Đại học Bách Khoa -
Đại học Đà Nẵng

Phân tích phổ &
Nhận dạng tín hiệu nguyên âm


Nhóm trình bày: 4


Nhóm HP: 20.12


GVHD: Ninh Khánh Duy

Thành viên nhóm

Phan Mạnh Cường
102200250

Đinh Huy Hoàng
102200255
Trường nhóm

Nguyễn Văn Hoài Nam
102200274

Nhóm 4

Nội dung chính

Bài 1: Phân tích đặc trưng phổ các nguyên âm của nhiều người nói

- Các bước tiến hành
- Kết quả chương trình
- Nhận xét

Bài 2: Nhận dạng nguyên âm không phụ thuộc người nói dùng đặc trưng phổ FFT

- Các bước tiến hành
- Kết quả chương trình
- Nhận xét

Bài 3: Nhận dạng nguyên âm không phụ thuộc người nói dùng đặc trưng phổ MFCC

- Các bước tiến hành
- Kết quả chương trình
- Nhận xét

Nhóm 4

Bài 1:
Phân tích đặc trưng phổ nguyên âm của nhiều người nói

Các bước tiến hành

Input

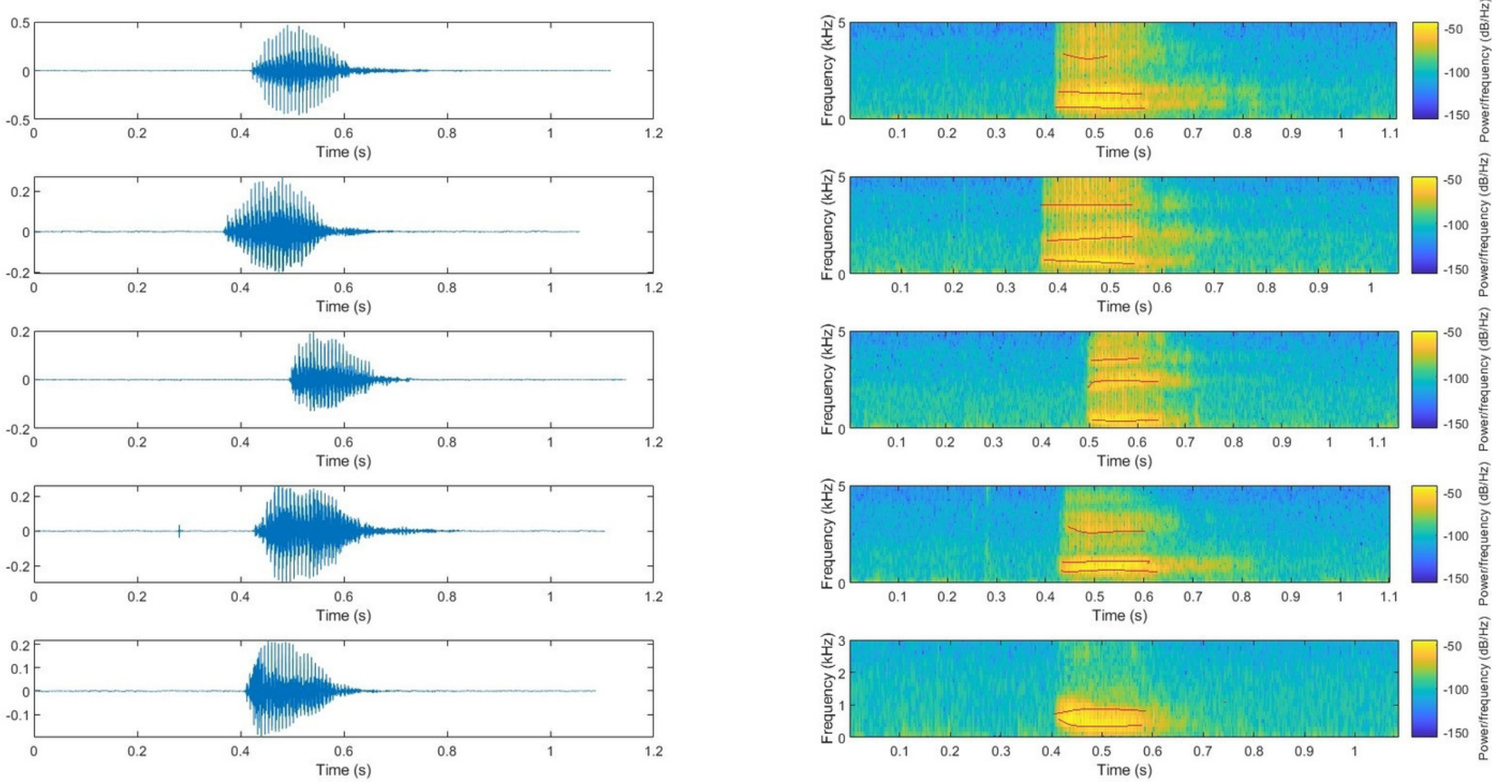
Tín hiệu tiếng nói
(8 người chọn ngẫu nhiên từ 21 người trong thư mục NguyenAmHuanLuyen-16k)

Xuất ảnh phổ băng rộng

Xác định bộ ba tần số formant (F1, F2, F3)

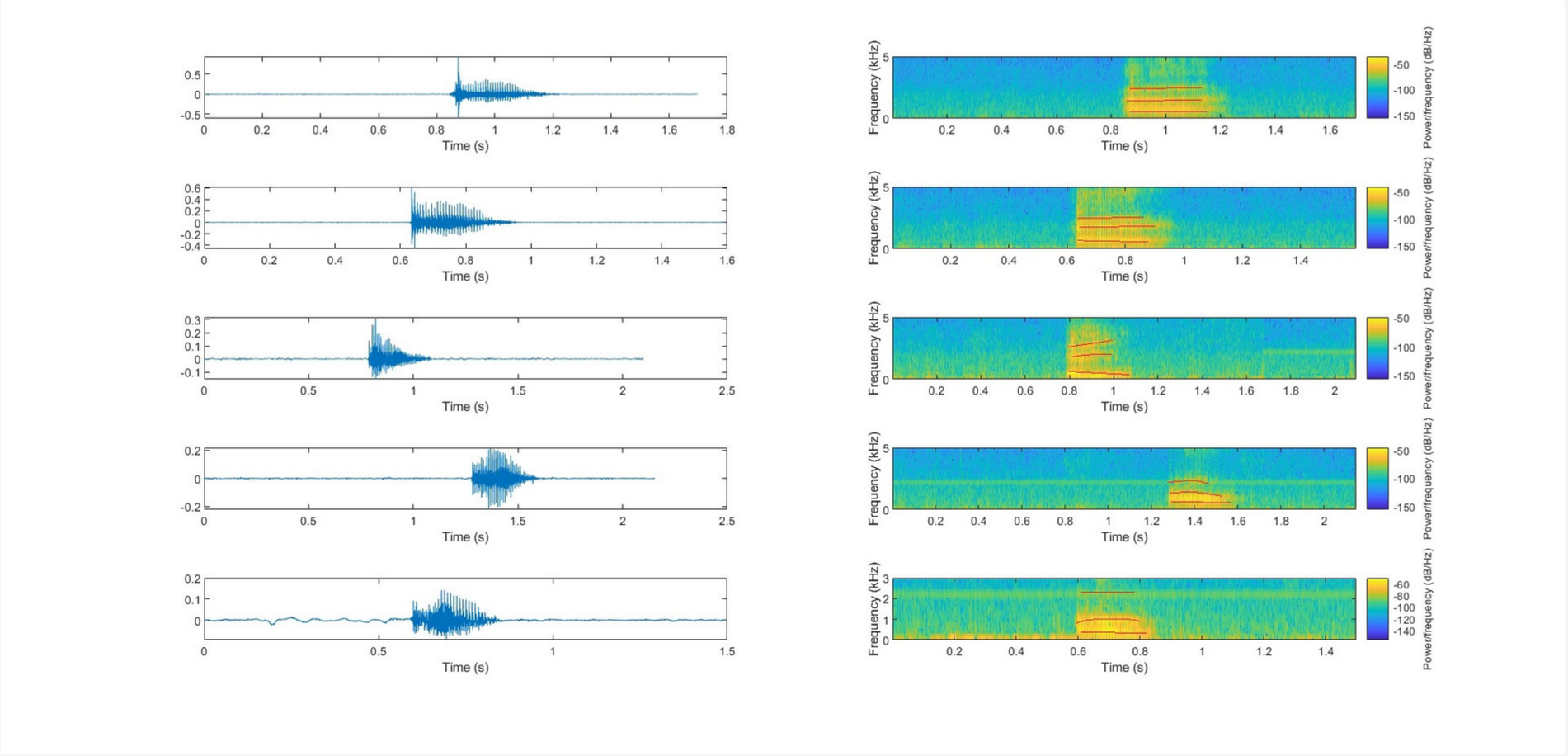
Xuất bảng dữ liệu & Nhận xét

Figure 1: 01MDA



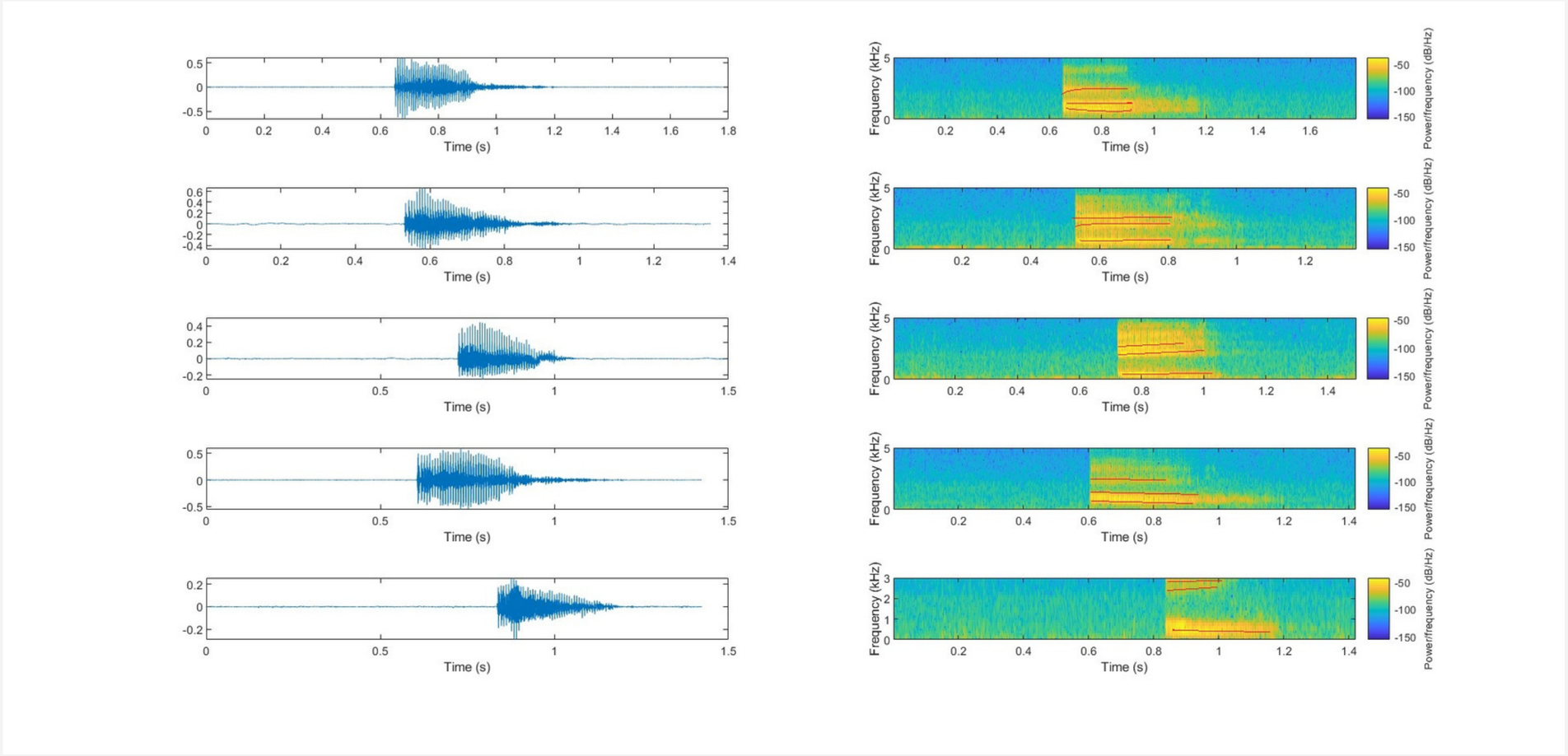
	/a/	/e/	/i/	/o/	/u/
F1	800	700	500	700	400
F2	1400	1900	2300	1000	800
F3	3000	3400	3400	2800	1700

Figure 2: 03MAB



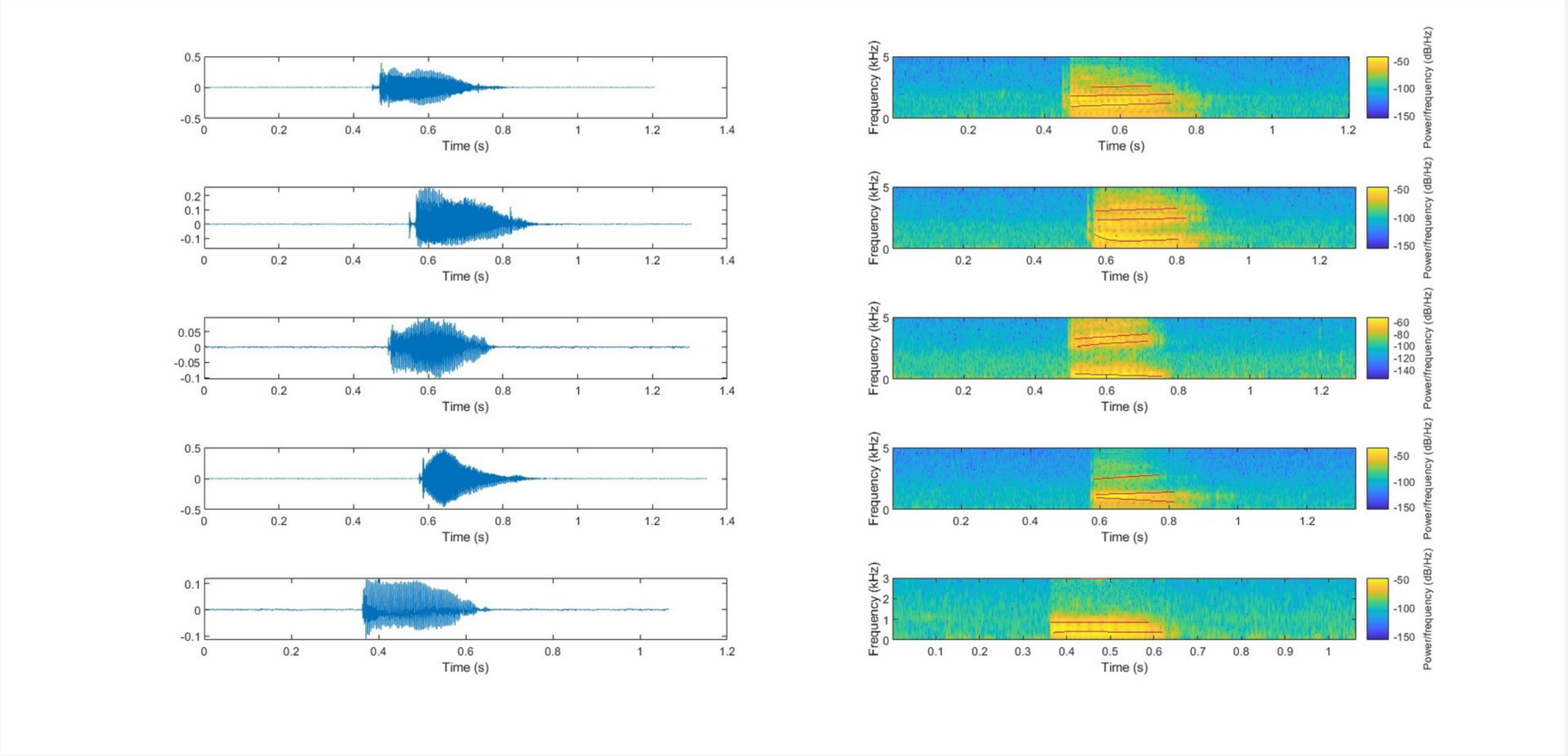
	/a/	/e/	/i/	/o/	/u/
F1	780	720	430	740	400
F2	1400	1700	1800	1200	900
F3	2300	2200	2600	2300	2200

Figure 3: 05MVB



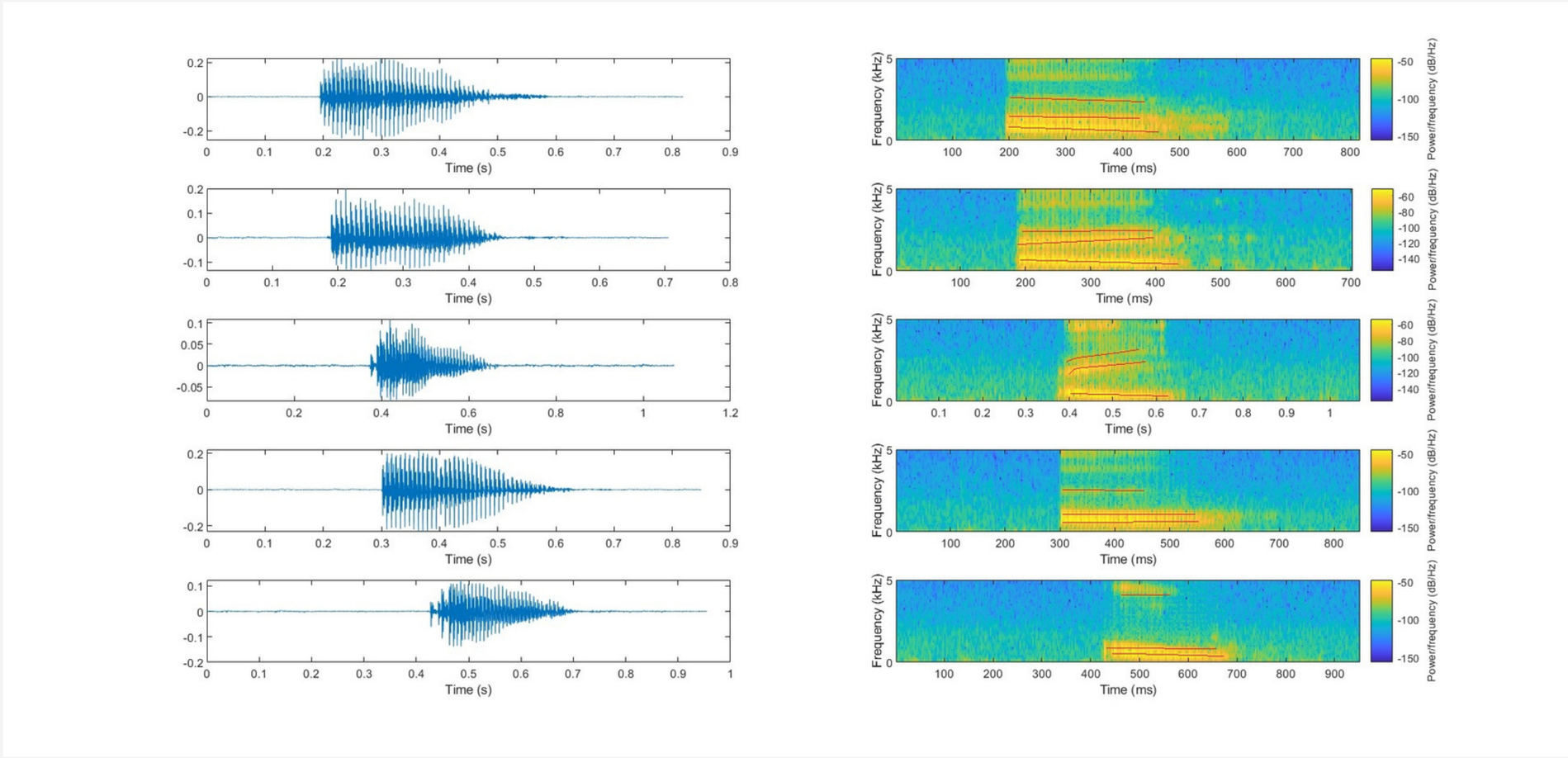
	/a/	/e/	/i/	/o/	/u/
F1	830	700	440	750	460
F2	1200	2100	2100	1000	2400
F3	2400	2700	2800	2300	2900

Figure 4: 07FTC



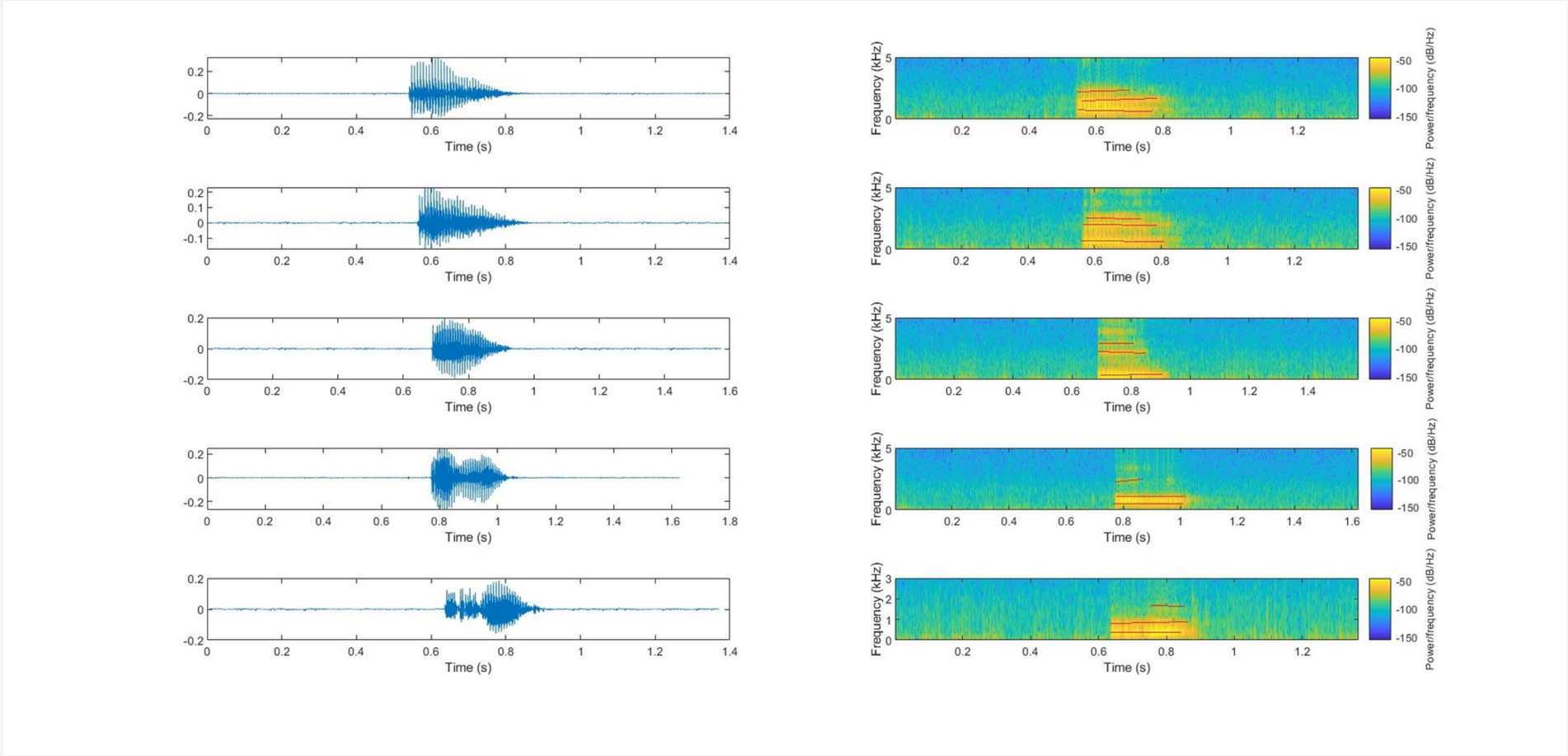
	/a/	/e/	/i/	/o/	/u/
F1	1000	900	370	950	400
F2	1600	2200	2900	1200	830
F3	2400	3300	3400	2500	3000

Figure 5: 09MPD



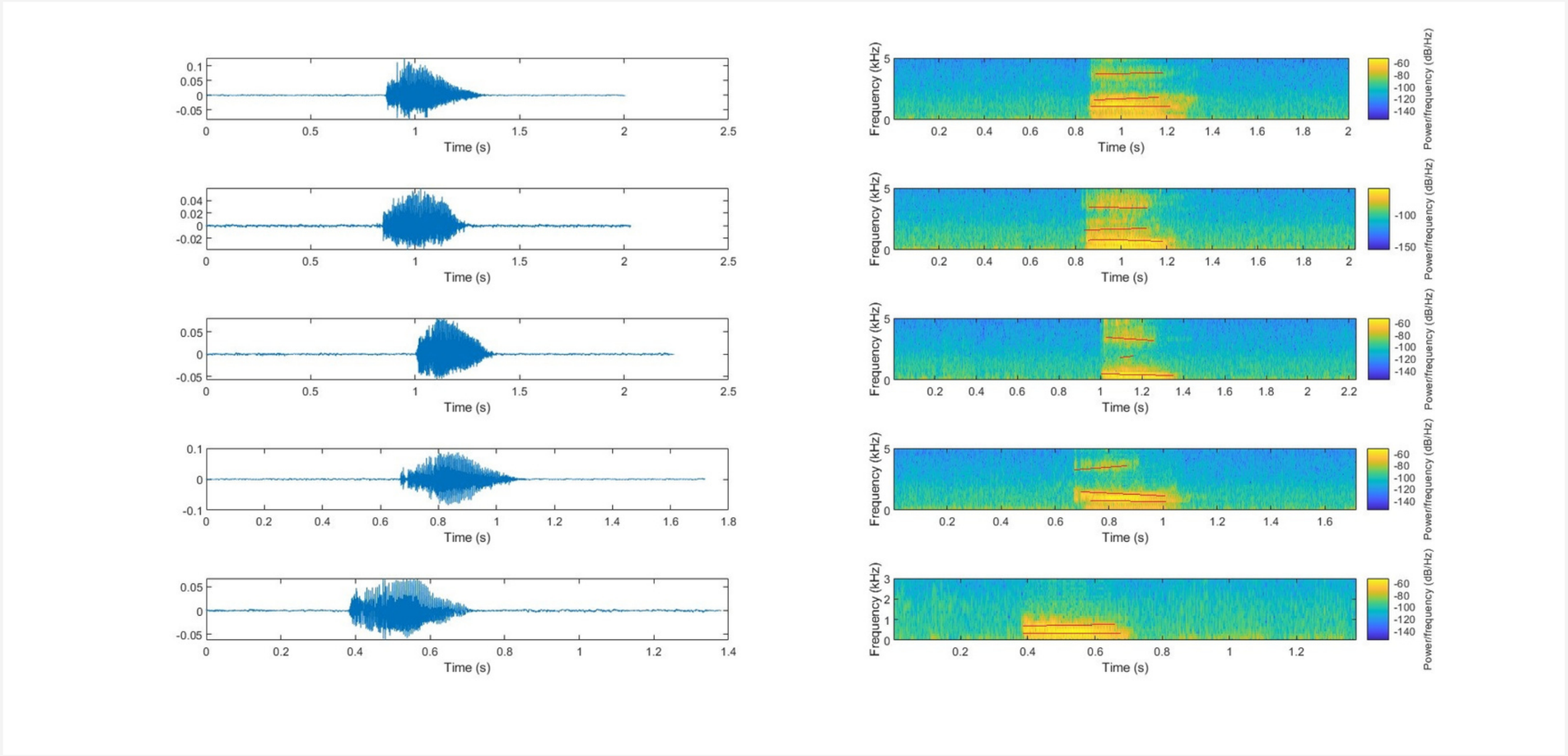
	/a/	/e/	/i/	/o/	/u/
F1	740	600	450	680	460
F2	1300	1900	2000	1000	800
F3	2500	2500	2800	2500	3500

Figure 6: 11MVD



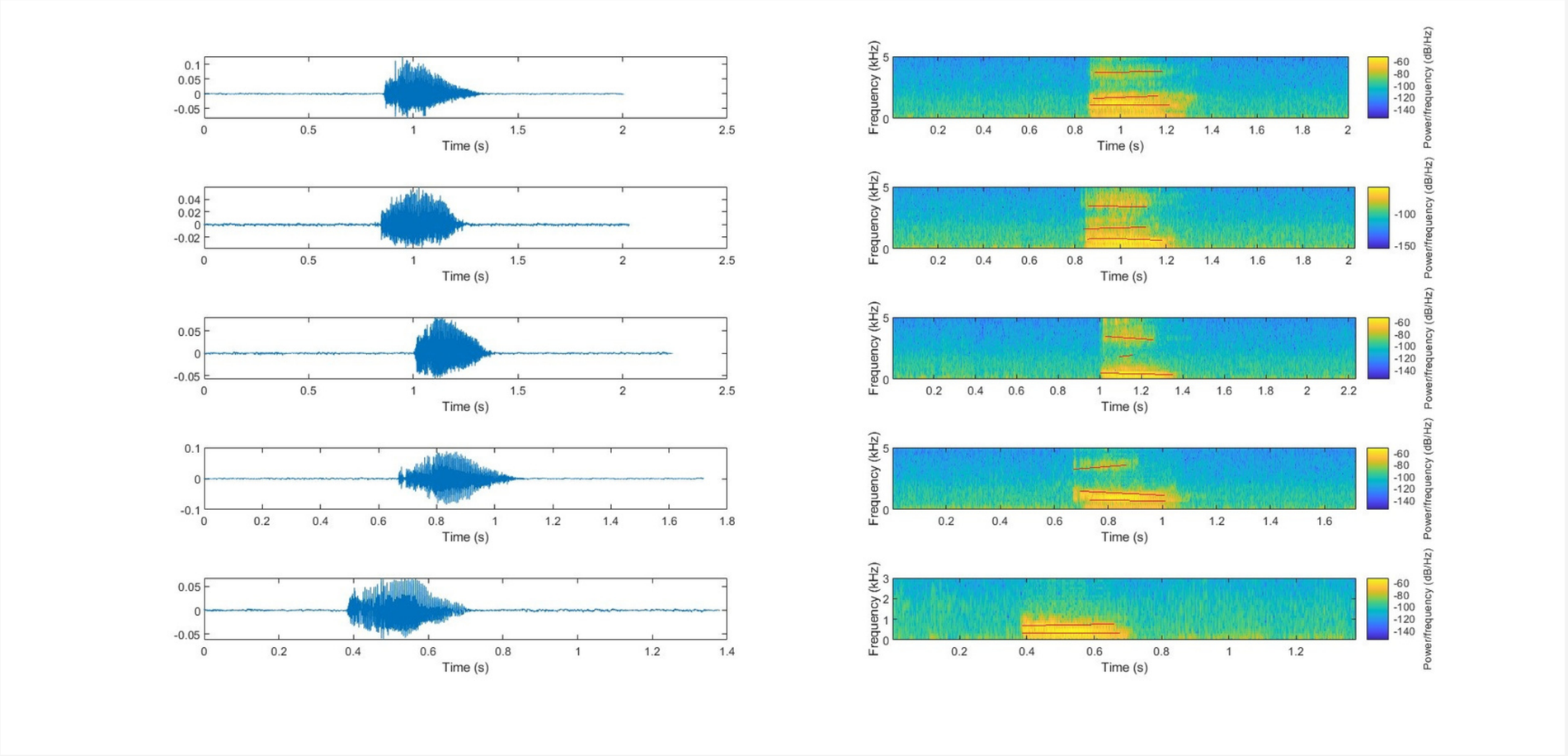
	/a/	/e/	/i/	/o/	/u/
F1	750	660	400	630	400
F2	1400	1800	2200	900	780
F3	2100	2500	2900	2300	1900

Figure 7: 14FHH



	/a/	/e/	/i/	/o/	/u/
F1	1100	700	400	1000	480
F2	1800	2000	2100	1200	1000
F3	3500	3600	3500	3500	2600

Figure 8: 16FTH



	/a/	/e/	/i/	/o/	/u/
F1	730	720	510	800	530
F2	1300	2500	2900	1100	900
F3	2100	3700	3700	3200	2500

Bảng số liệu tổng hợp

TT	ID	/a/			/e/			/i/			/o/			/u/		
		F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	01MDA	800	1400	3000	700	1900	3400	500	2300	3400	700	1000	2800	400	800	1700
2	03MAB	780	1400	2300	720	1700	2200	430	1800	2600	740	1200	2300	400	900	2200
3	05MVB	830	1200	2400	700	2100	2700	440	2100	2800	750	1000	2300	460	2400	2900
4	07FTC	1000	1600	2400	900	2300	3300	370	2900	3400	960	1200	2500	400	431	3000
5	09MPD	740	1300	2500	600	1800	2500	450	2000	2800	680	1000	2500	460	800	3500
6	11MVD	750	1400	2100	700	1800	2500	410	2200	2900	900	900	2300	390	780	1900
7	14FHH	1100	1800	3500	700	2000	3600	400	2200	3500	1000	1200	3200	480	1000	2600
8	16FTH	730	1300	2100	700	2500	3600	510	2900	3600	800	1100	3200	530	930	2500

Nhận xét

Từ bảng số liệu tổng hợp:

- Bộ 3 tần số formant tăng dần từ thấp đến cao.
- Mỗi người nói có bộ 3 tần số formant riêng biệt ứng với 5 nguyên âm.

- Khác biệt giữa 5 nguyên âm 1 người nói:
  - Nguyên âm /e/ và /i/ có bộ 3 tần số formant cao nhất (F2: 1900-2900 Hz và F3: 2500- 3700 Hz). Nhưng nguyên âm /i/ có tần số F1 nhỏ (khoảng 400-550 Hz).
  - Tiếp theo là nguyên âm /a/ và /o/ (F1: 700-850 Hz, F2: 1000-1600Hz, F3: 2000-3500 Hz).
  - Nguyên âm /u/ có bộ 3 tần số formant thấp nhất (đa số tần số F3 không vượt quá 3000 Hz và F1 & F2: 400-1000 Hz).

- Khác biệt giữa 1 nguyên âm nhiêu người nói:
  - Theo giới tính:
    - Hầu hết người nói giọng nữ có bộ 3 tần số formant cao hơn người nói giọng nam.
    - Các nguyên âm /a/, /e/, /o/, /i/:
      - F1: Nam(700-850 Hz) & Nữ(700-1000 Hz) (khoảng /i/ nhỏ hơn).
      - F2: Nam(1200-1800 Hz) & Nữ(1400-2000 Hz).
      - F3: Nam(2100-3000 Hz) & Nữ(2500-3600 Hz).

Nhóm 4

Bài 2:

Nhận dạng nguyên âm không phụ thuộc người nói dùng đặc trưng phổ FFT

Các bước tiến hành

Input

105 file huấn luyện

Xác định nguyên âm khoảng lặng

Xác định vùng ổn định

Trích xuất vector FFT

N_FFT (256, 512, 1024, 2048)

Vector đặc trưng 1 nguyên âm - người nói

Vector đặc trưng 1 nguyên âm nhiều người nói

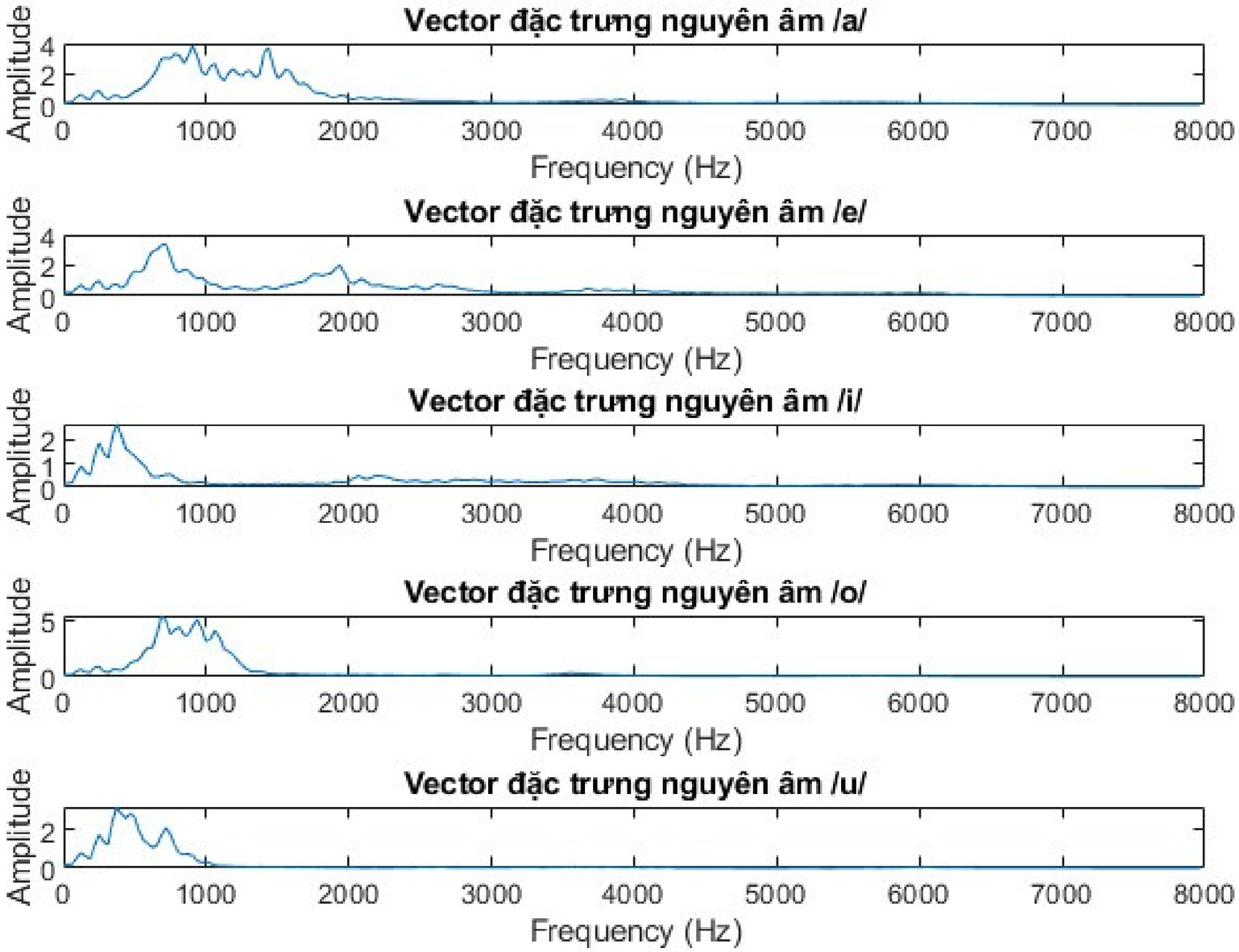
So khớp dựa vào Euclidean

Tính confusion matrix

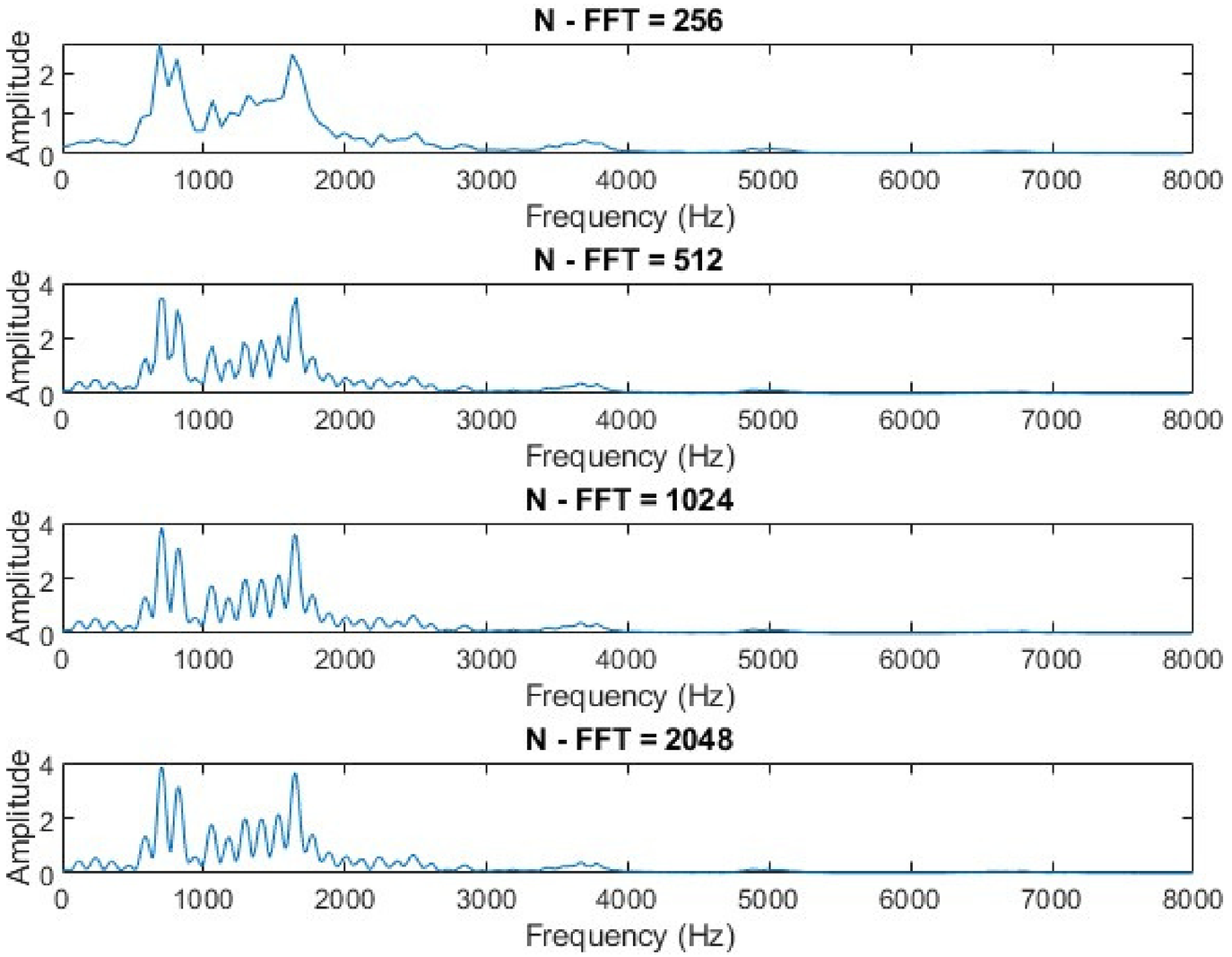
Nhóm 4

Kết quả chươnng trình

Vector đặc trưng phổ của 05 nguyên âm



Vector FFT của 1 khung tín hiệu với số chiều là N_FFT = 256, 512, 1024, 2048



Ma trận nhầm lẫn (confusion matrix)

N_FFT = 128

	/a/	/e/	/o/	/u/	/i/	%
/a/	13	4	3	1	0	62
/e/	0	14	5	2	0	67
/i/	0	0	15	0	6	71
/o/	0	4	3	13	1	62
/u/	0	0	8	0	13	62
Tổng:						66

Ma trận nhầm lẫn (confusion matrix)

N_FFT = 256

	/a/	/e/	/o/	/u/	/i/	%
/a/	14	3	4	0	0	67
/e/	0	16	3	2	0	76
/i/	0	0	15	0	6	71
/o/	1	4	3	12	1	57
/u/	0	0	8	0	13	62
Tổng:						67

Ma trận nhầm lẫn (confusion matrix)

N_FFT = 512

	/a/	/e/	/o/	/u/	/i/	%
/a/	13	4	4	0	0	62
/e/	0	16	4	1	0	76
/i/	0	0	16	0	0	76
/o/	1	2	3	14	1	67
/u/	0	0	9	0	12	57
Tổng:						68

Ma trận nhầm lẫn (confusion matrix)

N_FFT = 1024

	/a/	/e/	/o/	/u/	/i/	%
/a/	13	4	4	0	0	62
/e/	0	16	4	1	0	76
/i/	0	0	16	0	5	76
/o/	1	2	3	14	1	67
/u/	0	0	9	0	12	57
Tổng:						68

Ma trận nhầm lẫn (confusion matrix)

N_FFT = 2048

	/a/	/e/	/o/	/u/	/i/	%
/a/	13	4	4	0	0	62
/e/	0	16	4	1	0	76
/i/	0	0	16	0	5	76
/o/	1	2	3	14	1	67
/u/	0	0	9	0	12	57
Tổng:						68

Bảng thống kê độ chính xác nhận dạng tổng hợp (%) theo số chiều của vector đặc trưng

	/a/	/e/	/i/	/o/	/u/	Tổng
N_FFT = 128	62	67	71	62	62	66
N_FFT = 256	67	76	71	57	62	67
N_FFT = 512	62	76	76	67	57	68
N_FFT = 1024	62	76	76	67	57	68
N_FFT = 2048	62	76	76	67	57	68

Nhận xét

Từ ma trận nhầm lẫn (confusion matrix) của 5 giá trị N_FFT và bảng thống kê độ chính xác:

- Phương pháp dùng đặc trưng phổ FFT có độ chính xác trung bình (khoảng 66-68%).
- Nguyên âm /e/ và /i/ là nguyên âm được nhận dạng đúng nhất (nhận dạng đúng khoảng 75%).
- Nguyên âm /u/ là nguyên âm bị nhận dạng sai nhiều nhất (nhận dạng đúng khoảng 60%).
- Trong 5 giá trị N_FFT (128, 256, 512, 1024, 2048) thì các giá trị N_FFT từ 512 trở lên có độ chính xác cao hơn (68%).

Nhóm 4

Bài 3:

Nhận dạng nguyên âm không phụ thuộc người nói dùng đặc trưng phổ MFCC

Các bước tiến hành

Input

105 file huấn luyện

Xác định nguyên âm khoảng lặng

Xác định vùng ổn định

Trích xuất vector MFCC

N_MFCC (13,26,39)

Vector đặc trưng 1 nguyên âm - người nói

Vector đặc trưng 1 nguyên âm nhiều người nói

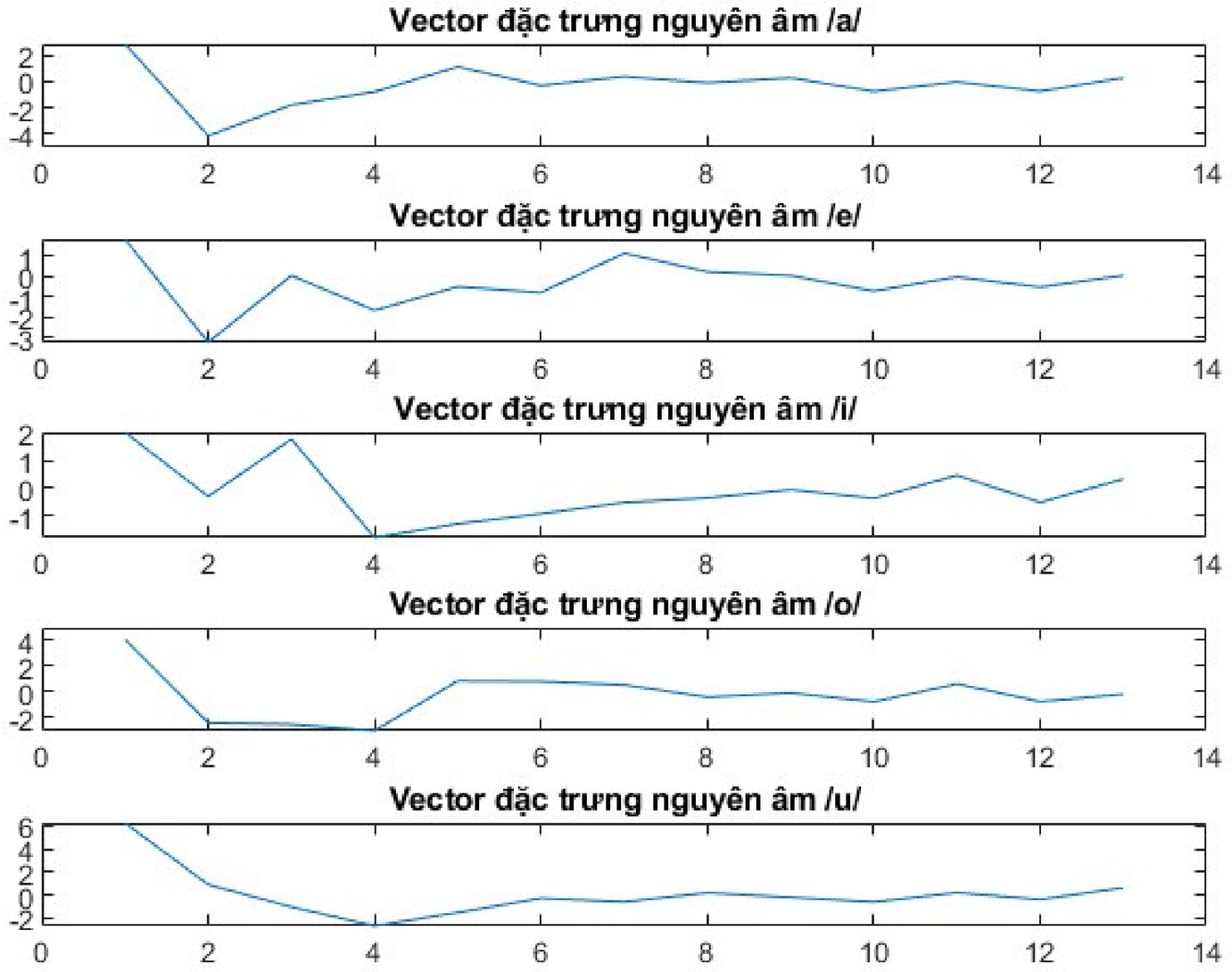
So khớp dựa vào Euclidean

Tính confusion matrix

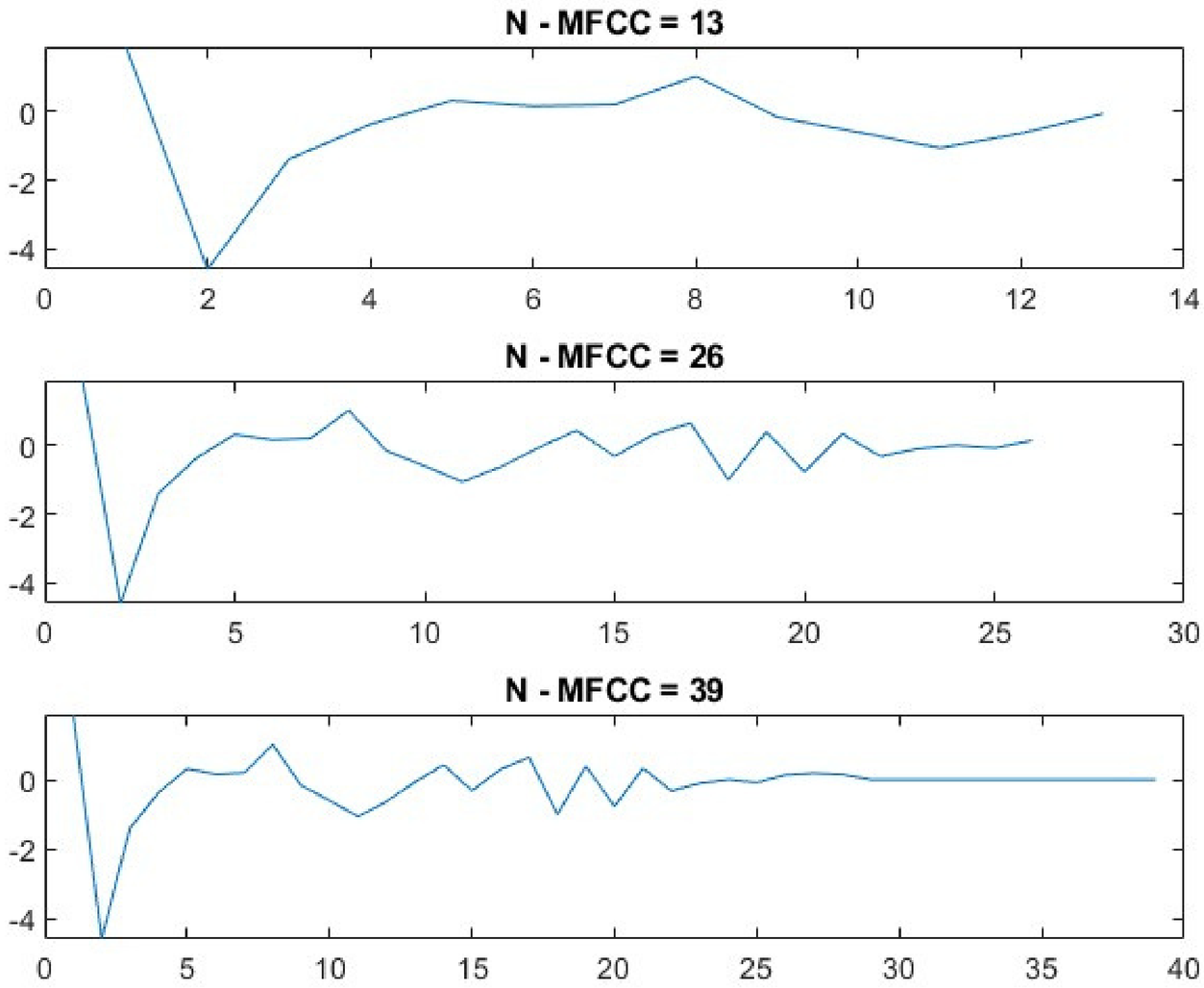
Nhóm 4

Kết quả chươnng trình

Vector đặc trưng phổ của 05 nguyên âm



Vector MFFC của 1 khung tín hiệu với số chiều là N_MFFC = 13, 26, 39



Ma trận nhầm lẫn (confusion matrix)

N_MFCC = 13

	/a/	/e/	/o/	/u/	/i/	%
/a/	19	2	0	0	0	90
/e/	4	17	0	0	0	81
/i/	0	0	21	0	0	100
/o/	1	0	0	18	2	86
/u/	0	0	0	0	21	100
Tổng:						91

Ma trận nhầm lẫn (confusion matrix)

N_MFCC = 26

	/a/	/e/	/o/	/u/	/i/	%
/a/	19	2	0	0	0	90
/e/	3	18	0	0	0	86
/i/	0	0	21	0	0	100
/o/	1	0	0	18	2	86
/u/	0	0	1	0	20	95
Tổng:						91

Ma trận nhầm lẫn (confusion matrix)

N_MFCC = 39

	/a/	/e/	/o/	/u/	/i/	%
/a/	19	2	0	0	0	90
/e/	3	18	0	0	0	86
/i/	0	0	21	0	1	100
/o/	0	0	0	18	2	86
/u/	0	0	1	0	20	95
Tổng:						91

Bảng thống kê độ chính xác nhận dạng tổng hợp (%) theo số chiều của vector đặc trưng

	/a/	/e/	/i/	/o/	/u/	Tổng
N_MFCC = 13	90	**81**	100	**86**	**100**	**91**
N_MFCC = 26	**90**	**86**	100	86	95	**91**
N_MFCC = 39	90	86	100	86	95	**91**

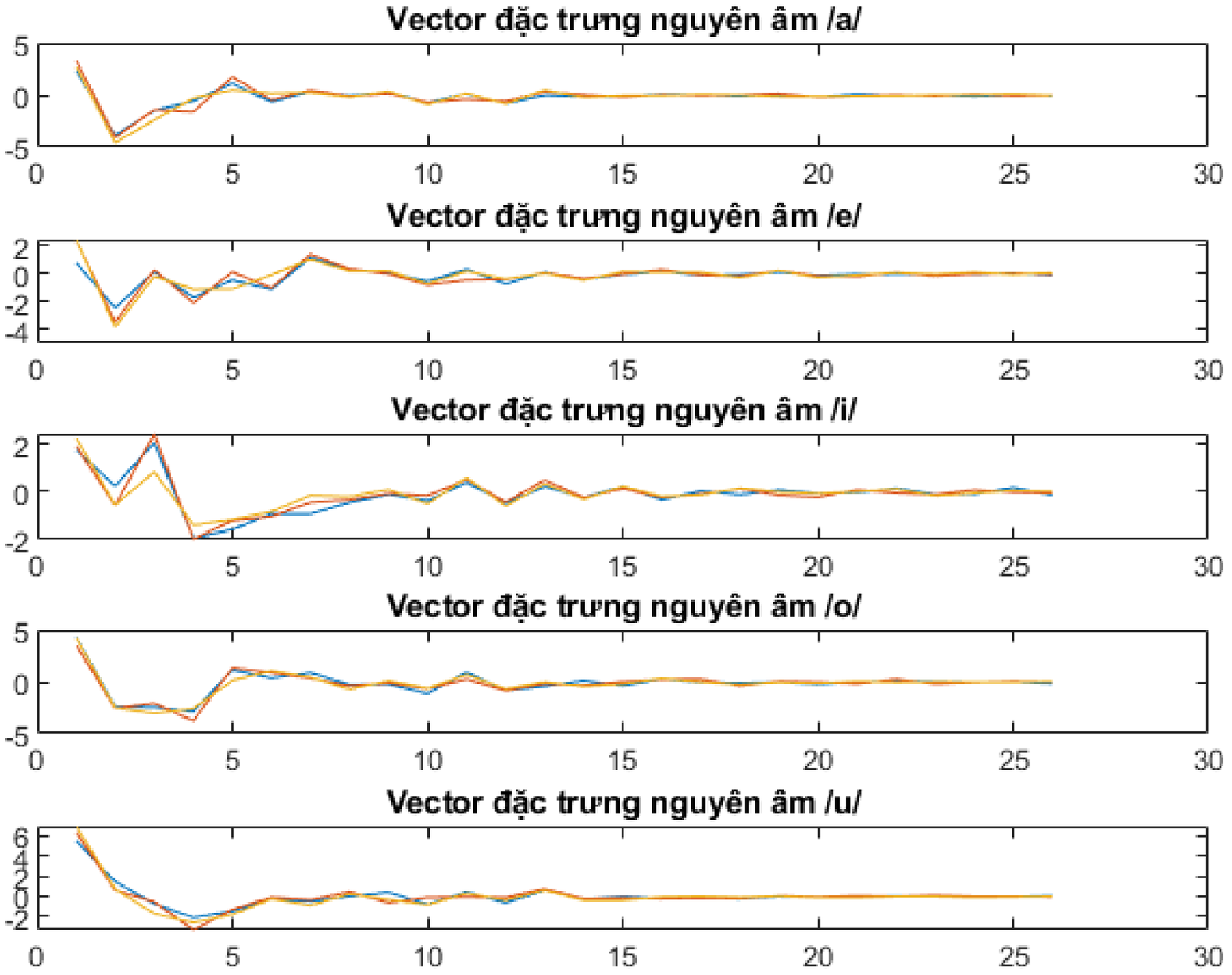
Nhận xét

Từ ma trận nhầm lẫn (confusion matrix) của 3 giá trị N_MFCC và bảng thống kê độ chính xác:

- Phương pháp dùng đặc trưng phổ MFCC có độ chính xác cao (khoảng 91%).
- Nguyên âm /i/ là nguyên âm được nhận dạng đúng nhất (nhận dạng đúng 100%).
- Nguyên âm /e/ là nguyên âm bị nhận dạng sai nhiều nhất (nhận dạng đúng khoảng 84%).
- Cả 3 giá trị N_MFCC (13, 26, 39) đều có độ chính xác 91%.

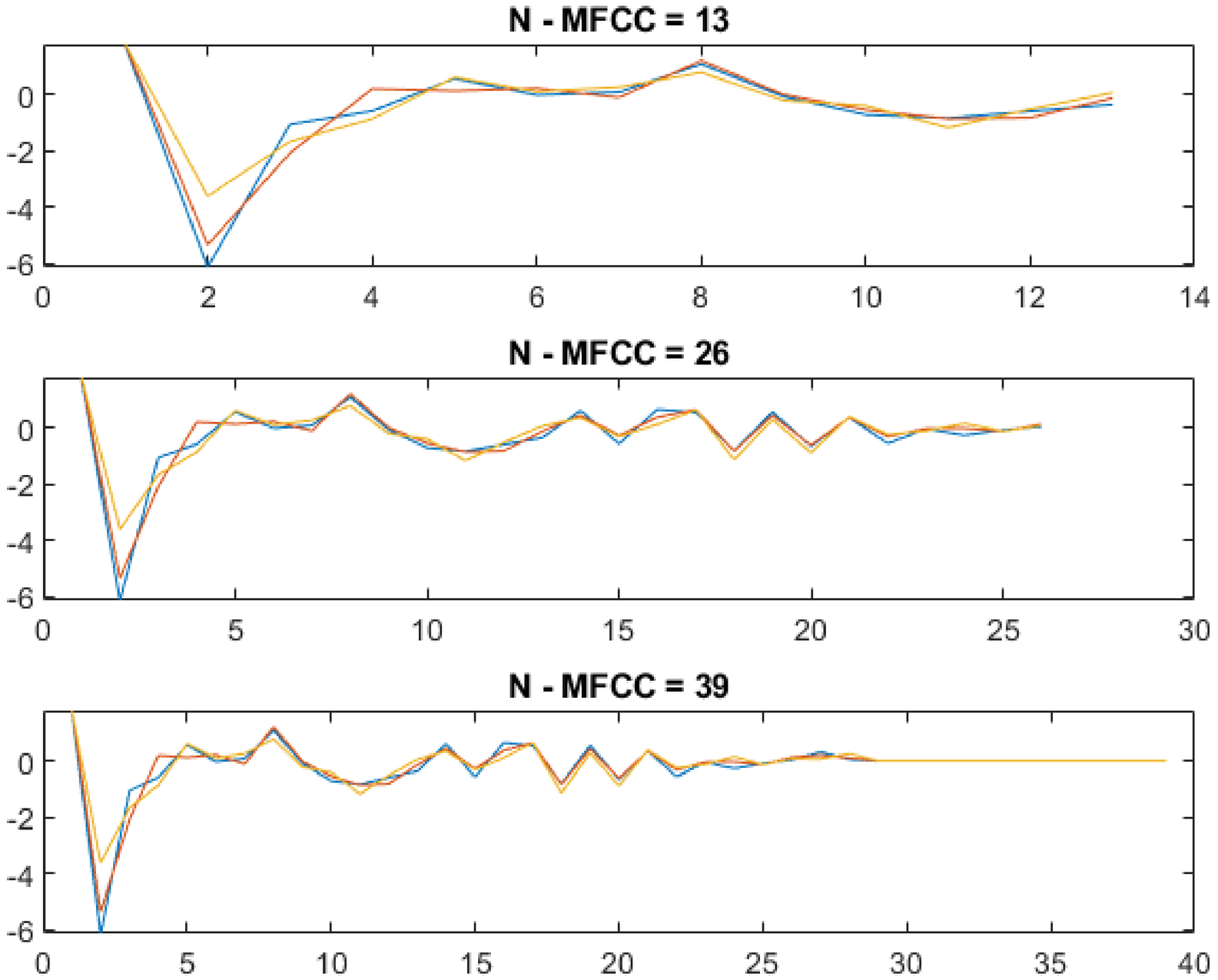
Vector đặc trưng phổ của 05 nguyên âm

K=3



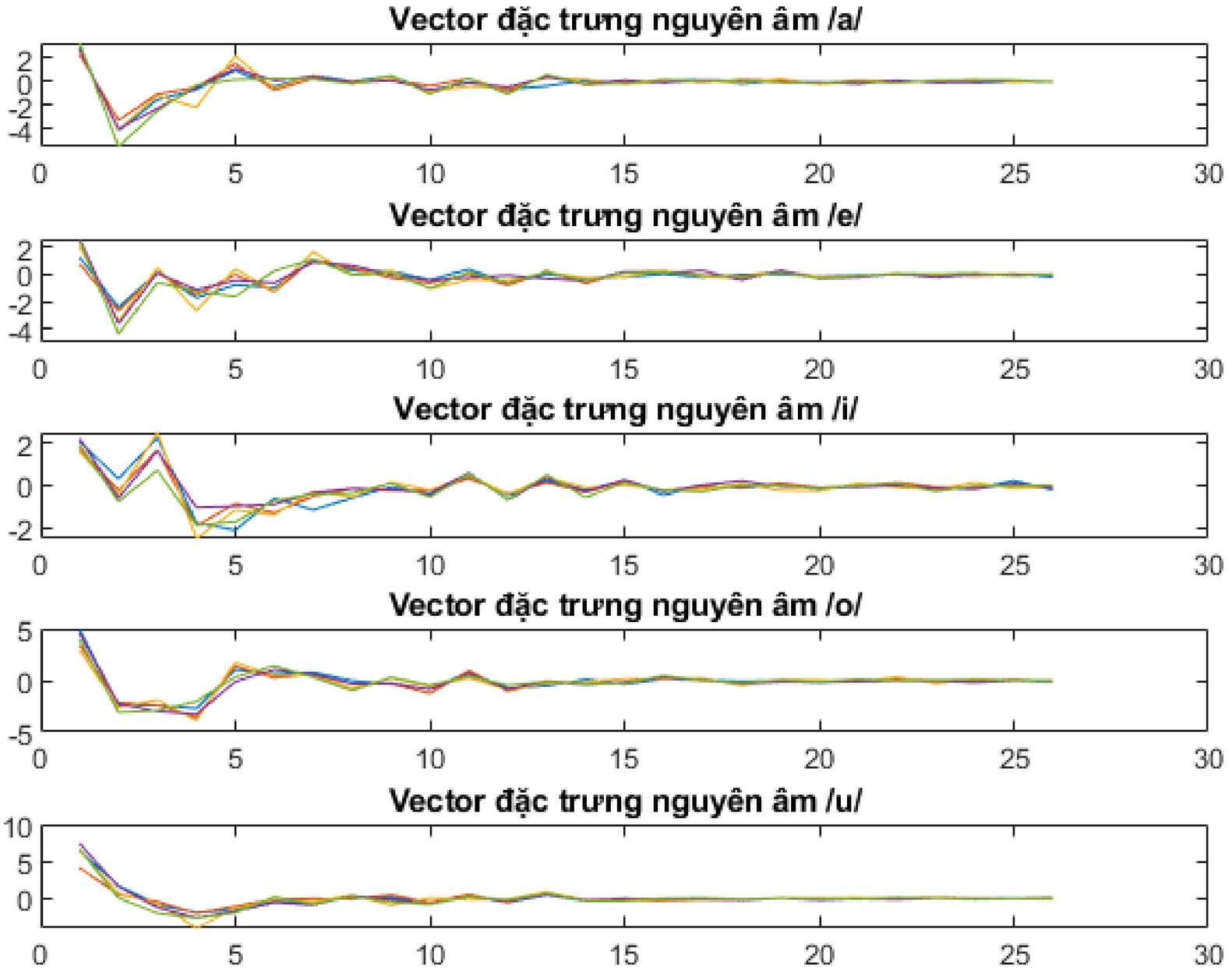
Vector MFFC của 1 khung tín hiệu với số chiều là N_MFFC = 13, 26, 39

K=3



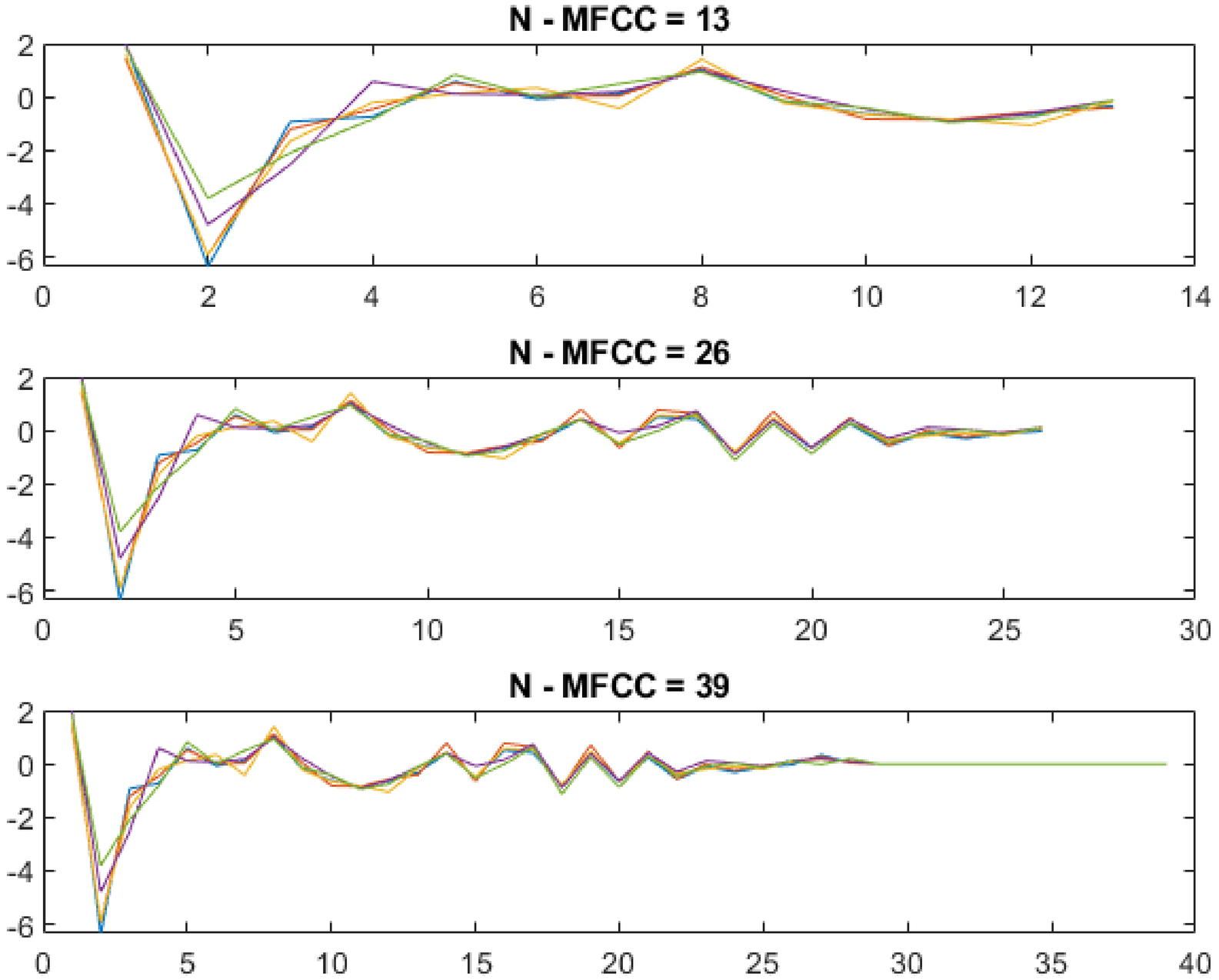
Vector đặc trưng phổ của 05 nguyên âm

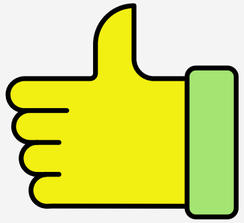
K=5



Vector MFFC của 1 khung tín hiệu với số chiều là N_MFFC = 13, 26, 39

K=5





THANK FOR LISTENING AND WATCHING