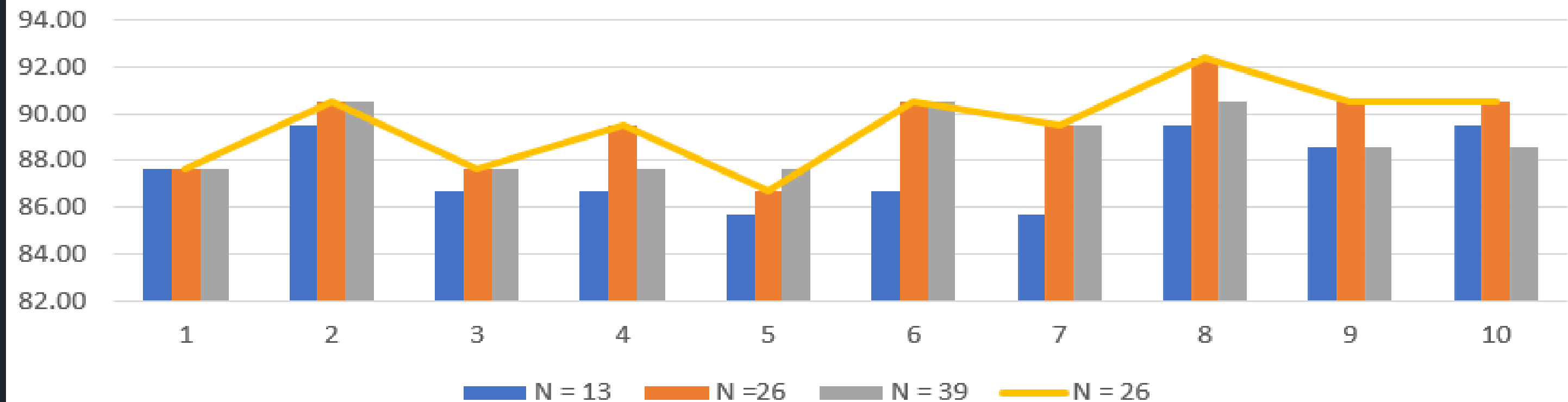
Bảng thống kê độ chính xác thuật toán MFCC										
N \ K	1	2	3	4	5	6	7	8	9	10
13	87.62	89.52	86.67	86.67	85.71	86.67	85.71	89.52	88.57	89.52
26	87.62	90.48	87.62	89.52	86.67	90.48	89.52	92.38	90.48	90.48
39	87.62	90.48	87.62	87.62	87.62	90.48	89.52	90.48	88.57	88.57



Bảng thống kê độ chính xác thuật toán MFCC

N \ K	1	2	3	4	5	6	7	8	9	10
13	87.62	89.52	86.67	86.67	85.71	86.67	85.71	89.52	88.57	89.52
26	87.62	90.48	87.62	89.52	86.67	90.48	89.52	92.38	90.48	90.48
39	87.62	90.48	87.62	87.62	87.62	90.48	89.52	90.48	88.57	88.57

Biểu đồ về độ chính xác thuật toán MFCC



MA TRẬN NHẦM LẦN

N = 26, K = 8	/a/	/e/	/i/	/o/	/u/
/a/	19	1	0	1	0
/e/	1	18	2	0	0
/i/	0	0	19	0	2
/o/	1	0	0	20	0
/u/	0	0	0	0	21

/a/ : 90.48%

/e/ : 85.71%

/i/ : 90.48%

/o/ : 95.23%

/u/ : 100%

=> Độ chính xác: 92.38%

MA TRẬN NHẦM LẦN

N = 26, K = 8	/a/	/e/	/i/	/o/	/u/
/a/	19	1	0	1	0
/e/	1	18	2	0	0
/i/	0	0	19	0	2
/o/	1	0	0	20	0
/u/	0	0	0	0	21

/a/ : 90.48%
/e/ : 85.71%
/i/ : 90.48%
/o/ : 95.23%
/u/ : 100%

=> Độ chính xác: 92.38%

NHẬN XÉT CHUNG:

- Có thể nhận dạng nguyên âm không phụ thuộc người nói dựa vào vector đặc trưng MFCC hoặc FFT
- Vector MFCC có ĐCX cao hơn khi kết hợp với phân cụm K-mean trong khi vector FFT có ĐCX cao hơn khi không phân cụm
- Việc tăng số chiều của MFCC không cải thiện được độ chính xác thuật toán rõ rệt
- Việc phân cụm K_mean có thể cải thiện độ chính xác đáng kể so với việc tăng số chiều của vector

CẢM ƠN THẦY ĐÃ LẮNG NGHE!

