

MỞ ĐẦU

1. LÍ DO CHỌN ĐỀ TÀI

Tiếng Việt là một ngôn ngữ thuộc loại hình đơn lập, trong đó mỗi âm tiết là một đơn vị hoàn chỉnh độc lập, là đơn vị cuối cùng trong việc phân xuất các đơn vị của lời nói, đồng thời âm tiết cũng là cơ sở để phân tích cấu trúc kết hợp các âm vị.

Các âm vị trong tiếng Việt có ba loại: nguyên âm, phụ âm và thanh điệu. Nguyên âm và phụ âm là những âm vị tuyến tính, chúng kết hợp với nhau theo trật tự trước sau trong quá trình phát âm. Chúng là những âm vị có thể phân tách ra thành từng đơn vị nhỏ hơn, nên còn gọi là những âm vị đoạn tính. Trái với loại âm vị trên, thanh điệu là âm vị phi tuyến tính, bao trùm toàn bộ âm tiết và gắn liền với âm tiết trong suốt quá trình phát âm. Thanh không thể chiết ra khỏi âm tiết được, mà nhất thiết phải gắn với âm tiết, vì thế nên thanh còn gọi là âm vị siêu đoạn tính.

Chính do vai trò của âm tiết như vậy nên từ lâu nhiều tác giả đã quan tâm nghiên cứu đơn vị cơ bản này, đặc biệt là bằng phương pháp thực nghiệm. Các công trình thực nghiệm đã được công bố nhiều hơn cả là nghiên cứu về *hệ thanh điệu và formant của nguyên âm đơn*.

Có thể coi Lê Văn Lý là người đầu tiên nghiên cứu thực nghiệm về thanh điệu. Những nghiên cứu thực nghiệm về *thanh điệu* tiếng Việt còn có thể kể: Nguyễn Hàm Dương (1962); Han Mieko(1968), Hoàng Cao cương (1982); Cao Xuân Hạo (1986), Vũ Kim Bảng (1984), Nguyễn Văn Lợi, Vũ Thanh Phương, Andrea Phạm Hoa...

Những nghiên cứu thực nghiệm về *formant* tiếng Việt có thể kể: Nguyễn Văn Ái (1973, 1974); Vũ Kim Bảng (2002).

Tuy nhiên, trong phạm vi âm tiết tiếng Việt, xét về các đặc trưng âm học của nó, còn có một số vấn đề chưa được đề cập tới, đó là: mối quan hệ hay vai trò của âm đầu và âm cuối đối với nguyên âm (formant) và thanh điệu (âm vực và đường nét). Những kết quả nghiên cứu này có giá trị cả về mặt lí thuyết cũng như ứng dụng thực tiễn.

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 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 Interface) truyền thống vào phiên bản điều hành Windows thế hệ mới.

Tại Việt Nam, nội dung nghiên cứu về *nhận dạng tiếng Việt* đã được đề cập tới từ năm 1981. Bắt đầu từ năm 1991 cho đến nay, trong các chương trình, đề tài về khoa học và công nghệ, vấn đề *nhận dạng và tổng hợp tiếng Việt* luôn là một nội dung được ưu tiên. Gần đây, nhiều đề tài các cấp về *nghiên cứu phát triển công nghệ nhận dạng và tổng hợp và xử lí ngôn ngữ tiếng Việt* đã có những kết quả đột phá trong lĩnh vực này.

Trong giai đoạn hiện nay, vấn đề xử lí tiếng Việt cho tổng hợp và nhận dạng là một nhu cầu cấp bách đòi hỏi có những nghiên cứu thực nghiệm cụ thể về ngữ âm tiếng Việt.

Từ những tiền đề trên chúng tôi lựa chọn đề tài: *Cấu trúc formant của nguyên âm tiếng Việt trong kết hợp với âm tắc và thanh điệu (Trên cơ sở khảo sát thực nghiệm)* làm đề tài luận án.

2. TỔNG QUAN

2.1. Những nghiên cứu về formant của nguyên âm trên thế giới

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 lý 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 kỹ 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* (regional characteristics), và *đặc trưng cá nhân của người nói* (subject characteristics).

Đặc điểm cấu trúc âm học của formant được nghiên cứu khảo sát sâu và kĩ hơn ở nhiều ngôn ngữ khác nhau mà điển hình là tiếng Anh. Việc lập bản đồ âm học các nguyên âm tiếng Anh đã thực hiện bằng cách dùng tần số formant của các nguyên âm làm cơ sở. Việc lập bản đồ nguyên âm này được tiến hành ngay sau khi xuất hiện máy phân tích phổ lời nói, do hai tác giả Peterson và Barney tiến hành vào năm 1952, các tác giả đã phân tích một tập

hợp gồm 10 nguyên âm đơn của mỗi người nói trong tổng số 76 người gồm cả nam, nữ và trẻ em [126, tr 175].

Các tác giả cũng chỉ ra rằng ở những người khác nhau có sự hiện thực hoá về mặt ngữ âm rất khác nhau trong phát âm các nguyên âm của họ, nhưng họ vẫn duy trì các đối lập hệ thống một cách bình thường. Vì vậy, hai người phát âm có thể thay đổi về hình dáng khoảng cách nguyên âm, và các giá trị formant đối với âm /E/, chẳng hạn, ở người này có thể gần với các giá trị formant đối với âm /Θ/ ở người kia; nhưng cả hai người đều sẽ thể hiện nét khu biệt âm học thoả đáng giữa /E/ và /Θ/ trong hệ thống nguyên âm của họ. Bên cạnh đó, những người khác nhau có trường độ riêng của bộ máy phát âm, và những khác biệt này có ảnh hưởng cố hữu lên mô hình các giá trị formant theo nhiều cách mà người phát âm cũng thực sự không thể kiểm soát nổi. Vì thế không có một phân biệt tuyệt đối về mặt âm học giữa các nguyên âm cận kề hoặc gần nhau đối với nhiều người khác nhau. Là người sử dụng ngôn ngữ mọi người hầu như không gặp khó khăn gì trong việc giải quyết vấn đề này, vì chúng ta có thể, chỉ cần một mẫu lời nói rất nhỏ của một người nói bình thường đối với hệ thống nguyên âm của người đó. Nói cách khác, chúng ta có thể điều chỉnh bản đồ này ở tất cả mọi người, nhưng vẫn có những vùng chung nhất định phụ thuộc vào giới tính, tuổi tác cũng như giọng nói của từng vùng phương ngữ [126, tr 178].

Khi nghiên cứu tần số của các formant phía trên, Peterson và Barney (1952) đã tiến hành thu âm các nguyên âm tự nhiên ở các từ bắt đầu bằng /h/ và kết thúc bằng /d/ của nhiều người khác nhau (đàn ông, phụ nữ và trẻ em), các tác giả đã phân tích cấu trúc formant của những nguyên âm này, và hình thành nên những nghiên cứu thắm nhận bằng cách dùng các bản ghi âm tương tự. Từ phân tích của hai ông, rồi đến công trình của Shepard (1972) sau này, đã chỉ ra rằng ở đâu có sự nhầm lẫn thắm nhận thì ở đó chúng có liên quan

mật thiết với sự gần gũi âm học được xác định bằng ba formant thấp nhất. Tư liệu của họ cũng cho thấy một mức thay đổi đáng kể giữa các nguyên âm được coi là điển hình và sự gồi lên nhau giữa các nguyên âm khác nhau. Các tác giả đã chỉ ra sự thay đổi của rất nhiều nguyên âm khác nhau khi các formant của chúng được dựng lên so sánh với nhau [126, tr 176].

Những tư liệu này đã hé mở một khía cạnh quan trọng trong hoạt động tiếp nhận nguyên âm, đặc biệt là bản chất hệ thống của formant liên quan đến các giá trị tuyệt đối của chúng. Họ đã lưu ý rằng có một sự đa dạng trong các thuộc tính âm học của các nguyên âm ở trẻ em, phụ nữ và đàn ông nảy sinh từ những khác biệt về độ dài của bộ máy phát âm, cũng như sự đa dạng do những khác biệt giữa các cá thể ở bộ máy phát âm và thay đổi mang tính thói quen của cơ quan phát âm. Kết quả là nhiều nhà nghiên cứu đã lập ra các thuật toán nhằm chuẩn hoá nét biến thể của tư liệu, trên cơ sở sự khác biệt về trường độ của bộ máy phát âm.

Thí nghiệm của Ladefoged và Maddieson chứng minh các tần số formant xác định đặc trưng âm vị học của hệ thống nguyên âm. Bằng việc dùng lời nói tổng hợp, họ đã chứng minh nếu một hệ thống nguyên âm trong một câu bị dịch chuyển thì người nghe sẽ thẩm nhận sự di chuyển mang tính hệ thống của nguyên âm vào trong đường giao âm học của một nguyên âm khác xét về mặt âm vị học.

Ví dụ: nguyên âm của từ *head* (cái đầu) có thể bị làm cho nghe thành nguyên âm của từ *hid* (trốn) nếu các tần số formant của tất cả các nguyên âm khác bị hạ xuống. Dường như người nghe có thể tiếp nhận một cách bình thường đối với một người nói ở một vài từ ban đầu mà họ nghe thấy [109, tr 74].

Bên cạnh những kết quả nghiên cứu cấu trúc formant của tiếng Anh cấu trúc formant của các nguyên âm tiếng Thụy Điển cũng được nghiên cứu rất kĩ

lượng trong các công trình của G. Fant và các cộng sự (1959). Việc khảo sát thực nghiệm được thực hiện với một nhóm 24 sinh viên nam tại Viện Kỹ thuật Hoàng gia, những người từng tham gia một khóa học truyền thông, họ đọc một danh sách các nguyên âm độc lập tiếng Thụy Điển bao gồm: /o, a, ǎ, e, i, y, u, o_ / với khoảng cách 1,5 giây giữa các âm. Tương ứng với bảng phiên âm quốc tế IPA của các ký tự được phát âm là /u:, o:, a:, e:, y:, i:/. Tần số formant được đo qua bộ lọc mẫu phổ tại một vị trí thời gian bằng 1/4 độ dài của nguyên âm. Mục đích trong nghiên cứu của Fant là tập hợp một dữ liệu liên quan đến tần số formant của nguyên âm tiếng Thụy Điển. Những nghiên cứu của Fant (1959) nhằm phục vụ những ứng dụng của công ty điện thoại Ericsson từ năm 1946 đến 1948. Thời gian đầu những nguyên âm trong nghiên cứu được thể hiện trong khoảng thời gian là 4 giây để thực hiện một dải phân tích tần số. Những nghiên cứu bước đầu của Fant có những hạn chế trong việc tạo ra giọng nói tổng hợp một cách đều đặn và giọng này được thể hiện giống như tiếng hát hơn là lời nói. Trong nghiên cứu tiếp theo, dữ liệu được bổ sung là giọng nói của trẻ em và phụ nữ, giới hạn của nghiên cứu hiện tại chỉ gồm các nguyên âm dài [92, tr 1]. Ngoài ra, trong công trình “Stops in CV”, Fant cũng đã mô tả ảnh hưởng của phụ âm tắc đến formant của nguyên âm tiếng Thụy Điển ở trường độ và vùng tần số giữa phụ âm và nguyên âm.

Năm 1998, Manjari Ohala và John J. Ohala đã tiến hành khảo sát sự tương quan giữa các chuyển tiếp (formant) trong cấu trúc VC phụ âm tính (trong đó V là các nguyên âm /a i u/ còn C là các phụ âm /p t^h □ t^h k/) với mức độ lẫn lộn của thính giác về đối lập vị trí trong tiếng Hindi. Bằng thủ pháp thử và sai, tác giả đã đánh giá vai trò của các chuyển tiếp formant VC độc lập với sự giải phóng âm tắc [124, tr1].

Năm 1999, James J.Hant và Abeer Alwan đã nghiên cứu và lập mô hình cho chuyển tiếp formant trong bối cảnh nhiễu. Đánh giá cao vai trò quyết

định của chuyển tiếp formant trong việc nhận ra vị trí cấu âm của các phụ âm, các tác giả đã đưa ra giả định nếu chuyển tiếp formant này “nhúng” vào trong một môi trường nhiễu, và tiến hành xác định ngưỡng của các trượt lướt thanh điệu (tone glides) và đường formant có thể ở các tần số khác nhau, có trường độ khác nhau. Kết quả nghiên cứu cho thấy ngưỡng là độc lập với mức tần số và chỉ phụ thuộc vào trường độ và tần số trung tâm của chuyển tiếp [96, tr 1896].

Năm 1958, trong bài viết *Effect of Third-Formant Transitions on Perception of Voiced Stop Consonants* đăng trên *The Journal of Acoustical Society of America*, Vol. 30, No. 2, 122-126, tác giả dựa trên lời nói tổng hợp (synthetic speech) đã tiến hành thực nghiệm và đi đến nhận xét “chuyển tiếp formant thứ ba là chìa khóa cho việc thâm nhận các âm /b d t/”, “sự đa dạng của chuyển tiếp formant thứ ba tương ứng cặp đôi với chuyển tiếp formant thứ hai ở vị trí đầu trước nguyên âm /i/ và /e/”. Từ kết quả miêu tả về chuyển tiếp và mức độ ổn định của formant thứ ba các tác giả đi đến nhận xét “sự tồn tại của locus formant thứ ba tương tự với locus đã được tìm thấy trước đó đối với formant thứ nhất và thứ hai” [103, tr 124].

Cũng năm 1958, P.C. Delattre, A. M. Liberman, F. S. Cooper qua khảo sát âm xát tiếng Anh Mỹ, đã cho rằng chuyển tiếp formant là các tương liên âm học của vị trí cấu âm. Tác giả cũng đã tổng kết những nghiên cứu đã có trước đó và cho rằng có hai loại tín hiệu âm học để nhận ra vị trí cấu âm của phụ âm: một là *những đặc trưng bản thể*, có ở âm tắc (stop) ,âm xát, và âm tắc xát, là điểm mà tỉ lệ tần số của tiếng ồn hay tiếng rít được tạo ra tại điểm rít, và hai là *những chuyển tiếp hay biến đổi* (sự thay đổi tần số tương đối nhanh) của các formant thứ hai và thứ ba. Trong bài viết này, các tác giả cũng khẳng định chuyển tiếp của formant thứ hai và thứ ba là những dấu hiệu nhận dạng vị trí cấu âm của phụ âm theo hướng chuyển dịch tần số [87, tr 773].

Năm 1955, nhóm tác giả này trong bài viết "*Acoustic Loci and Transitional Cues for Consonants*" cho rằng những nghiên cứu tổng hợp tiếng nói trước đây đã mô tả sự quá độ của formant thứ hai là những tiêu điểm cho nhận dạng các phụ âm tắc và mũi. Kết quả của những thí nghiệm này có thể đơn giản hóa nếu nó giả định rằng: mỗi phụ âm có một nét đặc trưng và vị trí tần số cố định, hoặc tiêu điểm cho formant thứ hai, tương ứng với vị trí cố định của cơ chế tạo phụ âm. Trên nền tảng đó, những dấu hiệu quá độ có thể được xem như "những sự biến động (sự vận động)" từ tiêu điểm tới trạng thái bền vững của nguyên âm. Thí nghiệm trong bài viết này đã cung cấp thêm một bằng chứng liên quan đến sự tồn tại và những vị trí là tiêu điểm của formant thứ hai cho những phụ âm tắc hữu thanh /b, d, g/. Ở đó xuất hiện một tiêu điểm cho /d/ tại 1800 cps và cho /b/ tại 720 cps. Một tiêu điểm cho /g/ có thể được chứng minh chỉ khi nguyên âm bên cạnh có formant thứ hai của nó trên khoảng 1200 cps. Đối với những phụ âm tắc hữu thanh, thời kì quá độ không thể bắt đầu tại tiêu điểm và đi từ điểm đó đến vị trí cố định của nguyên âm. Đúng hơn là, nếu chúng ta nghe phụ âm một cách thích hợp, phần đầu của đoạn quá độ phải là khoảng lặng. Và nghiên cứu chỉ ra rằng, trong tổng hợp lời nói phụ âm tắc hữu thanh được tổng hợp tốt nhất bằng cách tạo ra trường độ thời gian giữa các khoảng lặng cân bằng với trường độ của chính đoạn quá độ [88, tr 771].

Năm 1995, một nghiên cứu chuyên tiếp formant giữa âm tiết đầu có kết thúc là nguyên âm và âm tiết thứ hai có nguyên âm là âm đầu trong chuỗi kết hợp hai âm tiết trong tiếng Trung Quốc chuẩn cũng đã được nghiên cứu, (Yan Jingzhu). Trong nghiên cứu này 661 chuỗi kết hợp âm tiết tiếng Trung trong đó âm tiết đầu có kết thúc là nguyên âm và âm tiết thứ hai bắt đầu bằng một nguyên âm (VV) được một người bản ngữ nam nói tiếng Bắc Kinh tạo ra, âm thanh tiếng nói là chuỗi kết hợp âm tiết được phân tích bằng máy Kay và

Sonagraph 5500. Kết quả cho thấy một sự chuyển tiếp formant được hình thành trong kết hợp âm tiết. Tần số formant có giá trị tại điểm bắt đầu của sự chuyển tiếp cân bằng với formant của nguyên âm hạt nhân của âm tiết đầu [97, tr 42].

Đặc điểm địa phương ở formant được quan tâm nghiên cứu trên phương diện tìm các đặc trưng phương ngữ thể hiện trong cấu trúc của formant. Theo hướng này, người ta mới chỉ quan tâm đến những khác biệt vùng địa phương formant của một ngôn ngữ. Ví dụ bài viết "*Tần số formant của nguyên âm trong âm tiết /hVd/ trong lời nói của phụ nữ người miền Nam Australia*" của Andrew Butcher, năm 2006. Trong khoảng thời gian 5 năm (2002 - 2006) các nguyên âm được ghi âm bởi các sinh viên năm thứ hai chuyên ngành bệnh học lời nói tại trường đại học Flinder ở Adelaide. Trong nghiên cứu này, dữ liệu chỉ bao gồm những phụ nữ trẻ, những người lớn lên ở miền nam Australia (n = 92, độ tuổi trung bình là 22 tuổi), đa số họ được huấn luyện. Trước khi phát âm danh sách các từ có cấu trúc /hVd/ được chuẩn bị trước và ba dấu hiệu của mỗi nguyên âm là đọc từ những danh sách ngẫu nhiên. (năm 2006, n = 22 ghi âm cộng thêm vào từ "Who'll (ai sẽ)" theo trật tự tạo ra dấu hiệu của tha âm vị trước bên của /u/). Trong tất cả các trường hợp lời nói được ghi âm trực tiếp trên ổ cứng máy vi tính trong môi trường yên tĩnh, số hóa tại 22,5kHz, với độ phân giải 16 - bit. Tần số formant được đo được từ sự hiển thị LPC, những điểm đo được xác định bằng sự liên quan tới những dải phổ rộng, sử dụng chương trình SIL phân tích tiếng nói (phiên bản thuộc 1.5 đến 2.7). Formant được đo ở giữa của sự phân chia đều đặn F1 lớn đối với nguyên âm mở và F1 nhỏ cho những nguyên âm đóng. Nguyên âm đóng /i/ và /u/ trong trường hợp đặc biệt thường không đạt tới F1 nhỏ này cho đến khi nguyên âm kết thúc hoàn toàn. Kết quả miêu tả giá trị trung bình

formant thu được cho tất cả các nguyên âm bao gồm cả nguyên âm đơn và nguyên âm đôi của những người ở miền nam Australia [81, tr 450].

Đặc trưng cá nhân người nói thể hiện ở cấu trúc formant bao gồm giới tính, độ tuổi và cả những người mắc các tật phát âm như nói lắp, nói ngọng hay phẫu thuật thanh quản... Và mô hình chuyển tiếp formant đặc thù cá nhân thể hiện trong một vài tổ hợp âm học. "*Cấu trúc formant của nguyên âm được tạo ra bởi những người có tật nói lắp với tốc độ nói nhanh và bình thường*". Trong nghiên cứu này tác giả Martine Toda, Shinji Maeda, Micheal Aron, và Marie-Odile Berger đã phân tích phần cố định của hai formant F1 và F2 trong cấu trúc CV, chứa nguyên âm /a, i, u/ được phát âm với hai tốc độ nhanh và bình thường, bởi hai nhóm: được điều trị tật nói lắp và không được điều trị tật nói lắp. Quỹ tích bằng nhau được tính toán để quan sát những khả năng khác nhau trong đồng cấu âm giữa ba nhóm. Dữ liệu phân tích bộc lộ sự giảm bớt khoảng cách của nguyên âm cho những người mắc tật nói lắp trong lời nói bình thường. Khi tốc độ lời nói tăng lên, dễ nhận thấy khoảng cách nguyên âm không hề giảm đối với những nhóm người nói sau, tương phản với cách xem xét những người nói lắp với tốc độ nhanh [134, tr 357].

2.2. Những nghiên cứu về formant của nguyên âm tiếng Việt

Nghiên cứu formant của nguyên âm tiếng Việt đã được một tác giả người nước ngoài là Han Mieko (1968) đề cập đến khi thực hiện một loạt nghiên cứu âm vị học các ngôn ngữ châu Á. Cụ thể, trong công trình "*Studies in phonology of Asian Languages: Vietnamese Vowel*" 11 nguyên âm đơn tiếng Hà Nội đã được nghiên cứu và mô tả. F1 và F2 của 11 nguyên âm được đo, phân tích, cuối cùng tác giả đi đến kết luận rằng có sự chi phối ảnh hưởng bởi các yếu tố khác là thanh tính và phụ âm tính đến chất lượng của nguyên âm. Khi thực hiện một nghiên cứu về trường độ cố hữu của 11 nguyên âm đơn tiếng Việt, kết quả nghiên cứu của Han Mieko cho thấy có sự ảnh hưởng