ĐẠI HỌC ĐÀ NẴNG

TRƯỜNG ĐẠI HỌC BÁCH KHOA

----- -----

PHẠM ĐỨC THIỆN

PHÂN TÍCH FORMANT CÁC NGUYÊN ÂM CỦA NHIỀU NGƯỜI NÓI ỨNG DỤNG XỬ LÝ ĐỒNG HÌNH

LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2021

ĐẠI HỌC ĐÀ NẴNG

TRƯỜNG ĐẠI HỌC BÁCH KHOA

----- □ -----

PHẠM ĐỨC THIỆN

PHÂN TÍCH FORMANT CÁC NGUYÊN ÂM CỦA NHIỀU NGƯỜI NÓI ỨNG DỤNG XỬ LÝ ĐỒNG HÌNH

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 8480101

LUẬN VĂN THẠC SĨ KỸ THUẬT

Người hướng dẫn khoa học: **TS. Ninh Khánh Duy**

Đà Nẵng - Năm 2021

LỜI CẢM ƠN

Để hoàn thành đề tài luận văn này em đã nhận được sự hướng dẫn, giúp đỡ và động viên tận tình từ nhiều phía. Tất cả những điều đó đã trở thành một động lực rất lớn giúp em có thể hoàn thành tốt luận văn này. Với tất cả sự cảm kích và trân trọng, em xin được gửi lời cảm ơn đến tất cả mọi người.

Em xin chân thành cảm ơn thầy TS. Ninh Khánh Duy đã tận tình giúp đỡ em trong suốt thời gian thực hiện luận văn.

Em xin chân thành cảm ơn các thầy cô giảng viên của Khoa Công nghệ Thông tin - Trường Đại học Bách Khoa Đà Nẵng đã tận tình dạy bảo, giúp đỡ em trong suốt thời gian qua.

Để có được kết quả như ngày hôm nay, em rất biết ơn gia đình và bạn bè đã động viên, khích lệ, giúp đỡ và tạo mọi điều kiện thuận lợi nhất có thể trong suốt quá trình học tập cũng như quá trình thực hiện luận văn này.

Mặc dù em đã cố gắng nỗ lực để thực hiện luận văn này, song chắc chắn khó tránh khỏi những thiếu sót. Do đó, em rất mong nhận được sự thông cảm, góp ý và chỉ bảo tận tình của quý thầy cô.

Một lần nữa, em xin trân trọng cảm ơn!

LỜI CAM ĐOAN

Tôi xin cam đoan :

1. *Những nội dung trong luận văn này là do tôi thực hiện dưới sự hướng dẫn trực tiếp của TS. Ninh Khánh Duy.*
2. *Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng tên tác giả, tên công trình, thời gian, địa điểm công bố.*
3. *Các số liệu kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.*

Tác giả

Phạm Đức Thiện

TÓM TẮT ĐỀ TÀI

PHÂN TÍCH FORMANT CÁC NGUYÊN ÂM CỦA NHIỀU NGƯỜI NÓI ỨNG DỤNG XỬ LÝ ĐỒNG HÌNH

Học viên: Phạm Đức Thiện. Chuyên ngành: Khoa học Máy tính
Mã số: 8480101. Khóa: K37. Trường Đại học Bách khoa - ĐHĐN

Tóm tắt - Formant của tín hiệu tiếng nói là một trong các tham số quan trọng và hữu ích có ứng dụng rộng rãi trong nhiều lĩnh vực chẳng hạn như trong việc xử lý, tổng hợp và nhận dạng tiếng nói. Tuy nhiên việc xác định tần số formant không đơn giản bởi vì các đỉnh phổ của tín hiệu ra của bộ máy phát âm phụ thuộc vào nhiều yếu tố phức tạp. Một trong các kỹ thuật phổ biến để xác định tần số formant là phương pháp xử lý đồng hình (hoặc cepstrum). Trong luận văn này, tôi nghiên cứu ứng dụng phương pháp xử lý đồng hình nhằm xác định tần số formant trong phân tích nguyên âm của nhiều người nói (dữ liệu bao gồm 50 người nói nam và nữ phát ra năm nguyên âm /a/, /e/, /i/, /o/ và /u/). Từ đó khảo sát tính đặc trưng của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói. Luận văn được tổ chức thành 3 chương. Chương 1 dành để nghiên cứu tổng quan về xử lý tín hiệu tiếng nói. Chương 2 dành để nghiên cứu phương pháp xử lý đồng hình và ứng dụng nó để xác định tần số formant. Chương 3 dành để giới thiệu các bước triển khai, kết quả thu được và đánh giá thuật toán.

Từ khóa - tín hiệu tiếng nói, cepstrum, xử lý đồng hình, formant, nguyên âm.

FORMANT ANALYSIS OF VOWELS PRONOUNCED BY MANY SPEAKERS USING HOMOMORPHIC DECONVOLUTION METHOD

Abstract - The formant of the speech signal is one of the most important and useful parameters that has wide application in many fields such as processing, synthesis and recognition of speech. However, the determination of the formant frequency is not straightforward as the spectral peaks of the output signals from emitters depend on various complicated factors. One of the popular techniques for determining the formant frequency is the homomorphic deconvolution method (also known as cepstrum). In this thesis, I researched the homomorphic deconvolution method in terms of its application in determining the formant frequency in the analysis of vowels produced by many speakers (the data included 50 male and female speakers who produced five vowels /a/, /e/, /i/, /o/ and /u/). Subsequently, the characteristics of different vowels came from the same speaker and a vowel pronounced by many speakers were studied. The thesis is organized into 3 chapters. Chapter 1 is about the theoretical background of speech signal processing studies while the homomorphic deconvolution method and its applications in formant frequency determination was studied and showed in chapter 2. Chapter 3 is represented for the introduction of the implementation steps, results, and algorithm evaluation.

Key words - speech signal, cepstrum, homomorphic deconvolution, formant, vowel.

MỤC LỤC

DANH MỤC CÁC CHỮ VIẾT TẮT

Từ viết tắt	Diễn giải
VUI	Voice User Interface (Giao diện giọng nói người dùng)
GUI	Graphic User Intreface (Giao diện đồ họa người dùng)
STFT	Short-time Fourier transform (Biến đổi Fourier thời gian ngắn)
LPC	Linear Predictive Coding (Mã hóa dự đoán tuyến tính)
FFT	Fast Fourier Transform (Biến đổi Fourier nhanh)
IFFT	Inverse Fast Fourier Transform (Biến đổi ngược Fourier nhanh)
DFT	Discrete Fourier Transform (Biến đổi Fourier rời rạc)
STD	Standard Deviation (Độ lệch chuẩn)
CV	Coefficient of Variation (Hệ số biến thiên)

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH

MỞ ĐẦU

1. Lý do chọn đề tài

Thông tin tiếng nói là loại hình thông tin phổ biến nhất trong các hệ thống thông tin viễn thông hiện nay. Do vậy lĩnh vực nghiên cứu về tiếng nói và xử lý tiếng nói được rất nhiều nhà nghiên cứu trong ngành công nghệ thông tin quan tâm. Về cơ bản tiếng nói là một loại tín hiệu một chiều điển hình nên các kiến thức về xử lý tín hiệu hoàn toàn có thể áp dụng với tín hiệu tiếng nói. Đó cũng là một điều thuận lợi đối với những nhà nghiên cứu về xử lý tiếng nói vì lý thuyết và công nghệ xử lý tín hiệu đã có những bước phát triển to lớn và được ứng dụng rộng rãi trong thời gian gần đây. Xử lý tiếng nói bao gồm nhiều lĩnh vực như triệt nhiễu và nâng cao chất lượng tiếng nói, mã hóa và nén tiếng nói, tổng hợp tiếng nói, nhận dạng tiếng nói, ...

Trong vòng 50 năm qua, công nghệ thông tin đã phát triển nhanh chóng và mạnh mẽ. Trong xu hướng chung đó cùng với vai trò của mạng Internet và thông tin di động viễn thông nói riêng, sự phát triển các hệ thống tự động nhận dạng và tổng hợp tiếng nói như là một nhu cầu tất yếu. Trên thế giới đã có những bộ phần mềm thương mại thuộc lĩnh vực này dành cho tiếng Anh như: IBM Via Voice, Dragon Naturally Speaking, L&H Voice Xpress. Gần đây nhất, hãng Microsoft đã công bố việc tích hợp VUI (Voice User Interface) thay cho GUI (Graphic User Intreface) truyền thống vào phiên bản điều hành Windows thế hệ mới.

Tín hiệu tiếng nói là tín hiệu thay đổi theo thời gian. Nó có các đặc trưng cơ bản như nguồn kích thích (excitation), cường độ (pitch), biên độ (amplitude), ... Các tham số thay đổi theo thời gian của tín hiệu tiếng nói có thể kể đến là tần số cơ bản (fundamental frequency - pitch), loại âm (âm hữu thanh - voiced, vô thanh - unvoiced, tắc - fricative hay khoảng lặng - silence), các tần số cộng hưởng chính (formant), hàm diện tích của tuyến âm (vocal tract area), ...

Trong phạm vi tạo tiếng nói, những tần số cộng hưởng của tuyến âm được gọi là tần số formant hay đơn giản là formant. Những tần số này phụ thuộc vào dạng và kích thước của tuyến âm, do đó mỗi dạng tuyến âm được đặc trưng bằng một tổ hợp tần số formant. Các âm khác nhau được tạo bởi sự thay đổi dạng của tuyến âm. Cùng một người phát âm nhưng formant có thể khác nhau. Nếu chỉ căn cứ vào giá trị của formant để đặc trưng cho âm hữu thanh thì chưa chính xác mà phải dựa vào phân bố tương đối giữa các formant. Ngoài ra, nếu xác định formant trực tiếp từ phổ thì không chính xác mà phải dựa vào đường bao phổ, đây cũng chính là đáp ứng tần số của tuyến âm.

Formant của tín hiệu tiếng nói là một trong các tham số quan trọng và hữu ích có ứng dụng rộng rãi trong nhiều lĩnh vực chẳng hạn như trong việc xử lý, tổng hợp và nhận dạng tiếng nói. Formant thường được thể hiện trong các biểu diễn phổ chẳng hạn như trong biểu diễn spectrogram như là một vùng có năng lượng cao, và chúng biến đổi chậm theo thời gian theo hoạt động của bộ máy phát âm. Sở dĩ formant có vai trò quan trọng và là một tham số hữu ích trong các nghiên cứu xử lý tiếng nói là vì các formant có thể miêu tả được các khía cạnh quan trọng nhất của tiếng nói bằng việc sử dụng một tập rất hạn chế các đặc trưng. Chẳng hạn trong mã hóa tiếng nói, nếu sử dụng các tham số formant để biểu diễn cấu hình của bộ máy phát âm và một vài tham số phụ trợ biểu diễn nguồn kích thích, có thể đạt được tốc độ mã hóa thấp đến 2,4kbps.

Nhiều nghiên cứu về xử lý và nhận dạng tiếng nói đã chỉ ra rằng các tham số formant là ứng cử viên tốt nhất cho việc biểu diễn phổ của bộ máy phát âm một cách hiệu quả. Tuy nhiên việc xác định tần số formant không đơn giản chỉ là việc xác định các đỉnh trong phổ biên độ bởi vì các đỉnh phổ của tín hiệu ra của bộ máy phát âm phụ thuộc một cách phức tạp vào nhiều yếu tố chẳng hạn như cấu hình bộ máy phát âm, các nguồn kích thích, ...

Các phương pháp xác định formant liên quan đến việc tìm kiếm các đỉnh trong biểu diễn phổ, thường là từ kết quả phân tích phổ theo phương pháp STFT

(Short-time Fourier transform) hoặc mã hóa dự đoán tuyến tính (Linear Predictive Coding) [5]. Phương pháp xử lý đồng hình là một trong các kỹ thuật phổ biến để xác định tần số formant. Xuất phát từ những lý do trên, tôi thực hiện nghiên cứu ứng dụng phương pháp xử lý đồng hình xác định tần số formant trong phân tích nguyên âm của nhiều người nói nhằm khảo sát tính đặc trưng của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói [11], [13], [14].

2. Mục đích và ý nghĩa đề tài

2.1. Mục đích

Mục đích nghiên cứu đề tài:

- Nghiên cứu và cài đặt thuật toán xác định tần số formant trong phân tích nguyên âm của nhiều người nói dựa trên phương pháp xử lý đồng hình.
- Ứng dụng phân tích đặc trưng của các nguyên âm của nhiều người nói khác nhau.
- Phân tích ưu nhược điểm của phương pháp xử lý đồng hình.

2.2. Ý nghĩa khoa học và thực tiễn của đề tài

- Đóng góp và giải thích phương pháp xác định tần số formant trong lĩnh vực xử lý tín hiệu tiếng nói.
- Là cơ sở cho các nghiên cứu khác áp dụng để xác định tần số formant trên các phương pháp khác nhau trong tương lai.

3. Mục tiêu và nhiệm vụ

3.1. Mục tiêu

Mục tiêu chính của đề tài là nghiên cứu ứng dụng xử lý đồng hình nhằm xác định tần số formant trong phân tích nguyên âm của nhiều người nói và phân tích ưu nhược điểm của phương pháp.

3.2. Nhiệm vụ

Để đạt được mục tiêu, nhiệm vụ đặt ra của đề tài là:

- Nghiên cứu lý thuyết liên quan đến tần số formant.

- Nghiên cứu lý thuyết, công thức liên quan phương pháp xử lý đồng hình.

- Thực hiện phân tích, đánh giá kết quả xác định tần số formant trong phân tích nguyên âm của nhiều người nói dựa trên phương pháp xử lý đồng hình.

4. Đối tượng và phạm vi nghiên cứu

4.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài là tín hiệu tiếng nói, các thuật toán xử lý tín hiệu tiếng nói và các nguyên âm của nhiều người nói.

4.2. Phạm vi nghiên cứu

Phạm vi nghiên cứu của đề tài là:

- Phương pháp xác định tần số formant của tín hiệu tiếng nói.

- 05 nguyên âm chính: /a/, /e/, /i/, /o/, /u/.

- Người nói giọng miền Trung.

5. Phương pháp nghiên cứu

5.1. Phương pháp lý thuyết

Thu thập và nghiên cứu các tài liệu liên quan đến đề tài.

5.2. Phương pháp thực nghiệm

- Nghiên cứu và khai thác các công cụ, phần mềm hỗ trợ.

- Phân tích, đánh giá kết quả xác định tần số formant trong phân tích nguyên âm của nhiều người nói dựa trên phương pháp xử lý đồng hình.

6. Kết luận

6.1. Kết quả của đề tài

- Nghiên cứu và tính được tần số formant dựa trên phương pháp xử lý đồng hình.

- Đánh giá kết quả phương pháp, phân tích ưu nhược điểm.

- Đưa ra bảng thống kê đặc trưng formant của 05 nguyên âm chính của nhiều người nói giọng miền Trung, từ đó đánh giá khả năng dùng đặc trưng formant để phân biệt các người nói với nhau và phân biệt chất giọng theo tỉnh/thành phố.

6.2. Hướng phát triển của đề tài

- Nghiên cứu giải pháp để cải thiện độ chính xác của phương pháp xử lý đồng hình để xác định tần số formant.

- Nghiên cứu so sánh các phương pháp xác định tần số formant.

7. Bố cục của luận văn

Dự kiến luận văn được trình bày bao gồm các phần chính như sau:

MỞ ĐẦU

Nêu bối cảnh nghiên cứu, lý do chọn đề tài và mục tiêu nghiên cứu.

CHƯƠNG I: TỔNG QUAN VỀ XỬ LÝ TÍN HIỆU TIẾNG NÓI

Trình bày các khái niệm cơ bản của tiếng nói, quá trình hình thành tiếng nói, mô hình nguồn-bộ lọc của tín hiệu tiếng nói, các đặc tính cơ bản của tín hiệu tiếng nói và kỹ thuật xử lý tín hiệu tiếng nói ngắn hạn (short-time).

CHƯƠNG II: XÁC ĐỊNH TẦN SỐ FORMANT DÙNG XỬ LÝ ĐỒNG HÌNH

- Trình bày khái niệm tần số formant, đặc điểm cấu trúc formant của các nguyên âm.

- Trình bày lý thuyết, phân tích ưu/nhược điểm của phương pháp xử lý đồng hình nhằm xác định tần số formant.

CHƯƠNG III: TRIỂN KHAI VÀ ĐÁNH GIÁ THUẬT TOÁN

- Để áp dụng được thuật toán trên Matlab, trong chương này trình bày công cụ Matlab và các hàm liên quan đến xử lý tín hiệu tiếng nói [9], [15].

- Áp dụng phương pháp xử lý đồng hình xác định tần số formant trong phân tích nguyên âm của nhiều người nói.

- Đánh giá kết quả thu được dựa trên dữ liệu tiếng nói tự thu thập.

- Rút ra được ưu nhược điểm của phương pháp.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

PHỤ LỤC

TÀI LIỆU THAM KHẢO

CHƯƠNG I: TỔNG QUAN VỀ XỬ LÝ TÍN HIỆU TIẾNG NÓI

Ngày nay với sự phát triển của công nghệ, để có sự giao tiếp trở nên linh hoạt hơn, tiếng nói như là một công cụ hỗ trợ mạnh mẽ để thúc đẩy việc biểu diễn tiếng nói trong khoa học máy tính. Tiếng nói được sử dụng như là một dữ liệu được lưu trữ trong máy tính, qua đó có thể truyền đạt thông qua mạng truyền thông để phục vụ nhiều mục đích khác nhau để phục vụ lợi ích trong đời sống của con người. Trong các hệ thống xử lý tiếng nói, cần chú ý đến hai điểm: sự nguyên vẹn của nội dung thông điệp trong tín hiệu tiếng nói; biểu diễn tín hiệu tiếng nói phải tiện lợi cho việc truyền tải, lưu trữ hoặc trong một dạng linh động để có thể chuyển đổi thành tín hiệu tiếng nói mà không giảm nội dung của thông điệp.

Trong chương này, tôi trình bày tổng quan về khái niệm tiếng nói bao gồm nguồn gốc, quá trình hình thành, phân loại tiếng nói và biểu diễn tín hiệu tiếng nói; các đặc tính cơ bản của tín hiệu tiếng nói như cao độ, cường độ, trường độ và phổ. Cuối chương tập trung trình bày về kỹ thuật xử lý tín hiệu tiếng nói ngắn hạn.

1.1. Khái niệm tiếng nói

Tiếng nói thường xuất hiện dưới nhiều hình thức gọi là đàm thoại, việc đàm thoại thể hiện kinh nghiệm của con người. Đàm thoại là một quá trình gồm nhiều người, có sự hiểu hiết chung và một nghi thức luân phiên nhau nói. Những người có điều kiện thể chất và tinh thần bình thường thì rất dễ diễn đạt tiếng nói của mình, do đó tiếng nói là phương tiện giao tiếp chính trong lúc đàm thoại. Tiếng nói có rất nhiều yếu tố khác hỗ trợ nhằm giúp người nghe hiểu được ý cần diễn đạt như biểu hiện trên gương mặt, cử chỉ, điệu bộ. Vì có đặc tính tác động qua lại, nên tiếng nói được sử dụng trong nhu cầu giao tiếp nhanh chóng. Trong khi đó, chữ viết lại có khoảng cách về không gian lẫn thời gian giữa tác giả và người đọc. Sự biểu đạt của tiếng nói hỗ trợ mạnh mẽ cho việc

ra đời các hệ thống máy tính có sử dụng tiếng nói, ví dụ như lưu trữ tiếng nói như là một loại dữ liệu, hay dùng tiếng nói làm phương tiện giao tiếp qua lại. Nếu có thể phân tích quá trình giao tiếp qua nhiều lớp, thì lớp thấp nhất chính là âm thanh và lớp cuối cùng là tiếng nói diễn tả ý nghĩa muốn nói.

1.1.1. Nguồn gốc của tiếng nói

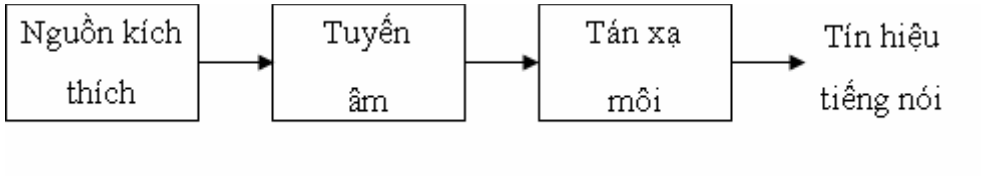
Âm thanh của lời nói cũng như âm thanh trong thế giới tự nhiên xung quanh, về bản chất đều là những sóng âm được lan truyền trong một môi trường nhất định (thường là không khí). Tai con người chỉ cảm thụ được những dao động có tần số từ khoảng 16Hz đến khoảng 20000Hz. Những dao động trong miền tần số này gọi là dao động âm hay âm thanh, và các sóng tương ứng gọi là sóng âm. Những sóng có tần số nhỏ hơn 16Hz gọi là sóng hạ âm, những sóng có tần số lớn hơn 20000Hz gọi là sóng siêu âm, con người không cảm nhận được (ví dụ loài dơi có thể nghe được tiếng siêu âm). Sóng âm, sóng siêu âm và hạ âm không chỉ truyền trong không khí mà còn có thể lan truyền tốt ở những môi trường rắn, lỏng, do đó cũng được sử dụng rất nhiều trong các thiết bị máy móc hiện nay.

1.1.2. Quá trình hình thành tiếng nói

1.1.2.1. Cấu tạo của hệ thống cấu âm

Tính chất phổ của tín hiệu tiếng nói thay đổi theo thời gian giống với sự thay đổi dạng của tuyến âm. Quá trình truyền âm qua tuyến âm làm mạnh lên ở một vùng tần số nào đó bằng cộng hưởng và tạo cho mỗi âm những tính chất riêng biệt gọi là quá trình phát âm. Âm được phát có nghĩa nó đã mang thông tin về âm vị được tán xạ ra ngoài từ môi. Trong một vài trường hợp, đối với những âm mũi (như /m/, /n/ trong tiếng Anh), tuyến mũi cũng tham gia vào quá trình phát âm và âm được tán xạ ra từ mũi.

Tóm lại, sóng tín hiệu được chế tạo bằng ba động tác: tạo nguồn âm (hữu thanh và vô thanh), phát âm khi truyền qua tuyến âm và tán xạ âm từ môi hoặc từ mũi (Hình 1.1).



Hình 1.1. Quá trình cơ bản tạo tín hiệu tiếng nói [5].

1.1.2.2. Cấu tạo của hệ thống tiếp âm

Không giống như các cơ quan tham gia vào quá trình tạo ra tiếng nói khi thực hiện các chức năng khác trong cơ thể như: thở, ăn, ngửi. Tai chỉ sử dụng cho chức năng nghe. Tai đặc biệt nhạy cảm với những tần số trong tín hiệu tiếng nói chứa thông tin phù hợp nhất với việc liên lạc (những tần số xấp xỉ 200 - 5600Hz). Người nghe có thể phân biệt được những sự khác biệt nhỏ trong thời gian và tần số của những âm thanh nằm trong vùng tần số này.

Tai gồm có ba phần: tai ngoài, tai giữa và tai trong. Tai ngoài dẫn hướng những thay đổi áp xuất tiếng nói vào trong màng nhĩ, ở đó tai giữa sẽ chuyển đổi áp xuất này thành chuyển động cơ học. Tai trong chuyển đổi những rung động cơ học này thành những luồng điện trong nơron thính giác dẫn đến não.

1.1.3. Phân loại tiếng nói

Về cơ bản chia tiếng nói thành 3 loại như sau:

• **Âm hữu thanh**: Là âm khi phát ra thì có thanh, ví dụ như nói /i/, /a/, hay /o/ chẳng hạn. Thực ra âm hữu thanh được tạo ra là do việc không khí qua thanh môn (thanh môn tạo ra sự khép mở của dây thanh dưới sự điều khiển của hai sụn chóp) với một độ căng của dây thanh sao cho chúng tạo nên dao động.

Trong luận văn này tiến hành nghiên cứu phân tích các âm hữu thanh của người nói và tập trung chủ yếu ở 5 nguyên âm: /a/, /e/, /i/, /o/, /u/.

• **Âm vô thanh**: Là âm khi tạo ra tiếng thì dây thanh không rung hoặc rung đôi chút tạo ra giọng như giọng thở, ví dụ /h/, /p/ hay /th/.

• **Âm bật**: để phát ra âm bật, đầu tiên bộ máy phát âm phải đóng kín, tạo nên một áp suất, sau đó không khí được giải phóng một cách đột ngột, ví dụ /ch/, /t/.

1.1.4. Biểu diễn tín hiệu tiếng nói

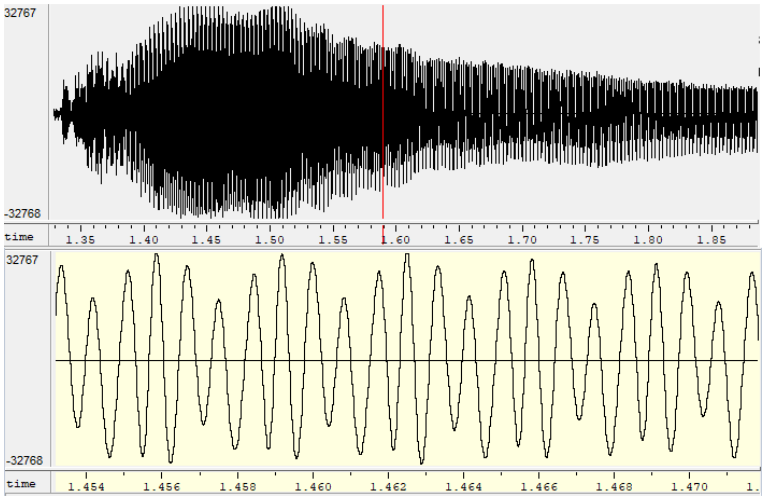
Có 3 phương pháp biểu diễn tín hiệu tiếng nói cơ bản là:

- Biểu diến dưới dạng sóng theo thời gian.
- Biểu diến trong miền tần số: phổ của tín hiệu tiếng nói.
- Biểu diễn trong không gian 3 chiều (Sonagram).

1.1.4.1. Dạng sóng theo thời gian

Phần tín hiệu ứng với âm vô thanh là không tuần hoàn, ngẫu nhiên và có biên độ hay năng lượng nhỏ hơn của nguyên âm (cỡ khoảng 1/3).

Ranh giới giữa các từ: là các khoảng lặng (Silent). Cần phân biệt rõ các khoảng lặng với âm vô thanh.

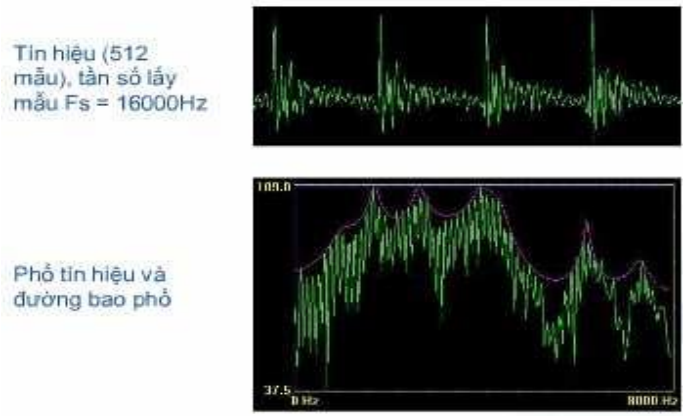


Hình 1.2. Dạng sóng theo thời gian [5].

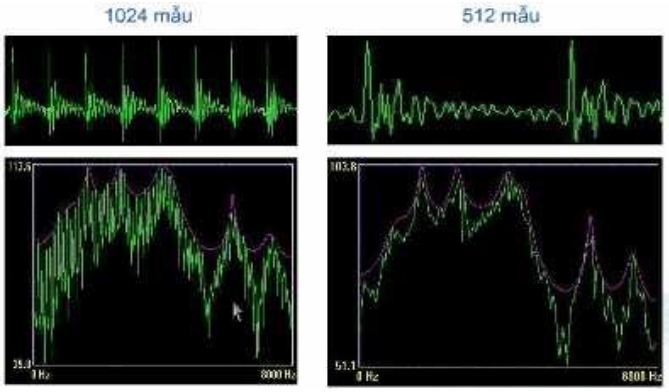
Âm thanh dưới dạng sóng được lưu trữ theo định dạng thông dụng trong máy tính là *.WAV với các tần số lấy mẫu thường gặp là: 8000Hz, 10000Hz, 11025Hz, 16000Hz, 22050Hz, 32000Hz, 44100Hz,…; độ phân giải hay còn gọi là số bít/mẫu là 8 hoặc 16 bít và số kênh là 1 (Mono) hoặc 2 (Stereo).

1.1.4.2. Phổ tín hiệu tiếng nói

Dải tần số của tín hiệu âm thanh là khoảng từ 0Hz đến 20KHz, tuy nhiên phần lớn công suất nằm trong dải tần số từ 0,3KHz đến 3,4KHz. Dưới đây là một số hình ảnh của phổ tín hiệu tiếng nói:



Hình 1.3. Phổ tín hiệu tiếng nói và đường bao phổ [5].



Hình 1.4. Phổ tín hiệu tiếng nói với số mẫu khác nhau [5].

1.1.4.3. Biểu diễn tín hiệu tiếng nói trong không gian ba chiều (Sonagram)

Để biểu diễn trong không gian 3 chiều, chia tín hiệu thành các khung cửa sổ (frame) ứng với các ô quan sát (Hình 1.5).

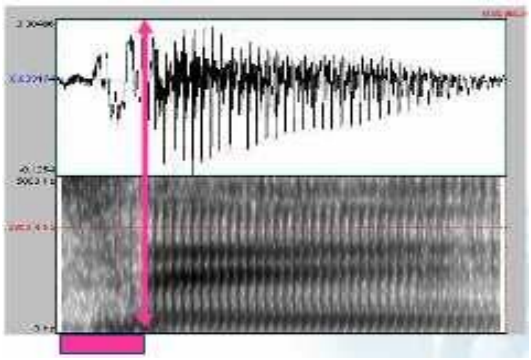
Hình 1.5. Chia tín hiệu thành các khung cửa sổ [5].

Độ dài một cửa sổ tương ứng là 10ms.

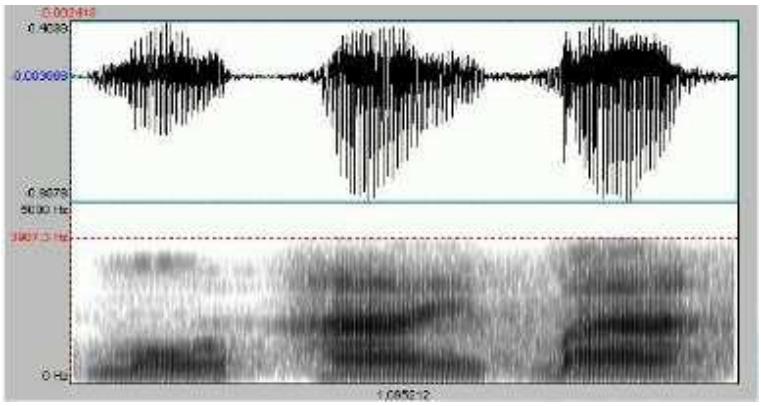
Vậy, nếu tần số Fs = 16000Hz thì có 160 mẫu trên một cửa sổ.

Các cửa số có đoạn chồng lẫn lên nhau (khoảng 1/2 cửa sổ).

Tiếp theo vẽ phổ của khung tín hiệu trên trục thẳng đứng, biên độ phổ biểu diễn bằng độ đậm, nhạt của màu sắc. Sau đó vẽ theo trục thời gian bằng cách chuyển sang cửa số tiếp theo.



Hình 1.6. Phổ của một khung cửa sổ [5].



Hình 1.7. Các khung cửa sổ liền nhau và spectrogram tương ứng [5].

Biểu diễn tín hiệu tiếng nói theo không gian 3 chiều là một công cụ rất mạnh để quan sát và phân tích tín hiệu. Ví dụ: theo phương thức biểu diễn này ta có thể dễ dàng phân biệt âm vô thanh và âm hữu thanh dựa theo các đặc điểm sau:

+ Âm vô thanh: Năng lượng tập trung ở tần số cao. Các tần số phân bố khá đồng đều trong 2 miền tần số cao và tần số thấp.

+ Âm hữu thanh: Năng lượng tập không đồng đều. Có những vạch cực trị.

1.2. Các đặc tính cơ bản của tín hiệu tiếng nói

1.2.1. Cao độ

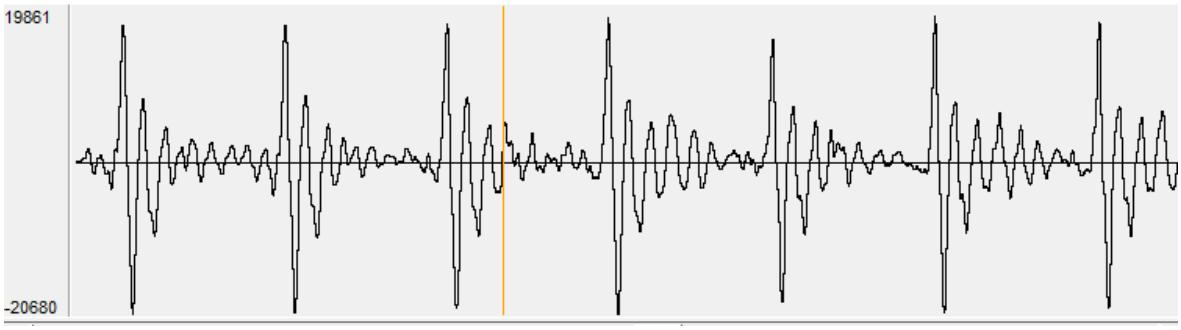
Hai thuộc tính sinh âm nổi bật giữ vị trí quan trọng trong việc miêu tả giọng nói: một là, cao độ là một chỉ tố xác định nổi bật nhất của sự nổi trội. Nó

thường được đo theo tỉ lệ, ví dụ, tần số dao động của thanh quản trong quá trình sinh âm, có thể được đo trực tiếp từ dạng sóng lời nói; và hai là, cường độ, như một yếu tố xác định chính của cường độ lời nói nói chung. Không thể gộp các kiểu sinh âm theo một cách hoàn toàn giống nhau, bởi vì mặc dù chúng có thể được phân loại một cách dễ dàng về mặt thính giác, nhưng chúng nằm trong mối quan hệ phức tạp hơn nhiều, gián tiếp hơn và không nhất quán với các giá trị âm học khác nhau.

Cao độ thể hiện bằng vùng tần số cơ bản phản ánh những sự khác nhau có tính chất sinh học về thanh quản, đặc biệt ở chiều dài và các cấu trúc cơ của các khe thanh ở nam giới, nữ giới và trẻ em. Tần số càng lớn âm phát ra càng cao. Một âm thanh có thể là tổ hợp của nhiều tần số, trong tiếng nói, tần số cơ bản là đáp ứng của sự rung động các dây thanh âm, tần số cơ bản thường được ký hiệu là F0. Tần số cơ bản phụ thuộc vào khối lượng và sự căng của đôi dây thanh. Dây thanh của phụ nữ, trẻ em thường mảnh hơn và căng hơn của đàn ông, người già do đó âm phát ra có tần số cao hơn.

1.2.2. Cường độ

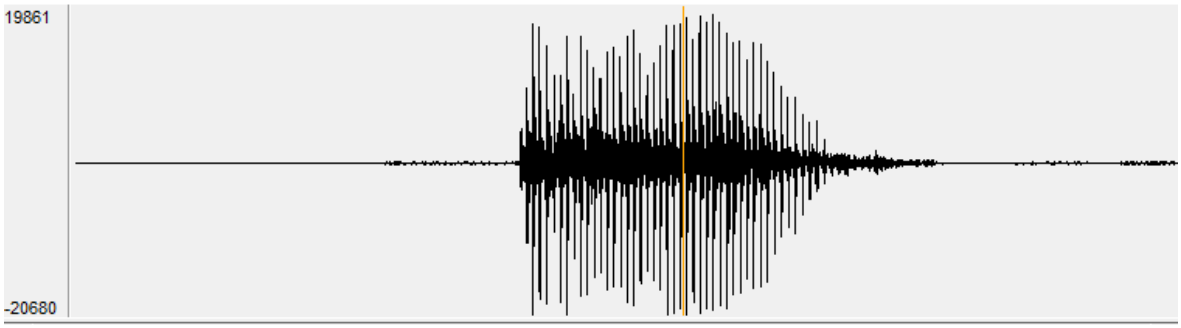
Cường độ là độ to hay nhỏ của âm thanh nói ra. Cường độ càng lớn thì âm thanh truyền càng xa trong môi trường truyền. Cường độ âm là số năng lượng mà sóng âm truyền đi trong một thời gian nhất định trên đơn vị diện tích cố định và vuông góc với phương truyền âm. Cường độ của âm thanh không ảnh hưởng đến những đặc điểm về phẩm chất, tức về âm sắc của nguyên âm. Cường độ của nguyên âm tùy thuộc trước hết vào mức độ to nhỏ của toàn câu nói, ngoài ra cũng tùy thuộc vào vị trí của nguyên âm đối với trọng âm từ và trọng âm câu. Nếu trọng âm là trọng âm lực thì nguyên âm có trọng âm sẽ mạnh hơn nguyên âm không có trọng âm, và ngược lại. Trong tiếng nói, cường độ của nguyên âm thường lớn cường độ của phụ âm. Trên đồ thị biểu diễn sóng tín hiệu (waveform), cường độ âm thanh tỉ lệ thuận với giá trị tuyệt đối của biên độ tín hiệu.

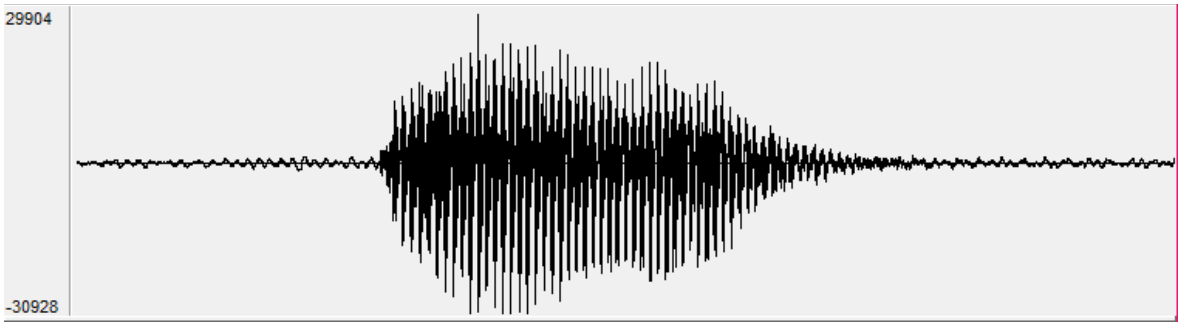


Hình 1.8. Đồ thị biểu diễn sóng tín hiệu của nguyên âm /a/ của một người nói.

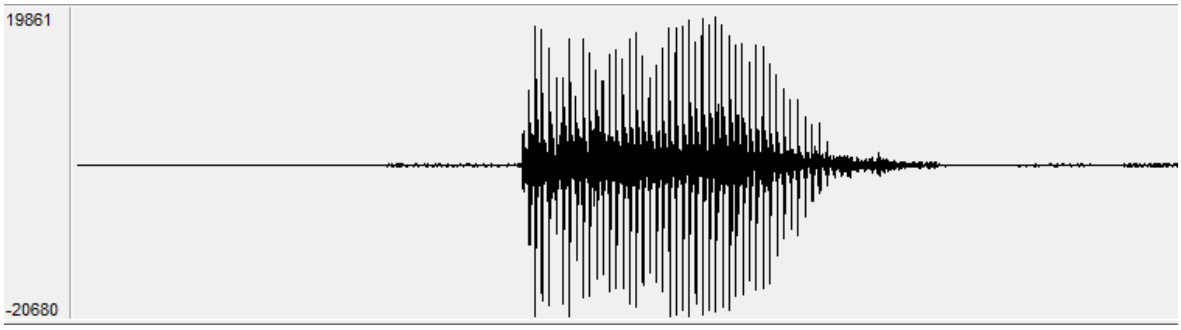
1.2.3. Trường độ

Trường độ là độ dài của âm thanh hay nói cách khác là thời gian diễn ra dao động sóng âm từ lúc bắt đầu đến khi kết thúc tạo nên sự tương phản giữa các bộ phận của lời nói. Nó là yếu tố tạo nên sự đối lập giữa nguyên âm này với nguyên âm khác trong một số ngôn ngữ. Đơn vị đo trường độ tính bằng mili giây (1 ms = 1/1000s). Không có quy luật chung về trường độ tất yếu cho mọi ngôn ngữ. Quy luật duy nhất có thể được xem là phổ biến đó là trường độ của nguyên âm phụ thuộc vào nhịp điệu nói. Đối với mỗi ngôn ngữ trường độ trung bình của một nguyên âm ở một vị trí nhất định là một đại lượng ít nhiều cố định. Trường độ thường lệ thuộc vào những điều kiện ngữ âm học, hay nói cách khác là phụ thuộc vào vị trí ngữ âm học. Trường độ trong âm tiết khép và trong âm tiết mở nhiều khi khác nhau, nó cũng có thể phụ thuộc vào tính chất của phụ âm đi sau (tắc, xát, hữu thanh, vô thanh), vào số lượng phụ âm đi sau, vào vị trí của trọng âm và vào số âm tiết có trong từ. Ngoài ra, trường độ của nguyên âm cũng lệ thuộc một phần vào phẩm chất của nó.

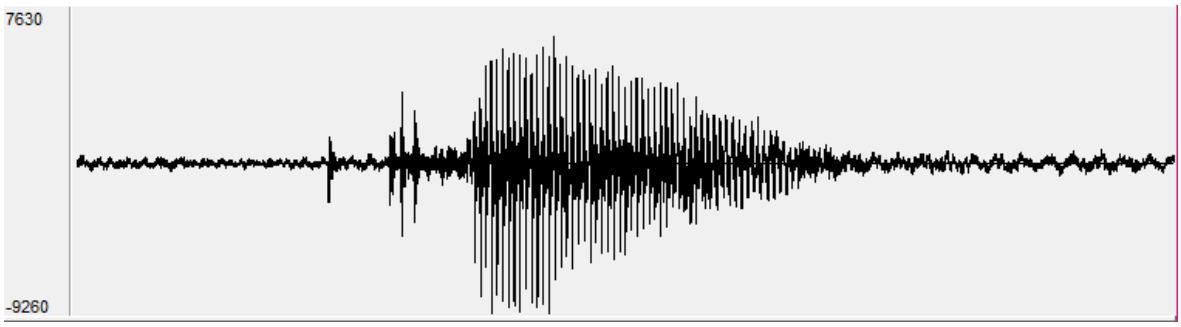




Hình 1.9. Nguyên âm /a/ ở hai thời điểm khác nhau của cùng một người nói.



Hình 1.10. Âm /a/ của một người nam.

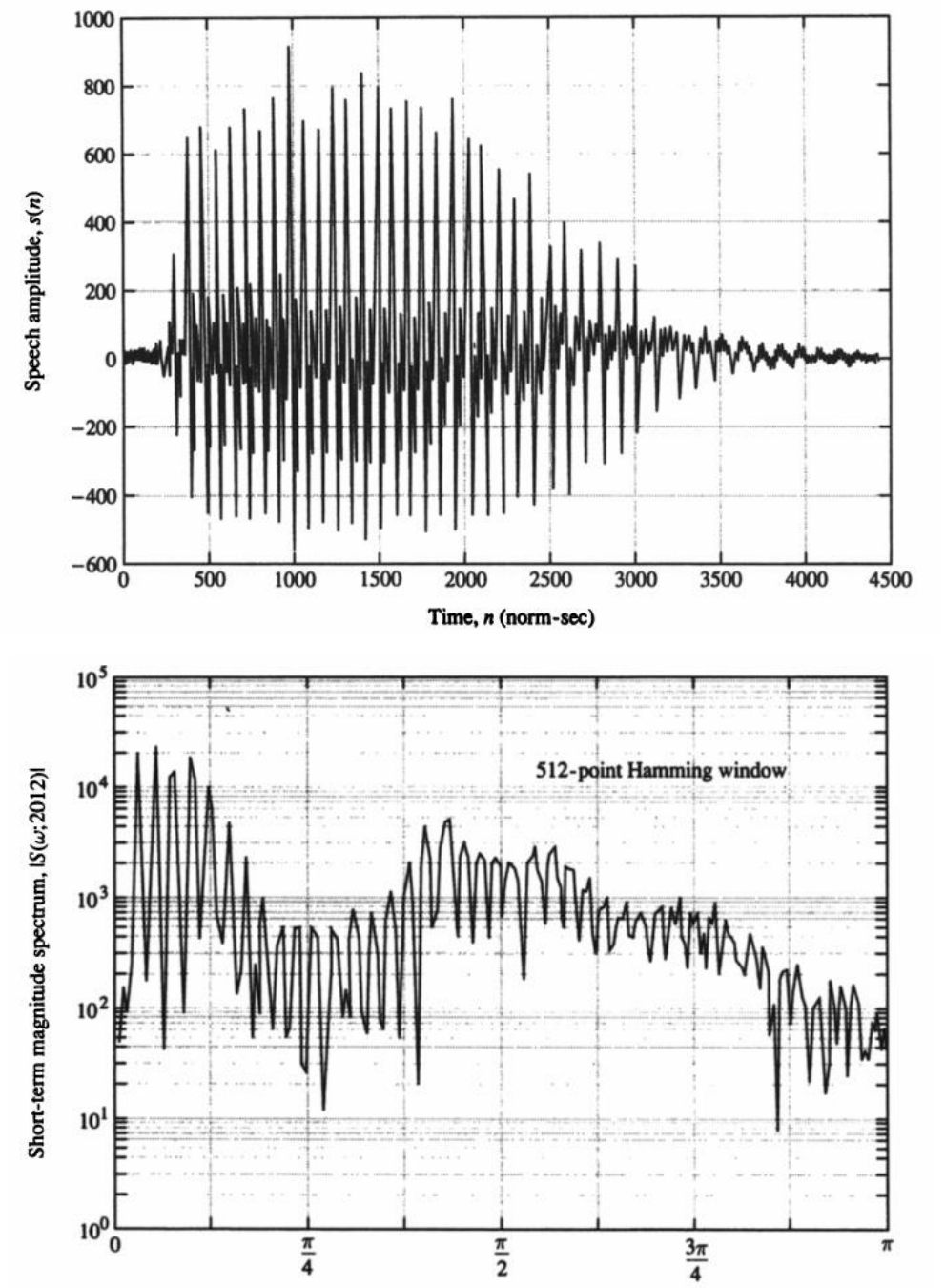


Hình 1.11. Âm /a/ của một người nữ.

1.2.4. Phổ

Trong phân tích tín hiệu tiếng nói, thay vì sử dụng trực tiếp tín hiệu tiếng nói trong miền thời gian, người ta thường hay sử dụng các đặc trưng phổ của tiếng nói. Điều này xuất phát từ quan điểm rằng tín hiệu tiếng nói cũng giống như các tín hiệu xác định khác có thể xem như là tổng của các tín hiệu hình sin với biên độ và pha thay đổi chậm. Hơn nữa, một nguyên nhân quan trọng không kém đó là việc cảm nhận tiếng nói của con người liên quan trực tiếp đến thông tin phổ của tín hiệu tiếng nói nhiều hơn trong khi các thông tin về pha của tín hiệu tiếng nói không có vai trò quyết định.

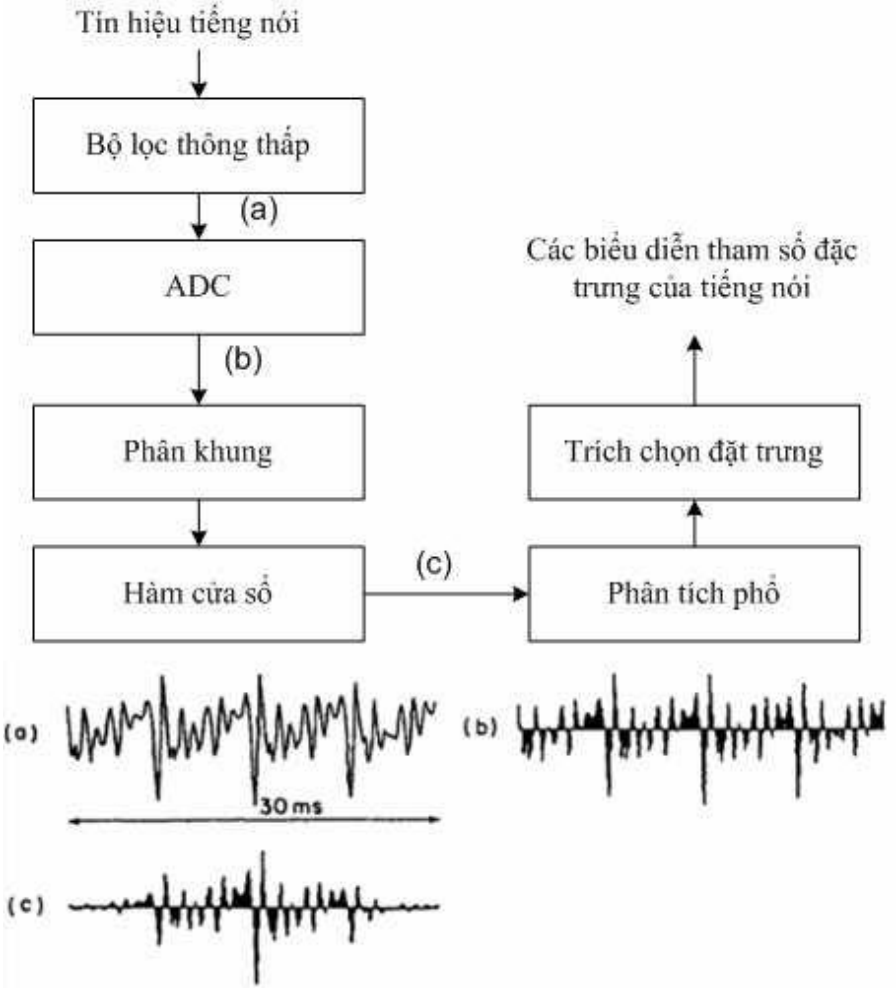
Mật độ phổ công suất trong một khoảng thời gian ngắn, tức là phổ ngắn hạn của tín hiệu tiếng nói, có thể được xem như là tích của hai thành phần: thành phần thứ nhất là đường biên phổ thay đổi một cách chậm chạp theo tần số; thành phần thứ hai là cấu trúc phổ mịn thay đổi rất nhanh theo tần số. Đối với các âm hữu thanh thì cấu trúc phổ mịn tạo thành các mẫu tuần hoàn, còn đối với các âm vô thanh thì không.



Hình 1.12. Khung tín hiệu và phổ tương ứng [5].

1.3. Xử lý tín hiệu tiếng nói ngắn hạn

Tín hiệu tiếng nói là tín hiệu thay đổi theo thời gian. Các tham số thay đổi theo thời gian của tín hiệu tiếng nói có thể kể đến là tần số cơ bản, loại âm (âm hữu thanh, vô thanh, …), các tần số cộng hưởng chính (formant), …



Hình 1.13. Mô hình tổng quát của việc xử lý tín hiệu tiếng nói [5].

Việc thực hiện phân tích tiếng nói ngắn hạn tức là xem xét tín hiệu tiếng nói trong một khoảng nhỏ thời gian xung quanh thời điểm đang xét n nào đó. Các khoảng này thường khoảng từ 10-30ms. Khoảng nhỏ tín hiệu dùng để phân tích thường được gọi là một khung (frame), các khung này có thể trùng nhau (overlap) một phần để đảm bảo các đặc tính của tín hiệu biến đổi trơn tru giữa hai khung liên tiếp. Việc chia khung này sẽ được lặp lại từ đầu đến cuối trên tín hiệu cần xử lý. Kết quả của việc xử lý trên mỗi khung có thể chỉ gồm một giá

trị số (ví dụ như giá trị năng lượng hoặc giá trị F0), có thể gồm nhiều giá trị số (ví dụ như các hệ số phổ). Việc chia tín hiệu tiếng nói thành các khung tín hiệu giúp ta xác định và xử lý được các tín hiệu tiếng nói có đặc tính hầu như không thay đổi, độc lập.

Một phép phân tích tín hiệu tiếng nói ngắn hạn tổng quát có thể biểu diễn là:

$$X_n(m) = \sum_{m=-\infty}^{\infty} T\{s(m)w(n-m)\} \qquad (1)$$

Trong đó: X_n biểu diễn tham số phân tích tín hiệu tiếng nói ngắn hạn tại thời điểm phân tích n. *T{}:* khung tín hiệu ngắn hạn, tuyến tính hoặc phi tuyến tính, có kết quả là các cửa sổ có trình tự và vị trí, thời gian tương ứng với mẫu chỉ số n. Giá trị X_n được tính là tổng giá trị các số khác không của khung tín hiệu *T{}* với mọi giá trị của m trong tập xác định của hàm cửa sổ.

1.4. Tổng kết chương

Trong xử lý tín hiệu tiếng nói, tiếng nói được biểu diễn trên miền thời gian và trên miền tần số. Tín hiệu tiếng nói được biểu diễn trên miền thời gian là đồ thị biểu diễn tín hiệu tiếng nói theo trục thời gian. Tín hiệu tiếng nói được biểu diễn trên miền tần số là đồ thị biểu diễn tín hiệu tiếng nói theo trục tần số.

Tiếng nói ở mỗi người đều có đặc trưng khác nhau. Các đặc trưng này được tạo nên từ cao độ, cường độ, trường độ. Ở mỗi người, các đại lượng này có sự khác biệt nên tiếng nói cảm nhận được đều khác nhau. Trong lĩnh vực xử lý tín hiệu tiếng nói, tần số formant (hay còn gọi là formant) là đặc trưng quan trọng của tín hiệu tiếng nói. Để tìm formant của tín hiệu tiếng nói, cần dùng đến kỹ thuật xử lý ngắn hạn chia tín hiệu tiếng nói thành nhiều khung nhỏ để xử lý.

Chương này đã tập trung tìm hiểu, nghiên cứu để làm rõ lý thuyết tổng quan về khái niệm tiếng nói; các đặc tính cơ bản của tín hiệu tiếng nói như cao độ, cường độ, trường độ và phổ; cuối cùng nói về kỹ thuật xử lý tín hiệu tiếng nói ngắn hạn. Qua chương 2 tôi sẽ trình bày tổng quan về tần số formant và ứng

dụng phương pháp xử lý đồng hình để xác định tần số formant cũng như nêu rõ thuật toán và đánh giá ưu nhược điểm của thuật toán.

CHƯƠNG II: XÁC ĐỊNH TẦN SỐ FORMANT DÙNG XỬ LÝ ĐỒNG HÌNH

Formant của tín hiệu tiếng nói là một trong các tham số quan trọng và hữu ích có ứng dụng rộng rãi trong nhiều lĩnh vực chẳng hạn như trong việc xử lý, tổng hợp và nhận dạng tiếng nói. Xác định được formant chính xác là tiền đề để tiến hành các nghiên cứu khác trong lĩnh vực này.

Trong chương này tôi sẽ trình bày tổng quan về tần số formant bao gồm khái niệm, đặc điểm cấu trúc formant của các nguyên âm và ứng dụng phương pháp xử lý đồng hình để xác định tần số formant, nêu rõ thuật toán tính formant và sau đó đánh giá ưu nhược điểm của thuật toán.

2.1. Tần số formant

2.1.1. Khái niệm

Formant được định nghĩa bởi Gunnar Fant (1960) "*những đỉnh quang phổ của phổ âm thanh được gọi là các formant*". Định nghĩa này được sử dụng rộng rãi trong nghiên cứu ngữ âm học và các xử lí âm thanh trong công nghệ [7]. Một nghiên cứu khác của Han Mieko (1966) đã định nghĩa: "Nguyên âm được mô tả bởi khoang cộng hưởng tương đối lớn trong so sánh với phụ âm. Khoang cộng hưởng này được mạnh thêm ở những vùng khác nhau theo khẩu hình đặc trưng của bộ máy phát âm của âm thanh lời nói. Những vùng có cộng hưởng tăng mạnh được gọi là các formant. Mỗi nguyên âm có một kiểu formant đặc trưng, và thực nghiệm đã chứng tỏ rằng hai formant đầu tiên mang hầu hết thông tin về phẩm chất của nguyên âm" [8].

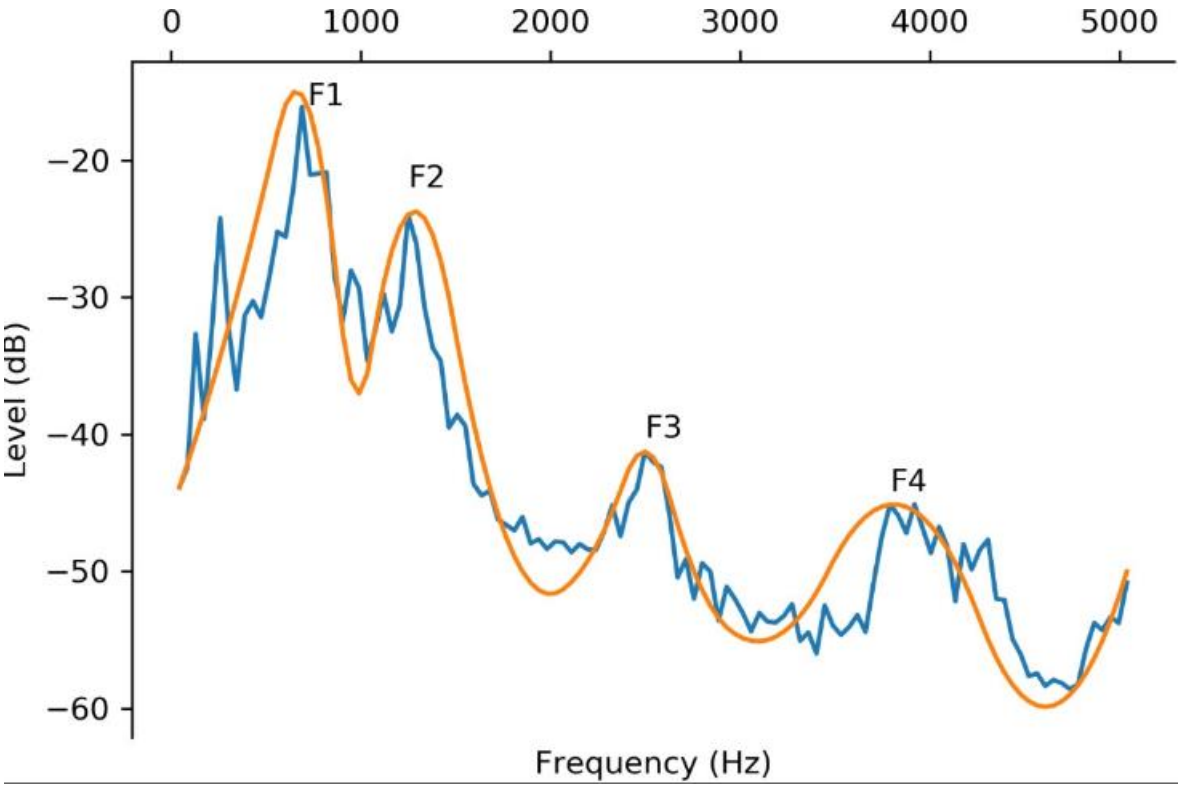
Có hai khoang cộng hưởng quan trọng nhất được ngăn cách và thay đổi do sự chuyển động của lưỡi tạo nên hai formant cơ bản của một nguyên âm:

F1: ứng với cộng hưởng vùng họng.

F2: ứng với cộng hưởng khoang miệng.

Khi người nói, các âm mũi sẽ có sự xuất hiện của formant F3, các formant khác F4, F5… liên quan đến các đặc trưng giọng nói riêng của mỗi cá nhân. Mỗi lần môi, lưỡi, hàm ở những vị trí khác nhau là một lần hộp cộng hưởng miệng và yết hầu thay đổi hình dáng, thể tích, lối thoát của không khí làm biến đổi âm sắc của âm thanh đi qua chúng.

Hiện nay, nhờ sự phát triển của các phần mềm tin học, người ta có thể đo được chính xác tần số formant, số lượng formant của các nguyên âm tương ứng với vị trí cấu âm khác nhau.



Hình 2.1. Đường bao phổ và các tần số formant F1, F2, F3, F4.

(Nguồn: https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/anae.14732)

Các tần số formant có thể được ước lượng từ các tham số dự đoán theo một trong hai cách. Cách thứ nhất là xác định trực tiếp bằng cách phân tích nhân tử đa thức dự đoán và dựa trên các nghiệm thu được để quyết định xem nghiệm nào tương ứng với formant. Cách thứ hai là sử dụng phân tích phổ và chọn các formant tương ứng với các đỉnh nhọn bằng một trong các thuật toán

chọn đỉnh đã biết. Trong phạm vi luận văn, tôi chọn cách thứ hai để xác định các tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình.

2.1.2. Đặc điểm cấu trúc formant của các nguyên âm

Formant và cấu trúc formant của nguyên âm là một trong những lĩnh vực nghiên cứu mang tính ứng dụng cao đã được thực hiện ở nhiều ngôn ngữ trên thế giới. Theo hướng nghiên cứu này, người ta đã thu được nhiều thành công và đã có kết quả nghiên cứu ứng dụng vào công nghệ xử lí tiếng nói: phần mềm tổng hợp tiếng nói thực hiện bằng phương pháp tổng hợp formant đã được tích hợp vào các tiện ích của điện thoại di động, hộp thư trả lời tự động, xếp hàng tự động... Những tiến bộ và điều kiện kĩ thuật hiện nay, đặc biệt là sự phát triển của công nghệ thông tin cho phép các nhà khoa học nghiên cứu một cách toàn diện và có hệ thống đặc trưng âm học của các ngôn ngữ nói chung và của nguyên âm nói riêng.

Vào thế kỉ thứ 19, các nhà khoa học đã nhận thấy vai trò của cộng hưởng trong bộ máy phát âm, đặc biệt là cấu trúc formant của nguyên âm trong việc tạo ra bộ máy phát âm nhân tạo, tiền đề cho những máy tổng hợp lời nói và phần mềm tổng hợp có thể bắt chước giọng nói của con người trong những năm sau này. Các nhà ngữ âm học và cả kĩ sư tin học trong lĩnh vực công nghệ tiếng nói đã quan tâm đến ba khía cạnh của formant trong lời nói tự nhiên: đặc điểm cấu trúc formant, đặc điểm địa phương, và đặc trưng cá nhân của người nói.

Một nghiên cứu vào năm 1973 của Nguyễn Văn Ái khi nghiên cứu thực nghiệm về mặt vật lí, với việc phân tích 400 ảnh sóng âm của 9 nguyên âm đơn tiếng Việt được lấy từ 10 cộng tác viên. Tác giả đi đến kết luận: Ngoài đặc điểm cấu âm đầu tiên chung của các nguyên âm là sự sản sinh tiếng thanh cơ bản trong thanh quản ra, tất cả các nguyên âm tiếng Việt đều được hình thành cuối cùng ở khoang miệng (không có sự tham gia của khoang mũi, vì khi cấu âm, phần sau của ngạc mềm nâng lên đóng chặt đường thông lên khoang mũi). Do đó, chính hiện tượng cấu âm ở khoang miệng: độ mở của miệng, góc độ của

hàm và vị trí của lưỡi là nguyên nhân trực tiếp có tính chất quyết định đến hiệu quả âm học của từng nguyên âm [1]. Sau đó 01 năm, Nguyễn Văn Ái đã phân tích 11 nguyên âm đơn tiếng Việt. Kết quả cho thấy: số lượng formant của mỗi nguyên âm không giống nhau hoàn toàn, thường có từ 2 đến 5 formant [2].

Năm 2002, những kết quả nghiên cứu về hệ formant của 9 nguyên âm đơn tiếng Hà Nội đọc tách rời đã được tác giả Vũ Kim Bảng trình bày trong giới hạn phạm vi nghiên cứu là các cộng tác viên người Hà Nội và kết quả nghiên cứu được tính theo giới tính. Việc trình bày giá trị khách quan của formant tính bằng Hz trong mối tương quan với giá trị cảm nhận tính bằng đơn vị Bark cho phép đưa ra các nhận xét về sự phân bố của hệ thống nguyên âm đơn tiếng Hà Nội. Các đặc trưng âm học khác của tiếng Việt được nghiên cứu theo trình tự âm tố (nguyên âm, phụ âm), âm tiết bao gồm cả thanh điệu và chuỗi lời nói góp phần làm sáng tỏ đặc điểm đơn lập của tiếng Việt [3].

2.1.3. Một số phương pháp xác định formant

Các phương pháp xác định formant liên quan đến việc tìm kiếm các đỉnh trong các biểu diễn phổ, thường là từ kết quả phân tích phổ theo phương pháp STFT (Short-time Fourier Transform), mã hóa dự đoán tuyến tính LPC (Linear Predictive Coding) và xử lý đồng hình (Homomorphic deconvolution hay còn gọi là Cepstrum).

2.1.3.1. Xác định formant từ phân tích STFT

Do tín hiệu tiếng nói là tín hiệu không dừng, nên không thể áp dụng phép phân tích Fourier thông thường. Song, nếu chia tín hiệu tiếng nói ra thành từng đoạn đủ nhỏ theo thời gian, thì tín hiệu tiếng nói trong mỗi đoạn có thể xem là tín hiệu dừng, và do đó có thể lấy biến đổi Fourier trên từng đoạn tín hiệu này. Đây là nguyên lý của STFT, còn gọi là biến đổi Fourier cửa sổ hóa.

Các phân tích STFT tương tự và rời rạc đã trở thành một công cụ cơ bản cho nhiều phát triển trong phân tích và tổng hợp tín hiệu tiếng nói.

Dễ dàng thấy STFT trực tiếp chứa các thông tin về formant ngay trong biên độ phổ. Do đó, nó trở thành một cơ sở cho việc phân tích các tần số formant của tín hiệu tiếng nói.

2.1.3.2. Xác định formant từ phân tích LPC [11]

Các tần số formant có thể được ước lượng từ các tham số dự đoán theo một trong hai cách. Cách thứ nhất là xác định trực tiếp bằng cách phân tích nhân tử đa thức dự đoán và dựa trên các nghiệm thu được để quyết định xem nghiệm nào tương ứng với formant. Cách thứ hai là sử dụng phân tích phổ và chọn các formant tương ứng với các đỉnh nhọn bằng một trong các thuật toán chọn đỉnh đã biết.

Một lợi điểm khi sử dụng phương pháp phân tích LPC để phân tích formant là tần số trung tâm của các formant và băng tần của chúng có thể xác định được một cách chính xác thông qua việc phân tích nhân tử đa thức dự đoán. Một phép phân tích LPC bậc p được chọn trước, thì số khả năng lớn nhất có thể có các điểm cực liên hợp phức là p/2. Do đó, việc gán nhãn trong quá trình xác định xem điểm cực nào tương ứng với các formant đơn giản hơn các phương pháp khác. Ngoài ra, với các điểm cực bên ngoài thường có thể dễ dàng phân tách trong phân tích LPC vì băng tần của chúng thường rất lớn so với băng tần thông thường của các formant tín hiệu tiếng nói.

Tín hiệu lời nói có thể được định nghĩa là:

$$s(n) = -\sum_{i=1}^{N_{LP}} a_{LP}(i) \times s(n-i) + e(n) \qquad (2)$$

Trong đó N_{LP}, a_{LP} và $e(n)$ lần lượt là số lượng hệ số trong mô hình (thứ tự của dự đoán), hệ số LPC và sai số trong mô hình (sự khác biệt giữa giá trị dự đoán và giá trị đo thực tế). Công thức (2) có thể được viết bằng ký hiệu biến đổi Z như một phép toán lọc tuyến tính:

$$E(z) = H_{LP}(z) \times S(z) \qquad (3)$$

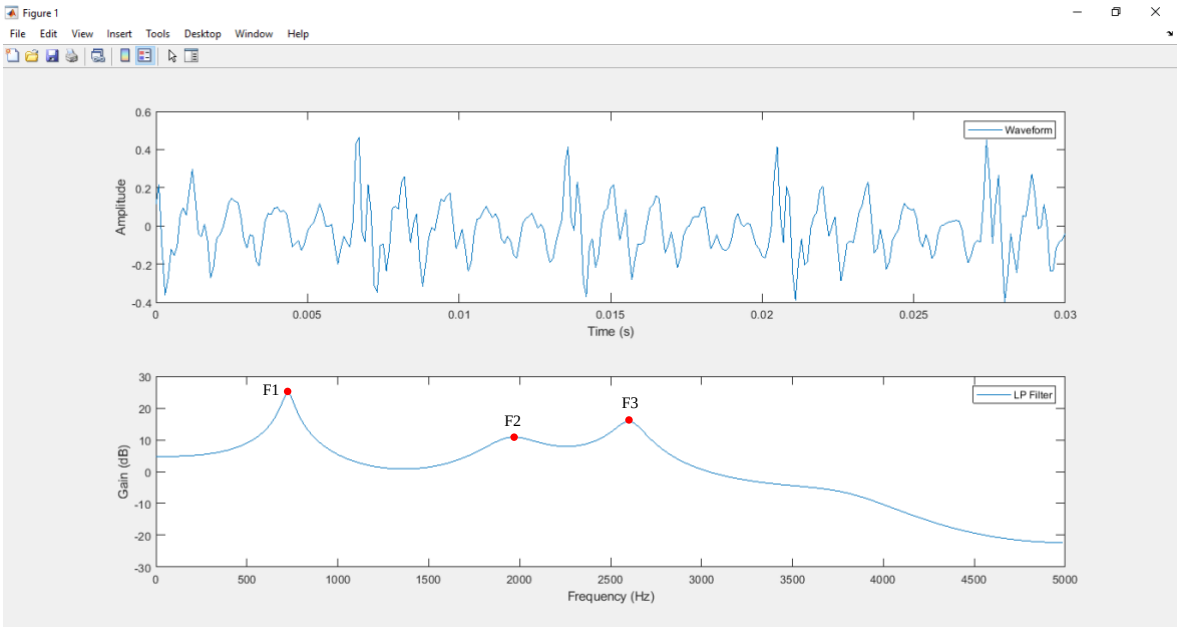
Trong đó, $E(z)$ và $S(z)$ lần lượt là biến đổi Z của tín hiệu lỗi và tín hiệu lời nói. $H_{LP}(z)$ được định nghĩa là một bộ lọc LPC ngược:

$$H_{LP}(z) = \sum_{i=0}^{N_{LP}} a_{LP}(i) \times z^{-1} \qquad (4)$$

Hoặc

$$H_{LP}(z) = 1 + \sum_{i=1}^{N_{LP}} a_{LP}(i) \times z^{-1} \qquad (5)$$

Tần số formant có thể được ước tính từ phổ làm mịn LPC. Từ phổ này, các cực đại cục bộ được tìm thấy và những băng thông nhỏ có liên quan đến các formant. Sau đó, phương pháp chọn đỉnh được sử dụng để xác định các formant. Kết quả formant F1, F2, F3 tại các đỉnh của đường bao phổ sinh ra từ các hệ số LPC được minh họa ở Hình 2.2.



Hình 2.2. Kết quả formant F1, F2, F3 tại các đỉnh của đường bao phổ sinh ra từ các hệ số LPC.

2.2. Ứng dụng phương pháp xử lý đồng hình xác định tần số formant

2.2.1. Khái quát phương pháp xử lý đồng hình

Các tần số của ba formant đầu tiên (F1, F2 và F3) chứa đầy đủ thông tin để nhận dạng nguyên âm cũng như các âm hữu thanh khác. Một đặc điểm chung cho gần như tất cả các dạng phổ là dẫn xuất của đường bao phổ thông qua một số hoạt động làm mịn. Trong số các phương pháp áp dụng trong phân tích tiếng nói để xác định formant, một phương pháp dựa trên làm mịn phổ thu được bằng hoạt động làm mịn cepstral. Cepstrum đôi khi được gọi là "xử lý đồng hình" [16]. Phương pháp này tách các thành phần tín hiệu chập bằng cách biến đổi tín hiệu tiếng nói $s(t)$ thành một miền mà tích chập trở thành một tổng đơn giản với công thức như sau:

$$s(t) = g(t) \otimes h(t) \qquad (6)$$

Trong đó: $\otimes$ là tích chập, $g(t)$ và $h(t)$ lần lượt là nguồn kích thích và tuyến âm.

Sau đó, lấy phép biến đổi Fourier (Fast Fourier Transforms - FFT) cả hai vế công thức (1), được:

$$S(w) = G(w) \times H(w) \quad (7)$$

Trong đó: S, G, H đại diện cho phổ phức tạp của s, g, h theo thời gian.

Biên độ phổ của tín hiệu có thể được viết dưới dạng:

$$|S(w)| = |G(w)| \times |H(w)| \qquad (8)$$

Lấy logarit cả hai vế của (3) được:

$$\ln|S(w)| = \ln|G(w)| + \ln|H(w)| \qquad (9)$$

Vậy, một tích chập theo thời gian đã được biến đổi thành tổng các thành phần có biên độ log trong miền tần số. Cuối cùng để tách các thành phần g và h ta áp dụng phép biến đổi Fourier ngược (Inverse Fast Fourier Transform - IFFT) cho phổ log, được công thức:

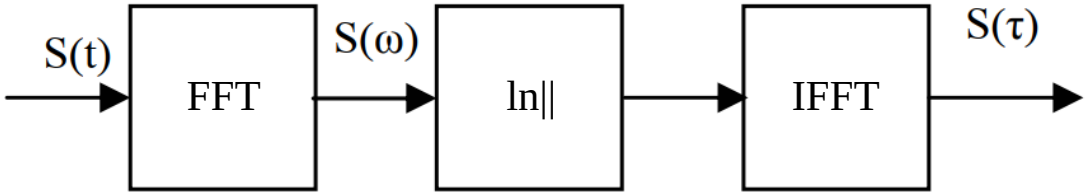
$$F^{-1}\{\ln|S(w)|\} = F^{-1}\{\ln|G(w)|\} + F^{-1}\{\ln|H(w)|\} \qquad (10)$$

Trong đó F{} biểu thị biến đổi Fourier (FFT) và F^{-1} là nghịch đảo của biến đổi Fourier (IFFT).

Hoặc có thể biểu diễn cepstrum dưới dạng công thức đơn giản như sau:

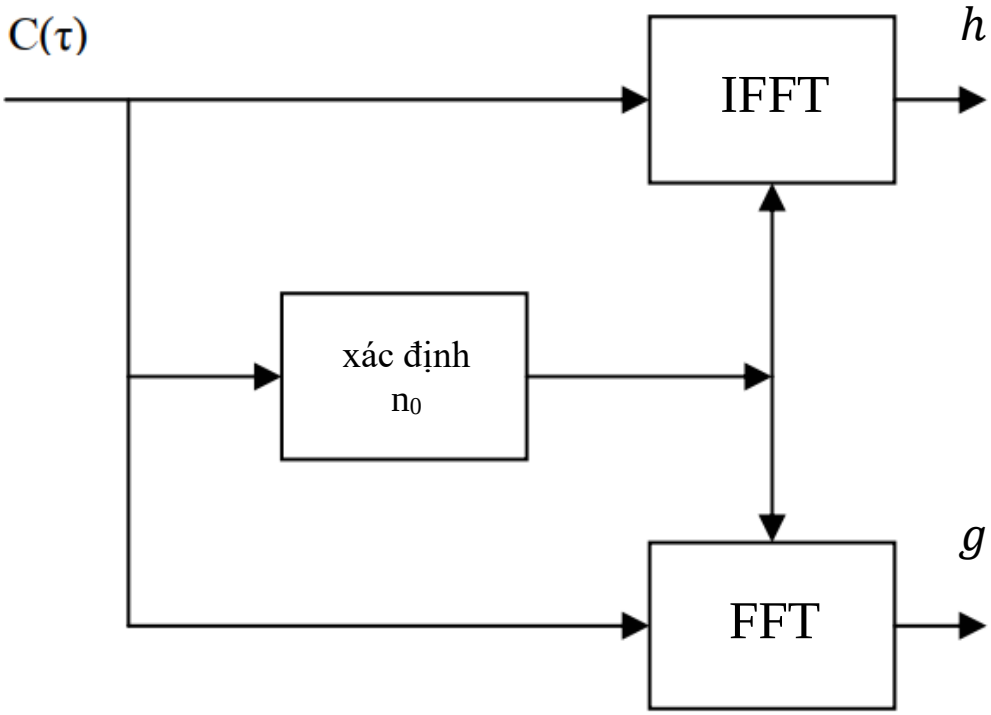
$$c(n) = FFT^{-1}(Log\left(FFT(s(n))\right)) \quad (11)$$

Biến đổi cuối cùng (IFFT) đưa hàm trở lại miền thời gian, nhưng nó không giống với thời gian của tín hiệu ban đầu. Trên thực tế, nó là một thước đo tốc độ thay đổi của các biên độ quang phổ. Miền này được gọi là cepstrum và trục thời gian thường được gọi là trục "quefrency" [10], [11].



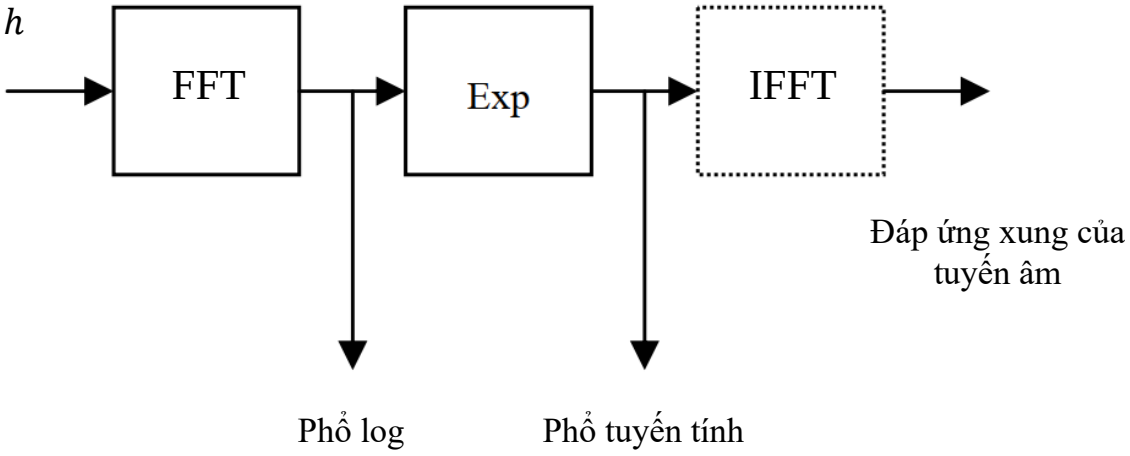
Hình 2.3. Mô hình tổng quát của phương pháp xử lý đồng hình [10].

Trên thực tế, các thuật ngữ bậc thấp của cepstrum chứa thông tin liên quan đến tuyến âm. Sự đóng góp này trở nên không quan trọng so với mẫu n_0 (n_0 tương ứng với tần số cơ bản F_0). Các "đỉnh" có thể nhìn thấy một cách tuần hoàn ngoại trừ n_0 phản ánh các xung của nguồn. Hai thành phần g và h được phân tách sau khi xác định được n_0 bằng cửa sổ đơn giản như sau:



Hình 2.4. Cửa sổ tạm thời áp dụng cho cepstrum [10].

Sau đó làm mịn đường cepstrum của tuyến âm (h) như sau:

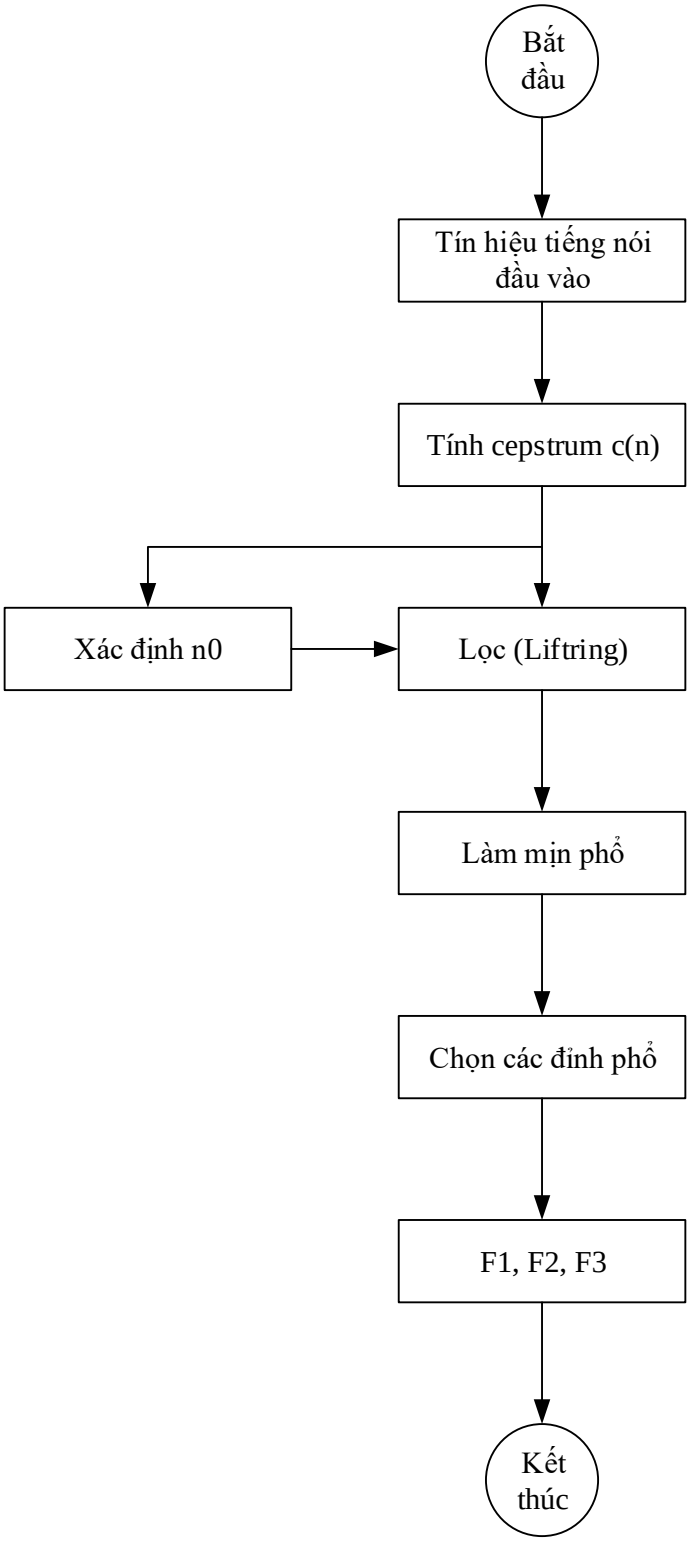


Hình 2.5. Làm mịn phổ [10].

Sau khi tính toán làm mịn phổ có thể trích xuất các biên độ tương ứng với độ cộng hưởng của tuyến âm (formant). Điều này có thể dễ dàng thu được bằng cách xác định vị trí đỉnh phổ từ dải tần số tương ứng với ba formant đầu tiên.

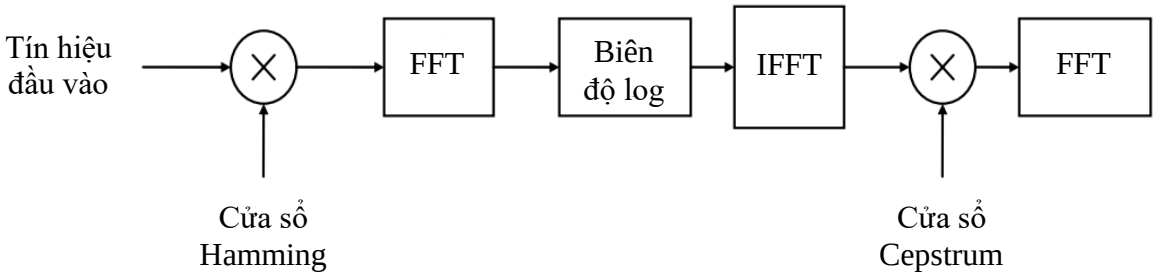
2.2.2. Thuật toán, sơ đồ khối xác định formant

Với những phân tích trong phần 2.2.1, thuật toán tính formant của một khung tín hiệu dựa trên phương pháp xử lý đồng hình như Hình 2.5:



Hình 2.6. Thuật toán tính formant [10].

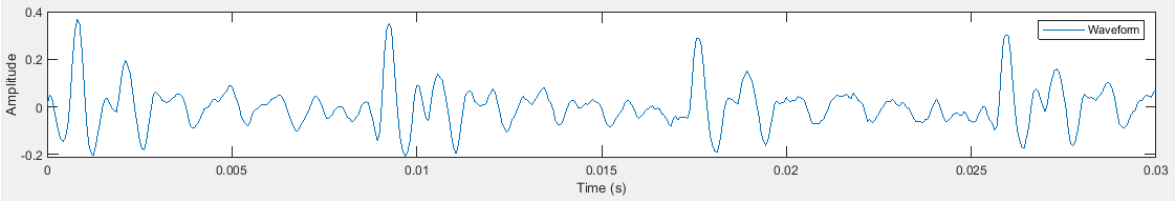
Trong phạm vi luận văn, tôi sử dụng sơ đồ khối tính formant như sau:



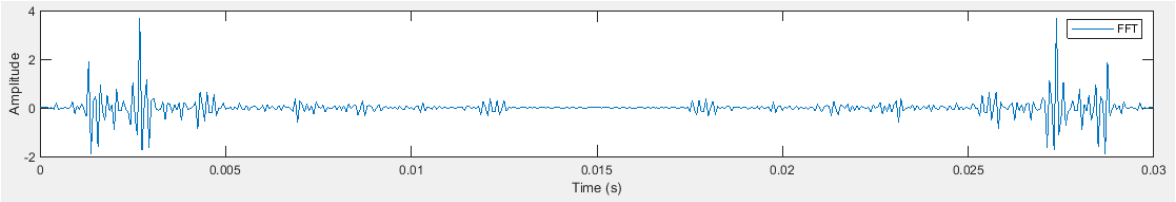
Hình 2.7. Sơ đồ khối tính formant [6].

Tín hiệu tiếng nói được cung cấp làm đầu vào cho một hệ thống bao gồm kích thích tuần hoàn (g) biến đổi với phản ứng xung động của tuyến âm (h), đây là một chức năng thay đổi chậm. Sau đó tín hiệu tiếng nói đầu vào được chia nhỏ thành các khung tín hiệu ngắn (có độ dài 30 ms) để xử lý. Trong luận văn thực hiện phân khung bằng hàm cửa sổ Hamming [12]. Khối FFT lấy DFT (Discrete Fourier Transform) của một tín hiệu và sẽ thu được phổ của tín hiệu. Sau đó lấy logarit, thấy rằng sự kích thích tuần hoàn là một chức năng thay đổi nhanh và phản ứng xung động của tuyến âm là một chức năng bao phổ thay đổi chậm. Khi thực hiện IFFT thấy rằng chức năng thay đổi chậm của các cụm tuyến âm gần điểm gốc và chức năng thay đổi nhanh là các xung đều đặn ở xa điểm gốc. Nên có thể thiết kế một cửa sổ cepstrum, cho phép thông tin formant (chức năng thay đổi chậm) đi qua. Đầu ra FFT của cửa sổ cepstrum là một phổ chỉ có chức năng thay đổi chậm. Nếu theo dõi các đỉnh của quang phổ này có thể tìm thấy các tần số formant F1, F2, F3. Chức năng thay đổi nhanh của âm thanh giờ đây đã bị cô lập, và do đó, khả năng trùng lặp formant đầu tiên với tần số cao độ sẽ bị loại bỏ [6].

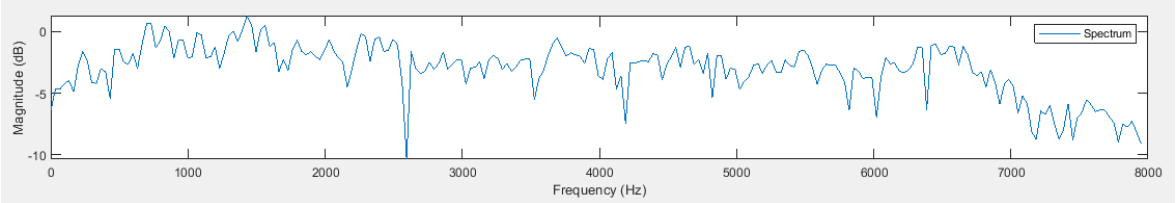
Tín hiệu âm hữu thanh đầu vào được thể hiện trong Hình 2.7. Đầu ra của FFT cửa sổ tín hiệu là một phổ như Hình 2.8. Hình 2.9 mô tả phổ log sau khi lấy logarit của khối FFT tín hiệu. Bây giờ lấy IFFT một lần nữa sẽ thu được phổ như Hình 2.10. Cuối cùng, đường bao phổ của cepstrum được mô tả trong Hình 2.11 và kết quả formant F1, F2, F3 được thể hiện trên Hình 2.12.



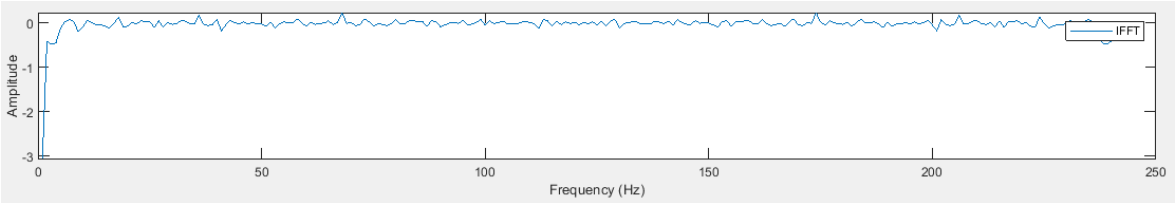
Hình 2.8. Tín hiệu âm hữu thanh đầu vào.



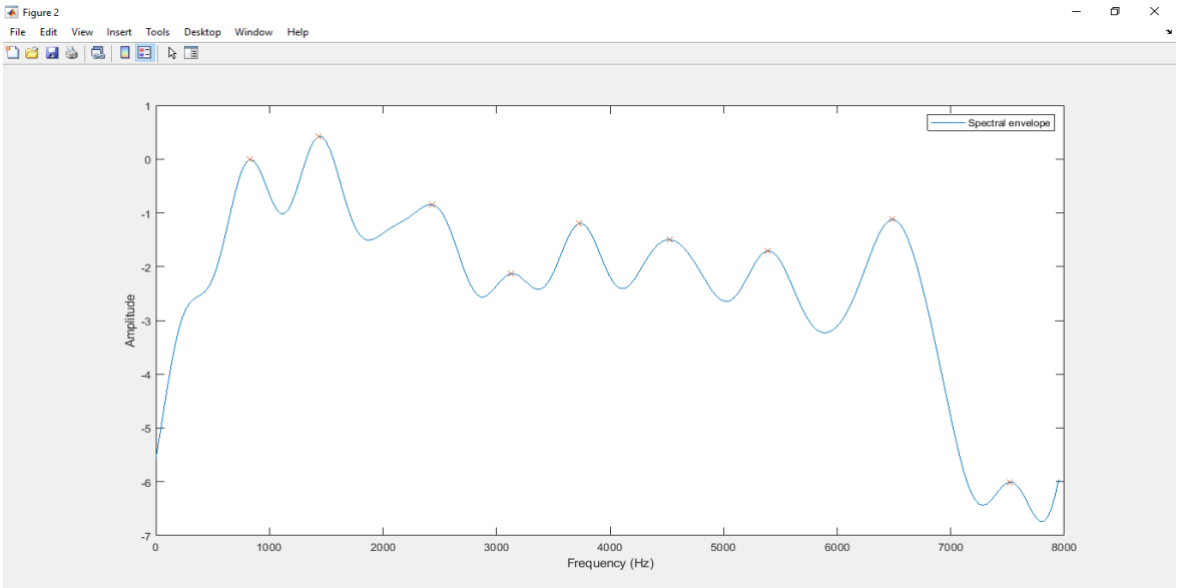
Hình 2.9. FFT cửa sổ tín hiệu.



Hình 2.10. Phổ log của cửa sổ tín hiệu.



Hình 2.11. IFFT của cửa sổ tín hiệu.



Hình 2.12. Đường bao phổ của tín hiệu.

Hình 2.13. Kết quả formant F1, F2, F3.

2.2.3. Ưu, nhược điểm của phương pháp xử lý đồng hình

2.2.3.1. Ưu điểm

- Cepstrum là một phương pháp hữu ích để trích xuất riêng rẽ tần số cơ bản và tần số formant của tín hiệu tiếng nói.

- Kết quả từ thuật toán cepstrum cho thấy rằng một phạm vi rộng các giá trị ước tính của tần số formant đã thu được [11].

2.2.3.2. Nhược điểm

- Thuật toán Cepstrum cho thấy một số hạn chế trong việc xác định các formant đặc biệt là ở tần số cao [11].

- Đôi lúc các đỉnh phổ chưa thực sự rõ ràng dẫn đến nhầm lẫn trong việc xác định các tần số formant.

2.3. Tổng kết chương

Chương này đã tập trung tìm hiểu, nghiên cứu tổng quan về tần số formant, một số phương pháp xác định formant như STFT (Short-time Fourier Transform), mã hóa dự đoán tuyến tính LPC (Linear Predictive Coding) và ứng dụng phương pháp xử lý đồng hình để xác định tần số formant, nêu rõ thuật toán xử lý đồng hình tính formant và sau đó đánh giá ưu nhược điểm của thuật toán.

Để đánh giá được thuật toán xử lý đồng hình trong việc tính tần số formant của tín hiệu tiếng nói thu được, trong chương 3 sẽ nói về môi trường cài đặt thuật toán cùng chương trình và thống kê thông tin dữ liệu người nói. Sau đó khảo sát tính đặc trưng formant của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói từ các thông số thống kê như giá trị trung bình (Mean), độ lệch chuẩn (STD), hệ số biến thiên (CV%) và sai số phần trăm.

Sau đó tôi sẽ đánh giá độ chính xác của thuật toán dựa trên sai số phần trăm từ việc so sánh thuật toán tính formant F1, F2, F3 tự động hiệu chỉnh thủ công ứng dụng phương pháp xử lý đồng hình và phương pháp LPC so với dữ liệu chuẩn được đo thủ công trên phần mềm Wavesurfer.

CHƯƠNG III: TRIỂN KHAI VÀ ĐÁNH GIÁ THUẬT TOÁN

Trong chương này, tôi tiến hành cài đặt thuật toán tính formant ứng dụng phương pháp xử lý đồng hình trên phần mềm Matlab [6]. Sau đó khảo sát tính đặc trưng formant của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói để đánh giá độ hiệu quả của thuật toán dựa trên các thông số thống kê Mean, STD và CV%.

Đồng thời, tôi so sánh thuật toán tính formant tự động ứng dụng phương pháp xử lý đồng hình và phương pháp LPC [11] so với dữ liệu chuẩn được đo thủ công trên phần mềm Wavesurfer dựa trên sai số phần trăm để đánh giá độ chính xác của thuật toán.

3.1. Môi trường phát triển

Tôi cài đặt thuật toán và tiến hành thực nghiệm trên máy tính có cấu hình:

- Hệ điều hành: Windows 10 Professional x64
- Bộ nhớ trong: 4GB
- Bộ vi xử lý: Intel® Core™ i5-2520M CPU @ 2.50GHz

Phần mềm được sử dụng: Matlab - Phiên bản R2019a.

3.2. Dữ liệu thử nghiệm

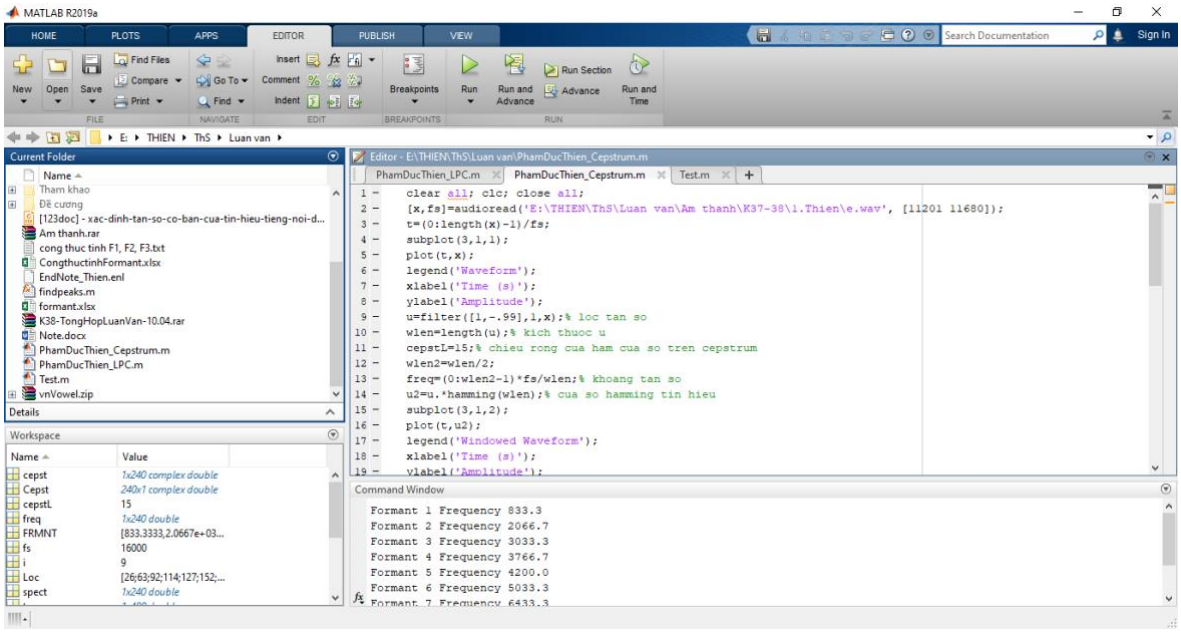
Việc khảo sát tính đặc trưng của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói là cần thiết để đánh giá hiệu quả của thuật toán. Tôi đã thu thập tín hiệu tiếng nói của năm nguyên âm /a/, /e/, /i/, /o/, /u/ trong điều kiện không tạp âm của 50 người nói trưởng thành: trong đó có 39 nam và 11 nữ, chủ yếu ở vùng miền Trung. Các tín hiệu được thu ở tần số lấy mẫu 16000 Hz trong điều kiện phòng thí nghiệm, đơn kênh (mono), và lưu trong các file .wav theo định dạng PCM của Microsoft.

Bảng 3.1. Thống kê dữ liệu người nói.

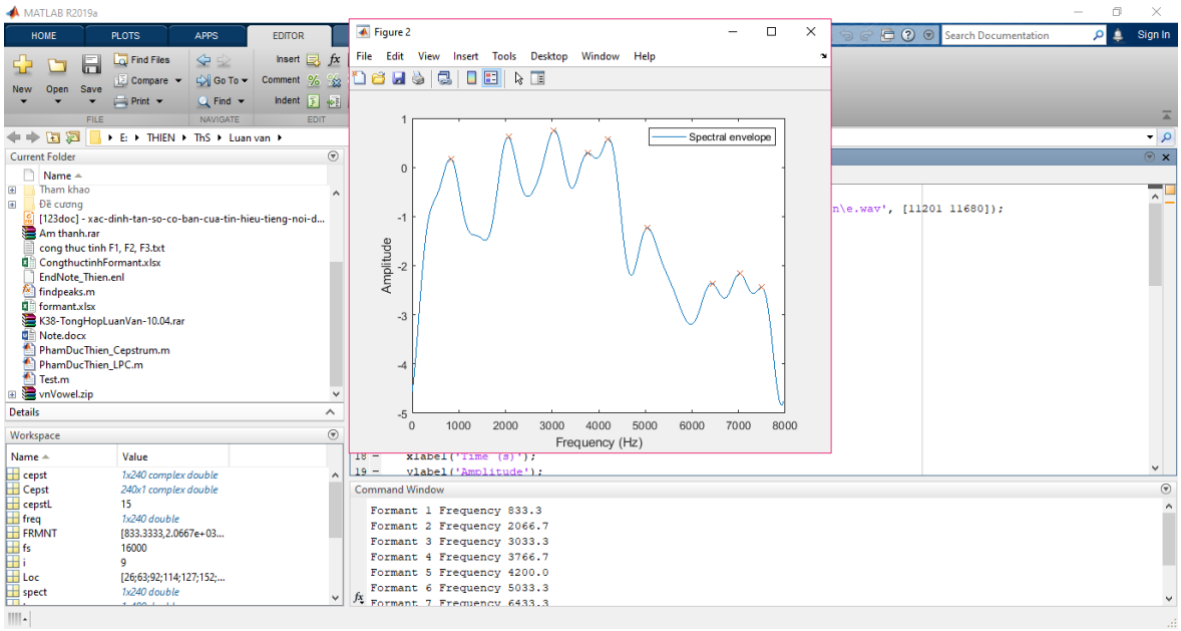
Giới tính / Tỉnh/Thành	Nam	Nữ	**Tổng**
Đà Nẵng	18	4	**22**
Quảng Nam	14	3	**17**
Quảng Trị	1	1	**2**
Thừa Thiên Huế	2	0	**2**
Hà Tĩnh	1	0	**1**
Nghệ An	0	1	**1**
Nam Định	1	0	**1**
Đắk Lắk	1	0	**1**
Bình Định	1	0	**1**
Hưng Yên	0	1	**1**
Quảng Bình	0	1	**1**
Tổng	**39**	**11**	**50**

3.3. Chương trình

Sau khi mở phần mềm Matlab, điều chỉnh đường dẫn đến thư mục chứa 05 file âm thanh /a/, /e/, /i/, /o/, /u/ của từng người nói để xác định 03 tần số formant F1, F2, F3.



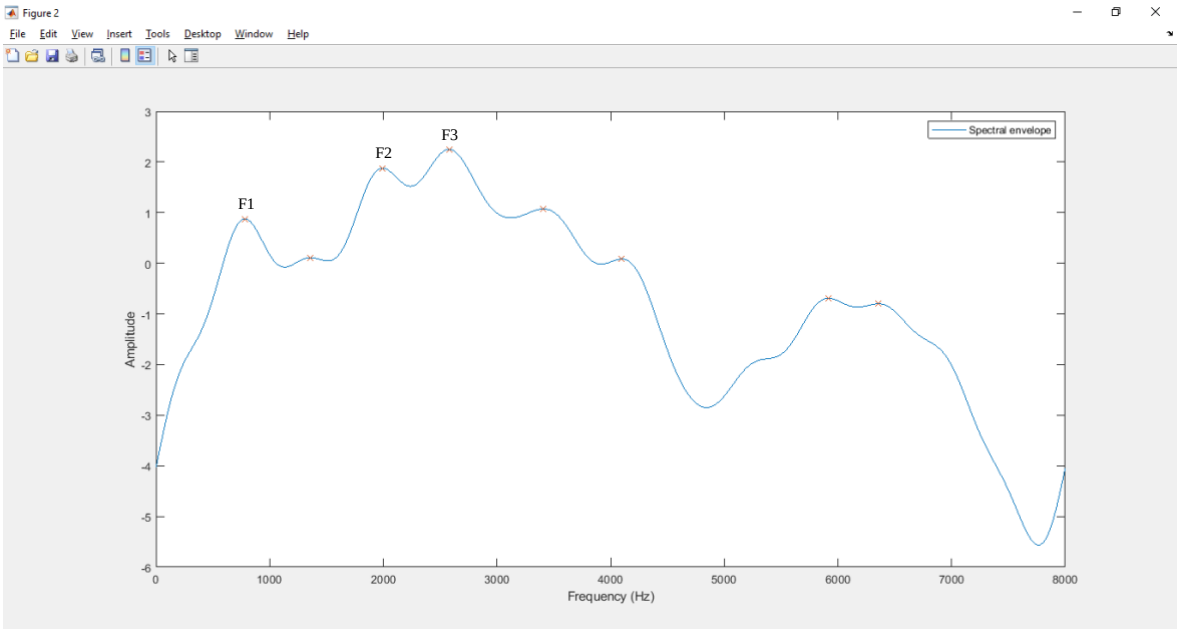
Hình 3.1. Giao diện chương trình Matlab.



Hình 3.2. Kết quả đường bao phổ và các tần số formant tìm được bởi chương trình.

3.4. Hiệu chỉnh kết quả tính formant tự động

Trong phạm vi luận văn, thuật toán tính tự động 03 tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình thường cho ra 03 kết quả đầu tiên tương ứng với 03 tần số formant F1, F2, F3 (như Hình 3.2). Tuy nhiên đôi lúc lại cho ra nhiều hơn 03 tần số (do có thêm vài đỉnh nhỏ tuy nhiên thuật toán vẫn lấy tần số các đỉnh đó) nên phải kết hợp thêm chọn thủ công từ đường bao phổ để xác định 03 formant F1, F2, F3 trong các tần số đó. Ví dụ như ở Hình 3.3 kết quả cho ra 04 tần số formant đầu tiên (F1: 783 Hz, F2: 1360 Hz, F3: 1994 Hz, F4: 2579 Hz) tuy nhiên nhìn trên đường bao phổ có thể thấy đỉnh tương ứng tần số thứ 2 (F2: 1360 Hz) không rõ ràng nên sẽ được loại bỏ. Vậy nên từ 04 tần số formant đầu tiên chọn ra 03 tần số F1, F2, F3 lần lượt là 783 Hz, 1994 Hz và 2579 Hz.



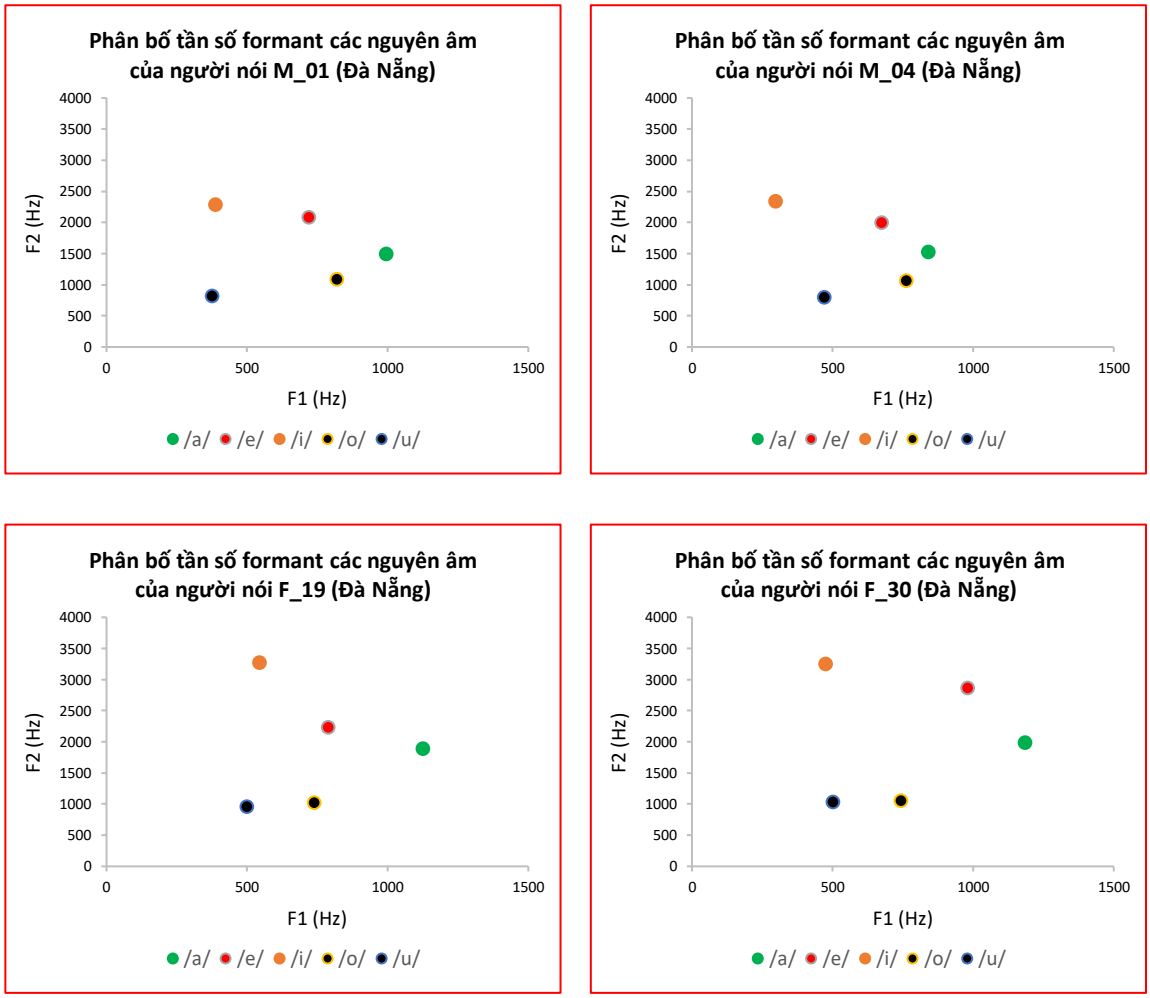
Hình 3.3. Chọn 03 formant F1, F2, F3 từ 04 formant tự động đầu tiên.

Việc tính toán kết quả đo tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình được thực hiện trên độ dài khung tín hiệu là 30 ms (tương đương 0,03 s cho mỗi khung tín hiệu). Vì tần số formant đo được của tín hiệu thu được là một dãy các giá trị trên các khung tín hiệu âm hữu thanh, nên sau khi xác định được tần số formant từ việc đo và hiệu chỉnh kết quả tính tự động, tôi thực hiện đo 03 lần tần số formant F1, F2, F3 trên 03 khung khác nhau của tín hiệu và sau đó lấy giá trị trung bình để xác định tần số formant cuối cùng. Tôi cũng sử dụng cách này để xác định tần số formant F1, F2, F3 của tín hiệu cho các phương pháp khác trong luận văn (phương pháp LPC và đo thủ công trên phần mềm Wavesurfer).

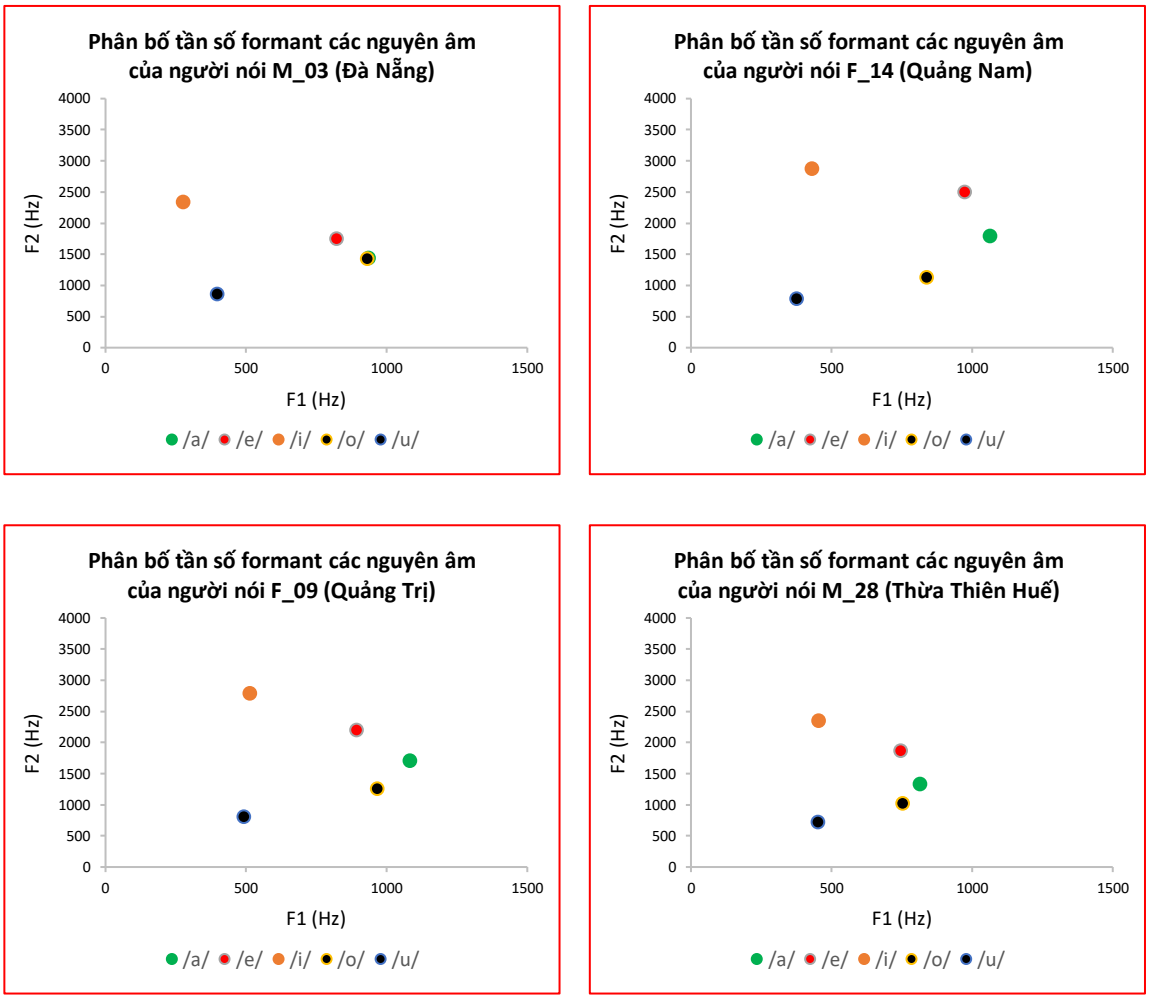
Sau khi xác định được tần số formant F1, F2, F3 của 05 nguyên âm của từng người nói (Phụ lục 1), tôi tập trung phân tích các thông số thống kê của tần số formant, cụ thể là phân tích Mean, STD và CV%, từ đó phân tích đặc trưng của 05 nguyên âm của cùng một người nói và một nguyên âm của 50 người nói.

3.5. Phân tích formant các nguyên âm của một người nói

Sau khi có được bảng số liệu tần số formant của 50 người nói (Phụ lục 1), tôi tiến hành thống kê phân bố tần số F1, F2 (do F1 và F2 đặc trưng cho các nguyên âm hơn so với F3) của 05 nguyên âm /a/, /e/, /i/, /o/, /u/ của từng người nói theo hai loại: người nói theo một vùng (chọn ra 04 người nói đến từ một vùng, cụ thể là Đà Nẵng) và người nói từ nhiều vùng khác nhau (chọn ra 04 người nói đến từ 04 vùng: Đà Nẵng, Quảng Nam, Quảng Trị và Thừa Thiên Huế), cho ra kết quả ở Hình 3.4 và 3.5 như sau:



Hình 3.4. Phân bố tần số formant 05 nguyên âm của mỗi người nói từ một vùng (Đà Nẵng).



Hình 3.5. Phân bố tần số formant 05 nguyên âm của mỗi người nói từ nhiều vùng.

Từ Hình 3.4 và 3.5 cho thấy đối với sự phân bố tần số formant F1, F2 của 05 nguyên âm có sự khác biệt rõ ràng đáng kể giữa mỗi người nói đến từ một vùng (Đà Nẵng) và giữa mỗi người nói đến từ nhiều vùng (04 vùng Đà Nẵng, Quảng Nam, Quảng Trị và Thừa Thiên Huế). Mỗi người nói sẽ có tần số formant F1, F2 riêng biệt của 05 nguyên âm. Tuy nhiên để phân biệt giữa các người nói với nhau dựa trên đặc trưng formant của 05 nguyên âm là chưa đủ, cần phải phân tích thêm các nguyên âm khi kết hợp cùng với phụ âm và thanh điệu thì mới có thể thấy được đặc trưng formant phân biệt giữa các người nói với nhau.

3.6. Phân tích formant một nguyên âm của nhiều người nói

3.6.1. Các thông số thống kê

3.6.1.1. Giá trị trung bình (Mean)

Mean là bình quân toán học đơn giản của một tập hợp gồm hai hoặc nhiều số, với công thức như sau:

$$\bar{x} = \frac{\sum x}{n} \qquad (12)$$

Trong đó:

$\bar{x}$: giá trị trung bình

$\sum x$: tổng các giá trị x

n: tổng số phần tử x

3.6.1.2. Độ lệch chuẩn (Standard Deviation - STD)

Độ lệch chuẩn là một đại lượng thống kê dùng để đo mức độ phân tán của một tập dữ liệu đã được lập thành bảng tần số. Có thể tính ra độ lệch chuẩn bằng cách lấy căn bậc hai của phương sai. Khi hai tập dữ liệu có cùng giá trị trung bình cộng, tập nào có độ lệch chuẩn lớn hơn là tập có dữ liệu biến thiên nhiều hơn. Có công thức độ lệch chuẩn như sau:

$$s = \sqrt{\sum_{i=1}^{n} \frac{1}{(n\ -\ 1)} (x_i\ -\ \bar{x})^2} \qquad (13)$$

Trong đó:

s: độ lệch chuẩn

x_i: giá trị thành phần thứ i

$\bar{x}$: giá trị trung bình

n: tổng số phần tử x

3.6.1.3. Hệ số biến thiên (Coefficient of Variation - CV%)

Hệ số biến thiên là tỉ lệ của độ lệch chuẩn so với giá trị trung bình, tính theo %. Nó là một thống kê hữu ích trong việc so sánh mức độ biến thiên của

chuỗi dữ liệu này với chuỗi dữ liệu khác, cho dù giá trị trung bình của chúng rất khác nhau. Ta có công thức như sau:

$$CV = \frac{s}{\bar{x}} \times 100\% \qquad (14)$$

Trong đó:

CV: hệ số biến thiên (%)

s: độ lệch chuẩn

$\bar{x}$: giá trị trung bình

3.6.2. Phân tích nhiều người nói

Từ số liệu tần số formant của 50 người nói (Phụ lục 1). Tôi tiến hành tính Mean, STD, CV% của tần số formant của từng nguyên âm /a/, /e/, /i/, /o/, /u/ của 50 người nói, thu được các bảng từ Bảng 3.2 đến Bảng 3.6 như sau:

Bảng 3.2. Mean, STD, CV% formant nguyên âm /a/ của 50 người nói.

/a/	**F1**	**F2**	**F3**
Mean (Hz)	917	1517	2745
STD	124,3	196,5	485,2
CV (%)	13,5	13,0	17,7

Bảng 3.3. Mean, STD, CV% formant nguyên âm /e/ của 50 người nói.

/e/	**F1**	**F2**	**F3**
Mean (Hz)	742	2049	2940
STD	112,2	285,4	437,9
CV (%)	15,1	13,9	14,9

Bảng 3.4. Mean, STD, CV% formant nguyên âm /i/ của 50 người nói.

/i/	**F1**	**F2**	**F3**
Mean (Hz)	431	2398	3297
STD	61,9	384,2	515,7
CV (%)	14,3	16,0	15,6

Bảng 3.5. Mean, STD, CV% formant nguyên âm /o/ của 50 người nói.

/o/	F1	F2	F3
Mean (Hz)	758	1073	2805
STD	97,2	114,4	462,1
CV (%)	12,8	10,7	16,5

Bảng 3.6. Mean, STD, CV% formant nguyên âm /u/ của 50 người nói.

/u/	F1	F2	F3
Mean (Hz)	452	810	2683
STD	55,9	107,5	732,5
CV (%)	12,4	13,3	27,3

Từ các số liệu Mean, STD, CV% tần số formant từng nguyên âm của nhiều người nói thể hiện trên Bảng 3.2 đến Bảng 3.6 cho thấy:

- Đối với Mean (Hz):

Nguyên âm /i/ là nguyên âm có mean F2, F3 cao nhất trong cả 05 nguyên âm (lần lượt là 2398 Hz, 3297 Hz) nhưng có tần số F1 thấp nhất (431 Hz) và có vùng tần số F1 từ 277 - 553 Hz, vùng tần số F2 từ 1830 - 3410 Hz, vùng tần số F3 từ 2428 - 4565 Hz.

Tiếp theo là nguyên âm /e/ mean F1, F2, F3 lần lượt là 742 Hz, 2049 Hz, 2940 và có vùng tần số F1 từ 388 - 981 Hz, vùng tần số F2 từ 1686 - 2890 Hz, vùng tần số F3 từ 2330 - 3781 Hz.

Thứ ba là nguyên âm /a/ mean F1, F2, F3 lần lượt là 917 Hz, 1517 Hz, 2745 Hz và có vùng tần số F1 từ 743 - 1286 Hz, vùng tần số F2 từ 1242 - 2071 Hz, vùng tần số F3 từ 1865 - 3773 Hz. Trong đó nguyên âm /a/ có tần số F1 cao nhất trong 05 nguyên âm.

Thứ tư là nguyên âm /o/ mean F1, F2, F3 lần lượt là 758 Hz, 1073 Hz, 2805 Hz và vùng tần số F1 từ 483 - 1019 Hz, vùng tần số F2 từ 822 - 1419 Hz, vùng tần số F3 từ 1972 - 3943 Hz.

Cuối cùng nguyên âm có mean F1, F2, F3 gần như thấp nhất trong 05 nguyên âm là nguyên âm /u/ với lần lượt là 452 Hz, 810 Hz, 2683 Hz và vùng

tần số F1 từ 306 - 557 Hz, vùng tần số F2 từ 529 - 1044 Hz, vùng tần số F3 từ 1433 - 4322 Hz. Tuy nhiên nguyên âm /u/ lại có độ rộng vùng tần số F3 lớn nhất (2889 Hz).

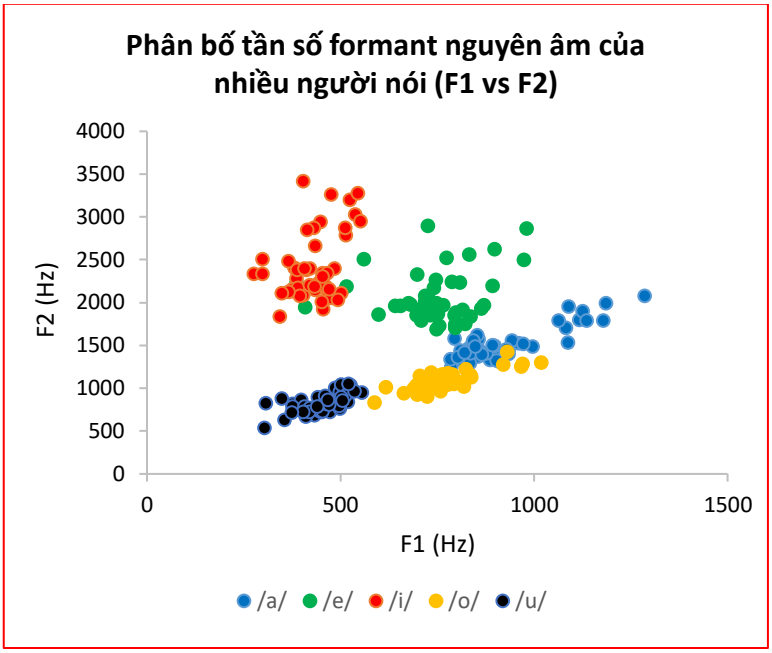
Độ rộng vùng tần số formant của cả 05 nguyên âm đều tăng dần từ formant bậc thấp đến bậc cao, cụ thể nhìn chung vùng tần số F1 dao động khoảng 250 - 1300 Hz (1050 Hz), vùng tần số F2 dao động khoảng 500 - 3450 Hz (2950 Hz) và vùng tần số F3 dao động khoảng 1450 - 4600 Hz (3150 Hz), tùy theo các điều kiện khác nhau thì sẽ có vùng tần số khác nhau cho từng nguyên âm.

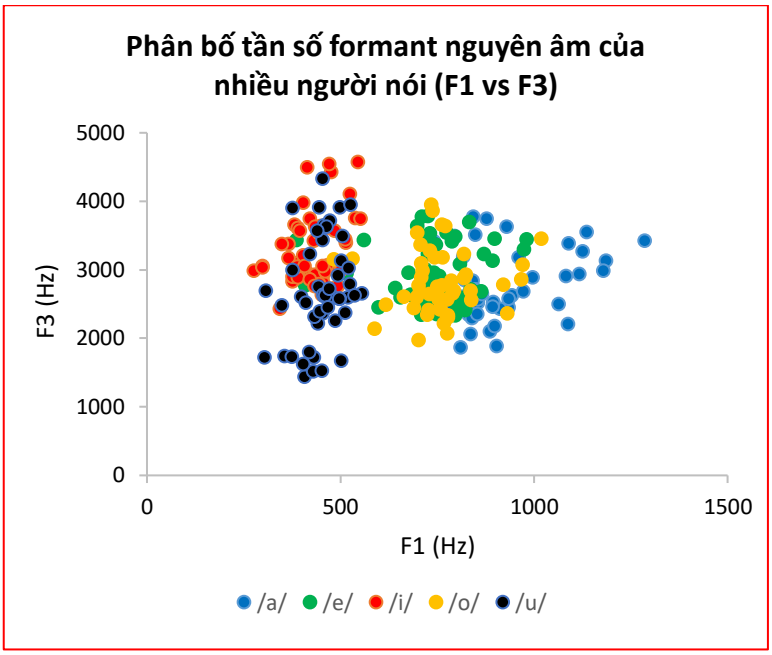
Những kết quả thống kê tần số formant 05 nguyên âm trên khá tương đồng với một nghiên cứu về vùng tần số formant nguyên âm tiếng Việt của Nguyễn Văn Ái vào năm 1973 [1] và luận án tiến sĩ kỹ thuật của Ngô Minh Dũng vào năm 2010 [4], ở đây tôi so sánh 02 tần số formant F1, F2 và thu được bảng như sau:

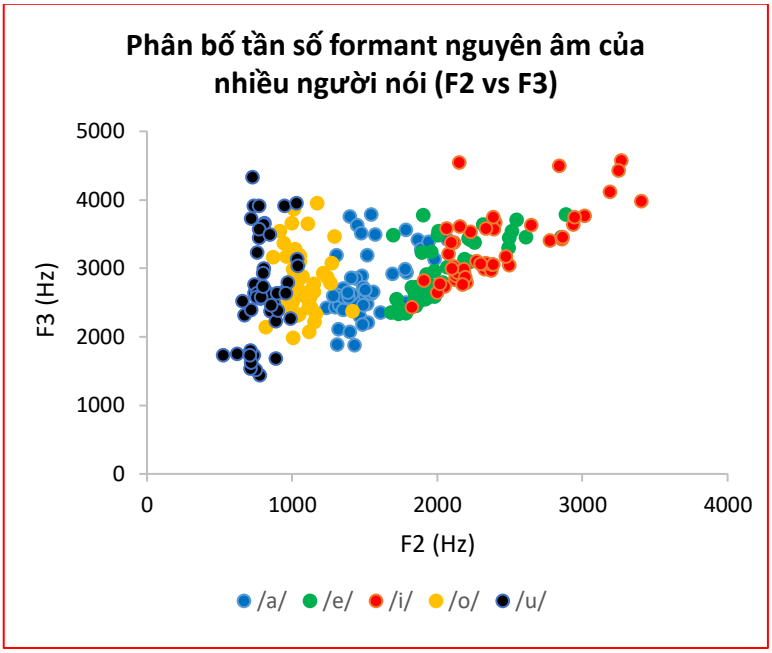
Bảng 3.7. Vùng tần số F1, F2 05 nguyên âm.

Nguyên âm	**Nghiên cứu trong luận văn**				**Nguyễn Văn Ái (1973)**		**Ngô Minh Dũng (2010)**	
	Mean F1 (Hz)	**Vùng tần số F1 (Hz)**	**Mean F2 (Hz)**	**Vùng tần số F2 (Hz)**	**Vùng tần số F1 (Hz)**	**Vùng tần số F2 (Hz)**	**Vùng tần số F1 (Hz)**	**Vùng tần số F2 (Hz)**
/a/	917	743 - 1286	1517	1242 - 2071	707 - 1000	1190 - 1410	600 - 1200	1200 - 1800
/e/	742	388 - 981	2049	1686 - 2890	354 - 595	2000 - 2830	300 - 600	1600 - 2200
/i/	431	277 - 553	2398	1830 - 3410	250 - 420	2380 - 3360	300 - 600	1600 - 2200
/o/	758	483 - 1019	1073	822 - 1419	354 - 595	707 - 1000	300 - 800	700 - 1200
/u/	452	306 - 557	810	529 - 1044	250 - 420	595 - 840	300 - 800	700 - 1200

Có thể quan sát rõ hơn phân bố tần số formant nguyên âm của nhiều người nói qua Hình 3.6 dưới đây:


Phân bố tần số formant nguyên âm của nhiều người nói (F1 vs F2)
4000
3500
3000
2500
2000
1500
1000
500
0
F2 (Hz)
0
500
1000
1500
F1 (Hz)
/a/ /e/ /i/ /o/ /u/


Phân bố tần số formant nguyên âm của nhiều người nói (F1 vs F3)
5000
4000
3000
2000
1000
0
F3 (Hz)
0
500
1000
1500
F1 (Hz)
/a/ /e/ /i/ /o/ /u/



Hình 3.6. Phân bố tần số formant nguyên âm nhiều người nói.

- Đối với độ lệch chuẩn: Các nguyên âm có tần số formant F1, F2, F3 lớn thì tương ứng độ lệch chuẩn hầu như đều lớn, trong đó nguyên âm /u/ có độ lệch chuẩn tần số formant F1, F2 thấp nhất (do có tần số formant F1, F2 nhỏ nhất) trong 05 nguyên âm tuy nhiên lại có độ lệch chuẩn tần số formant F3 lớn nhất với 732,5 Hz (tuy có mean F3 nhỏ nhất nhưng do có vùng tần số rộng nhất trong cả 05 nguyên âm nên sẽ có độ lệch chuẩn lớn). Qua đó có thể nhận định rằng nguyên âm có tần số formant càng lớn thì sẽ có độ lệch chuẩn càng lớn tương ứng.

- Đối với hệ số biến thiên CV%: Hầu hết tần số formant của cả 05 nguyên âm đều có CV dưới 20% (ngoại trừ tần số formant F3 ở nguyên âm /u/ với CV là 27,3 % do có độ lệch chuẩn lớn nhất, vùng tần số rộng nhất nên cũng sẽ có CV% cao nhất). Nguyên âm /o/ có độ biến thiên tần số formant thấp nhất trong cả 05 nguyên âm.

Với những kết quả phân tích mean, STD, CV% ở trên cho thấy mean tần số formant F1, F2, F3 của 05 nguyên âm có thể xếp theo thứ tự tăng dần từ nguyên âm /u/ → /o/ → /a/ → /e/ → /i/ (trong đó nguyên âm /i/ có tần số formant F1 thấp nhất và nguyên âm /a/ có tần số formant F1 cao nhất). Độ rộng vùng

tần số formant của cả 05 nguyên âm tăng dần từ formant bậc thấp đến bậc cao, tùy theo các điều kiện khác nhau thì sẽ có vùng tần số khác nhau cho từng nguyên âm. Nguyên âm có tần số formant càng lớn thì cũng sẽ có độ lệch chuẩn càng lớn tương ứng. Nhìn chung 05 nguyên âm đều có độ biến thiên tần số formant F1, F2, F3 khá thấp (< 20%) và khá ổn định. Sau đây tôi tiến hành phân tích tần số formant nguyên âm của nhiều người nói dựa theo giới tính và dựa theo địa phương.

3.6.3. Phân tích theo giới tính

Tôi tiến hành phân tích Mean, STD, CV% của tần số formant của từng nguyên âm /a/, /e/, /i/, /o/, /u/ của tất cả người nói theo giới tính, thu được các bảng từ Bảng 3.8 đến Bảng 3.12 như sau:

Bảng 3.8. Mean, STD, CV% formant nguyên âm /a/ của 50 người nói theo giới tính.

/a/	**F1**		**F2**		**F3**	
	Nam	**Nữ**	**Nam**	**Nữ**	**Nam**	**Nữ**
Mean (Hz)	876	1064	1437	1801	2616	3202
STD (Hz)	73,4	157,5	105,8	183,2	447,4	313,2
CV (%)	8,4	14,8	7,4	10,2	17,1	9,8

Bảng 3.9. Mean, STD, CV% formant nguyên âm /e/ của 50 người nói theo giới tính.

/e/	**F1**		**F2**		**F3**	
	Nam	**Nữ**	**Nam**	**Nữ**	**Nam**	**Nữ**
Mean (Hz)	721	815	1941	2433	2808	3406
STD (Hz)	101,1	123,2	197,6	206,9	396,4	188,4
CV (%)	14,0	15,1	10,2	8,5	14,1	5,5

Bảng 3.10. Mean, STD, CV% formant nguyên âm /i/ của 50 người nói theo giới tính.

/i/	F1		F2		F3	
	Nam	**Nữ**	**Nam**	**Nữ**	**Nam**	**Nữ**
Mean (Hz)	415	488	2218	3035	3126	3903
STD (Hz)	54,3	54,5	166,7	210,8	395,1	438,1
CV (%)	13,1	11,2	7,5	6,9	12,6	11,2

Bảng 3.11. Mean, STD, CV% formant nguyên âm /o/ của 50 người nói theo giới tính.

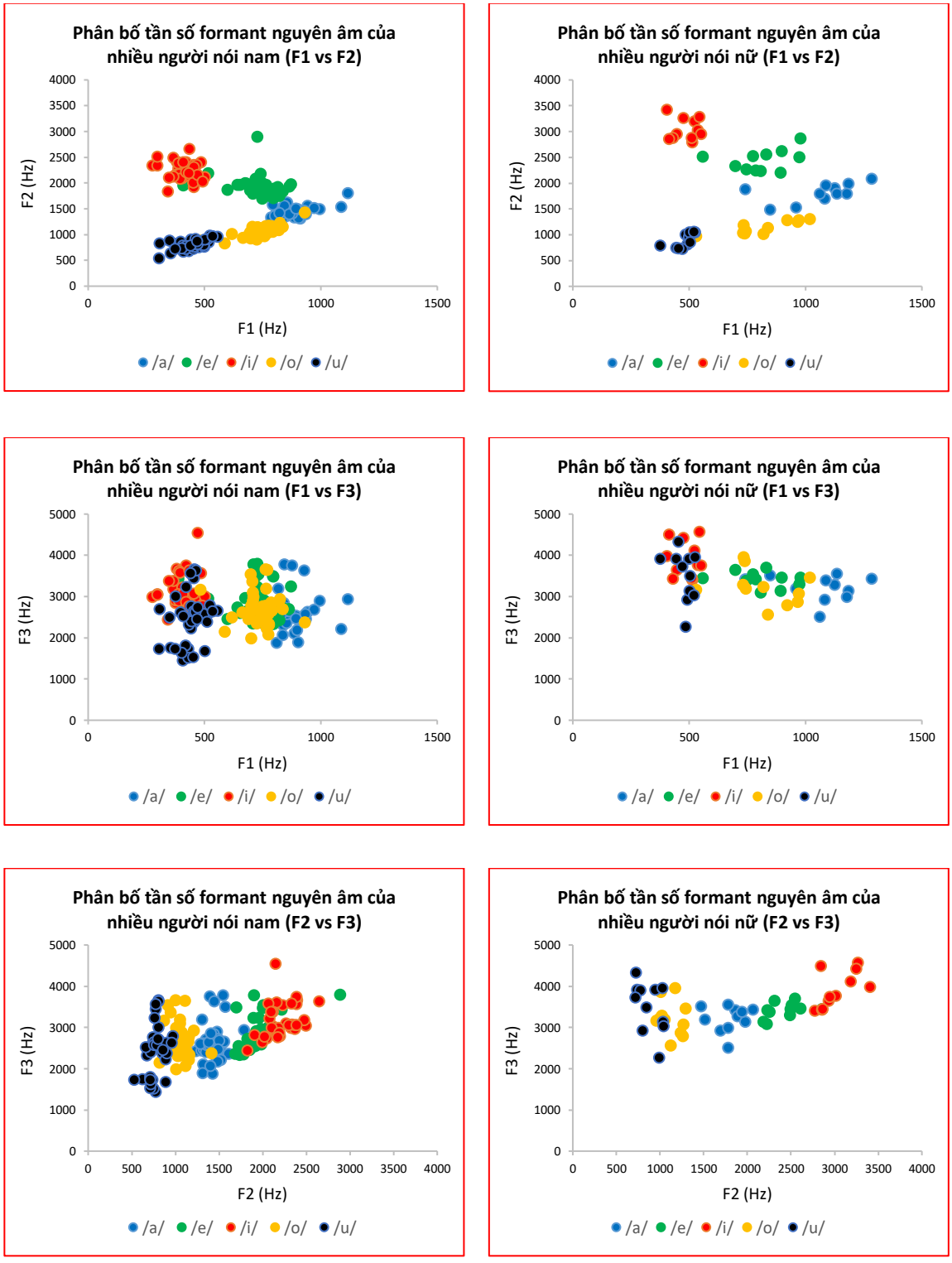
/o/	F1		F2		F3	
	Nam	**Nữ**	**Nam**	**Nữ**	**Nam**	**Nữ**
Mean (Hz)	740	821	1056	1132	2690	3213
STD (Hz)	72,9	143,4	107,1	124,3	408,1	421,8
CV (%)	9,9	17,5	10,1	11,0	15,2	13,1

Bảng 3.12. Mean, STD, CV% formant nguyên âm /u/ của 50 người nói theo giới tính.

/u/	F1		F2		F3	
	Nam	**Nữ**	**Nam**	**Nữ**	**Nam**	**Nữ**
Mean (Hz)	444	481	791	880	2453	3500
STD (Hz)	56,9	42,5	91,8	133,5	589,5	605,3
CV (%)	12,8	8,8	11,6	15,2	24,0	17,3

Kết quả từ Bảng 3.8 đến Bảng 3.12 cho thấy người nói giọng nữ hầu như có tần số formant cao hơn người nói giọng nam, kết quả trên tương đồng với một nghiên cứu vào năm 2004 (kết quả tần số formant của nam thấp hơn nữ) [10]. Độ rộng vùng tần số của giọng nữ hầu như ngắn hơn so với giọng nam (trong đó có 02 tần số formant có vùng tần số giọng nữ rộng hơn giọng nam, cụ thể: tần số F1, F2 của nguyên âm /a/ và tần số F1 nguyên âm /o/).

Hình 3.7 thể hiện sự phân bố tần số formant nguyên âm của người nói giọng nam và nữ:



Hình 3.7. Phân bố tần số formant nguyên âm của người nói nam và nữ.

- Đối với độ lệch chuẩn: Kết quả tương tự như mục 3.6.2 là nguyên âm có tần số formant càng lớn thì sẽ có độ lệch chuẩn càng lớn tương ứng.

- Đối với hệ số biến thiên CV%: Tần số formant cả 05 nguyên âm ở cả người nói giọng nam và nữ đều có CV dưới 20%, ngoại trừ tần số F3 nguyên âm /u/ ở giọng nam có CV là 24 %. Nhìn chung 05 nguyên âm đều có tần số formant ít biến thiên và khá ổn định ở cả nam và nữ.

Với những kết quả phân tích mean, STD, CV% ở trên cho thấy mean tần số formant F1, F2, F3 của 05 nguyên âm dựa theo giới tính có thể xếp theo thứ tự tăng dần từ nguyên âm /u/ → /o/ → /a/ → /e/ → /i/. Người nói giọng nữ có tần số formant cao hơn giọng nam và có độ rộng vùng tần số ngắn hơn so với giọng nam.

Kết luận chung: quan sát từ Hình 3.7 phân bố tần số formant nguyên âm của người nói nam và nữ cùng các kết quả phân tích các thông số thống kê cho thấy đặc trưng formant 05 nguyên âm giữa nam và nữ coi như không có sự khác biệt đáng kể.

3.6.4. Phân tích theo địa phương

Trong tổng số 50 người nói ở 11 tỉnh/thành tôi tập trung phân tích giọng nói ở thành phố Đà Nẵng (số lượng 22) và tỉnh Quảng Nam (số lượng 17), chiếm tỉ lệ 78% tổng số người nói (39/50). Những tỉnh/thành còn lại chủ yếu có từ 1-2 người nói, do số lượng quá ít nên khi phân tích sẽ không thấy được đặc trưng tần số formant của 05 nguyên âm mỗi người nói. Kết quả Mean, STD, CV% formant của từng nguyên âm thu được các bảng từ Bảng 3.13 đến Bảng 3.17 như sau:

Bảng 3.13. Mean, STD, CV% formant nguyên âm /a/ của người nói Đà Nẵng và Quảng Nam.

/a/	F1		F2		F3	
	Đà Nẵng	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**
Mean (Hz)	931	892	1550	1451	2788	2593
STD (Hz)	133,1	109,5	199,8	188,7	491,1	469,6
CV (%)	14,3	12,3	12,9	13,0	17,6	18,1

Bảng 3.14. Mean, STD, CV% formant nguyên âm /e/ của người nói Đà Nẵng và Quảng Nam.

/e/	F1		F2		F3	
	Đà Nẵng	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**
Mean (Hz)	768	710	2014	2079	2904	2964
STD (Hz)	79,0	149,1	299,1	304,6	466,4	429,1
CV (%)	10,3	21,0	14,8	14,7	16,1	14,5

Bảng 3.15. Mean, STD, CV% formant nguyên âm /i/ của người nói Đà Nẵng và Quảng Nam.

/i/	F1		F2		F3	
	Đà Nẵng	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**
Mean (Hz)	432	423	2351	2394	3266	3169
STD (Hz)	70,2	49,2	399,7	366,5	612,8	354,3
CV (%)	16,2	11,6	17,0	15,3	18,8	11,2

Bảng 3.16. Mean, STD, CV% formant nguyên âm /o/ của người nói Đà Nẵng và Quảng Nam.

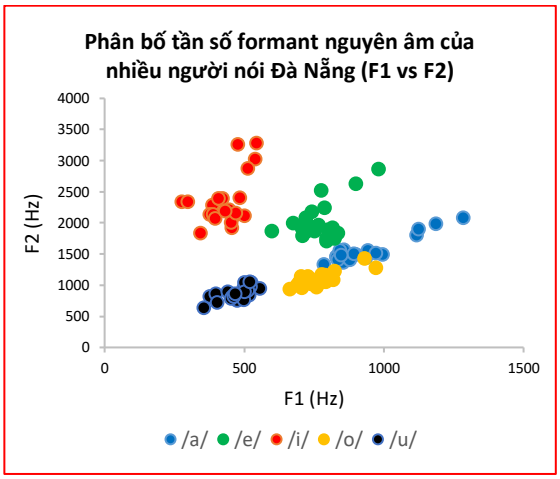
/o/	F1		F2		F3	
	Đà Nẵng	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**
Mean (Hz)	771	733	1084	1061	2879	2722
STD (Hz)	72,2	120,1	113,9	95,6	464,2	389,9
CV (%)	9,4	16,4	10,5	9,0	16,1	14,3

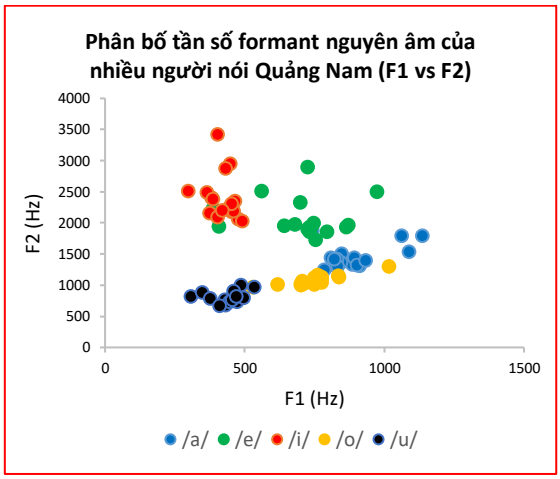
Bảng 3.17. Mean, STD, CV% formant nguyên âm /u/ của người nói Đà Nẵng và Quảng Nam.

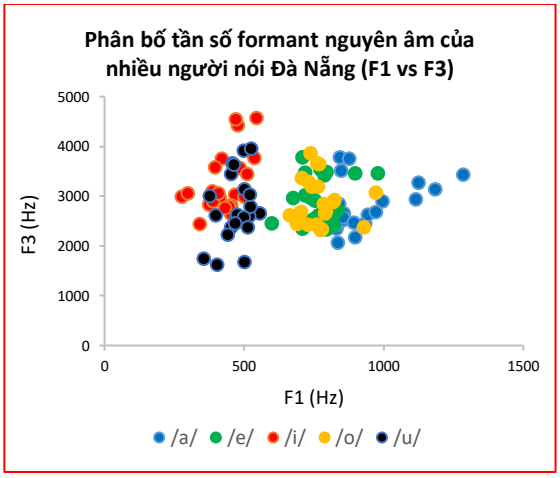
/u/	F1		F2		F3	
	Đà Nẵng	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**	**Đà Nẵng**	**Quảng Nam**
Mean (Hz)	473	440	859	794	2751	2590
STD (Hz)	51,9	55,0	108,3	92,9	667,0	570,5
CV (%)	11,0	12,5	12,6	11,7	24,2	22,0

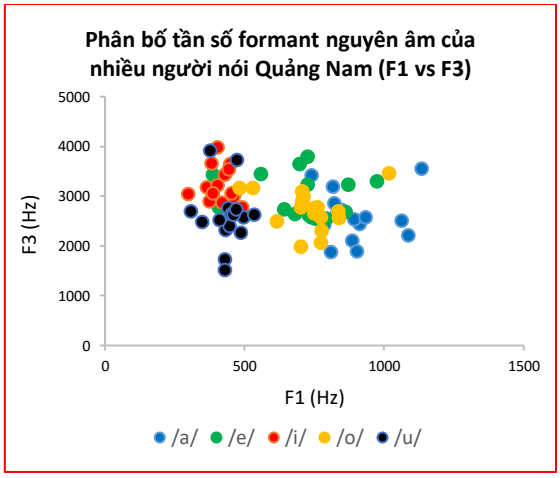
Kết quả từ Bảng 3.13 đến Bảng 3.17 cho thấy nhìn chung người nói giọng nói người Đà Nẵng có tần số formant cao hơn (có 12/15 tần số formant F1, F2, F3 05 nguyên âm giọng nói người Đà Nẵng cao hơn so với giọng người Quảng Nam) và cũng có vùng tần số formant rộng hơn (có 10/15 vùng tần số formant F1, F2, F3 rộng hơn) so với giọng nói người Quảng Nam.

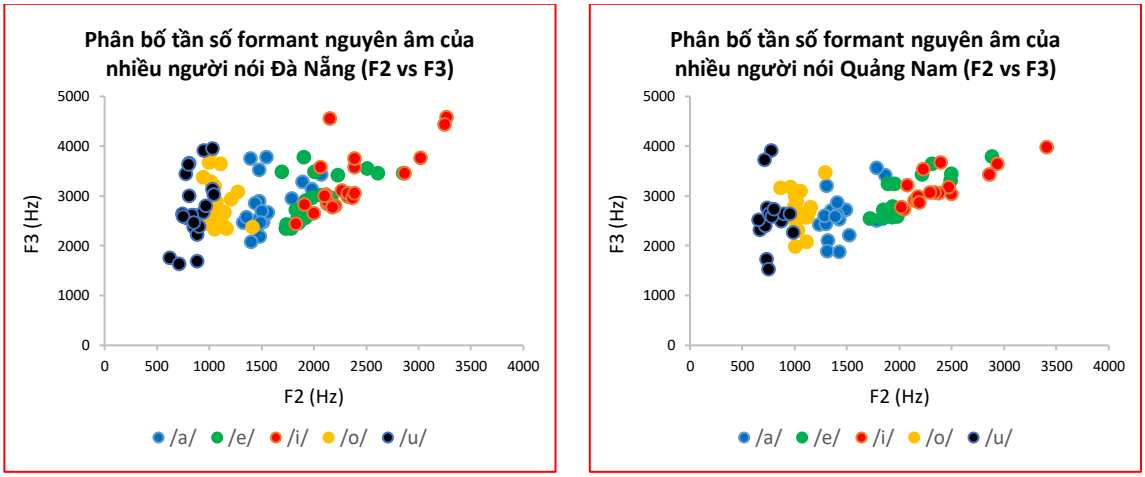
Sự phân bố tần số formant nguyên âm của người nói Đà Nẵng và Quảng Nam được thể hiện ở Hình 3.8 dưới đây:











Hình 3.8. Phân bố tần số formant nguyên âm của người nói giọng Đà Nẵng và Quảng Nam.

- Đối với độ lệch chuẩn: nguyên âm có tần số formant càng lớn thì sẽ có độ lệch chuẩn càng lớn tương ứng. Tần số bậc cao (F3) ở giọng người Đà Nẵng có độ lệch chuẩn cao hơn so với giọng người Quảng Nam.

- Đối với hệ số biến thiên CV%: Tần số formant cả 05 nguyên âm ở cả người nói giọng Đà Nẵng và Quảng Nam đều có CV dưới 20%, ngoại trừ tần số F3 nguyên âm /u/ (giọng Đà Nẵng, Quảng Nam có CV lần lượt là 24,2 % và 22 %) và tần số F1 nguyên âm /e/ của giọng người Quảng Nam (21 %) . Nhìn chung 05 nguyên âm đều có tần số formant ít biến thiên và khá ổn định ở giọng người Đà Nẵng và Quảng Nam.

Kết luận chung: Giọng nói người Đà Nẵng có tần số formant 05 nguyên âm cao hơn một chút và có vùng tần số rộng hơn so với giọng nói người Quảng Nam. Nguyên âm có tần số formant càng lớn thì cũng sẽ có độ lệch chuẩn càng lớn tương ứng. Giọng người Đà Nẵng và Quảng Nam đều có tần số formant ít biến thiên và khá ổn định với hầu hết CV đều nhỏ hơn 20 %.

3.7. Đánh giá độ chính xác của thuật toán

Để đánh giá được độ chính xác của thuật toán tôi sử dụng thông số thống kê sai số phần trăm để so sánh kết quả của thuật toán đo tần số formant F1, F2, F3 tự động kết hợp hiệu chỉnh thủ công ứng dụng phương pháp xử lý đồng hình

và phương pháp LPC [11] so với dữ liệu chuẩn tần số formant được đo thủ công trên phần mềm Wavesurfer.

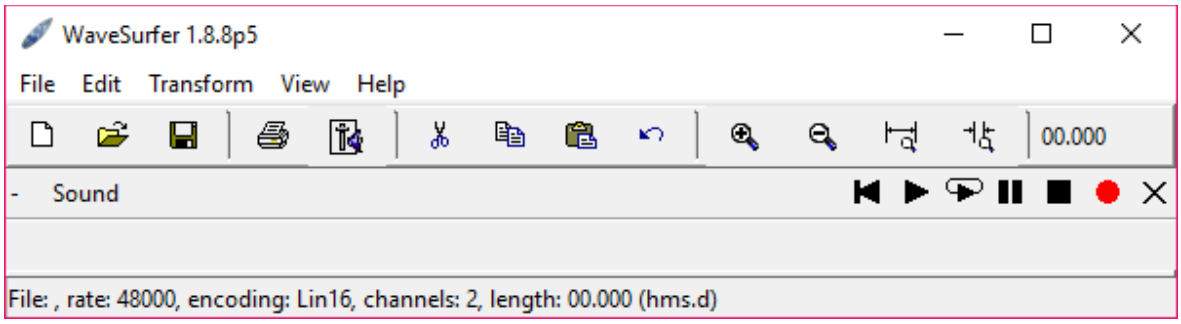
3.7.1. Sai số phần trăm (Percentage error)

Sai số phần trăm là phép đo sự khác biệt giữa giá trị đo lường so với giá trị đã biết. Việc tính toán sai số phần trăm liên quan đến việc sử dụng sai số tuyệt đối, đơn giản là hiệu của giá trị đo lường với giá trị đã biết. Sau đó, sai số tuyệt đối được chia cho giá trị đã biết, dẫn đến sai số tương đối, được nhân với 100 để có được sai số phần trăm. Ta có công thức như sau:

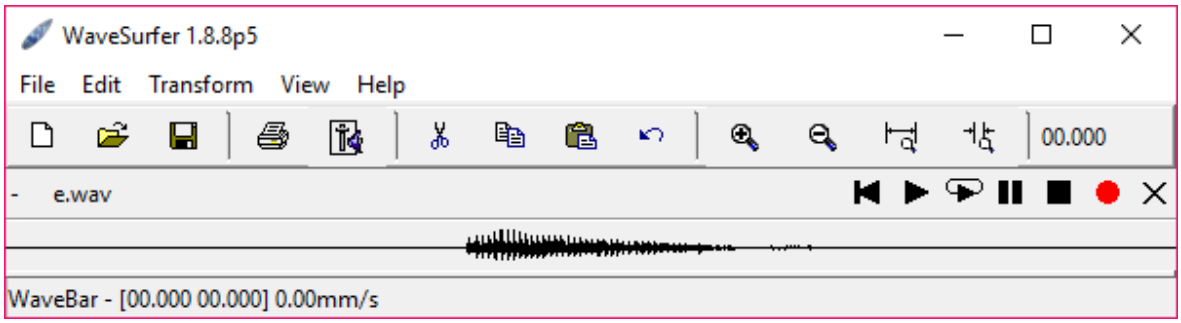
$$Percentage\ error = \frac{|giá\ trị\ đo\ lường - giá\ trị\ đã\ biết|}{giá\ trị\ đã\ biết} \times 100 \qquad (15)$$

3.7.2. Cách đo tần số formant thủ công

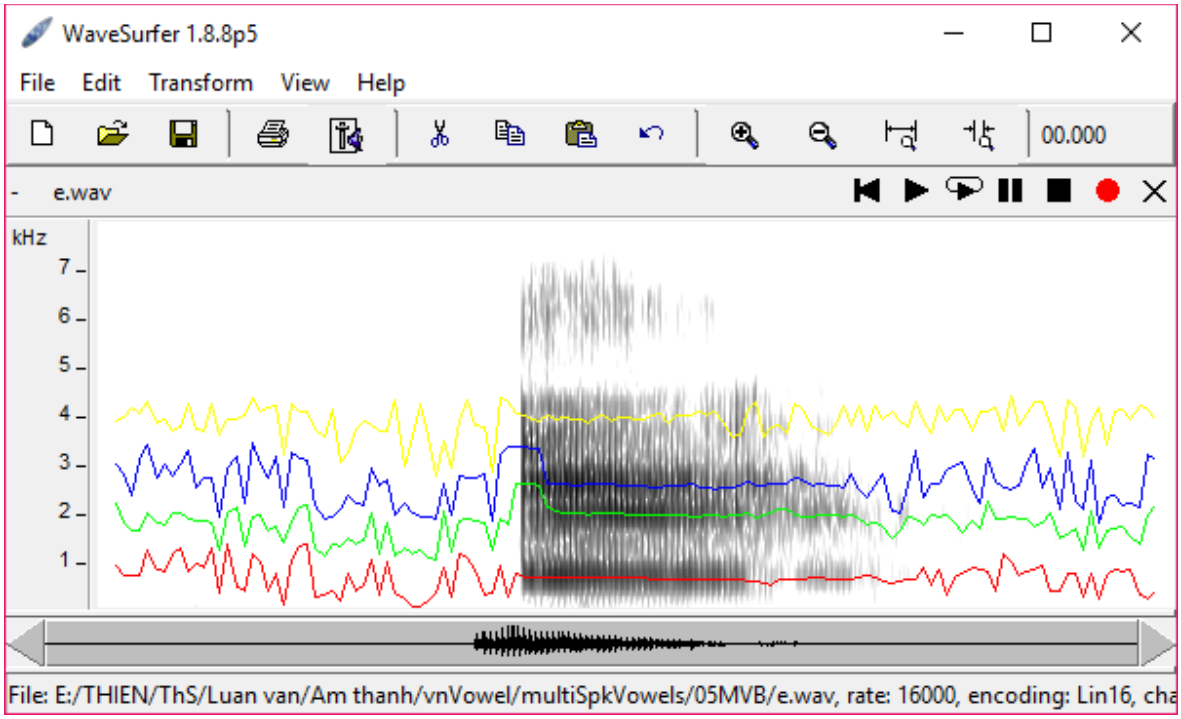
Để đo tần số formant F1, F2, F3 thủ công, tôi sử dụng phần mềm Wavesurfer phiên bản 1.8.8p5. Cách đo đã được nêu ở mục 3.4: thực hiện đo 03 lần tần số formant F1, F2, F3 trên 03 khung khác nhau (30 ms) của tín hiệu và sau đó lấy giá trị trung bình để xác định tần số formant cuối cùng.



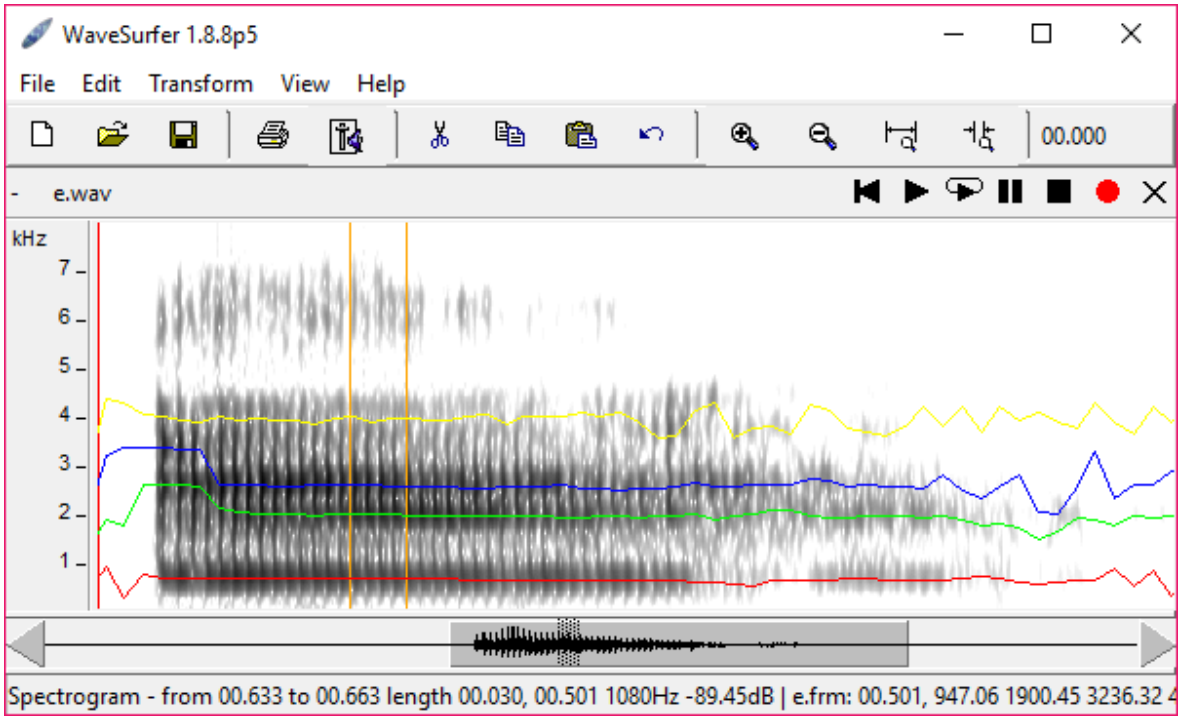
Hình 3.9. Giao diện phần mềm Wavesurfer.



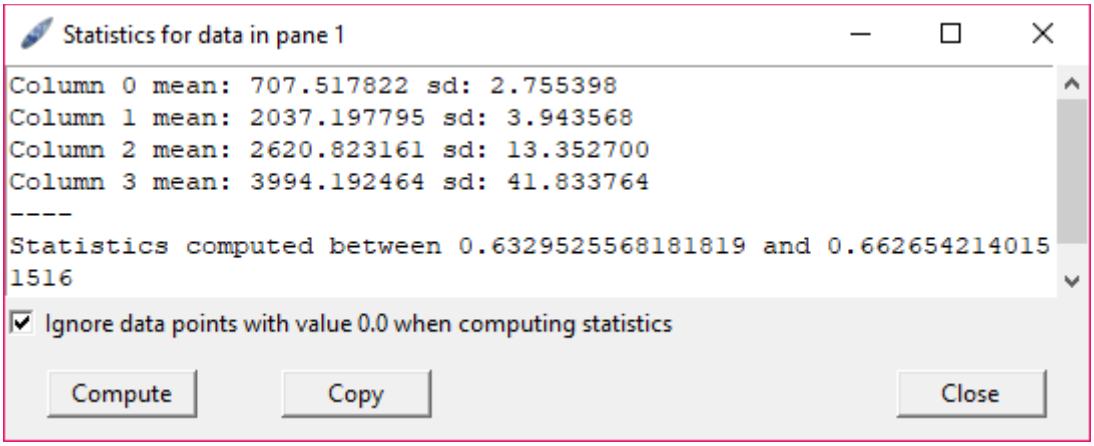
Hình 3.10. Import âm hữu thanh.



Hình 3.11. Tạo Formant Plot từ tín hiệu đầu vào.



Hình 3.12. Chọn 01 khung tín hiệu 30 ms.



Hình 3.13. Kết quả formant F1, F2, F3 đo thủ công.

3.7.3. Kết quả đánh giá

Trọng phạm vi luận văn, tôi phân tích mean tần số formant F1, F2, F3 05 nguyên âm của 50 người nói ở ba phương pháp xử lý đồng hình (Cepstrum), LPC và đo thủ công bằng phần mềm Wavesurfer, từ đó tính được sai số phần trăm giữa hai phương pháp Cepstrum, LPC so với đo thủ công. Kết quả thu được Bảng 3.18 như sau:

Bảng 3.18. Kết quả so sánh formant F1, F2, F3 giữa hai phương pháp xử lý đồng hình và LPC so với đo thủ công.

Nguyên âm	Formant	Cepstrum (Hz)	LPC (Hz)	Thủ công (Hz)	Sai số (%) Cepstrum và Thủ công	Sai số (%) LPC và Thủ công
/a/	F1	917	878	877	4,56	0,04
	F2	1517	1521	1510	0,46	0,69
	F3	2745	2757	2775	1,08	0,66
/e/	F1	742	692	712	4,21	2,71
	F2	2049	2010	2036	0,63	1,31
	F3	2940	2950	2953	0,44	0,10
/i/	F1	431	408	410	5,19	0,45
	F2	2398	2321	2376	0,91	2,29
	F3	3297	3349	3285	0,38	1,96
/o/	F1	758	728	744	1,85	2,09
	F2	1073	1088	1095	2,01	0,59
	F3	2805	2762	2797	0,28	1,27

Nguyên âm	Formant	Cepstrum (Hz)	LPC (Hz)	Thủ công (Hz)	Sai số (%) Cepstrum và Thủ công	Sai số (%) LPC và Thủ công
/u/	F1	452	425	440	2,76	3,47
	F2	810	830	819	1,03	1,36
	F3	2683	2690	2689	0,20	0,03

Từ Bảng 3.18 cho ra kết quả sai số phần trăm giữa thuật toán đo tần số formant F1, F2, F3 tự động kết hợp hiệu chỉnh thủ công ứng dụng phương pháp xử lý đồng hình và phương pháp LPC so với dữ liệu chuẩn được đo thủ công là rất nhỏ (hầu hết giá trị formant có sai số nhỏ hơn 5%, trong đó chỉ có 01 tần số F1 của nguyên âm /i/ ở phương pháp xử lý đồng hình cho kết quả sai số lớn hơn 5% (5,19 %) tuy nhiên có thể chấp nhận được do tần số F1 nhỏ) từ đó có thể kết luận rằng thuật toán xác định tần số formant F1, F2, F3 tự động kết hợp hiệu chỉnh thủ công ứng dụng phương pháp xử lý đồng hình có độ chính xác và tin cậy cao, ở phương pháp LPC cũng cho ra kết quả tương tự như vậy.

3.8. Tổng kết chương

Trong chương này, tôi đã tiến hành cài đặt thuật toán xác định formant F1, F2, F3 của 05 nguyên âm /a/, /e/, /i/, /o/, /u/ của người nói ứng dụng phương pháp xử lý đồng hình trên phần mềm Matlab. Qua đó đã phân tích đặc trưng formant của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói.

Với kết quả sai số phần trăm so với dữ liệu chuẩn được đo thủ công hầu như nhỏ hơn 5% thì có thể kết luận rằng thuật toán xác định tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình có độ tin cậy cao.

KẾT LUẬN

1. Những việc đã hoàn thành

Với mục tiêu chính của đề tài là nghiên cứu ứng dụng phương pháp xử lý đồng hình xác định tần số formant trong phân tích nguyên âm của nhiều người nói nhằm khảo sát tính đặc trưng của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói và phân tích ưu nhược điểm của phương pháp, tôi đã thực hiện được các việc sau:

- Nghiên cứu lý thuyết liên quan đến xử lý tín hiệu tiếng nói, các đặc tính cơ bản của tín hiệu tiếng nói và kỹ thuật xử lý tiếng nói ngắn hạn.

- Nghiên cứu lý thuyết về tần số formant và phương pháp xử lý đồng hình xác định tần số formant từ tín hiệu tiếng nói đầu vào.

- Cài đặt, phân tích thuật toán tính tần số formant ứng dụng phương pháp xử lý đồng hình trên phần mềm Matlab, sau đó tiến hành tính toán giá trị trung bình (Mean), độ lệch chuẩn (STD), hệ số biến thiên (CV%) để phân tích tính đặc trưng formant của các nguyên âm của cùng một người nói và một nguyên âm của nhiều người nói.

- Đánh giá độ chính xác của thuật toán từ việc so sánh kết quả của thuật toán xác định tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình và phương pháp LPC so với dữ liệu chuẩn được đo thủ công.

2. Các kết luận

Phương pháp xử lý đồng hình là một phương pháp đơn giản, hiệu quả để trích xuất riêng rẽ tần số cơ bản và tần số formant của tín hiệu tiếng nói.

Đặc trưng formant 05 nguyên âm chưa phân biệt rõ giữa các người nói với nhau, cần kết hợp nguyên âm cùng với phụ âm và thanh điệu thì mới có thể phân tích được đặc trưng formant giữa các người nói với nhau.

Đặc trưng formant 05 nguyên âm giữa nam và nữ không có sự khác biệt đáng kể. Người nói giọng nữ có giá trị trung bình tần số formant 05 nguyên âm cao hơn giọng nam.

Cả 05 nguyên âm đều có tần số formant F1, F2, F3 ít biến thiên và khá ổn định với hầu hết hệ số biến thiên đều nhỏ hơn 20 %.

Giọng nói người Đà Nẵng có giá trị trung bình tần số formant 05 nguyên âm cao hơn một chút và có vùng tần số formant rộng hơn so với giọng nói người Quảng Nam.

Thuật toán xác định tần số formant F1, F2, F3 ứng dụng phương pháp xử lý đồng hình cho kết quả tương tự so với dữ liệu chuẩn được đo thủ công nên có độ chính xác, tin cậy cao.

3. Hạn chế và hướng phát triển

3.1. Hạn chế

Dữ liệu tín hiệu tiếng nói chưa đa dạng, phong phú. Trong phạm vi luận văn tập trung phân tích tần số formant của 05 nguyên âm chính /a/, /e/, /i/, /o/, /u/ nên có thể chưa đánh giá được đầy đủ hiệu năng của thuật toán khi xử lý tín hiệu tiếng nói có tính chất biến đổi phức tạp hơn.

3.2. Hướng phát triển

Luận văn cần mở rộng dữ liệu tín hiệu tiếng nói, đa dạng về thành phần, số lượng tỉnh/thành. Việc phân tích các nguyên âm còn lại kết hợp các phụ âm, thanh điệu khác nhau sẽ khảo sát được tính đặc trưng formant các âm tiếng Việt giữa các người nói với nhau ứng dụng phương pháp xử lý đồng hình. Sau đó, luận văn sẽ mở rộng phân tích, so sánh tất cả phương pháp xác định tần số formant các âm tiếng Việt của người nói nhằm tìm ra phương pháp tối ưu nhất.

PHỤ LỤC 1

KẾT QUẢ TẦN SỐ FORMANT F1, F2, F3 CỦA 50 NGƯỜI NÓI ỨNG DỤNG PHƯƠNG PHÁP XỬ LÝ ĐỒNG HÌNH

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	M_01	Nam	Đà Nẵng	997	1482	2883	721	2073	3005	388	2273	3094	821	1076	2872	378	809	2990
2	M_02	Nam	Đà Nẵng	832	1456	2338	710	1785	2340	344	1830	2428	665	931	2606	454	854	2362
3	M_03	Nam	Đà Nẵng	936	1430	2445	824	1743	2413	277	2329	2983	931	1419	2362	399	854	2606
4	M_04	Nam	Đà Nẵng	843	1519	2473	676	1985	2949	299	2328	3049	765	1053	3182	472	783	2555
5	F_05	Nữ	Quảng Nam	743	1868	3409	561	2498	3435	449	2939	3635	532	965	3160	488	994	2256
6	M_06	Nam	Quảng Nam	818	1310	3186	410	1940	2772	299	2499	3035	709	1059	3085	309	815	2689
7	M_07	Nam	Quảng Nam	787	1242	2417	388	2218	3426	383	2401	3658	619	1006	2491	350	876	2478
8	M_08	Nam	Đà Nẵng	843	1440	2839	721	2007	3471	486	2394	3562	787	1057	2840	476	745	2625
9	F_09	Nữ	Quảng Trị	1084	1695	2909	895	2191	3126	515	2780	3394	968	1244	2859	494	802	2917
10	M_10	Nam	Đà Nẵng	831	1441	2406	797	1738	2330	377	2129	2827	776	1053	2317	444	891	2218
11	M_11	Nam	Quảng Nam	887	1325	2099	643	1952	2728	377	2151	2894	710	1020	2883	433	737	1719
12	M_12	Nam	Quảng Nam	914	1301	2427	749	1984	2578	455	2173	2883	778	1033	2298	458	793	2614
13	F_13	Nữ	Nghệ An	961	1519	3180	809	2229	3083	525	3192	4106	820	1009	3227	445	740	3908
14	F_14	Nữ	Quảng Nam	1064	1785	2495	976	2495	3293	432	2863	3420	839	1125	2558	378	778	3900
15	M_15	Nam	Đà Nẵng	1117	1793	2931	743	2168	2949	413	2375	2957	729	1133	2413	459	812	3647
16	M_16	Nam	Đà Nẵng	786	1329	2457	600	1856	2444	469	2117	3018	691	990	2439	456	778	3433
17	M_17	Nam	Đà Nẵng	856	1467	2511	708	1918	2544	391	2132	2885	708	948	3360	357	627	1741
18	M_18	Nam	Hà Tĩnh	841	1469	2293	656	1958	2591	424	2185	2928	725	893	2339	408	781	1433
19	F_19	Nữ	Đà Nẵng	1127	1893	3269	789	2233	3410	546	3271	4565	740	1016	3856	500	951	3908
20	F_20	Nữ	Quảng Nam	1137	1787	3549	700	2318	3635	406	3410	3973	1019	1297	3453	473	720	3714
21	M_21	Nam	Quảng Nam	850	1494	2712	742	1936	2563	445	2231	3526	703	994	2768	434	675	2311
22	F_22	Nữ	Đà Nẵng	1286	2071	3420	776	2516	3535	540	3020	3754	733	1025	3272	526	1032	3942
23	M_23	Nam	Đà Nẵng	945	1547	2621	800	1879	2508	456	1912	2812	791	1151	2652	557	945	2652
24	M_24	Nam	Quảng Nam	895	1429	2525	872	1961	3229	367	2478	3167	776	1102	2562	444	745	2752
25	M_25	Nam	Đà Nẵng	858	1564	2651	767	1960	2612	422	2389	3739	794	1041	2687	520	836	2597

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
26	M_26	Nam	Quảng Nam	905	1313	1882	728	1898	3226	405	2081	3206	483	871	3153	431	754	1513
27	M_27	Nam	Quảng Nam	846	1355	2696	756	1724	2542	478	2052	2722	763	1155	2766	447	722	2389
28	M_28	Nam	Huế	817	1319	2601	747	1856	2607	455	2337	3570	754	1013	2560	453	714	1526
29	M_29	Nam	Quảng Nam	812	1431	1865	682	1973	2632	389	2378	3044	703	1011	1972	411	662	2511
30	F_30	Nữ	Đà Nẵng	1187	1983	3126	981	2860	3445	478	3253	4419	744	1056	3178	503	1036	3128
31	M_31	Nam	Bình Định	855	1610	2348	517	2184	2944	435	2141	2928	768	1156	2212	422	763	3226
32	M_32	Nam	Nam Định	805	1353	2388	749	1686	2349	392	2161	3605	700	918	3534	420	716	1796
33	M_33	Nam	Quảng Nam	1089	1523	2201	865	1927	2680	466	2339	3063	776	1121	2064	456	761	2633
34	M_34	Nam	Đà Nẵng	900	1488	2170	753	1858	2534	454	2001	2636	760	961	2424	500	756	2571
35	F_35	Nữ	Quảng Bình	1180	1784	2981	747	2260	3366	416	2845	4488	735	1177	3943	456	729	4322
36	M_36	Nam	Đà Nẵng	856	1360	2568	711	1905	3771	397	2069	3571	764	1003	3654	404	719	1621
37	M_37	Nam	Quảng Nam	831	1413	2609	733	1927	2611	462	2185	2976	752	1111	2641	498	789	2570
38	M_38	Nam	Đà Nẵng	973	1507	2678	837	1828	2713	447	2204	2785	826	1216	2921	525	974	2784
39	M_39	Nam	Quảng Nam	835	1288	2592	727	2890	3781	456	2300	3056	713	1012	2982	461	902	2624
40	M_40	Nam	Đà Nẵng	837	1401	2060	722	1846	2506	502	2106	2988	779	1166	2330	514	905	2371
41	M_41	Nam	Đà Nẵng	844	1550	3773	757	1929	2900	409	2390	3051	706	1137	2676	464	808	3626
42	M_42	Nam	Quảng Nam	936	1393	2570	796	1846	2533	422	2197	2858	752	1008	2753	473	806	2721
43	M_43	Nam	Đà Nẵng	878	1400	3745	797	1699	3479	472	2153	4540	771	1113	3636	503	890	1672
44	M_44	Nam	Quảng Trị	867	1388	2638	697	1846	2645	367	2120	3372	589	822	2133	306	529	1722
45	M_45	Nam	Huế	930	1450	3623	733	2013	3529	436	2653	3624	731	1132	2644	441	777	3563
46	F_46	Nữ	Hưng Yên	1090	1945	3378	833	2551	3696	553	2947	3739	922	1267	2776	507	849	3485
47	M_47	Nam	Đắk Lắk	796	1575	3485	729	2039	2787	349	2099	3371	759	1115	2414	375	712	1723
48	M_48	Nam	Quảng Nam	822	1411	2851	733	1849	2714	494	2023	2762	837	1145	2688	537	962	2627
49	M_49	Nam	Đà Nẵng	895	1493	2456	817	1910	2581	433	2179	2757	697	1031	2589	469	858	2449
50	F_50	Nữ	Đà Nẵng	849	1478	3509	900	2616	3446	513	2867	3442	973	1275	3067	522	1044	3024

PHỤ LỤC 2

KẾT QUẢ TẦN SỐ FORMANT F1, F2, F3 CỦA 50 NGƯỜI NÓI ỨNG DỤNG PHƯƠNG PHÁP LPC

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	M_01	Nam	Đà Nẵng	987	1520	2890	760	2050	2995	403	2186	3043	815	1116	2831	379	857	2991
2	M_02	Nam	Đà Nẵng	865	1497	2304	711	1771	2391	488	1925	2536	716	934	2554	414	881	2358
3	M_03	Nam	Đà Nẵng	934	1456	2452	867	1718	2463	276	2431	3022	918	1403	2421	284	882	2594
4	M_04	Nam	Đà Nẵng	890	1551	2386	758	1910	2797	400	2452	3373	905	1101	2526	450	881	2495
5	F_05	Nữ	Quảng Nam	986	1688	3377	480	2499	3433	470	2796	3642	474	963	3109	481	1058	2374
6	M_06	Nam	Quảng Nam	866	1345	3191	538	1935	2769	307	2418	3196	695	1076	3088	303	828	2666
7	M_07	Nam	Quảng Nam	901	1219	2488	372	2083	3500	295	2467	3589	306	1007	2481	388	913	2314
8	M_08	Nam	Đà Nẵng	789	1469	2929	680	2007	3511	459	2289	3564	730	1016	2784	464	801	2563
9	F_09	Nữ	Quảng Trị	1111	1719	2969	682	2138	3164	509	2780	4433	679	927	2867	479	815	2917
10	M_10	Nam	Đà Nẵng	792	1435	2341	719	1657	2398	405	1753	3090	818	1070	2318	445	940	2197
11	M_11	Nam	Quảng Nam	881	1306	2634	650	1784	2930	351	2054	3107	773	1066	2333	403	820	1811
12	M_12	Nam	Quảng Nam	808	1252	2401	711	1968	2576	433	2125	2723	733	986	2398	430	700	2514
13	F_13	Nữ	Nghệ An	966	1522	3198	758	2076	3071	426	3083	4075	834	1021	3226	426	736	3872
14	F_14	Nữ	Quảng Nam	1107	1718	3247	958	2501	3329	326	2910	4423	1081	1116	2548	396	797	3823
15	M_15	Nam	Đà Nẵng	975	1811	2743	692	2091	2854	389	2245	2810	761	1212	2338	434	888	3630
16	M_16	Nam	Đà Nẵng	767	1363	2425	586	1820	2428	466	1956	2976	655	1034	2454	461	872	3460
17	M_17	Nam	Đà Nẵng	814	1495	2580	699	1906	2575	336	2097	2764	712	1003	3367	304	625	1805
18	M_18	Nam	Hà Tĩnh	729	1416	2282	654	1928	2631	393	2168	2953	632	909	2345	391	757	1697
19	F_19	Nữ	Đà Nẵng	1199	1922	3195	742	2145	3433	465	3208	4556	799	1135	3851	484	870	3922
20	F_20	Nữ	Quảng Nam	1174	1827	3726	661	2350	3619	363	3259	4599	937	1590	3750	394	687	3744
21	M_21	Nam	Quảng Nam	838	1510	2618	700	1839	2592	427	2027	3317	778	1143	2640	401	702	2407
22	F_22	Nữ	Đà Nẵng	1216	2100	3408	723	2459	3527	486	2818	3633	764	1163	3231	501	990	3985
23	M_23	Nam	Đà Nẵng	744	1606	2575	760	1802	2334	465	1819	2640	748	1149	2644	491	939	2638
24	M_24	Nam	Quảng Nam	790	1454	2498	785	1933	3335	384	2537	3214	752	1102	2571	399	779	2696
25	M_25	Nam	Đà Nẵng	778	1639	2477	733	1883	2723	464	2088	3661	761	1102	2585	469	880	2515

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
26	M_26	Nam	Quảng Nam	699	1117	1588	715	1867	3271	355	2066	3185	482	855	2568	414	769	1320
27	M_27	Nam	Quảng Nam	748	1429	2727	747	1803	2592	555	1971	2740	754	1183	2787	503	823	2230
28	M_28	Nam	Huế	704	1385	2595	685	1833	2584	411	2383	3435	695	1013	2519	372	695	1514
29	M_29	Nam	Quảng Nam	660	1489	1926	628	1952	2647	430	2038	2718	634	986	2006	468	926	2335
30	F_30	Nữ	Đà Nẵng	1139	1950	2947	870	2784	3448	461	3288	4454	651	1063	3125	458	1003	3104
31	M_31	Nam	Bình Định	740	1393	2512	483	2144	2888	377	2087	2855	752	1141	2210	397	756	2888
32	M_32	Nam	Nam Định	778	1310	2355	702	1692	2351	340	2164	3540	714	948	3595	336	693	2469
33	M_33	Nam	Quảng Nam	1073	1540	2307	792	1890	2691	434	2268	3041	739	1115	2079	437	668	2619
34	M_34	Nam	Đà Nẵng	838	1472	2243	676	1809	2493	419	1924	2581	746	993	2392	475	766	2549
35	F_35	Nữ	Quảng Bình	1139	1764	3621	692	2256	3449	327	2752	4461	733	1153	3944	431	636	4228
36	M_36	Nam	Đà Nẵng	840	1390	2469	690	1842	3760	341	1997	3537	700	1005	3608	359	715	1869
37	M_37	Nam	Quảng Nam	751	1452	2559	683	1893	2640	440	2103	2872	697	1143	2605	440	797	2549
38	M_38	Nam	Đà Nẵng	963	1511	2649	806	1754	2731	449	2150	3623	811	1238	2914	500	977	2764
39	M_39	Nam	Quảng Nam	774	1210	2794	694	2764	3818	453	1779	3018	694	1077	2891	446	918	2550
40	M_40	Nam	Đà Nẵng	795	1386	2022	698	1790	2396	430	2015	2929	729	1162	2250	526	968	2334
41	M_41	Nam	Đà Nẵng	777	1598	3655	681	1846	3018	362	2393	3645	697	1167	2681	421	814	3645
42	M_42	Nam	Quảng Nam	860	1458	2591	713	1812	2493	375	2122	2765	742	1017	2637	424	806	2685
43	M_43	Nam	Đà Nẵng	827	1426	3782	746	1809	3501	449	2185	4416	754	1103	3702	477	902	1951
44	M_44	Nam	Quảng Trị	850	1390	2323	619	1811	2694	282	2149	3316	609	856	2161	295	565	1624
45	M_45	Nam	Huế	898	1498	3703	649	2067	3580	383	2614	3563	695	1128	2570	384	756	2682
46	F_46	Nữ	Hưng Yên	1164	1977	3477	734	2504	3815	495	2909	3715	738	1355	2834	500	954	3475
47	M_47	Nam	Đắk Lắk	772	1602	3248	699	1967	2502	284	2062	3151	751	1151	2396	321	738	2979
48	M_48	Nam	Quảng Nam	653	1445	2469	652	1901	2732	459	1990	2862	743	1105	2700	449	943	2630
49	M_49	Nam	Đà Nẵng	765	1550	2458	707	1896	2527	435	2088	2711	722	1043	2588	438	862	2467
50	F_50	Nữ	Đà Nẵng	777	1465	3476	483	2555	3499	481	2686	3393	667	1063	3063	494	1142	3039

PHỤ LỤC 3

KẾT QUẢ TẦN SỐ FORMANT F1, F2, F3 CỦA 50 NGƯỜI NÓI ĐO THỦ CÔNG TRÊN PHẦN MỀM WAVESURFER

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	M_01	Nam	Đà Nẵng	974	1527	2876	716	2074	2961	403	2226	3002	793	1097	2843	375	833	2956
2	M_02	Nam	Đà Nẵng	891	1490	2380	702	1767	2344	345	1827	2479	663	925	2585	414	871	2363
3	M_03	Nam	Đà Nẵng	944	1441	2453	831	1703	2468	246	2308	2997	915	1318	2421	351	817	2583
4	M_04	Nam	Đà Nẵng	907	1577	2439	705	1952	2914	324	2277	2969	774	1063	3165	458	844	2498
5	F_05	Nữ	Quảng Nam	694	1759	3381	513	2500	3417	476	2834	3670	476	961	3153	463	935	2276
6	M_06	Nam	Quảng Nam	875	1343	3198	460	1938	2750	297	2404	2970	688	1085	3069	302	834	2669
7	M_07	Nam	Quảng Nam	604	1229	2390	369	2323	3372	330	2342	3606	472	961	2577	391	924	2557
8	M_08	Nam	Đà Nẵng	828	1432	2974	699	2011	3467	478	2347	3553	756	1054	2832	463	774	2578
9	F_09	Nữ	Quảng Trị	984	1737	2927	823	2133	3177	585	2752	3309	901	1208	2828	482	820	2891
10	M_10	Nam	Đà Nẵng	771	1470	2417	707	1735	2333	304	2073	2878	743	1034	2352	442	940	2240
11	M_11	Nam	Quảng Nam	887	1314	2145	593	1965	2663	385	2083	2935	689	1036	2827	384	751	1760
12	M_12	Nam	Quảng Nam	822	1235	2410	701	2015	2605	435	2190	2892	747	981	2348	438	790	2569
13	F_13	Nữ	Nghệ An	771	1554	3143	765	2232	3055	439	3091	4057	795	1031	3208	425	763	3894
14	F_14	Nữ	Quảng Nam	1120	1727	2509	896	2366	3320	360	2962	3493	849	1167	2582	395	824	3845
15	M_15	Nam	Đà Nẵng	1023	1815	2771	703	2207	2940	406	2313	2887	719	1148	2358	417	781	3687
16	M_16	Nam	Đà Nẵng	760	1335	2476	596	1845	2493	417	2111	2905	680	1007	2477	421	769	3180
17	M_17	Nam	Đà Nẵng	764	1523	2382	696	1898	2567	351	2128	2928	710	999	3258	382	658	1713
18	M_18	Nam	Hà Tĩnh	733	1333	2318	654	1940	2641	414	2183	2934	641	903	2360	407	778	1573
19	F_19	Nữ	Đà Nẵng	1241	1951	3644	777	2162	3408	480	3066	4551	759	1109	3827	490	820	3938
20	F_20	Nữ	Quảng Nam	1178	1861	3437	709	2268	3592	354	3388	3967	943	1285	3450	393	677	3713
21	M_21	Nam	Quảng Nam	782	1539	2646	689	1869	2562	444	2150	3452	736	1130	2681	437	759	2022
22	F_22	Nữ	Đà Nẵng	1263	2157	3395	736	2437	3636	515	2944	3713	793	1168	3212	528	988	3951
23	M_23	Nam	Đà Nẵng	863	1533	2599	760	1859	2577	453	1984	2806	779	1170	2650	547	958	2635
24	M_24	Nam	Quảng Nam	837	1409	2581	923	1806	2831	360	2482	3246	791	1134	2560	420	767	2697
25	M_25	Nam	Đà Nẵng	796	1517	2604	740	1917	2609	422	2343	3670	764	1086	2621	530	906	2545

TT	ID	Giới tính	Tỉnh/ Thành	/a/			/e/			/i/			/o/			/u/		
				F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
26	M_26	Nam	Quảng Nam	748	1212	1745	732	1892	3588	391	2085	3167	516	888	3255	430	782	1613
27	M_27	Nam	Quảng Nam	787	1324	2722	746	1848	2565	490	2088	2755	756	1159	2774	464	779	2469
28	M_28	Nam	Huế	768	1313	2696	710	1858	2628	452	2311	3547	721	1039	2517	399	735	1608
29	M_29	Nam	Quảng Nam	704	1453	1871	644	1963	2662	370	2419	3131	722	1044	2013	378	631	2614
30	F_30	Nữ	Đà Nẵng	1143	2006	3026	967	2907	3549	442	3141	4435	760	1226	3105	468	986	3141
31	M_31	Nam	Bình Định	798	1561	2574	508	2156	2949	409	2128	2904	766	1171	2192	380	718	3067
32	M_32	Nam	Nam Định	787	1322	2327	703	1689	2310	366	2144	3527	716	960	3543	370	719	1798
33	M_33	Nam	Quảng Nam	1014	1416	2291	775	1917	2700	457	2305	3076	765	1111	2081	489	907	2661
34	M_34	Nam	Đà Nẵng	860	1480	2202	713	1790	2489	421	2057	2761	740	978	2380	483	763	2567
35	F_35	Nữ	Quảng Bình	1136	1795	3132	694	2274	3427	416	2979	4450	753	1204	3940	444	683	4253
36	M_36	Nam	Đà Nẵng	841	1364	3754	676	1811	3763	376	2029	3564	695	1007	3607	393	726	1780
37	M_37	Nam	Quảng Nam	756	1379	2613	682	1931	2638	446	2180	2951	717	1133	2671	452	806	2571
38	M_38	Nam	Đà Nẵng	959	1511	2664	805	1752	2769	412	2302	2938	813	1228	2957	524	973	2758
39	M_39	Nam	Quảng Nam	790	1230	2758	691	2900	3724	450	2131	3001	713	1064	2913	488	909	2784
40	M_40	Nam	Đà Nẵng	817	1397	2085	710	1794	2400	423	2083	2983	726	1151	2289	530	936	2360
41	M_41	Nam	Đà Nẵng	817	1422	3784	703	1879	2997	374	2355	3042	720	1142	2702	441	813	3657
42	M_42	Nam	Quảng Nam	873	1320	2612	702	1875	2630	390	2186	2813	766	1039	2687	492	848	2730
43	M_43	Nam	Đà Nẵng	825	1491	3773	751	1830	3488	451	2126	4501	757	1110	3618	476	850	1861
44	M_44	Nam	Quảng Trị	818	1338	2445	646	1806	2688	304	2161	3325	634	882	2151	324	545	1723
45	M_45	Nam	Huế	929	1417	3634	697	2048	3577	402	2601	3587	724	1134	2572	418	822	3553
46	F_46	Nữ	Hưng Yên	1167	1927	3412	782	2597	3781	534	2929	3714	835	1298	2868	510	855	3512
47	M_47	Nam	Đắk Lắk	764	1608	3449	678	2025	2781	334	2114	3104	738	1096	2473	361	764	1725
48	M_48	Nam	Quảng Nam	831	1417	2654	774	1860	2724	465	2069	2802	803	1118	2702	503	982	2633
49	M_49	Nam	Đà Nẵng	838	1541	2531	757	1906	2571	417	2225	2793	765	1051	2526	466	865	2526
50	F_50	Nữ	Đà Nẵng	792	1466	3492	868	2590	3543	484	2841	3498	1006	1433	3073	556	956	3146

TÀI LIỆU THAM KHẢO

Tiếng Việt

[1] Nguyễn Văn Ái, (1973), *Tìm hiểu về vùng tần số fooc-man của các nguyên âm tiếng Việt bằng phương pháp thực nghiệm*, 4, Tạp chí Ngôn ngữ.

[2] Nguyễn Văn Ái, (1974), *Bàn về số lượng và sự phân bố fooc-man của các nguyên âm đơn tiếng Việt qua bản ghi Xô-na-gơ-rap*, 1, Tạp chí Ngôn ngữ.

[3] Vũ Kim Bảng, (2002), *Hệ formant của nguyên âm tiếng Hà Nội*, 15, Tạp chí Ngôn ngữ.

[4] Ngô Minh Dũng, (2010), *Nghiên cứu kỹ thuật nhận dạng người nói dựa trên từ khoá tiếng Việt*, Luận án Tiến sĩ Kỹ thuật.

[5] Phạm Văn Sự and Lê Xuân Thành, (2010), *Bài giảng Xử lý tiếng nói*, Học viện Công nghệ Bưu chính Viễn thông.

Tiếng Anh

[6] Apte S D, (2017), *Random Signal Processing*, CRC Press.

[7] Fant G, (1960), *Acoustic Theory of Speech Production*, Mouton & Co, The Hague, Netherlands.

[8] Han M S, (1966), *Studies in the Phonology of Asian Languages; IV, Vietnamese Vowels*, Los Angeles: Acoustic Phonetics Research Laboratory: University of Southern California.

[9] John G P and Dimitris G M, (1995), *Digital Signal Processing: Principles, Algorithirms & Applications*, Prentice Hall; United States of America, pp. 118.

[10] Kammoun M A, Gargouri D, Frikha M, and Hamida A B, (2004), "Cepstral method evaluation in speech formant frequencies estimation", *2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT'04*. 3, pp. 1612-1616.

[11] Kammoun M A, Gargouri D, Frikha M, and Hamida A B, (2006), "Cepstrum vs. LPC: A comparative study for speech formant frequencies estimation", *GESTS Int'l Trans. Communication and Signal Proce*. 9(1), pp. 87-102.

[12] Lawrence R R and Ronal W S, (1978), *Digital Processing Of Speech Signals*, Prentice Hall, pp. 118.

[13] Oppenheim A and Schafer R, (1968), "Homomorphic analysis of speech", *IEEE Transactions on Audio and Electroacoustics*. 16(2), pp. 221-226.

[14] Rahman M S and Shimamura T, (2005), "Formant frequency estimation of high-pitched speech by homomorphic prediction", *Acoustical science and technology*. 26(6), pp. 502-510.

[15] Vinay K I and John G P, (2012), *Digital Signal Processing Using MATLAB*, Cengage Learning, United States of America.

[16] Zazula D and Gyergyek L, (1992), "Complexity in signal processing using cepstral approach", *Electro technical Review*. 59(3-4), pp. 165-170.