

Mục lục

| | |
|---|----|
| Lời nói đầu..... | 2 |
| Phần 1: Tổng quan về tiếng nói..... | 3 |
| 1.1 Nhận thức chung..... | 3 |
| 1.2 Đặc tính vật lý của âm thanh..... | 4 |
| 1.2.1 Độ cao (Pitch)..... | 4 |
| 1.2.2 Cường độ..... | 5 |
| 1.2.2 Trường độ..... | 5 |
| 1.2.3 Âm sắc..... | 5 |
| 1.3 Đặc tính âm học của âm thanh..... | 6 |
| 1.3.1 Nguyên âm..... | 6 |
| 1.3.2 Phụ âm..... | 7 |
| 1.3.3 Tỷ suất thời gian..... | 8 |
| 1.3.4 Hàm năng lượng thời gian ngắn..... | 8 |
| 1.3.5 Tần số vượt qua điểm không..... | 9 |
| 1.3.6 Phát hiện điểm cuối..... | 9 |
| 1.3.7 Tần số cơ bản..... | 10 |
| 1.3.8 Formant..... | 12 |
| 1.4 Phân loại đơn giản sóng tiếng nói..... | 13 |
| Phần 2 Bộ máy phát âm và cơ chế phát âm..... | 13 |
| 2.1 Bộ máy phát âm..... | 14 |
| 2.2 Cơ chế phát âm..... | 14 |
| 2.3 Mô hình tạo tiếng nói..... | 16 |
| Phần 3 Xác định tần số cơ bản F_0 bằng hàm tự tương quan..... | 16 |
| 3.1 Tính hàm tự tương quan $R(k)$ của tín hiệu tiếng nói $x(n)$ | 16 |
| 3.2 Cải tiến..... | 17 |
| 3.2.1 Hạn chế biên độ tín hiệu vào..... | 17 |
| 3.2.2 Nội suy để tìm chính xác giá trị k cực đại..... | 18 |

Lời nói đầu

Trong thời đại ngày nay, cùng với sự phát triển mạnh mẽ của khoa học kỹ thuật thì vấn đề trao đổi thông tin đa phương tiện ngày càng trở nên cần thiết, từ lúc đầu chỉ giao tiếp, tương tác thông qua các văn bản giấy tờ, ngày nay nhu cầu sử dụng tiếng nói trong truyền thông, tương tác người máy càng trở nên cấp thiết hơn. Vì vậy mà một lĩnh vực kỹ thuật mới đã ra đời, đó là xử lý tiếng nói.

Mặc dù mới nhưng xử lý tiếng nói đã đạt được những thành tựu đáng kể. Các ứng dụng của xử lý tiếng nói đã và đang được áp dụng trong rất nhiều lĩnh vực khác nhau trong xã hội như nhận dạng, tổng hợp tiếng nói, tương tác người máy, truyền thông, dạy học, Nhờ có xử lý tiếng nói mà con người có thể tạo ra những máy móc thông minh hơn, có khả năng hiểu được tiếng nói con người và có thể giao tiếp với con người thông qua lời nói.

Xác định tần số cơ bản là một trong những vấn đề rất quan trọng của xử lý tiếng nói. Nó được sử dụng trong các hệ thống nhận dạng, tổng hợp, thẩm định ghi âm hay phát âm tiếng nói. Do sự quan trọng của nó, có nhiều giải pháp được đưa ra. Bài báo cáo này sẽ trình bày một phương pháp đơn giản và dễ áp dụng là dựa vào hàm tự tương quan.

Nội dung của báo cáo này gồm có 3 phần.

- Phần 1: Giới thiệu về tiếng nói, các đặc tính vật lý, âm học của âm thanh, phân loại đơn giản dạng sóng tiếng nói.

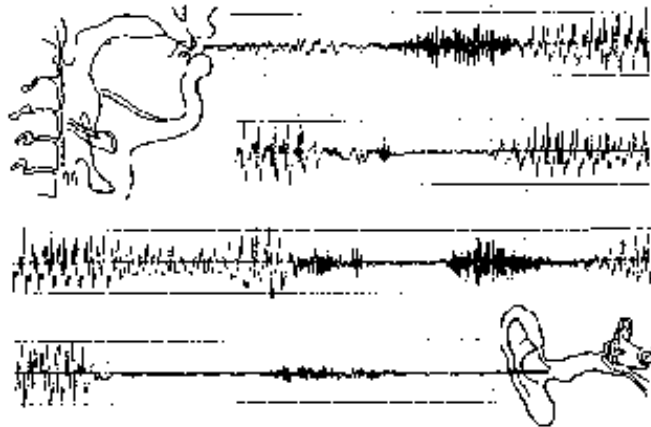
- Phần 2: Giới thiệu bộ máy phát âm và cơ chế phát âm của tiếng nói.
- Phần 3: Giới thiệu phương pháp xác định tần số cơ bản F_0 bằng hàm tự tương quan.

Phần 1 Tổng quan về tiếng nói

1.1 Nhận thức chung

Tiếng nói là một trong những phương tiện trao đổi thông tin của con người. Tiếng nói được tạo ra từ tư duy của con người dưới sự chỉ đạo của trung khu thần kinh, mệnh lệnh phát sinh tiếng nói được phát ra và hệ thống phát âm thực hiện nhiệm vụ tạo âm thanh.

Tiếng nói mà con người vẫn giao tiếp hàng ngày có bản chất là sóng âm thanh lan truyền trong không khí. Sóng âm thanh trong không khí là sóng dọc sinh ra do sự dẫn nở của không khí. Tín hiệu âm thanh là tín hiệu biến thiên liên tục về thời gian và biên độ, có dải tần số rất rộng. Tuy nhiên tai người chỉ có thể nhận biết được các sóng âm có tần số trong khoảng 20 - 20000 (Hz). Những sóng âm có tần số lớn hơn 20000 Hz gọi là sóng siêu âm. Những sóng có tần số nhỏ hơn 20 Hz gọi là sóng hạ âm. Thực tế người ta có thể hạn chế dải tần số của tín hiệu tiếng nói trong khoảng từ 300 đến 3500 Hz. Do đó, quá trình phân tích cũng như tổng hợp tiếng nói chỉ cần dùng một số nhất định các tham số cũng đủ để biểu diễn tín hiệu tiếng nói mà não người xử lý.



Hình 1.1 Mô phỏng quá trình truyền tiếng nói trong không khí

Về bản chất vật lý, sóng âm, sóng siêu âm, sóng hạ âm không khác gì nhau và cũng như các sóng cơ học khác. Sự phân biệt như trên là dựa vào khả năng cảm thụ các sóng cơ học của tai con người, do các đặc tính sinh lý của tai người quyết định. Vì vậy sóng âm thanh được phân biệt hai loại đặc tính là đặc tính vật lý và đặc tính âm học.

1.2 Đặc tính vật lý của âm thanh

Bản chất âm thanh tiếng nói là sóng cơ học nên có các tính chất cơ bản của sóng cơ học. Các tính chất của sóng cơ học mang một ý nghĩa khác khi xét trên góc độ là âm thanh tiếng nói.

Tín hiệu âm thanh tiếng nói là một tín hiệu ngẫu nhiên không dừng, tuy nhiên những đặc tính của nó tương đối ổn định trong những khoảng thời gian ngắn (vài chục mili giây). Trong khoảng thời gian nhỏ đó tín hiệu gần tuần hoàn, có thể coi như tuần hoàn.

1.2.1 Độ cao (Pitch)

Độ cao hay còn gọi là độ trầm bổng của âm thanh chính là tần số của sóng cơ học. Âm thanh nào phát ra cũng ở một độ cao nhất định. Độ trầm bổng của âm thanh phụ thuộc vào sự chấn động nhanh hay chậm của các phần tử trong không khí trong một đơn vị thời gian nhất định. Nói cách khác, độ cao của âm phụ thuộc vào tần số dao động. Đối với tiếng nói, tần số dao động của dây thanh quy định độ cao giọng nói của con người và mỗi người có một độ cao giọng nói khác nhau. Độ cao của nữ giới thường cao hơn so với nam giới và độ cao tiếng nói của trẻ em cao hơn so với nữ giới, điều này cũng tương tự đối với tần số của dây thanh.

1.2.2 Cường độ

Cường độ chính là độ to nhỏ của âm thanh, cường độ càng lớn thì âm thanh có thể truyền đi càng xa trong môi trường có nhiều. Nếu xét trên góc độ sóng cơ học thì cường độ chính là biên độ của dao động sóng âm, nó quyết định cho năng lượng của sóng âm.

Trong tiếng nói, cường độ của nguyên âm phát ra thường lớn hơn phụ âm. Do vậy chúng ta thường dễ phát hiện ra nguyên âm hơn so với phụ âm. Tuy nhiên đối với tai người giá trị tuyệt đối của cường độ âm I không quan trọng bằng giá trị tỉ đối của I so với một giá trị I_0 nào đó chọn làm chuẩn. Người ta định nghĩa mức cường độ âm L là logarit thập phân của tỉ số I/I_0 : (đơn vị mức cường độ là Ben - kí hiệu B)

$$L(B) = \lg \frac{I}{I_0}$$

1.2.3 Trường độ

Trường độ hay độ dài của âm phụ thuộc vào sự chấn động lâu hay nhanh của các phần tử không khí. Cùng một âm nhưng trong các từ khác nhau thì độ dài khác nhau.

1.2.4 Âm sắc

Âm sắc là bản sắc, sắc thái riêng của một âm, cùng một nội dung, cùng một độ cao nhưng khi nói mỗi người đều có âm sắc khác nhau.

1.3 Đặc tính âm học của âm thanh

Tín hiệu tiếng nói là tín hiệu tương tự biểu diễn cho thông tin về mặt ngôn ngữ và được mô tả bởi các âm vị khác nhau. Tùy theo từng ngôn ngữ cụ thể mà số lượng các âm vị nhiều hay ít. Thông thường số lượng các âm vị vào khoảng 20 – 30 và nhỏ hơn 50 đối với mọi ngôn ngữ. Đối với từng loại âm vị mà có các đặc tính âm thanh khác nhau. Các âm vị được chia thành hai loại nguyên âm và phụ âm. Tổ hợp các âm vị tạo nên âm tiết. Âm tiết đóng vai trò một từ trọn vẹn mang ngữ nghĩa.

1.3.1 Nguyên âm

Nguyên âm được tạo ra bằng sự cộng hưởng của dây thanh khi dòng khí được thanh môn đẩy lên. Khoang miệng được tạo lập thành nhiều hình dạng nhất định tạo thành các nguyên âm khác nhau. Số lượng các nguyên âm phụ thuộc vào từng ngôn ngữ nhất định. Mỗi nguyên âm được đặc trưng bởi 3 formant đầu tiên, các formant tiếp theo thường thì ít mang thông tin hơn.

Tiếng Việt có 14 nguyên âm trong đó có 11 nguyên âm đơn và 3 nguyên âm đôi.

| | | |
|----|-----|--------|
| 1 | i,y | ý chí |
| 2 | ê | ê chế |
| 3 | e | e dè |
| 4 | a | a ha |
| 5 | ă | mắt |
| 6 | ơ | bơ phờ |
| 7 | â | ân cần |
| 8 | ư | từ từ |
| 9 | ô | ô tô |
| 10 | o | co ro |
| 11 | u | lủ mù |

| | | |
|---|-----------------------------|---|
| 1 | ia,yê,ya,iê (đọc ia, yê) | kia kia, yêu kiều, khuya, tiên tiến |
| 2 | ua,uô (đọc ua) | tua rua, luôn |
| 3 | ưa,ươ (đọc ưa) | lừa thừa, lướt |

Trong 1 số tài liệu có tách nguyên âm ‘a’ và nguyên âm ‘o’ thành 2 nguyên âm nữa là:

a: a /ɛ/ : xanh xanh, anh ách

a /a/ : la đà, a ha

o: o /ɔ/ : co ro, xoong

o /ə/ : vòng lọng, tóc, học

Theo các tài liệu này thì tiếng việt có 16 âm tiết thay vì 14 âm tiết

1.3.2 Phụ âm

Phụ âm được tạo ra bởi các dòng khí hỗn loạn được phát ra gần những điểm co thắt của đường dẫn âm thanh do cách phát âm tạo thành. Dòng không khí tại chỗ đóng của vòm miệng tạo ra phụ âm tắc. Những phụ âm sát được phát ra từ chỗ co thắt lớn nhất và các âm tắc sát tạo ra từ khoảng giữa. Phụ âm có đặc tính hữu thanh và vô thanh tùy thuộc việc dây thanh có dao động để tạo thành cộng hưởng không. Đặc tính của phụ âm tùy thuộc vào tính chu kỳ của dạng sóng, phổ tần số, thời gian tồn tại và sự truyền dẫn âm.

Tiếng việt có 22 phụ âm

| | | | | | |
|----|------|-------------------|----|--------|--------------------|
| 1 | b | <i>bồng bênh</i> | 12 | tr | <i>trồng</i> |
| 2 | p | <i>ốp ép</i> | 13 | s | <i>sinh viên</i> |
| 3 | v | <i>vẫn vợ</i> | 14 | r | <i>rừng</i> |
| 4 | ph | <i>phôi pha</i> | 15 | ch | <i>chông</i> |
| 5 | m | <i>mơ màng</i> | 16 | nh | <i>nhọc</i> |
| 6 | đ | <i>đất đai</i> | 17 | ng,ngh | <i>ngô nghề</i> |
| 7 | t | <i>tin tưởng</i> | 18 | c,k,q | <i>con,kẹt,qua</i> |
| 8 | th | <i>thơ thần</i> | 19 | kh | <i>khúc</i> |
| 9 | d,gi | <i>duyên, giữ</i> | 20 | g,gh | <i>gỗ ghề</i> |
| 10 | n | <i>nóng</i> | 21 | h | <i>hả hê</i> |
| 11 | l | <i>long lanh</i> | 22 | x | <i>xa xôi</i> |

Ngoài ra, trong 1 số tài liệu, tiếng việt còn có thêm 1 phụ âm nữa, là phụ âm tắc thanh hầu. Những âm tiết không có âm đầu (như: **âm**, **êm**, **oai**, **uỳnh**) khi phát âm được bắt đầu bằng động tác khép kín khe thanh, sau đó mở ra đột ngột gây nên một tiếng bật. Động tác khép kín ấy có giá trị như một phụ âm nên người ta gọi là âm tắc thanh hầu, kí hiệu: */ʔ/*.

1.3.3 Tỷ suất thời gian

Trong khi nói chuyện, khoảng nói chuyện và khoảng nghỉ xen kẽ nhau. Phần trăm thời gian nói trên tổng số thời gian nói và nghỉ được gọi là tỷ suất thời gian. Giá trị này biến đổi tùy thuộc vào tốc độ nói và từ đó ta có thể phân loại thành nói nhanh, nói chậm hay nói bình thường.

1.3.4 Hàm năng lượng thời gian ngắn

Hàm năng lượng thời gian ngắn của tiếng nói được tính bằng cách chia tín hiệu tiếng nói thành nhiều khung chứa N mẫu và tính diện tích trung bình tổng các

mẫu tín hiệu trong mỗi khung. Các khung này được đưa qua một cửa sổ có dạng hàm như sau:

$$W(n) = \begin{cases} W(n) & \text{Với } 0 \leq n \leq N \\ 0 & \text{Với } n \geq N \end{cases}$$

Thông thường có ba dạng cửa sổ được sử dụng đó là cửa sổ chữ nhật, cửa sổ Hamming và cửa sổ Hanning.

❖ Cửa sổ chữ nhật:

$$W(n) = \begin{cases} 1 & \text{Với } 0 \leq n \leq N \\ 0 & \text{Với } n \geq N \end{cases}$$

❖ Cửa sổ Hamming :

$$W(n) = \begin{cases} 0.54 - 0.46 \cos(n) & \text{Với } 1 \leq n \leq N \\ 0 & \text{Với } n \geq N \end{cases}$$

❖ Cửa sổ Hanning:

$$W(n) = \begin{cases} 0.5 - 0.5 \cos(n) & \text{Với } 1 \leq n \leq N \\ 0 & \text{Với } n \geq N \end{cases}$$

Hàm năng lượng ngắn tại mẫu thứ m được tính theo công thức sau :

$$E_m = \sum_{n=0}^{N-1} \{x(n+m) * W(n)\}^2$$

Hàm năng lượng thời gian ngắn của âm hữu thanh thường lớn hơn so với âm vô thanh.

1.3.5 Tần số vượt qua điểm không.

Tần suất vượt qua điểm không là số lần biên độ tín hiệu tiếng nói vượt qua giá trị không trong một khoảng thời gian cho trước. Thông thường giá trị này đối với âm vô thanh lớn hơn âm hữu thanh do đặc tính ngẫu nhiên của âm vô thanh.

Do đo tần suất vượt qua điểm không là tham số quan trọng để phân loại âm hữu thanh và âm vô thanh.

1.3.6 Phát hiện điểm cuối.

Trong xử lý tiếng nói việc xác định khi nào bắt đầu xuất hiện tín hiệu tiếng nói và khi nào kết thúc quá trình nói rất cần thiết và quan trọng. Trong một môi trường nhiều tiếng ồn (nhiều) hoặc môi trường nhiều người nói thì việc phát hiện điểm kết thúc rất khó khăn. Có rất nhiều phương pháp để phát hiện điểm cuối của tiếng nói. Ví dụ như một phương pháp đơn giản sau :

Lấy một mẫu nhỏ của nền nhiễu trong khoảng thời gian yên lặng trước khi bắt đầu nói. Sử dụng hàm năng lượng thời gian ngắn để tính năng lượng cho mẫu. Ngưỡng của tiếng nói được chọn là giá trị giữa năng lượng yên tĩnh và năng lượng đỉnh. Ban đầu giả thiết điểm cuối xuất hiện tại điểm năng lượng tín hiệu vượt quá mức ngưỡng. Để tính đúng ước lượng này, người ta giả thiết và so sánh chúng với giá trị đó trong vùng yên tĩnh. Khi những thay đổi phát hiện được trong khi tính toán tần suất trên suất hiện ở ngoài ngưỡng giả thiết thì điểm cuối được giả thiết lại tại điểm mà sự thay đổi xảy ra.

1.3.7 Tần số cơ bản

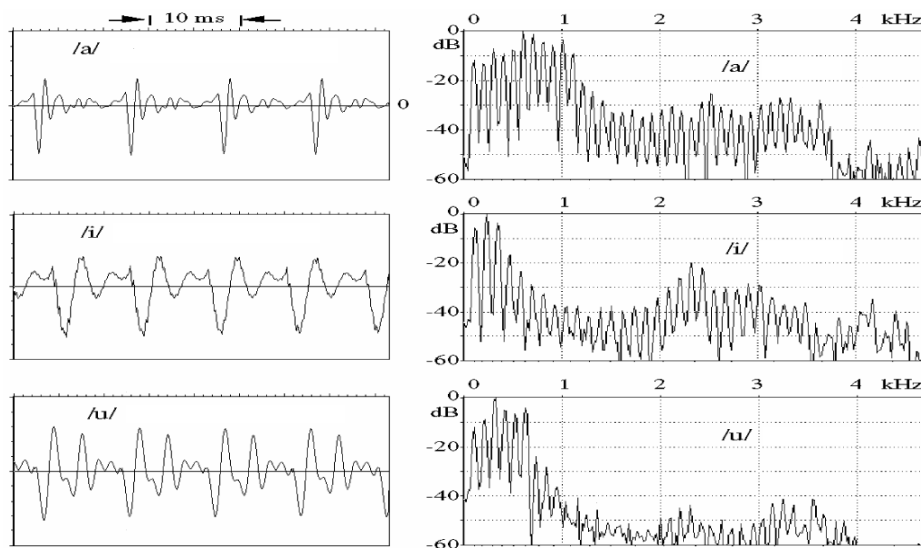
Dạng sóng của tiếng nói gồm hai phần: Phần gần giống nhiễu trong đó biên độ biến đổi ngẫu nhiên và phần tuần hoàn. Phần tín hiệu có tính chu kỳ chứa các thành phần tần số có dạng điều hòa. Tần số thấp nhất chính là tần số cơ bản và cũng chính là tần số dao động của dây dây thanh.

Đối với những người nói khác nhau, tần số cơ bản cũng khác nhau. Tần số cơ bản của trẻ em thường cao hơn so với người lớn và của nữ giới cao hơn so với nam giới, Sau đây là một số giá trị tần số cơ bản tương ứng với giới tính và tuổi:

| Người nói | Giá trị tần số cơ bản |
|-----------|-----------------------|
| Nam giới | 80 – 200 Hz |
| Nữ giới | 150 – 450 Hz |
| Trẻ em | 200 – 600 Hz |

Đối với hai âm có cùng cường độ, cùng độ cao sẽ được phân biệt bởi tính tuần hoàn. Một âm hữu thanh có tín hiệu gần như tuần hoàn khi được phân tích phổ sẽ xuất hiện một vạch tại vùng tần số rất thấp. Vạch này đặc trưng cho tính tuần hoàn cơ bản của âm hay đó chính là tần số cơ bản của âm. Trong giao tiếp bình thường tần số cơ bản thay đổi liên tục tạo nên ngữ điệu cho tiếng nói.

Hình dưới mô tả tín hiệu trên miền thời gian và phổ của chúng (trên miền tần số) của các nguyên âm a, i, u. đỉnh đầu tiên của các phổ tương ứng với tần số cơ bản F_0 .



Hình 1.2 Tín hiệu và phổ của tín hiệu

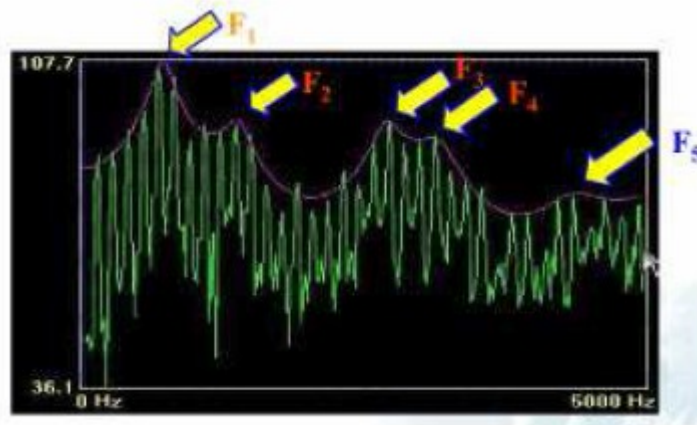
1.3.8 Formant

Trong phổ tần số của tín hiệu tiếng nói, mỗi đỉnh có biên độ cao nhất xét trong một khoảng nào đó (còn gọi là cực trị địa phương) xác định một formant. Ngoài tần số, formant còn được xác định bởi biên độ và dải thông của chúng. Về mặt vật lý các tần số formant tương ứng với các tần số cộng hưởng của tuyến âm. Trong xử lý tiếng nói và nhất là trong tổng hợp tiếng nói để mô phỏng lại tuyến âm người ta phải xác định được các tham số formant đối với từng loại âm vị, do đó việc đánh giá, ước lượng các formant có ý nghĩa rất quan trọng.

Tần số formant biến đổi trong một khoảng rộng phụ thuộc vào giới tính của người nói và phụ thuộc vào các dạng âm vị tương ứng với formant đó. Đồng thời, formant còn phụ thuộc các âm vị trước và sau đó. Về cấu trúc tự nhiên, tần số formant có liên hệ chặt chẽ với hình dạng và kích thước tuyến âm. Thông thường trong phổ tần số của tín hiệu có khoảng 6 formant nhưng chỉ có 3 formant đầu tiên ảnh hưởng quan trọng đến các đặc tính của các âm vị, còn các formant còn lại cũng có ảnh hưởng song rất ít. Các formant có giá trị tần số xê dịch từ vài trăm đến vài nghìn Hz.

Tần số formant đặc trưng cho các nguyên âm biến đổi tùy thuộc vào người nói trong điều kiện phát âm nhất định. Mặc dù phạm vi của các tần số formant tương ứng với mỗi nguyên âm có thể trùng lên nhau nhưng vị trí giữa các formant đó không đổi vì sự xê dịch của các formant là song song.

Ngoài formant, các âm mũi còn có các tần số bị suy giảm gọi là phản formant (anti-formant). Phản formant được tạo nên khi luồng khí đi qua khoang mũi. Các formant tương ứng với các điểm cực của hàm truyền đạt vì tại lân cận điểm cực giá trị hàm truyền đạt là rất lớn, tương tự vậy các anti-formant tương ứng với các điểm không của hàm truyền đạt.



Hình 1.3 Đường bao phổ và các Formant

Dựa trên hình 1.3 ta thấy có thể tính đến Formant thứ 5 (F_5) nhưng quan trọng nhất cần chú ý ở đây là các F_1 và F_2 . Cùng một người phát âm nhưng Formant có thể khác nhau. Nếu ta 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 là đáp ứng tần số của tuyến âm.

1.4 Phân loại đơn giản dạng sóng tiếng nói:

Tiếng nói của con người tạo ra bao gồm có hai thành phần đó là:

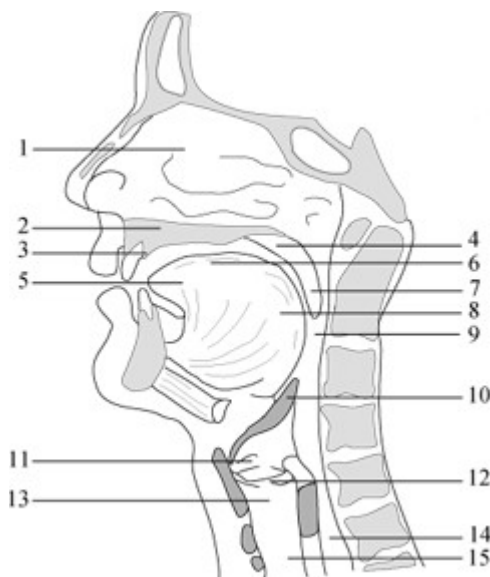
- ❖ Phần gần tuần hoàn mà hầu như lặp lại cùng chu kỳ được gọi là tiếng nói hữu thanh (voiced speech). Chu kỳ lặp lại đó gọi là chu kỳ cơ bản T_0 nghịch đảo của T_0 là tần số cơ bản F_0 . Âm hữu thanh được phát ra bởi một luồng khí cực mạnh từ thanh môn thổi qua dây thanh làm rung dây thanh, sự dao động của dây thanh tạo nên nguồn tuần hoàn. Nguồn tuần hoàn kích thích tuyến âm tạo nên âm hữu thanh. Vùng âm hữu thanh chiếm thành phần chủ yếu của sóng tiếng nói, chứa đựng lượng tin nhiều nhất và thời gian lớn nhất trong quá trình nói.
- ❖ Phần tín hiệu có dạng giống như tập âm nhiễu có biên độ ngẫu nhiên còn được gọi là tiếng nói vô thanh (unvoiced speech). Tiếng nói vô thanh

được tạo ra do sự co thắt theo một dạng nào đó của tuyến âm và luồng khí chạy qua chỗ thắt với tốc độ lớn tạo nên nhiễu loạn, ví dụ như lúc ta nói thì thảo (cần phân biệt thì thầm với thì thảo, theo từ điển tiếng Việt thì thảo là nói chuyện với nhau rất nhỏ tựa như gió thoảng qua tai còn thì thầm là nói chuyện với nhau không để người ngoài nghe thấy). Năng lượng do nguồn nhiễu loạn tạo ra sẽ kích thích tuyến âm tạo nên tiếng nói vô thanh, năng lượng của tiếng nói vô thanh nhỏ hơn so với tiếng nói hữu thanh.

Ta có thể phát hiện ra tiếng nói hữu thanh là khi nói dây thanh rung. Còn âm vô thanh khi nói dây thanh không rung. Nói thì thảo thì ở xa không nghe được do năng lượng của âm vô thanh rất nhỏ và tiếng thì thảo là do âm vô thanh tạo nên.

Phần 2 Bộ máy phát âm và cơ chế phát âm

2.1 Bộ máy phát âm



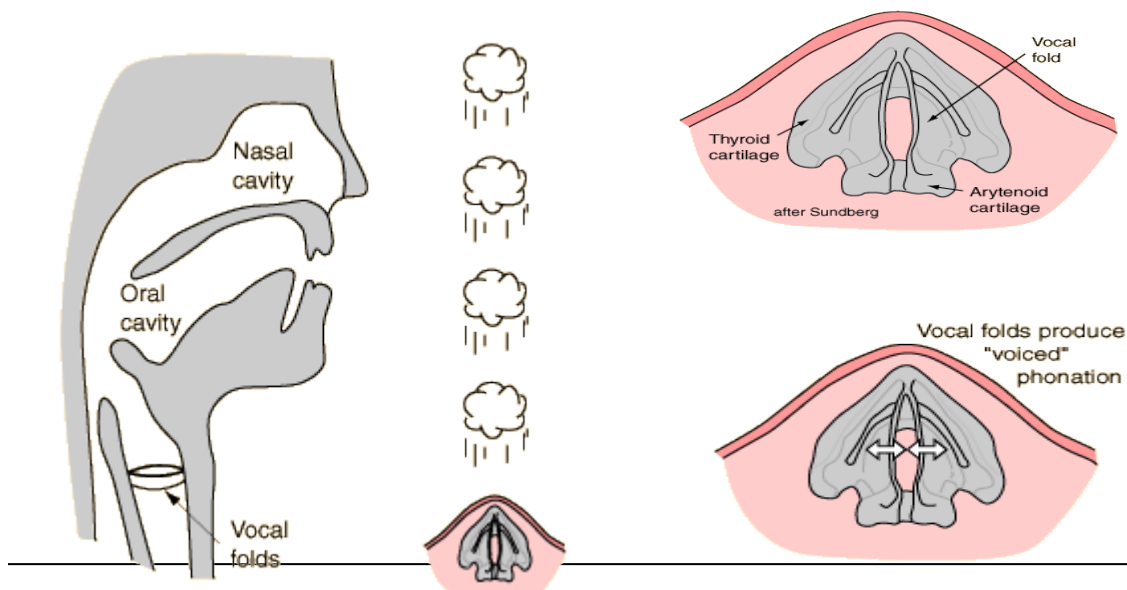
1. Hốc mũi
2. Vòm miệng trên
3. Ổ răng
4. Vòm miệng mềm
5. Đầu lưỡi
6. Thân lưỡi
7. Lưỡi gà
8. Cơ miệng
9. Yết hầu
10. nắp đóng của thanh quản
11. Dây thanh giả
12. Dây thanh
13. Thanh quản
14. Thực quản
15. Khí quản

Hình 3.1 Bộ máy phát âm của con người

2.2 Cơ chế phát âm

Hình 3.1 mô tả bộ máy phát âm của con người. Nguồn năng lượng chính nằm ở thanh môn, Tuyến âm sẽ được kích thích bởi nguồn năng lượng chính tại thanh môn. Tiếng nói được tạo ra sóng âm học do kích thích từ thanh môn phát ra đẩy không khí có trong phổi lên tạo thành dòng khí va chạm vào hai dây thanh trong tuyến âm. Hai dây thanh dao động sẽ tạo ra cộng hưởng, dao động âm sẽ được lan truyền theo tuyến âm (tính từ tuyến âm đến khoang miệng) và sau khi đi qua khoang mũi, môi sẽ tạo ra tiếng nói.

- Thanh quản chứa hai dây thanh có thể dao động tạo ra sự cộng hưởng đầu tiên của quá trình tạo thành âm thanh.
- Ống dẫn âm là một ống không đồng dạng bắt đầu từ môi, kết thúc bởi dây thanh hoặc thanh quản. Ống có độ dài khoảng 17cm đối với người bình thường.
- Khoang mũi cũng là ống không đồng dạng thuộc vùng cổ định bắt đầu từ mũi, kết thúc tại vòm miệng, đối với người bình thường khoang mũi có độ dài 12 cm.
- Khoang miệng là các nếp da chuyển động có thể điều khiển sự ghép âm thanh giữa khoang miệng và khoang mũi.



Hình 2.2. Mô tả dây thanh âm

Trong quá trình phát âm, nếu là âm mũi thì vòm miệng hạ thấp và dòng khí chỉ đi qua đường mũi, nếu là âm thường thì vòm miệng mở, đường mũi khép lại và dòng khí đi theo khoang miệng ra môi.

2.3 Mô hình tạo tiếng nói

Tiếng nói là một loại âm thanh, nhưng ngược lại, không phải bất kỳ âm thanh nào cũng là tiếng nói. Tiếng nói được phân biệt với các âm thanh khác bởi các đặc tính âm học có nguồn gốc từ cơ chế tạo tiếng nói.

Trong quá trình tạo tiếng nói có 2 loại nguồn âm:

- *Nguồn tuần hoàn*: là nguồn tương ứng với các âm mà khi phát ra làm cho dây thanh rung. Trong tiếng Việt nguồn tuần hoàn tương ứng với các nguyên âm và các phụ âm hữu thanh
- *Nguồn không tuần hoàn (nguồn tạp âm)*: là nguồn tương ứng với các âm mà khi phát ra không làm cho dây thanh rung. Trong tiếng Việt các nguồn không tuần hoàn tương ứng với các phụ âm vô thanh.

Với nguồn tuần hoàn thì vị trí nguồn tại chính thanh môn, còn nguồn không tuần hoàn thì có vị trí trong tuyến âm tức là từ thanh môn trở lên cho đến dưới môi.

Phần 3 Xác định tần số cơ bản F_0 bằng hàm tự tương quan

Xác định tần số cơ bản là một trong những vấn đề rất quan trọng của xử lý tiếng nói. Nó được sử dụng trong các hệ thống nhận dạng, tổng hợp, thẩm định ghi âm hay phát âm tiếng nói. Do sự quan trọng của nó, có nhiều giải pháp được đưa ra. Phần này sẽ trình bày 1 phương pháp đơn giản và dễ áp dụng là dựa vào hàm tự tương quan.

Hàm tự tương quan $R(k)$ sẽ đạt các giá trị cực khi tương ứng tại các điểm là bội của chu kỳ cơ bản của tín hiệu. Khi đó tần số cơ bản là tần số xuất hiện của các đỉnh của $R(k)$. Bài toán trở thành bài toán xác định chu kỳ hàm tự tương quan.

3.1 Tính hàm tự tương quan $R(k)$ của tín hiệu tiếng nói $x(n)$

$$r(k) = \sum_{n=1}^{N-1-k} x(n)x(n+k) \quad k = 0, 1, \dots, K$$

$F_s = 10\text{kHz}$, $N = 300$, $K = 150$. Tìm cực đại trong khoảng $(0, K)$

Dễ thấy rằng nếu tín hiệu $x(n)$ tuần hoàn với chu kỳ P thì hàm tự tương quan cũng tuần hoàn với chu kỳ P : $r(k) = r(k + P)$

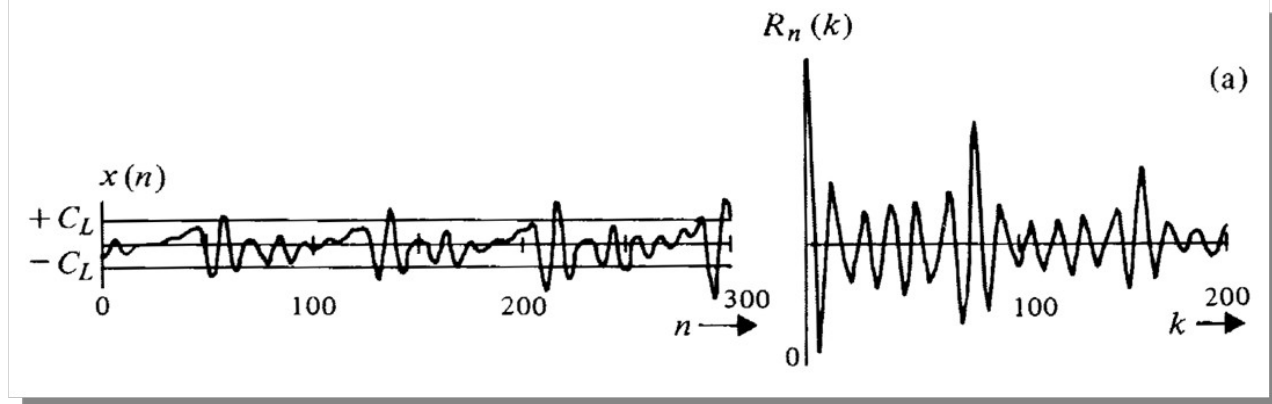
Hơn nữa hàm tự tương quan còn có những tính chất quan trọng sau:

- Là hàm chẵn $r(k) = r(-k)$
- $r(k)$ đạt giá trị cực đại tại 0: $r(k) \leq r(0)$ với mọi k
- Giá trị $r(0)$ chính bằng năng lượng của tín hiệu:

$$r(0) = \sum_{m=-\infty}^{\infty} x^2(m)$$

Dựa vào các tính chất trên ta có nhận xét: Hàm tự tương quan sẽ đạt giá trị cực đại tại các mẫu $0, \pm P, \pm 2P, \dots$ và bằng giá trị năng lượng của tín hiệu, các điểm cực đại được gọi là các đỉnh (peak). Như vậy việc xác định chu kỳ cơ bản của tín hiệu tiếng nói sẽ đưa về việc xác định chu kỳ của hàm tự tương

quan.



Hình 3.1: Tín hiệu theo thời gian và hàm tự tương quan $r(k)$

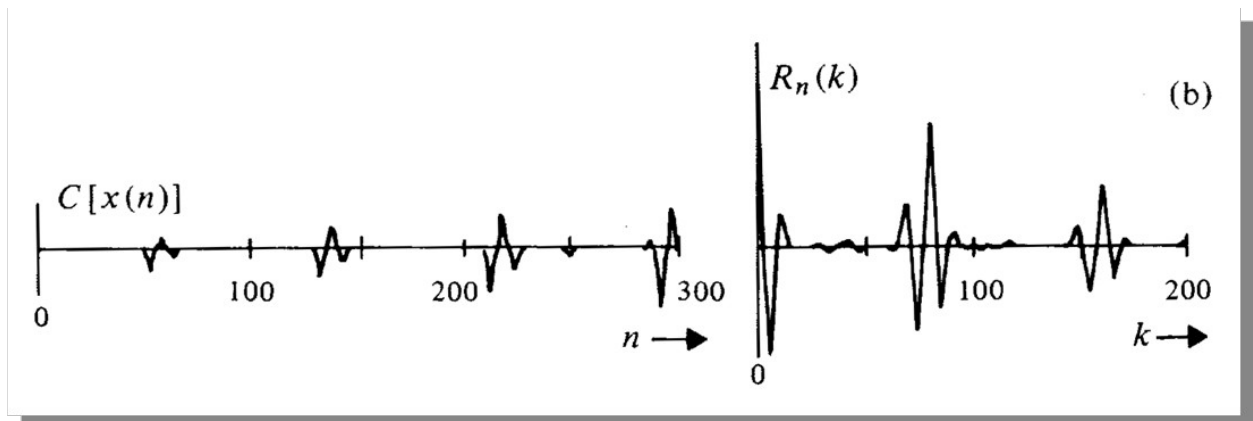
3.2 Cải tiến

3.2.1 Hạn chế biên độ tín hiệu vào

Để tính toán được dễ dàng hơn, ta hạn chế mức biên độ đối với tín hiệu đầu vào.

- Chọn ngưỡng cắt C_L
- Loại bỏ (coi như biên độ bằng 0) đối với những tín hiệu:

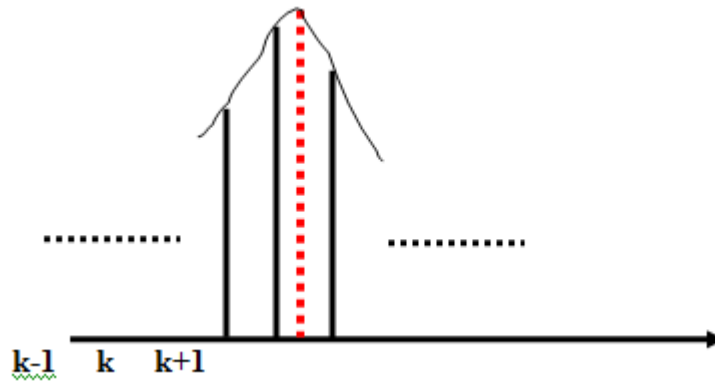
$$|x| < C_L$$



Hình 3.2 : Tín hiệu sau khi đã hạn chế và hàm tự tương quan tương ứng

3.2.2 Nội suy để tìm chính xác giá trị k cực đại

Do giá trị k với giá trị $r(k)$ lớn nhất tìm được ở trên chỉ là rời rạc, ta nội xây dựng hàm nội suy(chương trình sử dụng nội suy Lagrange) để tìm chính xác giá trị k cực đại trong khoảng $k-1 \rightarrow k+1$



Hình 3.3 : Nội suy $k-r(k)$ với 2 điểm lân cận là $k-1$ và $k+1$ để tìm k cực đại