HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
CƠ SỞ TẠI THÀNH PHỐ HỒ CHÍ MINH
KHOA VIỄN THÔNG II

ĐỒ ÁN
TỐT NGHIỆP ĐẠI HỌC

CHUYÊN NGHÀNH: ĐIỆN TỬ - VIỄN THÔNG
HỆ ĐẠI HỌC CHÍNH QUY
NIÊN KHOÁ: 2008-2013

Đề tài:

NGHIÊN CỨU KĨ THUẬT MÃ HOÁ TIẾNG NÓI TRONG DI ĐỘNG

Mã số đề tài: 12 408160072

NỘI DUNG:
- CHƯƠNG 1: GIỚI THIỆU SƠ LƯỢC VỀ XỬ LÝ TÍN HIỆU TRONG DI ĐỘNG
- CHƯƠNG 2: QUÁ TRÌNH TẠO TIẾNG NÓI
- CHƯƠNG 3: CÁC PHƯƠNG PHÁP CƠ SỞ MÃ HOÁ TIẾNG NÓI
- CHƯƠNG 4: MÃ HOÁ VÀ GIẢI MÃ TIẾNG NÓI TRONG HỆ THỐNG GSM
- CHƯƠNG 5: MÔ PHỎNG

Sinh viên thực hiện: **Nguyễn Đại Hoà**
MSSV: **408160072**
Lớp: **Đ08VTA2**
Giáo viên hướng dẫn: **Phạm Thanh Đàm**

MỤC LỤC

MỤC LỤC HÌNH

MỤC LỤC BẢNG

Lời mở đầu

Ngày nay, khi các phương tiện truyền thông phát triển và số lượng người sử dụng các phương tiện liên lạc ngày càng tăng lên thì mã hóa tiếng nói được nghiên cứu và ứng dụng càng rộng rãi trong các cuộc gọi điện thoại truyền thống, gọi điện thoại qua mạng di động, qua Internet hay qua vệ tinh, ... Mặc dù với sự phát triển của công nghệ truyền thông qua cáp quang đã làm cho băng thông không còn là vấn đề lớn trong các cuộc gọi điện truyền thống. Tuy nhiên, băng thông trong các cuộc gọi đường dài, các cuộc gọi quốc tế, các cuộc gọi qua vệ tinh hay các cuộc gọi di động thì cần phải duy trì băng thông ở một mức nhất định. Chính vì thế việc mã hóa tiếng nói là rất cần thiết, giúp giảm thiểu số lượng tín hiệu cần truyền đi trên đường truyền nhưng vẫn đảm bảo chất lượng cuộc gọi.

Xuất phát từ những yêu cầu ở trên, với mục đích tìm hiểu sâu hơn về kĩ thuật mã hoá tiếng nói, em đã quyết định thực hiện đề tài "Nghiên cứu kĩ thuật mã hoá tiếng nói trong di động".

Nội dung đề tài bao gồm 4 chương chính:

- Giới thiệu sơ lược về xử lý tín hiệu trong di động.

- Quá trình tạo tiếng nói.

- Các phương pháp cơ sở mã hoá tiếng nói.

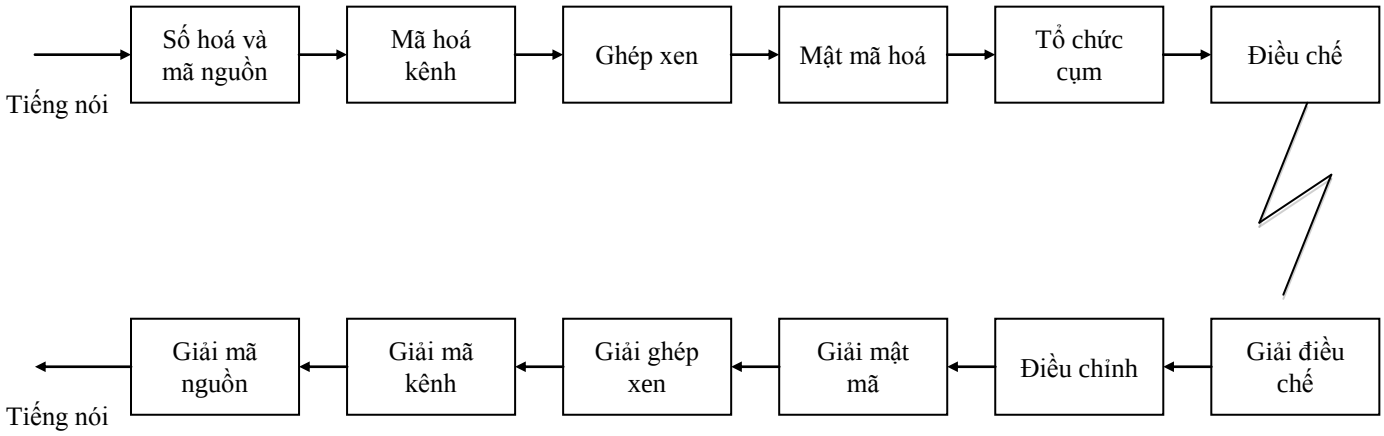
- Mã hoá và giải mã tiếng nói trong hệ thống GSM.

Để tăng tính thực tế của đề tài, em đã thực hiện chương trình mô phỏng mã hoá tiếng nói chạy trên PC bằng Matlab.

Em xin chân thành cảm ơn thầy Phạm Thanh Đàm đã hướng dẫn, tận tình giúp đỡ em hoàn thành đề tài này. Nhưng do thời gian và kiến thức có hạn nên luận văn thực hiện còn nhiều thiếu sót. Em rất mong sự nhận xét, đánh giá, đóng góp từ thầy cô và bạn bè.

CHƯƠNG 1: GIỚI THIỆU SƠ LƯỢC VỀ XỬ LÝ TÍN HIỆU TRONGDI ĐỘNG

Quá trình biến đổi và xử lý tín hiệu GSM được mô tả như sau:

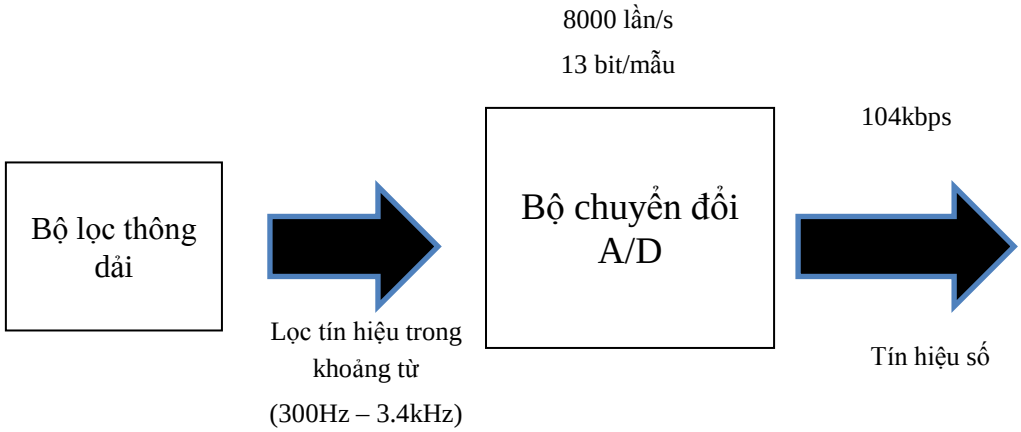


Hình 1.1 Quá trình biến đổi tín hiệu trong GSM

1.1 Số hoá và mã hoá tiếng nói

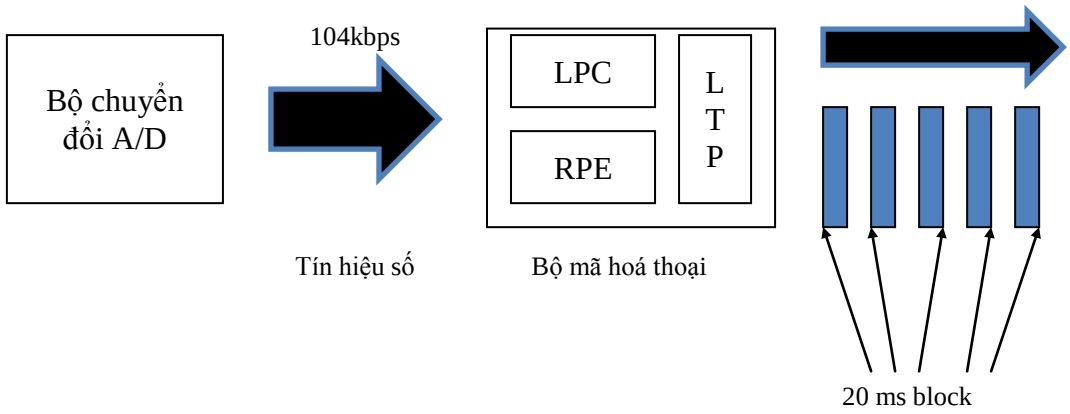
Đầu tiên, tiếng nói được microphone biến đổi sang tín hiệu điện ở dạng tương tự. Microphone bao gồm một màn mỏng và một cuộn dây đặt trong khe từ trường của một nam châm. Để giảm lượng dữ liệu cần thiết tương ứng với sóng âm, ta cho tín hiệu qua bộ lọc thông dải trong khoảng tần số từ 300 Hz đến 3.4 kHz. Sau đó, tín hiệu này được biến đổi sang tín hiệu số bằng bộ biến đổi A/D dùng kĩ thuật điều xung mã PCM với tần số lấy mẫu là 8kHz và mã hoá mỗi mẫu bằng 13 bit. Do đó, luồng tín hiệu số sau khi được biến đổi có tốc độ 104 kbps.

Tín hiệu số ở ngõ ra của bộ biến đổi A/D có tốc độ 104 kbps được nén lại bằng bộ mã hoá tiếng nói. Mã hoá tiếng nói là phương pháp nén tín hiệu thoại ở dạng số. Yêu cầu của mã hoá tiếng nói là phải đảm bảo thời gian thực và chất lượng có thể chấp nhận được. Trong GSM, người ta sử dụng mã Vocoder. Nguyên tắc của kỹ thuật này là thay vì truyền đi luồng số từ tiếng nói thì ta sẽ truyền đi thông số của cơ quan phát âm tại thời điểm phát ra tiếng đó. Như vậy, chuỗi bit truyền đi sẽ ngắn hơn nên tốc độ sẽ giảm xuống.



Hình 1.2 Biến đổi A/D

Tín hiệu số ở ngõ ra của bộ biến đổi A/D có tốc độ 104 kbps được chia thành từng đoạn có chiều dài 20 ms, như vậy mỗi đoạn chứa 2080 bit (tương ứng 160 mẫu). Để truyền đi chuỗi bit này, người ta sẽ thay thế thông số của bộ lọc có chiều dài 260 bit. Như vậy, 260 bit mỗi 20ms tương ứng với tốc độ truyền thật sự là 13 kbps.



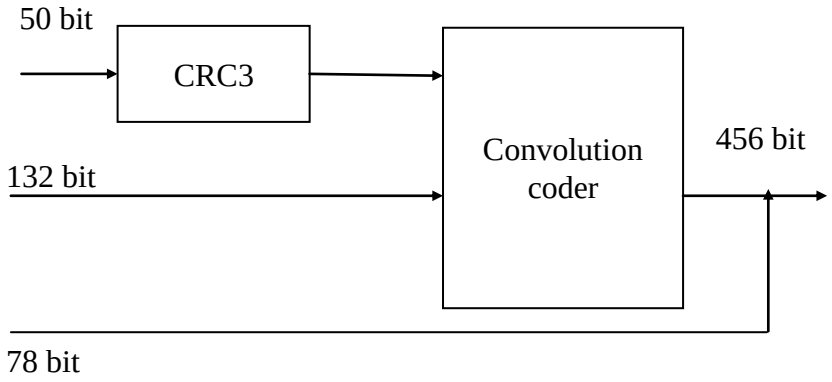
Hình 1.3 Mã hoá thoại

1.2 Mã hoá kênh

Mã kênh là thêm vào mỗi từ mã cần truyền một số bit dư thừa để làm tăng khoảng cách Hamming của bộ từ mã, nhằm mục đích là giúp cho đầu thu phát hiện và sửa được nhiều lỗi hơn.

Bộ mã hoá tiếng nói đưa các khối 260 bit/20ms đến bộ mã hoá kênh. Các bit này được chia thành 182 bit loại I (các bit được bảo vệ) và 78 bit loại II (các bit không được bảo vệ), dựa theo tầm quan trọng của các bit nhận được từ các thí nghiệm chủ quan. Các bit loại I được chia thành 2 loại, Ia và Ib.

50 bit đầu của loại I được bảo vệ bởi mã CRC để phát hiện lỗi và tạo thành 53 bit. Các bit thêm vào này được tính dựa trên đa thức tạo mã $g(x)= 1+x+x^3$. Sau đó, các bit loại I cùng với các bit chẵn lẻ (185 bit) được bổ sung thêm 4 bit đuôi bằng 0 và được mã hoá xoắn theo hai đa thức: $g_1(x)=1+x^3+x^4$ và$g_2(x)=1+x+x^3+x^4$ tạo thành 378 bit. Các bit nhóm II không được bảo vệ. Như vậy, đầu ra của mã hoá kênh sẽ là 456 bit tương ứng với 22,8 kbps.



Hình 1.4 Mã hoá kênh

1.3 Tổ chức cụm

Khi MS cần truy xuất vào mạng thì sẽ được hệ thống cung cấp cho một khe thời gian. Mỗi khe thời gian có độ dài 0,577 ms nhưng thông tin truyền đi trong khe này là chỉ chiếm có 0,546 ms. Thông tin trong khoảng thời gian này được gọi là cụm và khoảng thời gian còn lại hai đầu là thời gian bảo vệ dài 0,031 ms.

Tuỳ theo mỗi loại tín hiệu khác nhau mà các tổ chức cụm trong GSM khác nhau. Có 5 loại cụm trong thông tin di động GSM:

- *Cụm thường (Normal Burst)*

TB	57 bit thông tin	F	Chuỗi hướng dẫn	F	57 bit thông tin	TB	GP
3		1	26 bit	1		3	8.25

Cụm thường (NB)

TB: Tail bit (3 bit), là các bit đuôi, đặt ở đầu và cuối cụm.
Chuỗi hướng dẫn: 26 bit, dùng để xác định khe thời gian và giúp máy thu điều chỉnh tín hiệu thu.
Mỗi cụm thường chứa 114 bit thông tin và được chia thành hai gói, mỗi gói 57 bit, xen giữa hai gói là một chuỗi hướng dẫn chiều dài 26 bit. Ở hai đầu cụm sử dụng bit đuôi cho mỗi đầu.

- *Cụm điều chỉnh tần số (Frequency Correction Burst)*
  Cụm này chứa 142 bit cố định làm tín hiệu điều khiển, các bit khởi tạo và kết thúc cụm là 3 bit, được sử dụng cho kênh FCCH.

TB	142 bit thông tin	TB	GB
3		3	8.25

Cụm điều chỉnh tần số (FC)

- Cụm đồng bộ (Synchronization Burst)
  Được sử dụng để đồng bộ thời gian cho trạm di động. Cụm chứa 78 bit được mật mã hoá mang thông tin về FN (số khung) của TDMA và của BSIC (mã nhận dạng trạm gốc). Cụm SB được sử dụng để truyền kênh SCH.

TB	39 bit thông tin	Chuỗi đồng bộ	39 bit thông tin	TB	GB
3		64 bit		3	8.25

Cụm đồng bộ (SB)

- Cụm truy xuất (Access Burst)
  Được sử dụng cho các kênh điều khiển 1 chiều còn lại.

TB	Chuỗi đồng bộ	Các bit thông tin	TB	GP
3	41	36	3	68.25

Cụm truy xuất (AB)

- Cụm giả (Dummy Burst)
  Cụm DB có tổ chức giống như cụm NB nhưng thông tin trong cụm DB là thông tin giả, sử dụng các bit hỗn hợp. Được sử dụng trong các khe thời gian rỗi.

TB	Các bit hỗn hợp	Chuỗi hướng dẫn	Các bit hỗn hợp	TB	GP
3	58	26 bit	58	3	8.25

Cụm giả (DB)

1.4 Ghép xen

Ở thông tin di động, do tác động của fading nên các lỗi bit thường xảy ra từng cụm dài. Tuy nhiên, mã hoá kênh đặt biệt là mã hoá xoắn chỉ hiệu quả nhất khi phát hiện và sửa chữa các lỗi ngẫu nhiên đơn lẻ và cụm lỗi không quá dài. Để đối phó với vấn đề này người ta chia khối bản tin cần gởi thành các cụm ngắn rồi hoán vị các cụm

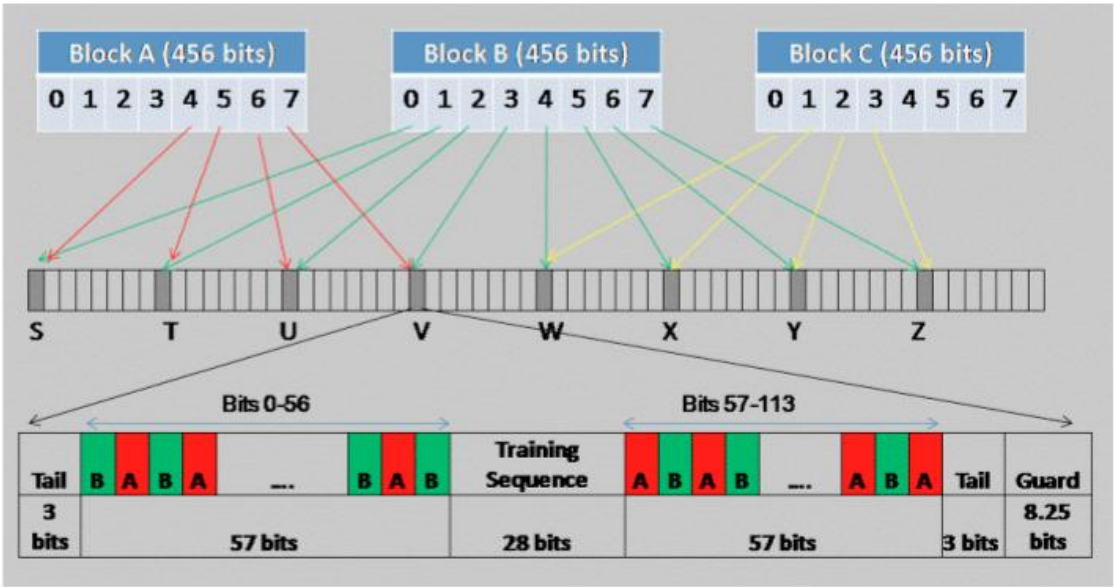
này với các cụm của khối bản tin khác. Do đó, khi xảy ra cụm lỗi dài mỗi bản tin chỉ mất đi một cụm nhỏ, phần còn lại của bản tin vẫn cho phép các dạng mã hoá kênh khôi phục lại được đúng sau khi đã sắp xếp lại các cụm của bản tin theo thứ tự như ở phía phát. Quá trình nói trên được gọi là ghép xen.

Các bit sau khi mã hoá có chiều dài 456 bit được tổ chức lại và được ghép xen theo 8 nửa cụm. Mỗi nửa cụm chứa 57 bit. Việc ghép xen lưu lượng được thực hiện theo các bước sau:

B1: Chia 456 bit thành 8 nhóm

- Nhóm 0: 1, 9 , 17 …….. 449
- Nhóm 1: 2, 10, 18 …….. 450
- Nhóm 2: 3, 11, 19 …….. 451
- Nhóm 3: 4, 12, 20 …….. 452
- Nhóm 4: 5, 13, 21 …….. 453
- Nhóm 5: 6, 14, 22 …….. 454
- Nhóm 6: 7, 15, 23 …….. 455
- Nhóm 7: 8, 16, 24 …….. 456

B2: Sau đó, các nhóm nói trên được ghép xen ở mức thứ 2. Ở ghép xen này ta thấy bốn nhóm đầu của một từ mã (cụ thể là nhóm 0, 1, 2, 3) được đặt vào vị trí đầu tiên của bốn cụm, bốn nhóm còn lại được đặt vào vị trí sau của bốn cụm tiếp theo. Phần còn lại của các cụm này được dùng để ghép tín hiệu của các từ mã lân cận. Như vậy, để truyền đi hết một từ mã 456 bit thì phải cần 8 cụm liên tiếp.



Hình 1.5 Ghép xen tín hiệu tiếng nói

1.5 Mật mã hoá

Mục đích của mật mã hoá là bảo mật tín hiệu trên đường truyền vô tuyến. Khi MS và BTS giao tiếp với nhau thì giữa chúng có chung một mật mã. Mỗi cuộc gọi khác nhau thì có mật mã khác nhau.

Trong GSM, để thực hiện mật mã, ở đầu phát tạo ra một chuỗi tín hiệu giả ngẫu nhiên để kết hợp với chuỗi tín hiệu cần truyền. Ở đầu thu muốn khôi phục lại tín hiệu thì máy thu phải biết chuỗi ngẫu nhiên ở đầu thu, do vậy chuỗi ngẫu nhiên được gọi là mật mã.

Mật mã hoá tín hiệu đạt được bằng cổng XOR giữa chuỗi ngẫu nhiên với 114 bit của cụm bình thường. Để giải mật mã, người ta thực hiện thao tác XOR giữa tín hiệu thu với chuỗi ngẫu nhiên giống đầu phát.

1.6 Điều chế

Điều chế là phép toán chuyển đổi từ một tín hiệu mang tin tức sang một tín hiệu khác mà không làm thay đổi về tin tức mang theo.

Điều chế số là quá trình trong đó các dữ liệu số được mã hoá vào trong sóng mang hình sin thích hợp với các đặc tính kênh truyền. Kỹ thuật truyền tín hiệu điều chế số còn gọi là kỹ thuật truyền tín hiệu dãy thông.

Dạng tổng quát của sóng mang hình sin s(t) là:

$$s(t) = A(t).\cos[\omega_0(t) + \Phi(t)] \quad (1.1)$$

Trong đó, A: biên độ

$\omega_0=2\pi f$:tần số góc

Φ: góc pha

Giải điều chế số là quá trình ngược lại với điều chế số nhằm phục hồi các luồng bit từ dạng sóng thu được càng ít lỗi càng tốt, mặc dù tín hiệu số có thể méo dạng hoặc nhiễu.

GSM sử dụng phương pháp điều chế khoá chuyển pha cực tiểu GMSK (Gaussian Minimum Shift Keying). Đây là phương pháp điều chế băng hẹp dựa trên kỹ thuật điều chế dịch pha. Để giải thích GSMK, trước hết chúng ta xét MSK bằng cách so sánh nó với PSK. Ta có thể trình bày sóng mang đã được điều chế đối với PSK và MSK như sau:

$$s(t) = A.\cos[\omega_0(t) + \psi(t) + \varphi_0] \quad (1.2)$$

Trong đó: A là biên độ không thay đổi.

$\omega_0=2\pi f$ (rad/s) là tần số góc của sóng mang

$\psi(t)$ là góc pha phụ thuộc vào luồng số mang lên điều chế

φ_0 là góc pha ban đầu

Đối với điều chế pha bốn trạng thái, ta được góc pha ψ(t) như sau: ψ(t) = nπ/2 với n= 0, 1, 2, 3 tương ứng với các cặp bit được đưa lên điều chế là {00, 01, 11, 10}.

Đối với điều chế MSK ta được góc pha ψ(t) như sau:

$$y_t = \sum_i k_i f_i(t - iT) \qquad (1.3)$$

Trong đó, chuỗi bit đưa lên điều chế là {…d_{i-1}, d_i, d_{i+1},…)

$k_i = 1$ nếu $d_i = d_{i-1}$

$k_i = -1$ nếu $d_i \neq d_{i-1}$

$f_i(t) = \frac{p}{2T_b}t$, T_b là khoảng thời gian của bit

Ta thấy, ở MSK nếu bit điều chế ở thời điểm xét giống như bit ở thời điểm trước đó, ψ(t) sẽ thay đổi tuyến tính từ 0 đến π/2, ngược lại nếu bit điều chế ở thời điểm xét khác với bit trước đó thì ψ(t) sẽ thay đổi tuyến tính từ 0 đến -π/2.

Sự thay đổi góc pha ở điều chế MSK cũng dẫn đến thay đổi tần số theo quan hệ sau ω = dφ(t)/dt. Trong đó: φ(t) = (ω$_0$(t) + ψ(t) + φ$_0$)

Nếu chuỗi bit đưa lên điều chế không đổi (toàn số 1 hoặc số 0) ta có tần số sau:

$\omega_1 = 2\pi f_1 = \omega_0 + \pi/(2T_b)$

Nếu chuỗi bit đưa lên điều chế thay đổi luân phiên (1, 0 , 1, 0, 1, 0, …) thì ta có tần số sau : $\omega_2 = 2\pi f_2 = \omega_0 - \pi/(2T_b)$

Để thu hẹp phổ tần của tín hiệu điều chế, luồng bit đưa lên điều chế được đưa qua bộ lọc Gauss. Ở GSM, bộ lọc Gauss được sử dụng tích dải thông chuẩn hoá BT=0.3 , trong đó, B là độ rộng băng tần.

Mục đích dùng GMSK là để tạo ra tín hiệu băng thông nhỏ, độ dịch tần nhỏ.

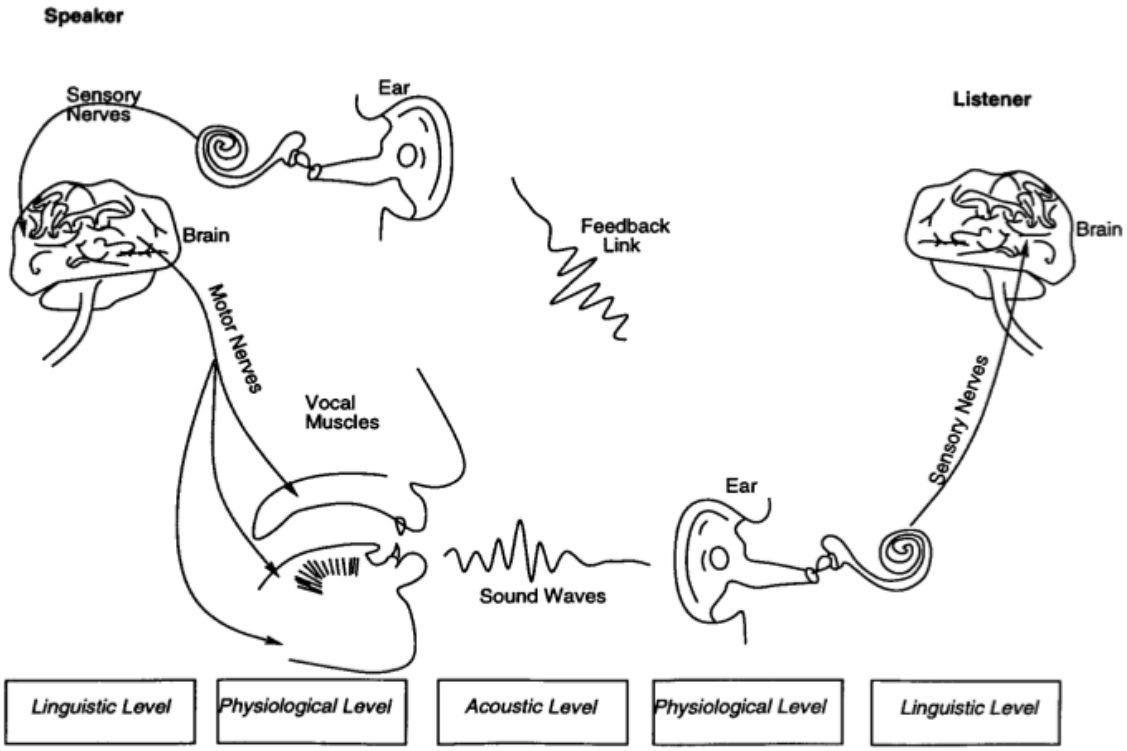
CHƯƠNG 2: QUÁ TRÌNH TẠO TIẾNG NÓI

Để hiểu được các phương pháp mã hoá thoại, điều đầu tiên là ta cần phải hiểu cấu trúc cơ quan phát âm và cơ quan thính giác của con người, hiểu về ngôn ngữ, sinh lý, các mức âm thanh cũng như việc ứng dụng nó vào trong các kĩ thuật mã hoá thoại hiện nay.

Mã hoá thoại có ưu điểm là được tạo ra dựa vào cấu trúc vocal tract (tuyến âm) của con người. Đặc điểm này cũng xác định và giới hạn cấu trúc của tín hiệu thoại.

2.1 Chuỗi thoại

Để rõ hơn ta xét quá trình hai người hội thoại với nhau, một người nói và một người nghe. Chuỗi thoại được tạo ra và truyền đến tai người nghe như trong hình 2.1. Đầu tiên, người nói sẽ sắp xếp các suy nghĩ của mình, xác định xem thử anh ta muốn nói gì và đặt những suy nghĩ đó vào trong một dạng ngôn ngữ bằng cách chọn các từ, cụm từ, nhóm từ chính xác và đặt chúng vào đúng cấu trúc ngữ pháp của ngôn ngữ mình nói.



Hình 2.1 Quá trình tạo thoại

Quá trình này kết hợp với não người nói, nơi sẽ đưa ra các lệnh dưới dạng các xung. Các xung này theo các dây thần kinh điều khiển cơ và cơ quan phát âm như lưới, môi, quai hàm và dây thanh chuyển động làm áp suất không khí xung quanh thay

đổi tạo ra sóng âm truyền trong không khí. Sóng âm này truyền đến tai người nghe và kích hoạt cơ quan thị giác. Cơ qua thính giác cũng tạo ra các xung thần kinh đưa đến não người nghe và não sẽ giúp nhận biết, hiểu được các thông tin từ người nói.

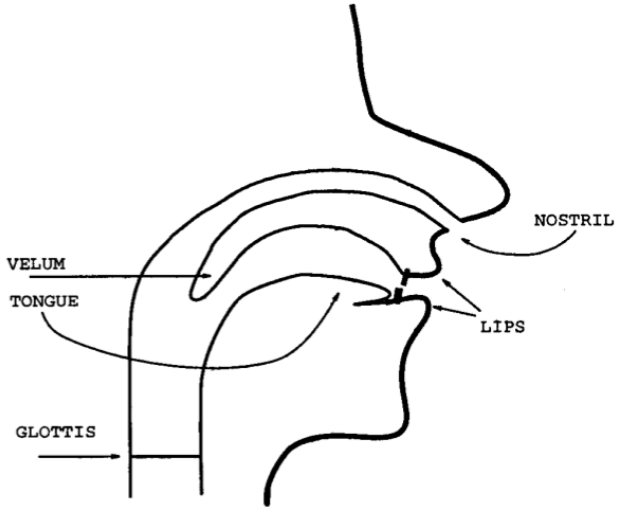
Các dây thần kinh thính giác của người nói cũng được hồi tiếp lại não. Não sẽ tiếp tục so sánh với âm thanh đã nói để có những điều chỉnh thích hợp. Sự hồi tiếp này là rất cần thiết để giúp cho người nói có thể dự đoán được người nghe có nghe rõ ràng và chính xác hay không ?

2.2 Phát âm

Do hoạt động và vị trí của cơ quan phát âm nên âm thanh của mỗi người khác nhau. Khi chúng ta nói khí từ phổi sẽ đi qua vocal tract và ra ngoài tạo thành tiếng nói.

Tín hiệu thoại là tín hiệu động có dạng sóng rất phức tạp. Bằng cách phân tích tín hiệu, người ta thấy rằng phân bố năng lượng theo tần số trong một đoạn thoại ngắn có nhiều dạng khác nhau. Năng lượng phân bố theo tần số được gọi là phổ công suất. Phổ công suất có thể tập trung ở tần số cao, tần số thấp hoặc ở hai bên một dải tần số nào đó. Cấu trúc của phổ có thể ngẫu nhiên hoặc xác định điều hoà. Phổ của của thoại luôn thay đổi làm cho mã hoá càng thêm phức tạp. Để khắc phục điều này, người ta sắp xếp thành các mức vật lý khác nhau. Bằng cách nghiên cứu cơ quan phát âm và hoạt động của nó, các dạng tín hiệu thoại khác nhau được xét riêng lẻ.

Hình 2.2 cho thấy sơ đồ đơn giản hoạt động của vocal tract. Không khí từ phổi đẩy vào khí quản, đi qua dây thanh và cuối cùng vào hốc mũi và miệng. Thanh môn cho phép một lượng không khí vừa đủ từ phổi đi qua hoặc có thể ngắt luồng không khí thành các xung tuần hoàn.



Hình 2.2 Phát âm của vocal tract

2.2.1 Kích thích

Tín hiệu thoại là do không khí từ phổi được biến đổi thành dạng năng lượng kích thích vocal tract rung và ta xem đây là tín hiệu kích thích trong bộ mã hoá. Dây thanh rung tạo ra các xung truyền đến mũi và miệng. Vì vậy, năng lượng kích thích ở nhiều tần số và cường độ của các tần số này phụ thuộc vào tốc độ chuyển động của vocal tract.

Tổng quát, kích thích được chia làm hai dạng: hữu thanh (voice) và vô thanh (unvoice). Âm thanh tạo ra do sự rung động của dây thanh được gọi là hữu thanh. Tất cả các nguyên âm và một số phụ âm là âm hữu thanh. Âm thanh được tạo ra không phải do sự rung của các dây thanh mà do không khí bị vocal tract co thắt thì được gọi là âm vô thanh, ví dụ như âm "s", "p". Đặc điểm của âm hữu thanh và âm vô thanh phụ thuộc vào:

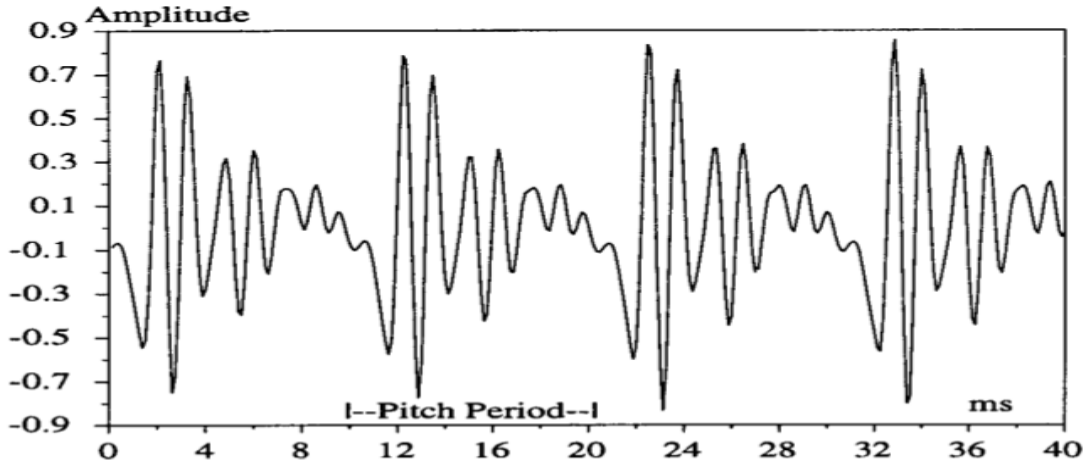
- Kích thước chia nhỏ luồng không khí từ phổi tạo thành các xung tựa tuần hoàn. Năng lượng để thực hiện điều này là kích thích âm hữu thanh như là các nguyên âm.

- Luồng không khí từ phổi đến mũi, giống như là nhiễu loạn tạo ra do sự co thắt vocal tract. Năng lượng để thực hiện quá trình này là kích thích âm vô thanh như âm "s".

Ngoài hai dạng trên còn có một dạng hỗn hợp của nó ví dụ như "z". Tuy nhiên, ta chỉ xét hai loại là hữu thanh và vô thanh dựa vào sự có mặt hay vắng mặt của kích thích tuần hoàn. Do đó, "z" cũng được xem là âm hữu thanh.

Pich

Tần số của kích thích tuần hoàn (hoặc tựa tuần hoàn) được gọi là pitch. Khoảng thời gian giữa điểm bắt đầu cũng như điểm kết thúc của dây thanh đến điểm tương ứng trong chu kì kế tiếp được gọi là chu kì pitch.



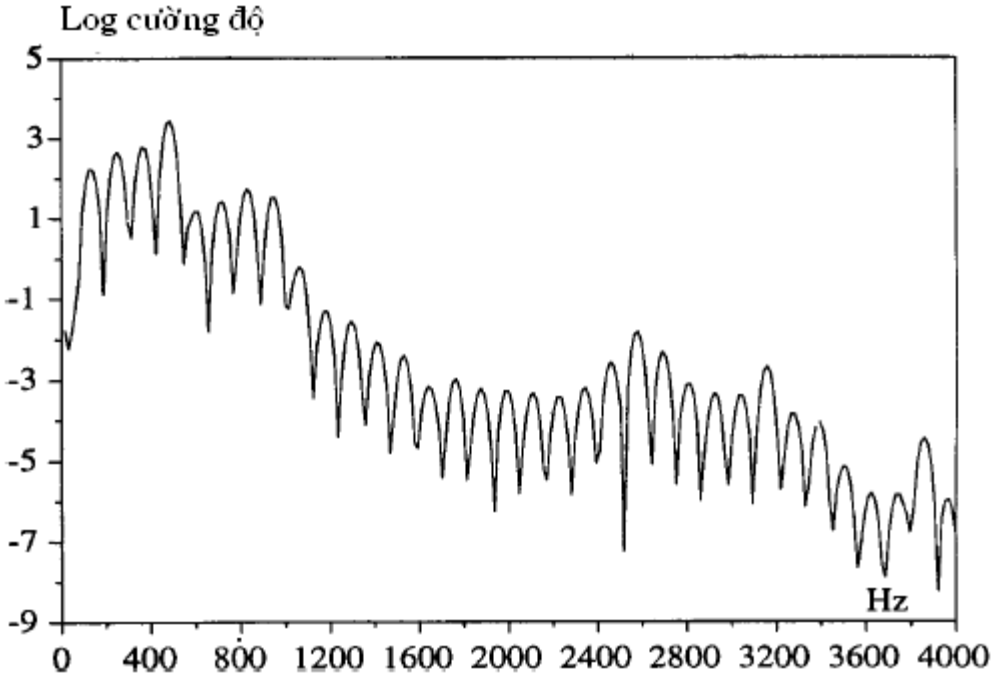
Hình 2.3 Dạng sóng tiếng nói của đoạn thoại (âm hữu thanh) ngắn

Hình 2.3 cho ta dạng sóng thời gian của một đoạn thoại dài 40 ms của âm hữu thanh. Trục x là trục thời gian (ms). Trục y là biên độ. Giá trị biên độ cao ở điểm bắt đầu xung pitch, chu kì pitch là 10 ms và tần số pitch là 1/10ms bằng 100 Hz.

2.2.2 Vocal tract

Kích thích là một trong hai hệ số quan trọng tác động đến tiếng nói. Cho kích thích là âm hữu thanh hoặc âm vô thanh, khi vocal tract thay đổi sẽ cho các âm thanh khác nhau. Khi hình dạng và vị trí của vocal tract thay đổi thì sẽ làm cho tần số cộng hưởng của vocal tract thay đổi theo.

Các tần số cộng hưởng này cho các đỉnh phổ nằm ở các tần số ứng với từng dạng vật lý của vocal tract. Tần số cộng hưởng được gọi là formant và vị trí tần số của chúng được gọi là tần số formant.



Hình 2.4 Log cường độ phổ của một đoạn thoại (âm hữu thanh) ngắn

Hình 2.4 cho phổ trong một đoạn ngắn của tín hiệu âm hữu thanh. Trục x từ 0 đến 4000 Hz. Trục y là log cường độ của đáp ứng tần số. Đỉnh hẹp cách đều nhau 120 Hz là hoà âm học pitch. Ba formant đầu tiên ở vị trí 400, 900, 2600 Hz.

Cách phát âm

Trong vocal tract, sự co thắt và ống dẫn không khí sẽ tạo nên cách phát âm. Để tạo ra các âm khác nhau thì kích thích được tạo ra bởi vocal tract phải khác nhau. Ví dụ nguyên âm được tạo ra bởi kích thích tuần hoàn và luồng không khí đi qua vocal tract có tốc độ không bị hạn chế. Tuy nhiên, tốc độ này không đều, nó còn phụ thuộc vào tần số formant. Ngược lại, âm vô thanh không có các thành phần tuần hoàn và được tạo ra do một số sự co thắt.

Phụ âm dừng hay còn gọi là âm bật, được tạo ra do áp suất luồng không khí bị chặn đột ngột. Phụ âm dừng có thể là âm hữu thanh như "b" hoặc âm vô thanh như âm "p".

Phụ âm mũi được tạo ra do luồng không khí qua vòm miệng, môi bị giảm để chuyển sang mũi như các âm "m", "n".

Vị trí phát âm

Cách phát âm xác định nhóm âm thanh và vị trí phát âm xác định chính xác điểm co thắt. Vị trí chính xác của vocal tract sẽ tạo nên âm thanh đặc trưng của từng người. Nguyên âm được phân biệt nhờ lưỡi tạo nên sự co thắt, ví dụ:

- Một nguyên âm trước như trong từ "beet"
- Một nguyên âm giữa như trong từ "bet"
- Một nguyên âm sau như trong từ "boot"

Trong từ "beet" lưỡi sẽ chạm lên phần trên của miệng và phần sau của răng, còn "boot" thì lưỡi lùi lại phía sau gần quai hàm tạo ra sự co thắt. Các âm "p", "t", "k" được tạo ra do vị trí khác nhau trong vocal tract nơi sự co thắt được thực hiện để dừng luồng không khí trước khi nói.

"p": đóng môi.

"t": lưỡi ở giữa hai hàm răng.

"k": lưỡi ở sau miệng

2.2.3 Âm vị

Chất lượng của kích thích, vị trí và cách phát âm sẽ tạo nên đặc điểm của âm vị. Vì vậy, mục đích của mã hoá thoại là nhằm giúp ta hiểu được các âm khác nhau trong cùng một ngôn ngữ.

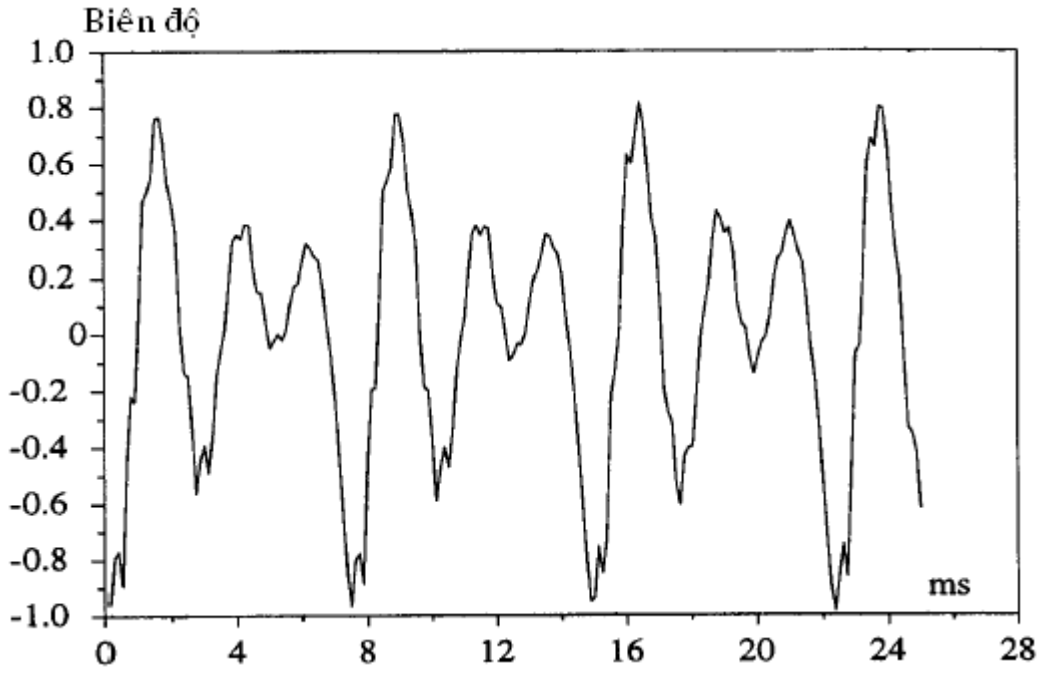
2.2.3.1 Nguyên âm

Nguyên âm là dạng âm hữu thanh có độ phát âm thay đổi không đáng để. Bảng 2.1 là danh sách các nguyên âm dựa trên độ co thắt và vị trí của lưỡi.

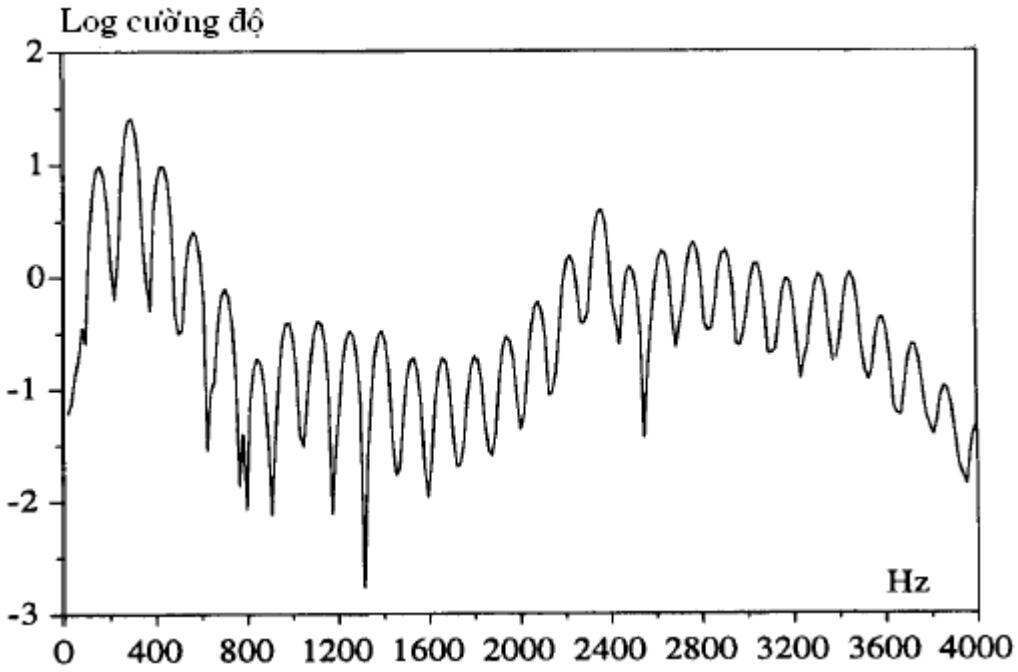
Vị trí / Co thắt	Trước	Giữa	Sau
Cao	/i/ beet	/ER/ bird	/u/ boot
Trung bình	/E/ bet	/UH/ but	/OW/ bought
Thấp	/ae/ bat		/a/ father

Bảng 2.1 Độ co thắt và vị trí lưỡi của các nguyên âm trong tiếng Anh

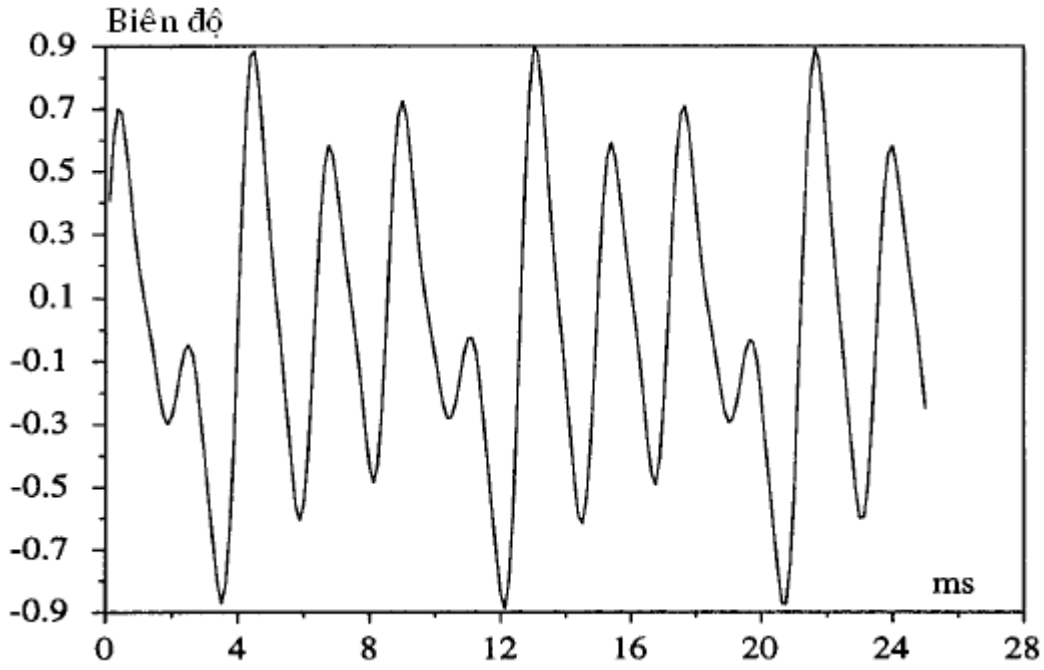
Hình 2.5 và 2.6 hiển thị dạng sóng log cường độ phổ của nguyên âm /I/ ("bit") và /U/ ("foot"). Dạng sóng thời gian cho thấy tần số của /I/ cao hơn nhiều so với /U/.



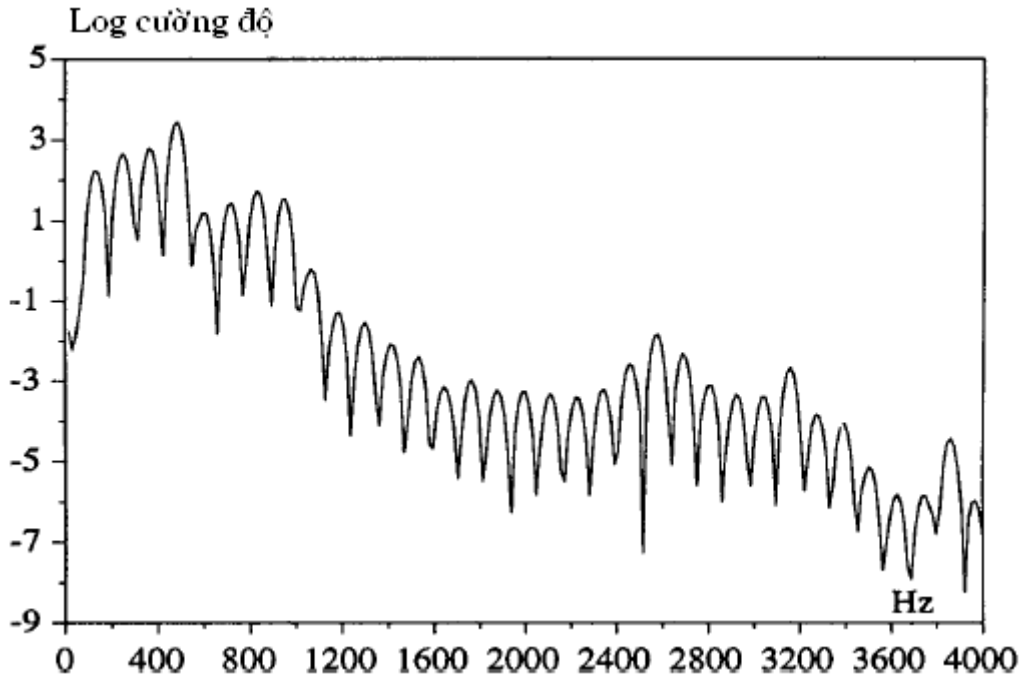
Hình 2.5(a) Dạng sóng thời gian của /I/ trong từ "bit"



Hình 2.5(b) Log cường độ phổ của /I/ trong từ "bit"



Hình 2.6(a) Dạng sóng thời gian của /U/ trong từ "foot"



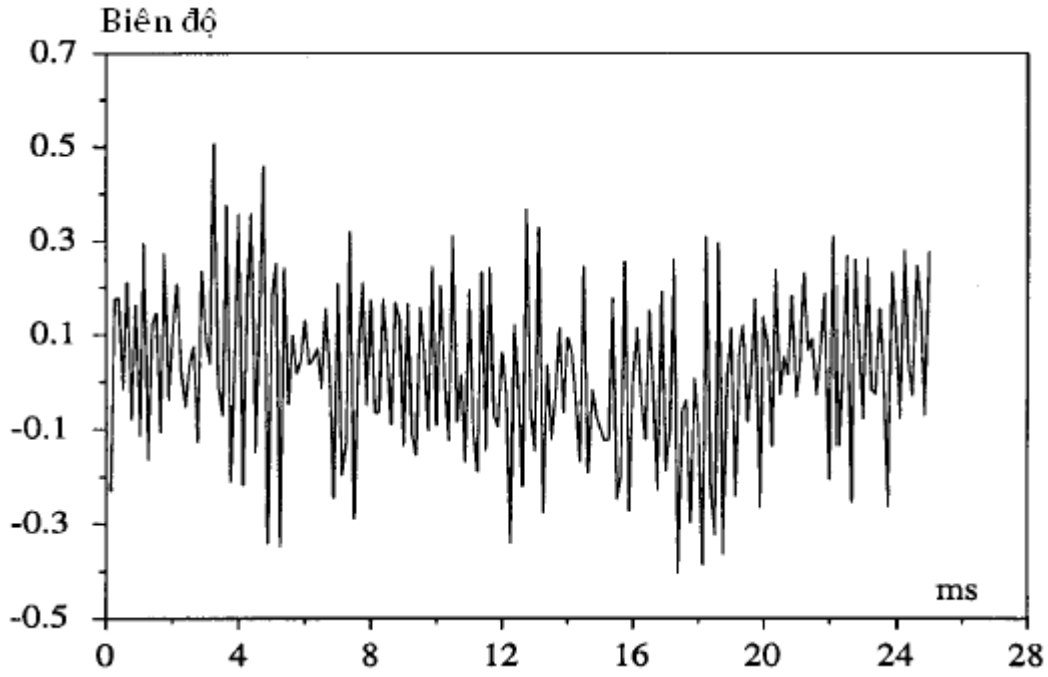
Hình 2.6(b) Log cường độ phổ của /U/ trong từ "foot"

2.2.3.2 Phụ âm xát

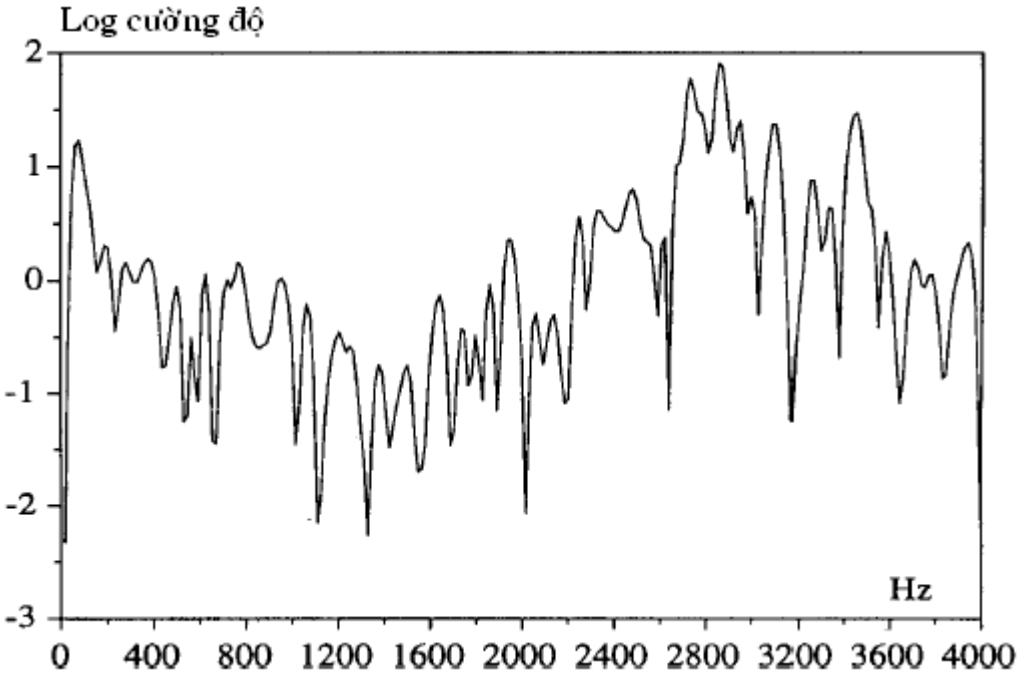
Phụ âm chủ yếu tạo nên do nhiễu loạn của luồng không khí được gọi là phụ âm xát. Phụ âm xát được tạo ra do luồng không khí bị vocal tract co thắt, bao gồm cả âm

hữu thanh lẫn âm vô thanh. Bảng 2.2 là danh sách phụ âm xát. Những từ liệt kê trong bảng cho ta ví dụ chung của âm vị.

Hình 2.7 là dạng sóng theo thời gian và log cường độ phổ của một mẫu /sh/. Âm là âm hữu thanh và dạng sóng thời gian giống như là nhiễu ngẫu nhiên. Phổ có dạng xác định, không bằng phẳng. Độ cao đỉnh phổ khoảng 2800 Hz.



Hình 2.7(a) Dạng sóng thời gian của /sh/ trong âm bắt đầu từ "shop"



Hình 2.7(b) Log cường độ phổ của /sh/ trong âm bắt đầu từ "shop"

Co thắt	Âm vô thanh	Âm hữu thanh
Răng/môi	/f/ fit	/v/ vat
Răng	/THE/ thaw	/TH/ that
Vòm miệng	/sh/ sap	/zh/ vision
Thanh môn	/h/ help	

Bảng 2.2 Vị trí co thắt và phụ âm xát trong tiếng Anh

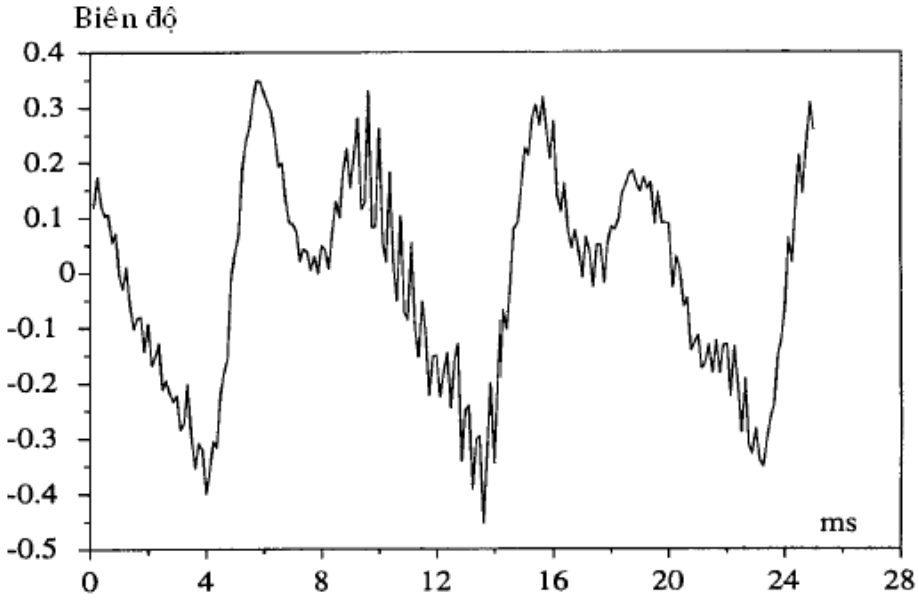
2.2.3.3 Phụ âm dừng

Phụ âm dừng hay âm bật là dạng không khí bị ngắt đột ngột do co thắt. Chúng chỉ là những âm ngắn xuất hiện nhanh. Tín hiệu dừng có thể là âm hữu thanh hay là âm vô thanh. Phụ âm dừng trong tiếng Anh được cho trong bảng 2.3. Sự co thắt xác định vị trí của môi, răng và vòm miệng. Bảng 2.3 là những từ thường gặp mà âm đầu tiên là phụ âm dừng.

Co thắt	Âm vô thanh	Âm hữu thanh
Môi	/p/ pat	/b/ bat
Răng	/t/ tap	/d/ dip
Sau vòm miệng	/k/ cat	/g/ good

Bảng 2.3 Vị trí co thắt và phụ âm dừng trong tiếng Anh

Hình 2.8 là giản đồ dạng sóng của /t/ khi phát âm "tap". Âm bật chủ yếu như một xung kim. Do chỉ dừng trong khoảng thời gian ngắn nên nó ảnh hưởng lớn đến các âm trước và sau. Nếu xuất hiện ở cuối một từ thì nó còn có thêm âm bật do không khí tạo ra.



Hình 2.8 Dạng sóng thời gian của /t/ khi phát âm từ "tap"

2.2.3.4 Phụ âm mũi

Âm mũi tạo ra do vocal tract đóng luồng không khí và đưa nó ra ngoài bằng mũi. Âm mũi là phụ âm âm hữu thanh. Bảng 2.4 liệt kê ba phụ âm mũi trong tiếng Anh. Do miệng đóng kín nên âm mũi có năng lượng thấp hơn so với các phụ âm âm hữu thanh khác. Luồng không khí đi qua hốc mũi, kết hợp với đóng miệng nên có phổ cũng khác với các dạng trước.

Co thắt	Âm hữu thanh
Môi	/m/ map
Răng	/n/ no
Sau vòm miệng	/ng/ hang

Bảng 2.4 Vị trí co thắt đối với phụ âm mũi trong tiếng Anh

2.3 Dạng bộ lọc nguồn

Để dễ dàng phân tích tín hiệu thoại, hầu hết bộ mã hoá tiếng nói đều có dạng vocal tract. Dạng này thường được dùng ở hầu hết các quá trình mã hoá và giải mã. Khi mã hoá, các kiểu thông số được xác định để miêu tả chính xác thoại ngõ vào. Đối với giải mã, cũng có cấu trúc tương tự và dựa vào các thông số này để tái tạo lại thoại ban đầu.

Một dạng tạo thoại thường được sử dụng nhất đó là dạng bộ lọc nguồn. Bộ lọc nguồn này có dạng giống như vocal tract. Nguồn tín hiệu cung cấp cho bộ lọc nguồn này là tín hiệu kích thích.

2.3.1 Vocal tract

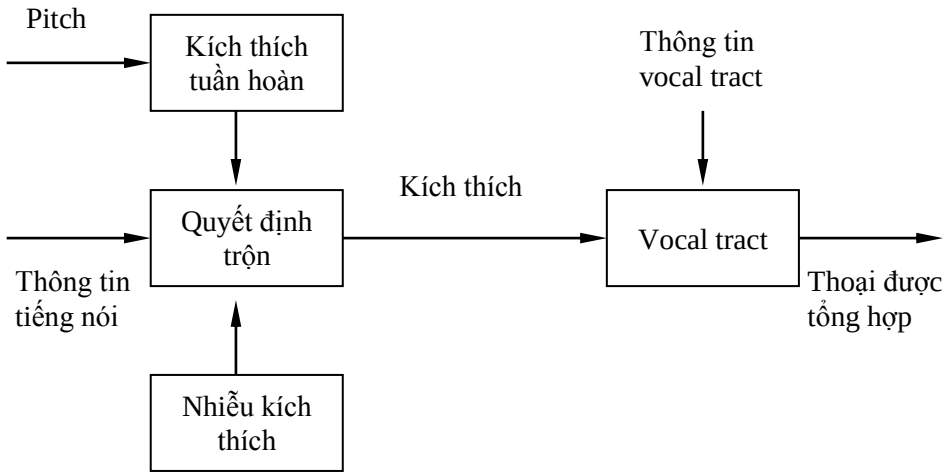
Cổ họng, mũi, lưỡi và miệng là hốc cộng hưởng không khí để tạo nên tiếng nói của con người. Vocal tract có cấu trúc khác nhau thì sẽ có các tần số cộng hưởng khác nhau. Tần số cộng hưởng cùng với tín hiệu kích thích là hai hệ số chính điều khiển vocal tract tạo ra các âm vị.

2.3.2 Kích thích

Đối với tiếng nói âm hữu thanh, dạng sóng tuần hoàn tạo kích thích đến vocal tract. Dạng sóng tuần hoàn từ các xung thanh môn sẽ làm cho dây thanh sẽ rung. Dạng đơn giản và hay dùng cho âm vô thanh là nhiễu trắng. Nhiễu trắng thường ngẫu nhiên và có phổ bằng phẳng ở mọi tần số có cùng công suất. Giả sử nhiễu trắng được tạo ra khi không khí đi qua bộ phận co thắt. Một số âm như âm /z/ được tạo ra vừa bởi một kích thích tuần hoàn và vocal tract co thắt không khí. Điều này được gọi là kích thích pha trộn. Vì vậy, nhiệm vụ chính của mã hoá thoại là phải phân biệt đâu là âm hữu thanh, âm vô thanh hay là pha trộn của nó.

2.3.3 Dạng bộ lọc nguồn tổng quát

Sơ đồ hình 2.9 chứng minh rằng luồng tín hiệu và thông tin của một bộ lọc nguồn tổng quát. Thông tin pitch thường được chứa trong giá trị chu kì pitch. Giá trị này thay đổi tuỳ theo sự thay đổi của tín hiệu thoại. Dựa vào chu kì pitch, khối "kích thích tuần hoàn" tạo ra một dạng sóng xung đại diện cho các xung thanh môn. Khối "nhiễu kích thích" có ngõ ra là nhiễu liên tục với đáp ứng phổ bằng phẳng. Hai kích thích này được cho vào bộ quyết định trộn. Thoại cũng sẽ cho vào một ngõ vào khác. Dựa vào các mức của thoại gốc, khối "quyết định trộn" kết hợp với "kích thích tuần hoàn" và "nhiễu kích thích" sẽ tạo ra tín hiệu kích thích phù hợp.



Hình 2.9 Dạng bộ lọc nguồn tổng quát

Thường có 2 dạng, bộ lọc nguồn sẽ kết hợp quyết định cứng âm hữu thanh/âm vô thanh đối với mỗi đoạn thoại. Trong trường hợp này, chức năng của khối "quyết định trộn" như một chuyển mạch với kích thích là âm hữu thanh/âm vô thanh. Thông tin vocal tract được cung cấp vào khối "vocal tract" để tạo ra một bộ lọc vocal tract. Bộ lọc sẽ làm cho phổ của kích thích giống như của tín hiệu thoại gốc. Thực tế, thông tin vocal tract được tạo ra bằng một số phương pháp bao gồm một dự đoán tuyến tính và giá trị Fourier. Kích thích được lọc bởi vocal tract để tạo ra thoại tổng hợp đến tai người nghe sao cho giống tín hiệu thoại ban đầu nhất.

CHƯƠNG 3: CÁC PHƯƠNG PHÁP CƠ SỞ MÃ HOÁ TIẾNG NÓI

3.1 Các phương pháp cơ sở mã hoá tiếng nói

Về cơ bản bộ mã hóa tiếng nói có 3 loại:

- Mã hóa dạng sóng (waveform).
- Mã hóa nguồn (source).
- Mã hóa lai (hybrid): là sự kết hợp của mã hoá dạng sóng và mã hoá nguồn.

Nguyên lý của mã hóa dạng sóng là tìm cách số hóa dạng sóng của tiếng nói theo cách thích hợp. Tái phía phát, bộ mã hóa sẽ nhận các tín hiệu nói tương tự liên tục và chuyển thành tín hiệu số trước khi truyền đi. Tại phía thu sẽ làm nhiệm vụ ngược lại để khôi phục tín hiệu tiếng nói. Khi không có lỗi truyền dẫn thì dạng sóng của tiếng nói khôi phục rất giống với dạng sóng của tiếng nói gốc. Ưu điểm của loại mã hóa này là: độ phức tạp, giá thành thiết kế, độ trễ và công suất tiêu thụ thấp. Bộ mã hóa dạng sóng đơn giản nhất là điều chế xung mã (PCM), điều chế Delta (DM)... Tuy nhiên, nhược điểm của bộ mã dạng sóng là không tạo được tiếng nói chất lượng cao, tốc độ dưới 16kbit/s.

Bộ mã hóa nguồn khắc phục được nhược điểm này. Nguyên lý của mã hóa là mã hóa kiểu phát âm (vocoder), ví dụ như bộ mã hóa bằng dự đoán tuyến tính (Linear Prediction Coding - LPC). Các bộ mã hóa này có thể thực hiện được tại tốc độ bit lớn hơn 1kbps. Hạn chế chủ yếu của mã hóa kiểu phát âm LPC là việc mô phỏng nguồn kích thích còn đơn giản nên tiếng nói tái tạo được là tiếng nói dạng tổng hợp, chất lượng không cao và khó có thể nhận ra giọng người nói chuyện. Vào năm 1982, Atal đã đề xuất một mô hình mới về kích thích, được gọi là kích thích đa xung .Trong mô hình này, không cần biết trước xem đó là âm hữu thanh hay vô thanh. Sự kích thích được mô hình hóa bởi một số xung có biên độ và vị trí được xác định bằng việc cực tiểu hóa sai lệch, có tính đến trọng số thụ cảm, giữa tiếng nói gốc và tiếng nói tổng hợp. Việc đưa ra mô hình này đã gây chú ý và đó là mô hình đầu tiên của một thế hệ mới của các bộ điều chế tiếng nói phân tích bằng tổng hợp (Analisis by Synthesis). Tín hiệu kích thích sẽ được tối ưu hóa một cách kỹ lưỡng và người ta sử dụng kỹ thuật mã hóa dạng sóng để mã hóa tín hiệu kích thích này một cách có hiệu quả.

Chỉ tiêu đánh giá thuật toán mã hoá:

- Hai mục tiêu quan trọng đặt ra là: tối thiểu hóa tốc độ bit và tối ưu hóa chất lượng. Hai mục tiêu này thường có mâu thuẫn với nhau. Tốc độ bit được tính bằng bps. Chất lượng được đánh giá ở việc được tái tạo lại dạng tương tự với một sai số cang nhỏ càng tốt. Việc lấy mẫu không ảnh hưởng đến chất lượng. Còn lượng tử hóa thì có thể gây ra những sai số làm mất mát thông tin so với tín hiệu ban đầu được gọi

là nhiễu lượng tử . Tỷ số tín hi ệu trên nhiễu (SNR) được dung đánh giá chất lượng tiếng nói. Nếu tỉ số này thấp người nghe sẽ thu được tiếng nói không tốt.

- Chất lượng chấp nh ận được có SNR khoảng trên 30 dB. Theo tính toán việc thêm 1 bit biểu diễn giá trị lượng tử sẽ làm tăng SNR lên khoảng 6dB, tương tự sẽ giảm 1 bit làm SNR giảm xuống 6dB.

- Người ta thường dùng m ột tiêu chuẩn gọ i là MOS (Mean Opinion Score) để so sánh chất lượng mã hoá tiếng nói , với thang giá trị từ 1 đến 5, cho ta biết m ột thuật toán điều chế đạt được chất lượng có gần với tiếng nói tự nhiên hay không.

3.1.1 Phương pháp mã hoá tiếng nói dạng sóng

Kiểu mã hóa này cố gắng mã hóa dạng sóng của tiếng nói m ột cách có hiệu quả, dạng đơn giản là điều chế xung mã PCM , ngoài ra còn có các thuật toán khác có thể làm giảm tốc đ ộ bit hơn nữa. Công nghệ mã hóa dạng sóng thườn g cho tiếng nói chất lượng tốt với băng thông 16kbps trở lên.

Để tránh hi ện tượng chồng phổ, tiếng nói tương tự được lọc trước khi số hóa để loại trừ các thành ph ần tần số cao không mong muốn . Phổ tiếng nói có thể gồm cả những thành phần t ần số tới 10 kHz, nhưng do hầu hết các t ần số tiếng nói tập trung vào khoảng từ (300 Hz – 3.4 kHz) nên tín hiệu tiếng nói được lọc đi để loại bỏ thành phần ngoài khoảng tần số ấy.Theo định luật lấy mẫu thì tần số lấy mẫu sẽ là 8 kHz. Hệ thống như vậy gọi là PCM (Pulse Code Modulation). Phổ biến hi ện nay người ta chọn tốc độ lấy mẫu là 8 kHz và số bit lượng tử n=8, tức là tốc đ ộ truyền sẽ là 64 kbps. Các bit mã hóa được truyền tuần tự trên đường truyền.

3.1.1.1 PCM (Pulse Code Modulation)

PCM đều (uniform PCM):Đầu vào của b ộ lượng tử là tín hi ệu tương tự đã được đưa qua b ộ lấy mẫu.Với một bộ lượng tử dùng N bit từ mã,miền giá trị lượng tử được chia thành 2^N mức, mỗi từ mã N bit tương ứng với 1 giá trị.Khoảng cách giữa các mức gọi là bước lượng tử (step size).Bộ lượng tử quyết định xem với mỗi giá trị đầu ra là giá trị lớn nhất của miền giá trị . Trong kiểu PCM đều ,các giá trị lượng tử cách đều nhau.Bước lượng tử phải được chọn sao cho đủ nhỏ để có thể tối thiểu nhiễu lượng tử,nhưng lại có thể đủ lớn để miền giá trị của cả b ộ lượng tử có đ ộ lớn thích hợp.Với một bộ lượng tử N bit có bước lượng tử là S,thì miền giá trị là $R=2^N*S$.

Nếu N không đủ lớn thì vi ệc cắt xén tín hi ệu vượt qua miền giá trị sẽ xảy ra nhiều hơn và đó là dĩ nhiên là một nguyên nhân khác của nhiễu lưỡng tử.

Phương pháp này có nhược điểm là SNR ,tức là chất lượng không chỉ phụ thu ộc vào bước lượng tử mà còn phụ thuộc và cả biên độ của tín hiệu lấy mẫu.

*Lượng tử hóa kiểu PCM đều :*Cần N cỡ 11 bit trở lên để có thể đảm bảo chất lượng tiếng nói.Điều này làm tốc đ ộ bit lớn nên chúng ít được sử dụng trong thực tế.

*Lượng tử hóa Logarithm (logarithmic PCM):*Mục tiêu của phương pháp này là duy trì một tỷ số SNR ít thay đổi trong toán phạm vi giá trị biên đ ộ.Thay vì lượng tử hóa giá trị tương tự của tín hiệu lấy mẫu,trước tiên ta tính toán hàm logarithm của từng giá trị rồi mới lượng tử hóa chúng .SNR sẽ chỉ phụ thu ộc vào bước lượng tử .Lượng tử logarithm là một quá trình nén, chúng làm giảm miền giá trị đầu vào một cách đáng kể tùy thuộc vào dạng hàm logarithm được dùng .Sau khi nén,một quá trình ngược lại là mũ hóa được sử dụng để tái tạo lại tín hi ệu nguyên thủy ban đầu .Toàn bộ chu trình được gọi là Companding(Compressing/expanding).

Hai tiêu chuẩn được dùng phổ biến hi ện nay là luậtμvà luật A.Lượng tử hoá theo luật μ sử dụng ở Bắc Mỹ và Nhật Bản, trong khi đó lượng tử hoá theo luật A được sử dụng ở châu Âu.

Các mẫu tín hiệu rời rạc theo biên độ được mã hoá nhị phân. Ví dụ, mã hoá theo luật A, người ta chia đường cong logarith thành 13 đoạn.

Bit thứ nhất là bit có trọng số lớn nhất, là bit đấu. Giá trị 1 chỉ thị tín hiệu dương và giá trị 0 chỉ thị tín hiệu âm.

Bit 2, 3, 4 xác định đoạn lượng tử hoá theo mỗi vùng âm và dương.

Bit 5, 6, 7, 8 là các bit có trọng số nhỏ nhất, xác định vị trí của giá trị lượng tử hoá trong đoạn.

3.1.1.2 DM(Delta Modulation)

Là một trong những phương pháp điều chế vi sai ,dựa trên tính chất là tín hi ệu tiếng nói tại thờ i điểm có ít nhiều phụ thu ộc vào tín hiệu ở các thời điểm trước đó ,vì thế ta có thể dự đoán tín hi ệu tại thời điểm hi ện tại,và chỉ cần lưu trữ giá trị khác bi ệt giữa giá trị thực và giá trị dự đoán của tín hi ệu,sự sai khác này,giúp tiết ki ệm băng thông để đạt hiệu quả cao.

Ý tưởng của phương pháp điều chế Delta là chỉ truyền đi giá trị thay đổi tuyệt đối của tín hi ệu.Dựa vào sự khác nhau của tín hi ệu tại thời điểm liền kề nhau mà ta tính được tín hiệu phải truyền trên đường dây .Phương pháp này chỉ sử dụng 1 bit để mã hóa tín hiệu sai khác đó,nghĩa là cho biết tín hi ệu tại thời điểm t+1 là lớn hơn hay nhỏ hơn tín hiệu tại thời điểm t.

3.1.1.3 DPCM(Differential PCM)

Đây là phương pháp cũng dựa trên nguyên tắc chỉ truyền đi sự khác nhau của tín hiệu tại hai thời điểm kề nhau là t và t +1.Khác với DM chỉ dùng 1 bit để giải mã,DPCM dùng N bit để có thể biểu diễn giá trị sai khác này .Chất lượng điều chế khá tốt với lượng bit cần dùng ít hơn so với PCM.

3.1.1.4 ADPCM (Adaptive Differential PCM)-G.726

Là phương pháp mở r ộng của DPCM.Người ta vẫn dùng m ột số bit nhất định để mã hóa sự sai khác giữa tín hi ệu tại 2 thời điểm kề nhau,nhưng bước lượng tử có thể được điều chỉnh tại các thời điểm khác nhau để tối ưu hóa vi ệc điều chế.

Với mục tiêu làm giảm tốc đ ộ bit hơn nữa mà chất lượng tín hi ệu tương đương,người ta sử dụng phương pháp thích nghi động giá trị của bước lượng tử trư ớc những thay đổi của biên đ ộ tín hiệu vào.Mục đích là duy trì miền giá trị lượng tử p hù hợp với miền giá trị của tín hi ệu vào .Đây được gọi là phương pháp Adaptive PCM(APCM).Thích nghi bước lượng tử có thể áp dụng cho cả kiểu lượng tử đều và không đều.Tiêu chuẩn thay đổi bước lượng tử dựa vào m ột số thống kê về tín hi ệu có liên quan đến biên đ ộ của nó.Có nhiều bước toán đ ể tính toán bước lượng tử. Thông thường có 2 kiểu là feedforward APCM và feedback APCM .Trong cả 2 kiểu người ta đều dựa trên những tính toán liên quan đến m ột khối (block) mẫu thu được trong m ột thời gianngắn,về năng lượng ,sự biến đổi và những đo đạ c khác.Ta còn gọi là block companding.Trong kiểu feedback ,việc tính toán bước lượng tử được thực hi ện trên mỗi câu khi nó được đưa vào xử lý (vẫn dùng giá trị bước lượng tử trước đó),thì cho ra kết quả là một giá trị bước lượng tử mới được dùng xử lý N mẫu tiếp theo.

Feedforward theo một cách tiếp c ận khác,dùng chính ngay giá trị bước lượng tử được tính toán ngay trên N mẫu để xử lý N mẫu đó .Như vậy qua trình xử lý phải cần tới môt b ộ đệm để chứa khối dữ li ệu lấy mẫu .Trong khi kiểu feedback có ưu điểm là rất nhạy cảm với nhiễu lượng tử vì nó có tính toán bước lượng tử và sử dụng ngay cho chính block mà từ đó nó thực hiện phép tính.

3.1.2 Phương pháp mã hóa tiếng nói kiểu Vocoder

Vocoder là kiểu điều mã hóa nói dựa trên các tham số mô phỏng b ộ máy phát âm,khác với mã hóa dạng sóng của tiếng nói tươ ng tự , gọi là mã hóa nguồn (Vocoder).Nguyên lý dựa trên vi ệc cho rằng tuyến âm thanh thay đổi từ từ,trạng thái và cấu hình của chúng tại bất cứ thời điểm nào có thể được mô phỏng m ột cách gần đúng bằng m ột tập nhỏ các tham số .Nhờ vi ệc tuyến âm có tốc đ ộ thay đổi từ từ cho phép mỗi t ập tham số có thể đại di ện cho trạ ng thái của nó qua m ột khoảng thời gian 25 ms.Hầu hết các Vocoder biểu diễn đ ặc tính của nguồn kích thích và tuyến âm chỉ bằng m ột tập tham số .Nó gồm khoảng 10 đến 15 hệ số của b ộ lọc để định nghĩa các đặc tính cộng hưởng củ a tuyến âm ,1 tham số 2 giá trị đơn giản để chỉ ra nguồn phát âm là vô thanh hay hữu thanh ,1 tham số chỉ ra năng lượng kích thích và 1 tham số chỉ ra chu kì cơ bản (âm sắc,chỉ có với hữu âm thanh).Trạng thái của tuyến âm được suy ra bằng cách phân tích dạng sóng tiếng nói trong khoảng thời gian 10 đến 25ms và tính toán ra một tập mới cá c tham số (một khung dữ li ệu) tại phần c uối của khoảng thời

gian đó.Khung dữ liệu này được truyền đi và sau đó dùng để điều khiển việc tổng hợp lại tiếng nói.Vocoder có khả năng chuyển giữa 2 kiểu nguồn kích thích là nguồn xung đối âm hữu thanh và nhiễu trắng với âm vô thanh.Bên phía tổng hợp sẽ dùng 1 trong 2 nguồn này cho đi qua bộ lọc gồm các hệ số của khung dữ liệu để tổng hợp tiếng nói.

Ngoài việc đạt được tốc đ ộ bit thấp,Vocoder còn có ưu điểm là phân tích được các tham số nguồn kích thích .Bit biểu thị âm s ắc, âm lượng và âm h ữu thanh/âm vô thanh. Bản thân nó là các bit trong khung dữ liệu, nên các sự thay đổi của chúng có thể được sửa đổi trước ho ặc trong khi tổng hợp .Vì thế ta có thể biến m ột âm thanh hữu thanh thành một lời thì thầm khi thiết đ ặt lại giá trị của bit âm hữu thanh/âm vô thanh.Cũng có thể thay đổi bản thân câu nói bằng cách sửa đổi các tham số c ộng hưởng.

Nhược điểm của phương pháp này là cho tiếng nói có dạng tổng hợp , khó có khả năng nhận dạng được người nói.

Mô tả bộ máy phát âm của con người: Khi chúng ta nói,âm thanh được tạo ra như sau:

- Không khí được đẩy vào phổi qua tuyến âm (vocal track) và miệng tạo thành câu nói.
- Đối với âm hữu thanh thì d ây thanh (vocal cords) rung lên. Tốc đ ộ rung của dâ y thanh nhanh hay chậm quyết định âm s ắc (pitch) của tiếng nói . Phụ nữ và trẻ em thường có giọng thanh (âm sắc cao-dao động nhanh hơn),trong khi nam giới thường có giọng trầm (dao động chậm).
- Âm thanh được tạo ra không phải do sự rung của các dây thanh mà do không khí bị dây thanh co thắt thì được gọi là âm vô thanh.
- Hình dạng của tuyến âm quyết định âm thanh tạo ra . Khi ta nói, tuyến âm thay đổi hình dạng để tạo ra các tiếng khác nhau ,nói chung là hình dạng của tuyến âm thay đổi một cách từ từ,thường là từ 10ms đến 100ms.
- Lượng không khí từ phổi quyết định âm lượng (gain) của tiếng nói.

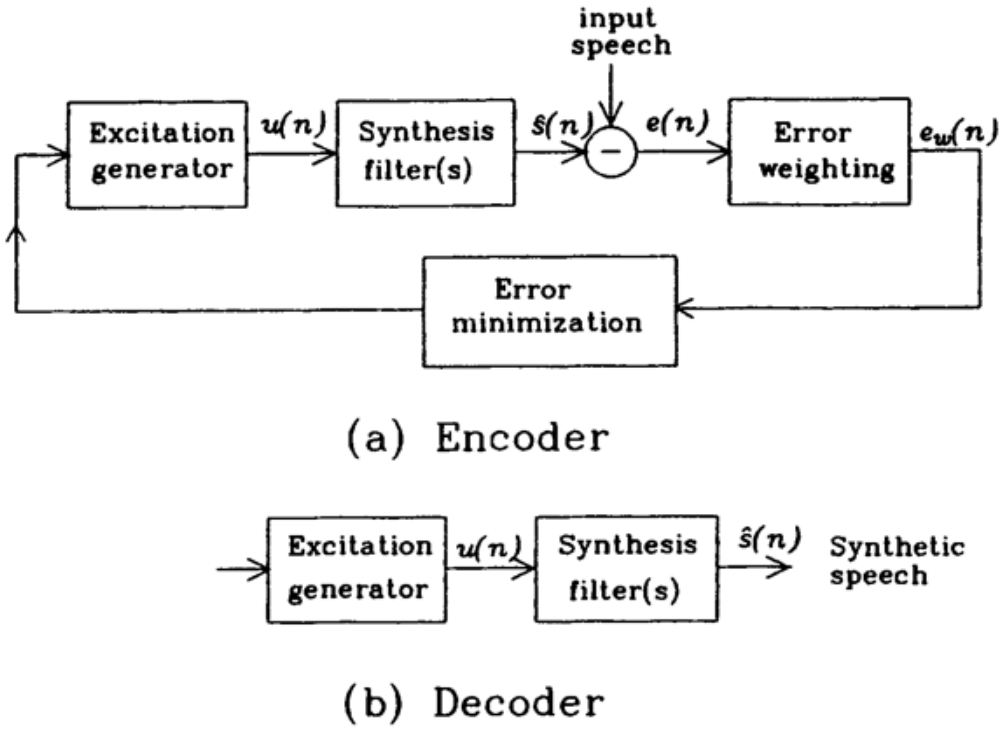
3.1.3 Phương pháp mã hóa lai (Hybrid)

Mã hóa dạng sóng nói chung không cho phép đạt chất lượng tiếng nói tốt ở tốc độ bit dưới 16Kbps.Mặt khác mã hóa vocoder có thể đạt được tốc đ ộ bit rất thấp ,tuy nhiên phương pháp này tổng hợp lại tiếng nói nên có nhược điểm là rất khó nh ận diện được người nói và thường xuyên g ặp vấn đề với nhiễu nền .Mã hóa lai cố gắng t ận dụng ưu điểm của cả hai phương pháp điều chế trên.Nó mã hóa tiếng nói ở tốc độ thấp, mà lại cho kết quả tiếng nói tái tạo lại tốt ,có thể nh ận dạng được người nói . Băng thông yêu cầu thường nằm trong khoảng 4.8 kbps đến 16kbps.

Vấn đề cơ bản đối với Vocoder là nguồn kích thích được mô phỏng một cách đơn giản:tín hiệu tiếng nói được coi là vô thanh hay hữu thanh ,nó làm cho tiếng nói nhận được có dạng được nhân tạo hơn là vẻ tự nhiên.Các phương pháp mã hóa lai có gắng cải thiện điều này bằng cách thay đổi nguồn kích thích tiếng nói theo các cách khác .

Mã hoá lai phổ biến nhất là mã hoá phân tích bằng tổng hợp AbS (Analysis by Synthesis), RPE-LTP, CELP, ACELP, CS-CELP, … Hầu hết các tiêu chuẩn mã hoá tiếng nói trong liên lạc di động đều sử dụng mã hoá kết hợp mã hoá lai AbS. Do đó, phần này sẽ trình bày chi tiết mã hoá lai AbS.

3.1.3.1 Mã hoá phân tích AbS



Hình 3.1 Mô hình chung bộ mã hoá phân tích bằng tổng hợp AbS

Cấu trúc cơ bản của mô hình chung bộ mã hoá tiếng nói phân tích bằng tổng hợp AbS được mô tả như hình 2.1. Mô hình trên bao gồm ba phần chính. Phần đầu tiên là bộ lọc tổng hợp, thường được gọi là bộ lọc tương quan ngắn hạn bởi các hệ số được tính ra dựa trên dự đoán một mẫu tiếng nói bằng các mẫu tiếng nói trước đó (thường là 8 đến 16 mẫu, do đó gọi là ngắn hạn). Bộ lọc tổng hợp cũng có thể là bộ lọc tương quan dài hạn nối tầng bộ lọc tương quan ngắn hạn. Các đoạn tiếng nói hữu thanh có dạng sóng tuần hoàn và sự tuần hoàn này có thể được khai thác để trợ giúp cho quá trình dự đoán tiếng nói. Cũng như các bộ dự đoán ngắn hạn là các bộ dự đoán tuyến tính nhưng trong khi bộ dự đoán ngắn hạn thực hiện việc dự đoán dựa trên các mẫu kề

nhau trước đó thì bộ dự đoán dài hạn dựa trên các mẫu từ một hay nhiều chu kì pitch trước đó (do đó, gọi là dài hạn). Phần thứ hai của mô hình là bộ tạo xung kích thích, tạo ra chuỗi kích thích đưa vào bộ lọc tổng hợp để tạo ra tiếng nói tái tạo bên phía thu. Cuối cùng là bộ giảm thiểu sai số cung cấp thông tin cần thiết cho bộ tạo tín hiệu kích thích. Trong phần sau, ta sẽ trình bày về bộ lọc tổng hợp LPC và tổng hợp pitch cũng như cách tính toán các thông số.

a, Dự đoán ngắn hạn STP (Short Term Predictor)

Dự đoán ngắn hạn mô hình hoá đường bao phổ ngắn hạn của tiếng nói. Đường bao phổ ngắn hạn của đoạn tiếng nói có độ dài L mẫu có thể được mô hình hoá bởi bộ lọc số toàn điểm cực có dạng sau:

$$H(z) = \frac{1}{1 - P_s(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{3.1}$$

với $$P_s(z) = \sum_{k=1}^{p} a_k z^{-k} \tag{3.2}$$

là bộ dự đoán ngắn hạn. Trong đó, các hệ số a_k được tính toán theo phương pháp dự đoán tuyến tính (LP). Tập các hệ số a_k được gọi là các tham số LPC hay còn gọi là các hệ số dự đoán, p là số lượng các hệ số dự đoán hay còn gọi là bậc dự đoán. Như vậy, ý tưởng của phân tích tuyến tính là các mẫu tiếng nói có thể xấp xỉ bằng tổ hợp tuyến tính của các mẫu tiếng nói trong quá khứ (8-16 mẫu)

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n - k) \tag{3.3}$$

Trong đó, s(n) là mẫu tiếng nói tại thời điểm lấy mẫu n, $\tilde{s}(n)$ là mẫu tiếng nói dự đoán tại thời điểm n. Sai số giữa giá trị dự đoán và giá trị thực e(n) là:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n - k) \tag{3.4}$$

Biến đổi Z biểu thức (3.4) ta có:

$$E(z) = S(z) - \sum_{k=1}^{p} a_k S(z) z^{-k} = S(z)\left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) = S(z)A(z) \tag{3.5}$$

với $$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{3.6}$$

là nghịch đảo của H(z). Vì vậy, A(z) được gọi là bộ lọc đảo.

Các hệ số dự đoána_k được tính bằng cực tiểu hoá sai số bình phương trung bình trên đoạn ngắn (10-20 ms) của dạng sóng tiếng nói.

$$E = \sum_n e^2(n) = \sum_n \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2 \quad (3.7)$$

Để tìm các giá trị a_k mà E cực tiểu, ta đặt $\partial E / \partial a_i = 0$ với i=1, ..., p.

$$\frac{\partial E}{\partial a_i} = \sum_n \left\{ 2 \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right] s(n-i) \right\} = 0 \quad (3.8)$$

$$\Leftrightarrow \sum_n s(n)s(n-i) = \sum_n \sum_{k=1}^{p} a_k s(n-k)s(n-i) \quad (3.9)$$

$$\Leftrightarrow \sum_n s(n)s(n-i) = \sum_{k=1}^{p} a_k \sum_n s(n-k)s(n-i) \quad (3.10)$$

Đặt: $f(i,k) = \sum_n s(n-i)s(n-k)$ (3.11)

(3.10) được biến đổi thành: $\sum_{k=1}^{p} a_k f(i,k) = f(i,0)$, i=1, …, p (3.12)

Có hai phương pháp để thực hiện điều này, đó là phương pháp tự tương quan và phương pháp hiệp phương sai. Phần sau chỉ trình bày về phương pháp tự tương quan.

Phương trình (3.12) được áp dụng chỉ trong trường hợp nếu mô hình tiếng nói là quá trình ngẫu nhiên dừng. Tất nhiên tín hiệu tiếng nói không là như thế trong khoảng dài của thời gian, cho phép tính dừng là xác thực chỉ trong khoảng ngắn tín hi ệu tiếng nói.

Giả sử các đoạn thoại tiến đến 0 khi nằm ngoài giới hạn cho trước $0 \le n \le L-1$, với L là độ dài của khung phân tích STP. Điều này tương đương với nhân tín hiệu tiếng nói đầu vào với cửa sổ w(n) có độ dài hữu hạn và bằng 0 nằm ngoài khoảng trên. Ta xét công thức (3.7) trong khoảng $0 \le n \le L+p-1$:

$$f(i,k) = \sum_{n=0}^{L+p-1} s(n-i)s(n-k) \quad , \quad \begin{matrix} i = 1,...,p \\ k = 1,...,p \end{matrix} \quad (3.11)$$

Đặt $m = n - i$:

$$f(i,k) = \sum_{m=0}^{L-1-(i-k)} s(m)s(m+i-k) \quad (3.12)$$

$f(i,k)$ chính là hàm tự tương quan tín hiệu của s(m) với độ dịch i-k:

$$f(i,k) = R(i-k) \quad (3.13)$$

với $R(j) = \sum_{n=0}^{L-1-j} s(n)s(n+j) = \sum_{n=j}^{L-1} s(n)s(n-j)$ (3.14)

Như vậy, công thức (3.12) có thể viết lại thành:

$$\sum_{k=1}^{p} a_k R(|i-k|) = R(i) \tag{3.15}$$

Biểu diễn dưới dạng ma trận, ta có:

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & R(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & R(p-3) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \\ R(p) \end{bmatrix} \tag{3.16}$$

Do có cấu trúc Toepliz (là ma trận đối xứng), nên phương pháp đệ quy Levinson-Durbin được dùng để giải quyết với giải thuật như sau:

$E(0) = R(0)$

For $i=1$ to p do

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} a_j^{i-1} R(i-j)}{E(i-1)} \tag{3.17}$$

$a_i^i = k_i$

For $j=1$ to $i-1$ do

$$a_j^i = a_j^{(i-1)} - k_i a_{i-j}^{i-1} \tag{3.18}$$

$$E(i) = (1-k_i^2)E(i-1) \tag{3.19}$$

Kết quả cuối cùng của giải thuật: $a_j = a_j^{(p)}$, j = 1, …, p (3.20)

E(i) ở biểu thức (3.19) là lỗi dự đoán của bộ dự đoán bậc i.

k_i là hệ số phản xạ và nằm trong khoảng $-1 \le k_i \le 1$.

Ví dụ cho p=2, khi đó:

$$\begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \end{bmatrix}$$

Đối với i=1:

$E(0) = R(0)$

$$k_1 = \frac{R(1)}{R(0)}$$

$$a_1^{(1)} = k_1 = \frac{R(1)}{R(0)}$$

$$E(1) = (1 - k_1)^2 E(0) = \frac{R^2(0) - R^2(1)}{R(0)}$$

Đối với i=2:

$$k_2 = \frac{R(2) - a_1 R(1)}{E(1)} = \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)}$$

$$a_2^{(2)} = k$$

$$a_1^{(2)} = a_1^{(1)} - k_2 a_1^{(1)} = \frac{R(1)R(0) - R(1)R(2)}{R^2(0) - R^2(1)}$$

Kết quả:

$a_1 = a_1^{(2)}$ và $a_2 = a_2^{(2)}$

Như đã đề cập ở phần trước, các mẫu tiếng nói *s(n)* bằng 0 nằm ngoài đoạn $0 \le n \le L - 1$. Sự cắt xén thình lình của các khung tiếng nói có khả năng tạo ra sự thay đổi lớn trong lỗi dự đoán tại điểm bắt đầu và kết thúc của khung tiếng nói được phân tích. Vấn đề này được giải quyết bằng cách sử dụng cửa sổ Hamming, có tác động thu hẹp đối với các rìa của một khối trong khi nó không có tác động nào trong các dải giữa của nó:

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{L - 1}) \quad , 0 \le n \le L - 1 \qquad (3.21)$$

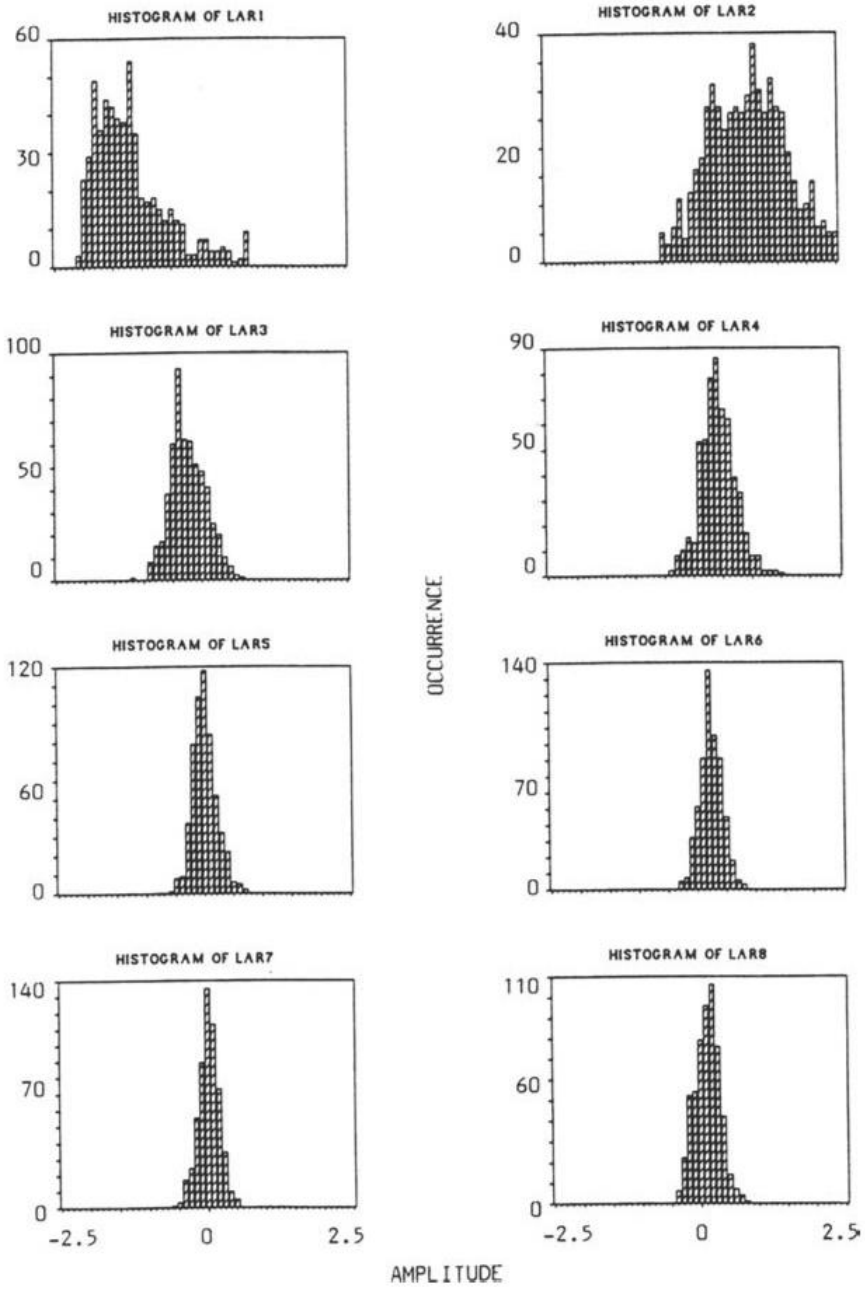
L là độ dài khung phân tích LPC. Độ dài của cửa sổ Hamming được sử dụng thường dài hơn độ dài của khung thoại. Các cửa sổ chồng lên nhau sẽ tạo hiệu ứng mượt trong phân tích LPC, có nghĩa là sẽ làm giảm sự thay đổi đột ngột các hệ số phân tích LPC giữa các khung được phân tích.

Hệ số phản xạ: Trong thực tế, các hệ số dự đoán a_k không được tính toán trực tiếp. Thay vào đó, một số hệ số phản xạ được tính từ các hệ số tự tương quan của khối tiếng nói. Các hệ số phản xạ k_i thu được trong quá trình giải công thức (3.12) bằng giải thuật Levinson-Durbin. Khi $|k_i|$ tiến đến 1 thì các điểm cực của hàm truyền *H(z)* cũng tiến đến vòng tròn đơn vị. Sự thay đổi nhỏ về k_i dẫn đến sự thay đổi lớn về phổ. Do đó, các hệ số phản xạ được biến đổi thành tập các hệ số khác gọi là các tỷ số vùng logarit LAR. Vì các tỷ số vùng logarit LAR được nén giản theo luật logarit có các tính chất lượng tử tốt hơn các hệ số k_i.

$$LAR(i) = \log\frac{1 - k_i}{1 + k_i} \qquad (3.22)$$

Hàm mật độ xác suất (PDF) các tham số LAR của bộ lọc bậc tám được trình bày như hình 2.2. Ta thấy rằng dải động của các tham số LAR(i) giảm khi i tăng. Do đó,

các bit được ấn định cho các tham số LAR càng nhiều khi bậc của LAR càng nhỏ. Điều này, lý giải trong trường hợp lượng tử hoá 8 LAR trên khối 20 ms tiếng nói bằng 6 bit cho LAR(1) và LAR(2), 5 bit cho LAR(3) và LAR(4), 4 bit cho LAR(5) và LAR(6) và 3 bit cho LAR(7) và LAR(8).

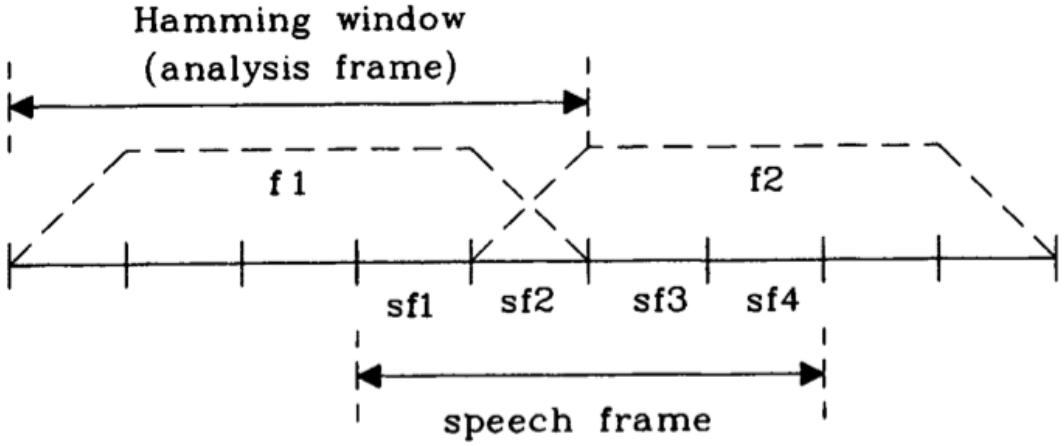


Hình 3.2 Đồ thị hàm mật độ xác suất của 8 hệ số LAR đầu tiên

Nội suy các tham số LPC: Như đã nói ở phần trước, độ dài khung kích thích thường nhỏ hơn độ dài khung LPC. Khung LPC được chia thành nhiều khung con, và

các tham số kích thích được cập nhật ở mỗi khung con này. Hình 2.3 sẽ chỉ ra mối quan hệ giữa khung, khung con, và cửa sổ Hamming được sử dụng để tính ra các tham số LPC.

Mỗi khung thoại bao gồm 160 mẫu (20 ms), khung con gồm 40 mẫu (5 ms) và cửa sổ Hamming gồm 200 mẫu (25 ms). Trong ví dụ này, các tham số LPC sẽ được truyền đi mỗi 20 ms. Để làm giảm bớt các thay đổi đột biến trong bản chất đường bao tín hiệu tiếng nói quanh rìa khung phân tích LPC, nội suy của các tham số LPC giữa các khung kế cận nhau được sử dụng để thu được các thông số cho mỗi khung con, bằng cách cập nhật chúng mỗi 5 ms trong khi truyền chúng mỗi 20 ms.



Hình 3.3 Mối quan hệ giữa khung, khung con và cửa sổ Hamming

Các hệ số dự đoán $a_i$không được sử dụng trong nội suy, bởi các tham số nội suy trong trường hợp này không đảm bảo cho bộ lọc tổng hợp được ổn định. Nội suy được sử dụng để biến đổi các tham số ở các bộ lọc cần sự ổn định, ví dụ như LARs.

Gọi f_n là các tham số LPC trong khung hiện tại, f_{n-1} là các tham số ở khung kề trước đó, thì tham số LPC được nội suy sf_k tại khung con k được tính như sau:

$$sf_k = d_k f_{n-1} + (1 - d_k) f_n \qquad (3.23)$$

với δ_k thuộc đoạn [0,1], $\delta_k$giảm dần theo chỉ số của khung con.

Ở ví dụ dưới đây, $\delta_k = 0.75, 0.5, 0.25$ và 0 tương ứng với $k= 1, ..., 4$. Với những giá trị này, tham số LPC được nội suy trong bốn khung con như sau:

$$sf_1 = 0.75 f_{n-1} + 0.25 f_n$$

$$sf_2 = 0.5 f_{n-1} + 0.5 f_n$$

$$sf_3 = 0.25 f_{n-1} + 0.75 f_n$$

$sf_4 = f_n$

b, Dự đoán dài hạn LTP (Long Term Predictor)

Lọc tiếng nói bằng b ộ lọc đảo *A(z)*có xu hướng loại bỏ nhiều đ ộ thừa bằng cách trừ mỗi mẫu tiếng nói m ột giá trị dự đoán của nó dùng bởi *p*mẫu trong quá khứ. Tín hiệu nhận được được gọi là dư thừa dự đoán ngắn hạn và nói chung nó sẽ có lượng chu kì nhất định liên quan đến chu kì pitch của tiếng nói gốc khi nó được phát âm . Tính chu kì này thể hi ện mức dư thừa nữa mà ta có thể loại bỏ bằng b ộ dự đoán pitch hay còn gọi là bộ dự đoán dài hạn. Dạng tổng quát của bộ lọc dự đoán dài hạn như sau:

$$\frac{1}{P(z)} = \frac{1}{1 - P_l(z)} = \frac{1}{1 - \sum_{k=-m_1}^{m_2} G_k z^{-(a+k)}} \tag{3.24}$$

Trong đó:

$$P_l(z) = \sum_{k=-m_1}^{m_2} G_k z^{-(a+k)} \tag{3.25}$$

là bộ dự đoán dài hạn; m_1,m_2 xác định số điểm trích bộ dự đoán; α là chu kì pitch hay gọi là độ trễ LTP và G_k là hệ số khuếch đại LTP. Các tham số αvà G_kđược xác định bằng cực tiểu hoá sai số còn dư bình phương trung bình sau khi dự đoán dài hạn và ngắn hạn trên chu kì N mẫu. Đối với dự đoán 1 điểm trích, sai số dự đoán LTP *e(n)* được cho bởi:

$$e(n) = r(n) - Gr(n - a) \tag{3.26}$$

ở đây, r(n) là phần dư tạo nên sau dự đoán ngắn hạn. Phần dư bình phương trung bình E là:

$$E = \sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} \left[r(n) - Gr\left(n - a\right) \right]^2 \tag{3.27}$$

$\P E / \P G = 0$nên:

$$G = \frac{\sum_{n=0}^{N-1} r(n) r(n - a)}{\sum_{n=0}^{N-1} \left[r(n - a) \right]^2} \tag{3.28}$$

Thế G vào(3.27), ta có

$$E = \sum_{n=0}^{N-1} r^2(n) - \frac{\left[\sum_{n=0}^{N-1} r(n)r(n-a)\right]^2}{\sum_{n=0}^{N-1}\left[r(n-a)\right]^2} \quad (3.29)$$

Cực tiểu sai số E, tức là tối đa biểu thức thứ hai ở vế phải đa thức (3.29). Nghĩa là cực đại hoá tương quan chéo giữa STP dư *r(n)* hiện tại và phiên bản trễ của nó. Giá trị α được chọn là giá trị lớn nhất.

Sự ổn định của bộ lọc tổng hợp pitch *1/P(z)* không phải lúc nào cũng ổn định. Đối với dự đoán 1 điểm trích, điều kiện ổn định là $|G| \le 1$. Do đó, để bảo đảm tính ổn định của bộ lọc thì đặt $|G| = 1$ khi $|G| \ge 1$.

3.2. Ứng dụng các phương pháp cơ sở mã hóa âm thanh trong truyền thông.

3.2.1 . Các yêu cầu đối với một bộ mã hóa âm thoại

Trong hầu hết các bộ mã hóa âm thoại, tín hiệu được xây dựng lại sẽ khác với tín hiệu nguyên thủy. Nguyên nhân là do khi cố gắng làm tăng chất lương âm thoại sẽ dẫn đến việc làm giảm các đặc tính tốt khác của hệ thống. Các yêu cầu lý tưởng của một bộ mã hóa thoại bao gồm:

Tốc độ bit thấp: đối với chuỗi bit mã hóa có tốc bit tỉ lệ thuận với băng thông cần cho truyền dữ liệu. Tốc độ bit thấp sẽ làm tăng hiệu suất của hệ thống. Tuy nhiên yêu cầu này lại xung đột với các các đặc tính tốt khác của hệ thống như chất lượng âm thoại. Tốc độ thoại càng cao thì đòi hỏi tốc độ bit càng cao, để bảo đảm âm thoại tại phía nhận được phát ra với tốc độ bằng với tốc độ của môt người bình thường nói chuyện lưu loát.

Chất lượng thoại cao : tín hiệu âm thoại đã giải mã phải có chất lượng có thể chấp nhận được đối với ứng dụng cần đạt . Có rất nhiều khía cạnh về mặt chất lượng bao gồm tính dễ hiểu , tự nhiên, dễ nghe và cũng như có thể nhận dạng người nói là nam hay nữ, già hay trẻ, …

Cường độ mạnh ở trong kênh truyền nhiễu : đây là yếu tố quan trọng đối với các hệ thống truyền thông số với các nhiễu ảnh hưởng mạnh đến chất lượng của tín hiệu thoại.

Kích thước bộ nhớ thấp và độ phức tạp tính toán thấp : nhằm mục đích sử dụng được bộ mã hóa âm thoại trong thực tế. Chi phí thực hiện liên quan đến việc triển khai hệ thống phải thấp , bao gồm cả chi phí cho bộ nhớ cần thiết để hỗ trợ khi hệ thống hoạt động cũng như các yêu cầu tính toán.

Độ trễ mã hóa thấp : trong quá trình xử lý mã hóa và giải mã thoại , độ trễ tín hiệu luôn luôn tồn tại . Việc trễ quá mức sẽ sinh ra nhiề u vấn đề trong vi ệc thực hi ện trao đổi tiếng nói hai chiều trong thời gian thực.

Khả năng cắt bỏ khoảng l ặng: khi nói chuyện không phải âm thoại đư ợc phát ra liên tục mà có những khoảng l ặng. Đó là những lúc đừng lại lấy hơi h ay là lúc nghe người khác nói. Những khoảng lặng này nếu có thể được nhận ra và cắt bỏ có thể giúp làm giảm tốc độ bit hệ thống mã hóa âm thoại.

3.2.2. Các tham số liên quan đến chất lượng thoại

Các tham số truyền dẫn cơ bản liên quan đến chất lượng thoại là:

- Tham số đánh giá cường đ ộ âm lượng /tổn hao tổng thể (OLR-Overall Loudness Rating)
- Trễ: thời gian truyền dẫn tín hi ệu giữa hai đầu cuối gây ra những khó khăn trong vi ệc hội thoại. Trễ bao gồm: trễ chuyển mã thoại, trễ mã hóa kênh, trễ mạng và trễ xử lý tín hiệu thoại để loại bỏ tiếng vọng và giảm nhiễu ở chế độ Handsfree.
- Tiếng vọng (echo).
- Cắt ngưỡng (clipping): là hiện tượng mất phần đầu ho ặc phần cuố i của cụm tín hi ệu thoại, do quá trình xử lý khoảng lặng bị sai.
- Các tính chất liên quan đến độ nhạy tần số.
- Xuyên âm (sidetone loss).
- Nhiễu nền...

3.2.3. Các phương pháp đánh giá chất lượng thoại cơ bản

Việc đánh giá chất lượng thoại trong mạng có thể được thực hi ện bằng cách đánh giá các tham số truyền dẫn có ảnh hưởng đến chất lượng thoại và xác định tác đ ộng của các tham số này đối với chất lượng tổng thể . Tuy nhiên, việc đánh giá từng tham số rất phức tạp và tốn kém . Hiện nay, việc đánh giá chất lượng thoại được dựa trên một tham số chất lượng tổng thể là MOS (Mean Opinion Score). Những phương pháp sử dụng MOS đều mang tính chất chủ quan do chúng phụ t huộc vào quan điểm của người sử dụng dịch vụ. Tuy vậy, chúng ta có thể phân chia các phương pháp đánh giá chất lượng thoại ra làm hai loại cơ bản:

- Các phương pháp đánh giá chủ quan : việc đánh giá theo quan điểm của người sử dụng về mức chất lượng được thực hiện trong thời gian thực.
- Các phương pháp đánh giá khách quan : sử dụng m ột số mô hình để ước lượng mức chất lượng theo thang điểm MOS.

3.2.3.1. Phương pháp đánh giá chủ quan (MOS)

Kỹ thuật này đánh giá chất lượng thoại sử dụng đối tượng là một số lượng lớn người nghe, sử dụng phương pháp thống kê để tính điểm chất lượng. Điểm đánh giá bình quân của nhiều người được tính là điểm Mean Opinion Scoring (MOS).Phương thức đánh giá theo MOS có thể được thực hiện theo các bài kiểm tra hội thoại hai chiều hoặc bài nghe một chiều. Các bài kiểm tra nghe một chiều sử dụng các mẫu thoại chuẩn. Người nghe nghe mẫu truyền qua một hệ thống và đánh giá chất lượng tổng thể của mẫu dựa trên thang điểm cho trước.

3.2.3.2. Các phương pháp đánh giá khách quan

- Các phương pháp so sánh: dựa trên việc so sánh tín hiệu thoại truyền dẫn với một tín hiệu chuẩn đã biết. Tín hiệu dùng để so sánh cũng có thể dùng chính tín hiệu âm thoại đầu vào. So sánh có thể dựa trên dạng sóng âm thanh của hai tín hiệu hoặc so sánh đựa trên các thông số đặc trưng cho âm thoại.
- Các phương pháp ước lượng tuyệt đối: dựa trên việc ước lượng tuyệt đối chất lượng tín hiệu thoại.
- Các mô hình đánh giá truyền dẫn : phương pháp này xác định giá trị chất lượng thoại mong muốn dựa trên những hiểu biết về mạng. Ví dụ: mô hình ETSI Model.

CHƯƠNG 4: MÃ HOÁ VÀ GIẢI MÃ TIẾNG NÓI TRONG HỆ THỐNG GSM

4.1 Các bộ mã hoá tiếng nói dự tuyển cho hệ thống GSM

Việc chọn bộ mã hoá và giải mã tiếng nói (speech codec) thích hợp nhất cho hệ thống GSM từ một tập các bộ mã hoá dự truyển đã được dựa trên các phép thử so sánh khái quát giữa một loạt các điều kiện hoạt động. Các so sánh khắt khe về chất lượng tiếng nói, sức kháng lỗi kênh, độ trễ hệ thống cũng như độ phức tạp.

4.1.1 SBC- APCM

SBC-APCM là codec mã hoá băng con với PCM thích nghi theo khối. Codec này sử dụng các bộ lọc gương cầu phương QMF () để phân tách tín hiệu lối vào thành 16 băng con rộng 250 Hz, hai băng cao nhất trong số đó không được truyền đi. Ấn định bit thích nghi đã được sử dụng trong các băng con trên cơ sở tỷ lệ công suất của một loạt băng tạo thành nên thông tin biên cần truyền đi. Tốc độ truyền dẫn tổng cộng của các tín hiệu băng con là 10 kbps, thông tin biên là 3kbps mà chúng được bảo vệ bởi độ dư thừa 3kbps của mã sửa lỗi hướng đi FEC (Forward Error Correction).

4.1.2 SBC-ADPCM

SBC-ADPCM là codec mã hoá băng con với PCM delta thích nghi. Trong sơ đồ này, tiếng nói lối vào đã được chia thành 8 băng con, trong số đó chỉ có 6 băng được truyền đi. Các tín hiệu băng con đã được mã hoá bằng mã vi sai với đánh giá ngược và thích nghi để đổi lại với SBC-APCM đã được đề nghị, trong đó đánh giá thuận và thích nghi đã được sử dụng. Ấn định bit của các băng con được đặt cố định, do vậy không có thông tin biên nào được truyền đi, nhờ đó làm cho hệ thống thích nghi với tạp nhiễu nhiều hơn và thế không cần mã FEC. Tốc độ mã của codec này chỉ 15 kbps.

4.1.3 MPE-LTP

MPE-LTP (Multi-Pulse Excited LPC codec with Long Term Predictor) là codec dự doán tuyến tính kích thích đa xung với bộ dự đoán dài hạn. Việc thực bộ mã hoá và giải mã tiếng nói cụ thể được sử dụng trong thử nghiệm đểso sánh đòi hỏi tốc độ truyền dẫn 13.2 kbps và mã hoá FEC được gắn vào đó với tốc độ 2.8 kbps nữa đã được sử dụng để bảo vệ các bit quan trọng nhất của bộ mã hoá và giải mã tiếng nói.

4.1.4 RPE-LTP

RPE-LTP (Regular Pulse Excited - Long Term Prediction) là codec LPC kích thích xung đều. Bộ mã hóa tiếng nói này dựa trên nền tảng kích thích xung đều (regular pulse excitation) với dự đoán dài hạn và liên quan tới 2 bộ mã hóa tiếng nói khác là: RELP (Residual Excited Linear Prediction) và MPE -LPC (Multi Pulse Excited LPC). Lợi thế của RELP là không quá phức tạp do sử dụng mã hóa dải tần gốc. Bộ mã hóa MPE-LTP phức tạp hơn nhưng nó cung cấp mức độ hiệu quả cao hơn.

Bộ mã hóa RPE-LTP cho một kết quả khá tốt , cân bằng giữa hi ệu năng và tính phức tạp.

Bốn codec này đã được so sánh với nhau về chất lượng tiếng nói, khả năng kháng tạp nhiễu, các trễ xử lý và độ phức tạp tính toán của chúng. Từ kinh nghiệm với hệ thống tham chiếu điều tần (FM), hai tỷ lệ lỗi bit chỉ tiêu đã được đề nghị mà tại đó các só sánh về chất lượng được thực hiện. Điểm số ý kiến trung bình MOS (Mean Opinion Score) tính trung bình trên một thang điểm 5 trên nhiều điều kiện thử nghiệm khác nhau đã được tìm ra là:

Codec	Bit rate (kbps)	MOS
FM	-	1.95
SBC-APCM	16	3.14
SBC-ADPCM	15	2.92
MPE-LTP	16	3.27
RPE-LPC	13	3.54
RPE-LTP	13	~ 4.0

Bảng 4.1

Các kết quả này đã nhấn mạnh tín vượt trội của các bộ codec kích thích xung và tầm quan trọng của bộ dự đoán dài hạn LTP. Codec RPE, do thể hiện các đặc tính ưa chuộng nhất, đã được cải thiện hơn nữa bằng cách áp dụng một LTP; codec RPE-LTP bảo đảm một MOS bằng khoảng 4.0 điểm trên một dải rộng điều kiện hoạt động.

4.2 Bộ mã hoá tiếng nói RPE-LTP

Sơ đồ bộ mã hoá RPE-LTP được thể hiện như trên hình 4.1. Trong đó, có các bộ phận chức năng sau:

- Tiền xử lý
- Lọc phân tích STP
- Lọc phân tích LTP
- Tính toán RPE

4.2.1 Tiền xử lý

Tín hiệu tiếng nói đã lấy mẫu đầu tiên được cho qua một bộ lọc để loại bỏ bất kì sai lệch DC nào có thể tồn tại rồi cho qua bộ lọc tiền nhấn.

Mô hình toán học của b ộ tạo tiếng nói trong b ộ mã hóa chỉ ra rằng năng lượng suy giảm dần với tần số tăng dần . Do đó, việc tiền nhấn được áp dụng để nâng độ chính xác tính toán bằng cách nhấn phần tần số cao công suất thấp của phổ tiếng nói.

Điều này có thể thực hiện được bằng bộ lọc một cực với hàm truyền dạng:

$$H(z) = 1 - c_1 z^{-1} \quad (4.1)$$

trong đó, $c_1 \sim 0.9$.

4.2.2 Lọc phân tích STP

Tiếng nói đã được tiền nhấn được phân đoạn thành các khối 160 mẫu tương ứng với khoảng thời gian 20 ms trong một bộ đệm.

Đối với mỗi một đoạn gồm L=160 mẫu, chín hệ số tự tương quan được tính từ s(k) theo công thức sau:

$$ACF(i) = \sum_{k=0}^{L-1-i} s(i)s(k+i) \quad , \quad i = 0,1,...,8 \quad (4.2)$$

Từ các hệ số tự tương quan của tiếng nói *ACF*(*i*), tám hệ số phản xạ được tính theo thuật toán lặp Schur, là phương pháp tương đương với thuật toán Levision-Durbin được sử dụng để giải phương trình then chốt LPC để tìm các hệ số phản xạ *r(i)*, cũng như các hệ số lọc STP. Tuy nhiên, thuật toán Schur chỉ đưa đến các hệ số phản xạ *r(i)* mà thôi.

Các hệ số phản xạ *r(i)* được tính nằm trong khoảng

$$-1 \le r(i) \le 1, \quad i = 1,...,8 \quad (4.3)$$

Các hệ số phản xạ *r(i)* được biến đổi thành các tỷ số vùng logarit *LAR(i)*, bởi vì các *LAR(i)* được nén-giản theo luật logarit có các tính chất lượng tử hoá tốt hơn các hệ số *r(i)*.

$$LAR(i) = \lg\frac{1+r(i)}{1-r(i)} \quad , \quad i = 1,...,8 \quad (4.4)$$

Tuy nhiên, để làm đơn giản hoá việc thực thi thời gian thực, một xấp xỉ tuyến tính kiểu từng đoạn với 5 đoạn được sử dụng

$$LAR(i) = \begin{cases} r(i) & ;|r(i)| < 0.675 \\ sign[r(i)].[2|r(i)| - 0.675] & ;0.675 \le |r(i)| < 0.950 \\ sign[r(i)].[8|r(i)| - 6.375] & ;0.950 \le |r(i)| \le 1 \end{cases} \quad (4.5)$$

Các tham số lọc *LAR(i)*, $i = 1, 2, \ldots, 8$ có các dải động khác nhau và các hàm mật độ xác suất có hình dáng khác nhau. Điều này lý giải cho việc mã hoá các cặp *LAR* thứ nhất, thứ hai, thứ ba, thứ tư tương ứng với 6 bit, 5 bit, 4 bit, 3 bit.

$$LAR_c(i) = Nint\{A(i).LAR(i) + B(i)\} \quad (4.6)$$

với $Nint(z) = int\{z + sign(z).0,5\}$ (4.6a)

Trong đó, hàm *Nint(z)* được định nghĩa là giá trị nguyên gần nhất của z và các hệ số *A(i), B(i)* cùng với các giá trị $LAR_c(i)$ tương ứng với *LAR(i)* được cho theo bảng 3.2.

LAR No i	A(i)	B(i)	Min $LAR_c(i)$	Max $LAR_c(i)$
1	20.000	0.000	-32	+31
2	20.000	0.000	-32	+31
3	20.000	4.000	-16	+15
4	20.000	-5.000	-16	+15
5	13.637	0.184	- 8	+ 7
6	15.000	-3.500	- 8	+ 7
7	8.334	-0.666	- 4	+ 3
8	8.824	-2.235	- 4	+ 3

Bảng 4.2 Lượng tử các hệ số $LAR_c(i)$

Các hệ số $LAR_c(i)$ được biến đổi về *LAR''(i)* như sau:

$$LAR''(i) = \frac{|LAR_c(i) - B(i)|}{A(i)} \qquad (4.7)$$

Để làm giảm bớt các thay đổi đột biến trong bản chất đường bao tín hiệu tiếng nói quanh các rìa khung phân tích STP, các tham số *LAR''* được nội suy tuyến tính thành *LAR'*. Trong mỗi khối chứa 160 mẫu tiếng nói, bộ lọc phân tích ngắn hạn được thực hiện với 4 chuỗi tham số khác nhau theo bảng 4.3:

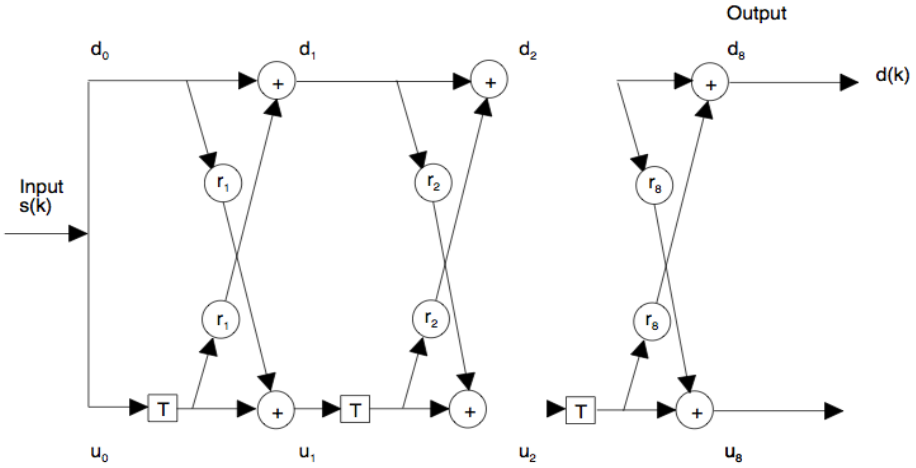
k	$LAR'_J(i)$ =
0… 12	$0.75*LAR''_{J-1}(i) + 0.25*LAR''_J(i)$
13…26	$0.50*LAR''_{J-1}(i)+0.50*LAR''_J(i)$
27…39	$0.25*LAR''_{J-1}(i) +0.75*LAR''_J(i)$
40..159	$LAR''_J(i)$

Bảng 4.3 Nội suy các tham số LAR (J=khối hiện tại)

Các hệ số phản xạ *r'(i)* được giải mã tại chỗ được tính bằng cách biến đổi *LAR'(i)* thành *r'(i)* như sau:

$$r'(i) = \begin{cases} LAR'(i) & ;|LAR'(i)| < 0.675 \\ sign[LAR'(i)].[0.005|LAR'(i)| + 0.337500] & ;0.675 \le |LAR'(i)| < 1.225 \\ sign[LAR'(i)].[0.125|LAR'(i)| + 0.796875] & ;1.225 \le |LAR'(i)| \le 1.625 \end{cases} \qquad (4.8)$$

Các hệ số phản xạ *r'(i)* được dùng để tính STP dư *d(k)*bằng bộ lọc phân tích ngắn hạn có cấu trúc mắt cáo được mô tả như hình 4.2.



Hình 4.2 Bộ lọc phân tích ngắn hạn

$d_0(k) = s(k)$ (4.8a)

$u_0(k) = s(k)$ (4.8b)

$d_i(k) = d_{i-1}(k) + r'_i \cdot u_{i-1}(k-1)$, i=1, …, 8 (4.8c)

$u_i(k) = u_{i-1}(k-1) + r'_i \cdot d_{i-1}(k)$ (4.8d)

$d(k)=d_8(k)$ (4.8e)

4.2.3 Lọc phân tích LTP

Tín hiệu STP dư từ việc lọc ngắn hạn có độ dài 160 mẫu, tương ứng với 20 ms được phân chia thành 4 đoạn con chứa 40 mẫu tương ứng với 5 ms.

Ta kí hiệu:

j = 0, …, 3 là số thứ tự đoạn con

$d(k_j+k)$ là tín hiệu dư thừa mỗi đoạn

với j = 0, …, 3; $k_j = k_0 + j.40$ (k_0 là giá trị đầu tiên của khung chứa 160 mẫu)và k = 0, …, 39

Sai số dự đoán LTP được tối thiểu hoá bởi độ trễ λ mà nó cực đại hoá tương quan chéo giữa STP dư hiện tại và giá trị của nó đã nhận được và được nhớ đệm với độ trễ λ. Cụ thể, STP dư có độ dài L =160 mẫu được chia thành bốn đoạn con với độ dài N=40 mẫu và đối với mỗi đoạn con thì tham số khuếch đại (gain) và độ trễ (lag) cho bộ lọc dự đoán dài hạn LTP được xác định bằng cách tính tương quan chéo giữa đoạn hiện đang xử lý và một đoạn dài 40 mẫu được trượt đi một cách liên tục của đoạn STP dư dài 120 mẫu trước đó.

$$R_j(l) = \sum_{i=0}^{39} d(k_j + i) \cdot d'(k_j + i - l) \quad , \quad \begin{matrix} j = 0,...,3 \\ k_j = k_0 + j.40 \\ l = 40,...,120 \end{matrix} \tag{4.9}$$

Giá trị tương quan lớn nhất được tìm thấy tại độ trễ $\lambda = N_j$ mà tại đó đoạn con hiện đang xử lý giống nhất với quá khứ của mình. Điều này có khả năng đúng với chu kì pitch hoặc tại bội của chu kì pitch. Do đó, hầu hết độ dư thừa có thể tách ra khỏi STP dư.

$$R_j(N_j) = \max\{R_j(l); l = 40,...,120\} \quad , \quad \text{j = 0, ..., 3} \tag{4.10}$$

Hệ số khuếch đại b_j được tính bằng cách chuẩn hoá hệ số tương quan chéo tại độ trễ N_j.

$$b_j = \frac{R_j(N_j)}{s_j(N_j)} \tag{4.11}$$

$$s_j(N_j) = \sum_{i=0}^{39} d'^2(k_j + i - N_j) \quad , \text{ j = 0, ..., 3} \tag{4.12}$$

Một khi tham số LTP là N_j (độ trễ) và b_j (độ lợi) đã tìm được, chúng được mã hoá thành N_{cj} và b_{cj}.

N_j có giá trị trong đoạn (40, ..., 120) nên chỉ cần dùng 7 bit để mã hoá N_{cj} là đủ.

b_{cj} được mã hoá với 2 bit như sau:

$$b_{cj} = \begin{cases} 0 & b_{cj} \le DLB(0) \\ 1 & DLB(0) < b_{cj} \le DLB(1) \\ 2 & DLB(1) < b_{cj} \le DLB(2) \\ 3 & DLB(2) < b_{cj} \end{cases} \quad , \tag{4.13}$$

Trong đó, *DLB(i)*, (i=0, 1..., 2) là mức quyết định được cho theo bảng 3.4 và b_{cj} là hệ số khuếch đại được mã hoá.

i	Decision Level DLB(i)	Quantizing Level QLB(i)
0	0.2	0.10
1	0.5	0.35
2	0.8	0.65
3		1.00

Bảng 4.4 Bảng lượng tử cho tham số khuếch đại LTP

Các tham số LTP được mã hoá (N_{cj}vàb_{cj}) được giải mã tại chỗ thành cặp (N_j' vàb_j') như sau.

$$N_j' = N_{cj} \tag{4.14}$$

$b_j' = QLB(b_{cj})$, $j=0, ..., 3$

với QLB(i), i=0, ..., 3 là mức lượng tử được tính theo bảng 3.4.

Với các tham số LTP vừa tính được, LTP dư được tính bằng sai lệch giữa STP dư và ước lượng của nó (tính được nhờ sự trợ giúp của các tham số LTP đã giải mã được tại chỗ N_j' vàb_j') như sau:

$$e(k_j+k) = d(k_j+k) - d''(k_j+k) \quad , \quad \begin{matrix} j = 0,...,3 \\ k_j = k_0 + j.40 \\ k = 0,...,39 \end{matrix} \tag{4.15}$$

$$\text{với} \quad d''(k_j+k) = b_j'.d'(k_j+k-N_j') \quad , \begin{matrix} j = 0,...,3 \\ k_j = k_0 + j.40 \\ k = 0,...,39 \end{matrix} \tag{4.16}$$

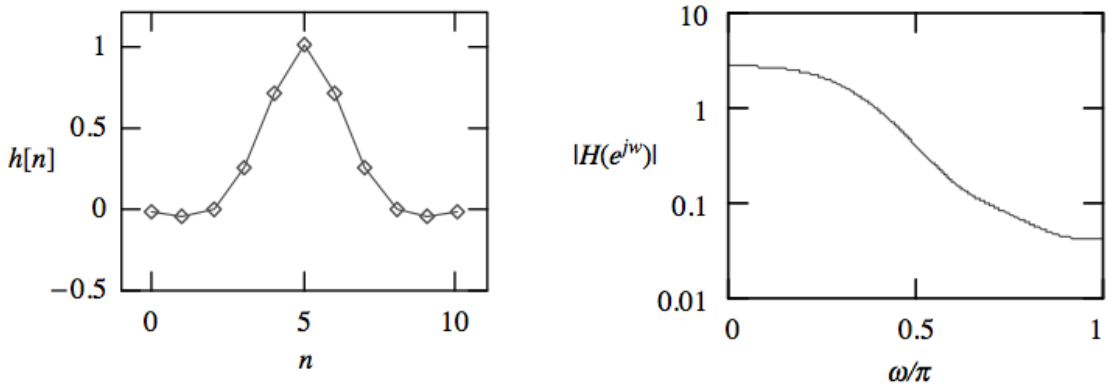
Ở đây, $d'(k_j+k-N_j')$ biểu diễn một đoạn đã biết rồi của quá khứ của $d'(k_j+k)$, được trữ trong bộ nhớ đệm tìm kiếm.

Cuối cùng, nội dung của bộ nhớ đệm tìm kiếm được cập nhật bằng cách sử dụng LTP dư đã được giải mã tại chỗ $e'(k_j+k)$ và STP dư đã được ước lượng $d''(k_j+k)$ để tạo nên $d'(k_j+k)$ như dưới đây:

$$d'(k_j+k) = e'(k_j+k) + d''(k_j+k) \quad , \quad \begin{matrix} j = 0,...,3 \\ k_j = k_0 + j.40 \\ k = 0,...,39 \end{matrix} \tag{4.17}$$

4.2.4 Tính toán RPE

Tín hiệu dư thừa dài hạn được lọc bởi bộ lọc trọng số. Đồ thị đáp ứng xung và đáp ứng tần số như hình 3.3. Bộ lọc trọng số là bộ lọc đáp ứng xung hữu hạn 11 điểm, về cơ bản là một bộ làm trơn, có tác dụng làm trơn sự thay đổi giữa các mẫu, loại bỏ nhiễu tần số cao, và làm cho sự chuyển tiếp giữa các đoạn con trở nên mềm mại hơn. Do đó, chất lượng tiếng nói tổng hợp được cải thiện.



Hình 4.3 Đáp ứng xung (trái) và đáp ứng tần số (phải) của bộ lọc trọng số

Phép chập giữa 40 mẫu trong chuỗi *e(k)* và 11 mẫu trong chuỗi *h(n)* tạo nên 40+11-1=50 mẫu được mô tả theo công thức 3.18, lưu ý rằng ta chỉ tính 40 mẫu của phép chập.

$$x(k) = \sum_{i=1}^{10} h(i) * e(k + 5 - i) \qquad (4.18)$$

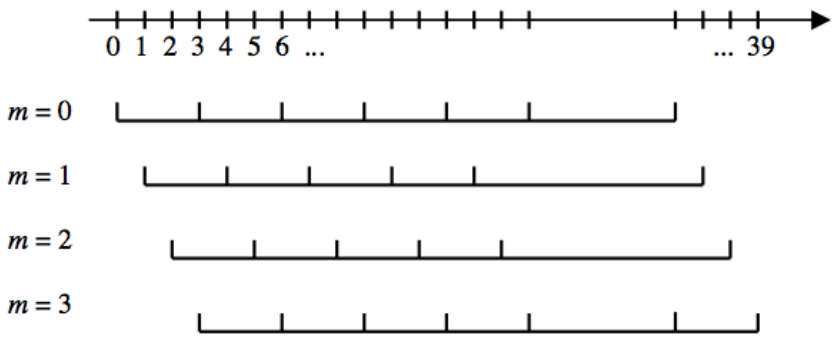
với k= 0, ..., 39

e(k+5-i) = 0 khi k+5-i < 0 hoặc k+5-i > 39

Giai đoạn kích thích xung đều bao gồm vi ệc giảm 40 mẫu dư thừa dài hạn xuống thành 4 bộ chuỗi con 13 bit thông qua sự kết hợp của kỹ thu ật đan xen và chia nhỏ mẫu.

$$x_m(i) = x(k_j + m + 3*i) \quad ; i = 0, \ldots, 12$$
$$m = 0, \ldots, 3 \qquad (4.19)$$

Ta có thể minh hoạ (4.19) bằng hình sau:



Hình 4.4 Vị trí các mẫu trong 4 chuỗi con

Năng lượng của bốn chuỗi con đã được chiết ra sẽ được tính toán, và chuỗi dự tuyển có năng lượng lớn nhất sẽ được chọn để biểu diễn một cách tốt nhất LTP dư.

$$E_M = \max_m \sum_{i=0}^{12} x_m^2(i) \quad ; m = 0, \ldots, 3 \qquad (4.20)$$

Theo 4 vị trí m của lưới ban đầu có thể có, 2 bit là đủ để mã hoá dịch trượt ban đầu của lưới đối với mỗi đoạn con.

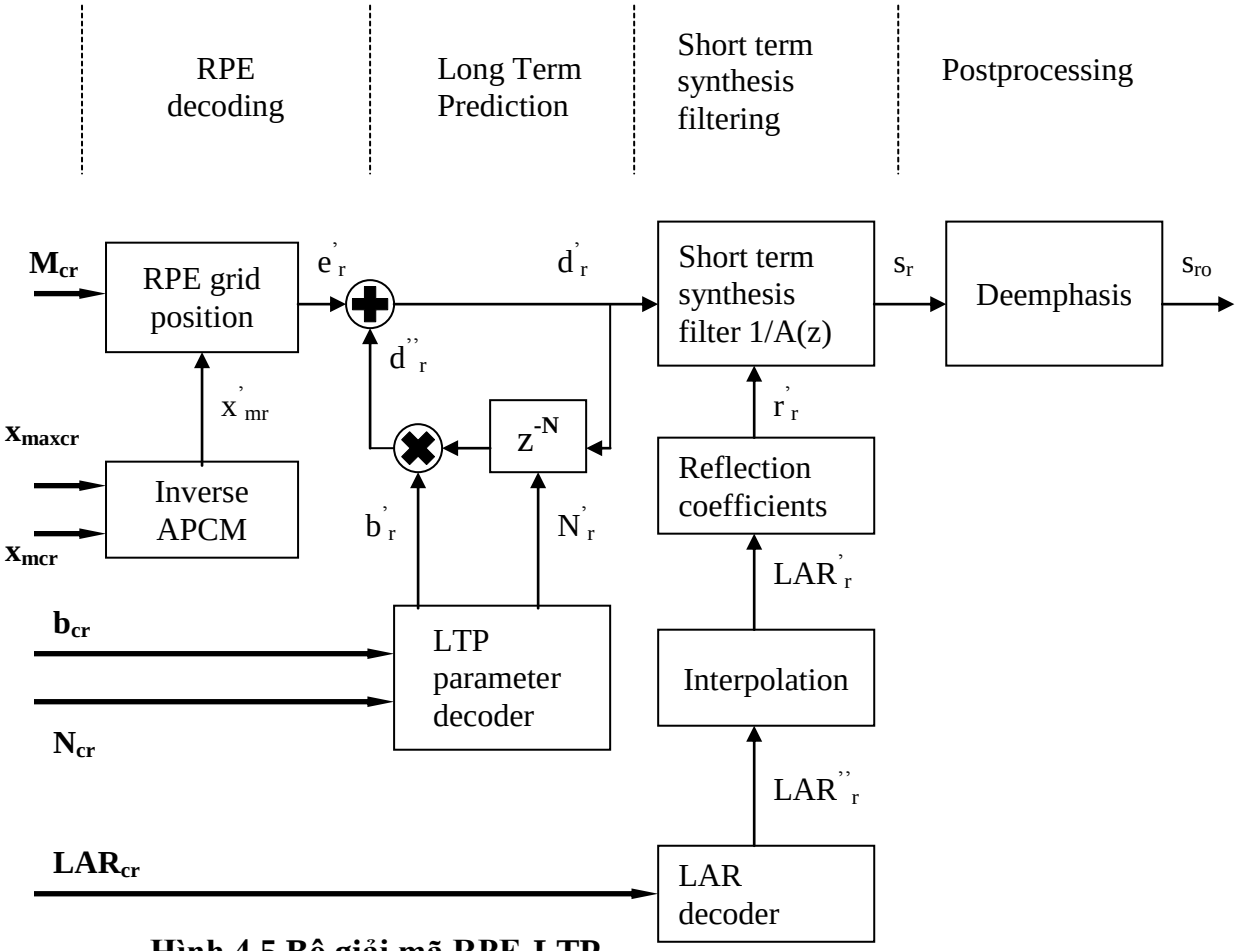
Các biên độ xung được chuẩn hoá theo biên độ cao nhất của khối và được lượng tử hoá bằng 3 bit. Và giá trị cực đại của khối được lượng tử hoá bằng 6 bit.

Các vị trí lưới, biên độ xung và giá trị cực đại của chuỗi được giải mã tại chỗ để cho ra LTP dư *e'(k)*, trong đó các xung bị thiếu trong chuỗi được điền với giá trị 0.

4.3Bộ giải mã tiếng nói RPE-LTP

Sơ đồ khối bộ giải mã RPE-LTP được trình bày trong hình 3.3, thể hiện một cấu trúc ngược hình thành bởi các bộ phận chức năng:

- Giải mã RPE
- Lọc tổng hợp LTP
- Lọc tổng hợp STP
- Hậu xử lý



Hình 4.5 Bộ giải mã RPE-LTP

4.3.1Giải mã RPE

Trong bộ giải mã, lưới vị trí *M*, các giá trị cực đại kích thích của đoạn con và các biên độ xung kích thích được lượng tử nghịch đảo và các biên độ xung kích thích được tính toán bằng cách nhân các biên độ đã giải mã được với các trị cực đại khối tương ứng của chúng. Mô hình LTP dư e'_r đã được tái tạo lại bằng việc định vị chính xác các biên độ xung theo theo lượng dịch *M* ban đầu.

4.3.2Lọc tổng hợp LTP

Đầu tiên, các tham số lọc LTP (khuếch đại b_{cr}và độ trễ N_{cr}) được khôi phục tạo ra b'_r và N'_rvà chúng được dùng để xây dựng bộ lọc tổng hợp LTP. Sau đó, tín hiệu LTP dư đã khôi phục được e_r' được sử dụng để kích thích bộ lọc tổng hợp LTP này để khôi phục một đoạn mới có độ dài *N=40* của STP dư đã được ước lượng d_r''. Để làm vậy, một đoạn trong quá khứ của STP dư đã tái tạo được *d'* được sử dụng, được làm trễ

đúng đi N_r' mẫu và được nhân với b'_r để có được STP dư được ước lượng d''_r. theo 3.16.

Rồi sau đó, d''_r được sử dụng để tính toán đoạn con gần đây nhất của STP dư đã được tái tạo theo 3.17.

4.3.3Lọc tổng hợp STP

Các tham số LAR''$_r$ được giải mã bằng cách sử dụng bộ giải mã LAR từ các LAR''$_{cr}$ mà nó nhận được. Và một lần nữa lại được nội suy tuyến tính về phía các rìa của khung phân tích giữa các tham số của các khung lân cận nhằm tránh các thay đổi đột ngột trong đặc điểm của đường bao phổ tiếng nói. Cuối cùng, tập tham số đã nội suy đã được biến đổi tạo thành các hệ số phản xạ r'$_r$, trong đó tính ổn định của bộ lọc tổng hợp STP được bảo đảm nếu các hệ số phản xạ được khôi phục rơi ra ngoài vòng tròn đơn vị được phản xạ ngược vào trong vòng tròn đơn vị nhờ thực hiện lấy giá trị nghịch đảo của chúng. Công thức biến đổi LAR'$_r$(i) trở lại thành r'$_r$ được cho như sau

$$r_r'(i) = \frac{10^{LAR_r'(i)} - 1}{10^{LAR_r'(i)} + 1} \qquad (4.18)$$

4.3.4Hậu xử lý

Quá trình hậu xử lý được thiết lập bởi việc giải nhấn bằng cách sử dụng bộ lọc H(z) trong biểu thức 3.1.

Như vậy, đối với một khoảng thời gian 20 ms, tương đương với việc mã hoá 160 mẫu, các bit được phân bố trong mã hoá tiếng nói RPE-LTP được trình bày theo bảng 3.5.

Tham số	Tên tham số	Kí hiệu	Số lượng bit	Bit
STP	Log. Area ratios 1 - 8	LAR 1	6	b1-b6
		LAR 2	6	b7-b12
		LAR 3	5	b13-b17
		LAR 4	5	b18-b22
		LAR 5	4	b23-b26
		LAR 6	4	b27-b30
		LAR 7	3	b31-b33
		LAR 8	3	b34-b36
Đoạn con thứ 1				
LTP	Độ trễ LTP	N1	7	b37-b43
	Khuếch đại LTP	b1	2	b44-b45

RPE	Vị trí lưới RPE	M1	2	b46-b47
	Giá trị cực đại khối RPE	Xmax1	6	b48-b53
	Xung RPE thứ 1	x1(0)	3	b54-b56
	Xung RPE thứ 2	x1(1)	3	b57-b59
	...	...		...
	Xung RPE thứ 13	x1(12)	3	b90-b92
Đoạn con thứ 2				
LTP	Độ trễ LTP	N2	7	b93-b99
	Khuếch đại LTP	b2	2	b100-b101
RPE	Vị trí lưới RPE	M2	2	b102-b103
	Giá trị cực đại khối RPE	Xmax2	6	b104-b109
	Xung RPE thứ 1	x2(0)	3	b110-b112
	Xung RPE thứ 2	x2(1)	3	b113-b115
	...	...		...
	Xung RPE thứ 13	x2(12)	3	b146-b148
Đoạn con thứ 3				
LTP	Độ trễ LTP	N3	7	b149-b155
	Khuếch đại LTP	b3	2	b156-b157
RPE	Vị trí lưới RPE	M3	2	b158-b159
	Giá trị cực đại khối RPE	Xmax3	6	b160-b165
	Xung RPE thứ 1	x3(0)	3	b166-b168
	Xung RPE thứ 2	x3(1)	3	b168-b171
	...	...		...
	Xung RPE thứ 13	x3(12)	3	b202-b204
Đoạn con thứ 4				
LTP	Độ trễ LTP	N4	7	b205-b211
	Khuếch đại LTP	b4	2	b212-b213
RPE	Vị trí lưới RPE	M4	2	b214-b215
	Giá trị cực đại khối RPE	Xmax4	6	b216-b221
	Xung RPE thứ 1	x4(0)	3	b222-b224
	Xung RPE thứ 2	x4(1)	3	b225-b227
	...	...		...
	Xung RPE thứ 13	x4(12)	3	b258-b260

Bảng 4.5 Vị trí bit các tham số ngõ ra của bộ mã hoá tiếng nói RPE-LTP trong khung thoại 20ms

Tóm lại, tổng số bit truyền dẫn trong một khung là 36+ 4x(2+7+2+6+13x3) = 260 bit.

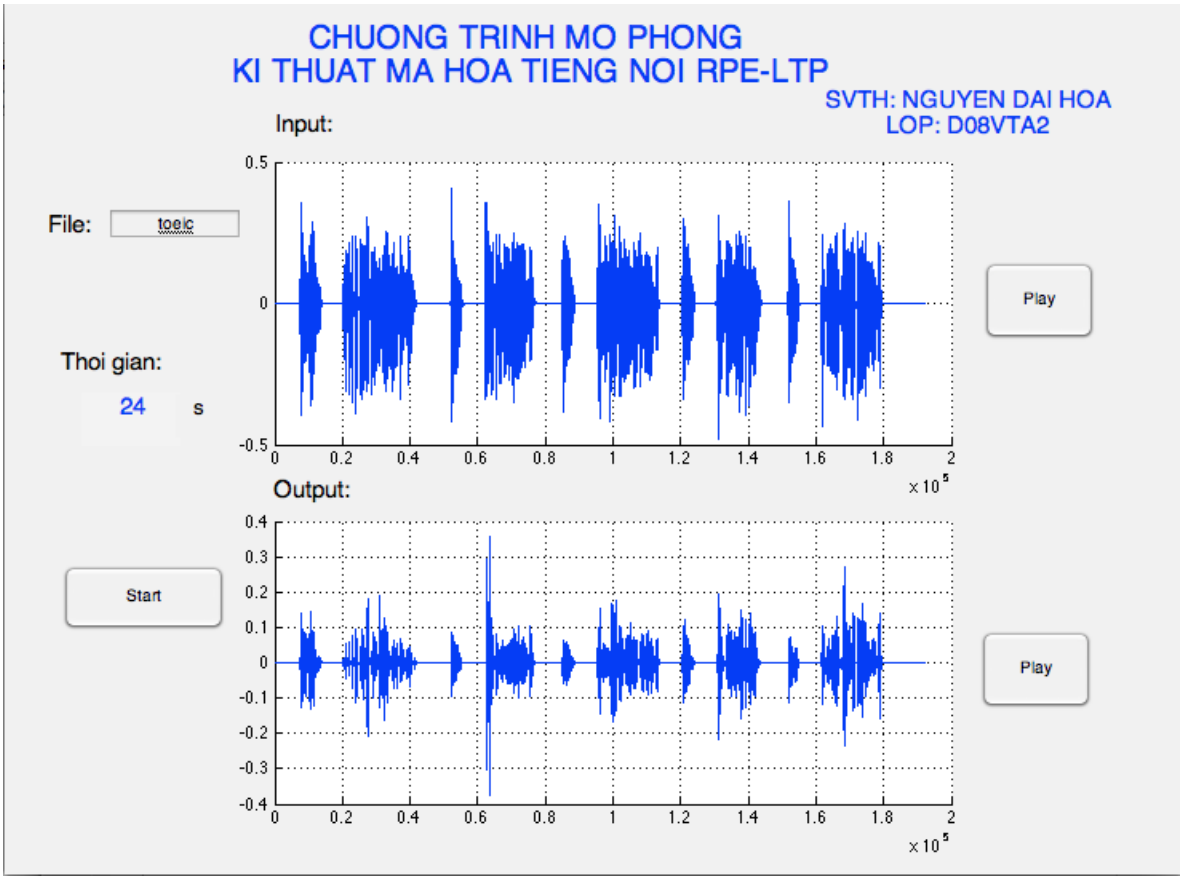
CHƯƠNG 5: MÔ PHỎNG

Matlab là một môi trường tính toán số và lập trình, được thiết kế bởi công ty MathWorks, Inc. Matlab cho phép tính toán số với ma trận, vẽ đồ thịhàm số hay biểu đồ thông tin, thực hiện thuật toán, tạo các giao diện người dùng và liên kết với những chương trình máy tính viết trên nhiều ngôn ngữ lập trình khác.

Chương trình mô phỏng quá trình nén và giải nén tiếng nói được viết trên Matlab, dựa trên kĩ thuật mã hoá RPE-LTP đã trình bày ở chương trước.

Người sử dụng sẽ chọn file tiếng nói được mã hoá PCM 13 bit ở đầu vào. Chương trình sẽ mô phỏng quá trình nén và giải nén, cuối cùng ta sẽ thu được tiếng nói giải nén ở ngõ ra.

So sánh kết quả ngõ vào và ngõ ra ta thấy kết quả chất lượng vẫn đảm bảo tốt. Giao diện chương trình mô phỏng như sau:



Hình 5.1 Giao diện chương trình mô phỏng

Trong đó:

File là tín hiệu tiếng nói ngõ vào.

Thời gian là độ dài thời gian tín hiệu tiếng nói ngõ vào.

Start là nút bắt đầu thực hiện chương trình mã hoá và giải mã tiếng nói.

Sau khi click vào *Start*, đợi một thời gian, ta sẽ thu được đồ thị dạng sóng của tiếng nói ngõ vào và ngõ ra.

Nhấn nút *Play* tương ứng để nghe file tiếng nói ban đầu và file tiếng nói sau khi thực hiện mã hoá và giải mã.

Kết luận

Về căn bản chúng ta có thể thấy bộ mã hoá tiếng nói trong GSM là một bộ mã hoá tiếng nói dạng lai (hybrid) giữa LPC vocoder và mã hoá dạng sóng. Trong đó mô hình lọc từ cấu hình vocoder được giữ nguyên song các tham số kích thích lại được cải thiện. Điều này nghĩa là phần chủ yếu của các tham số được truyền đi liên quan tới chuỗi kích thích. Bộ mã hoá lai đã san được hố ngăn cách giữa các bộ mã hoá vocoder và các bộ mã hoá dạng sóng.

Quy trình mã hoá tiếng nói trong bộ mã hoá tiếng nói có thể tóm tắt lại như sau. Tín hiệu tiếng nói lối vào được chia thành từng khung 20 ms để biến đổi thành tín hiệu số. Các bước cơ bản của quá trình mã hoá bao gồm: Lọc dự đoán tuyến tính LPC, Lọc dự đoán dài hạn LTP và mã hoá kích thích xung đều RPE. Các thông số được mã hoá do vậy cũng bao gồm bit mã của các thông số LPC, LTP và RPE.

Về mặt thực hành, em cũng đã cố gắng mô phỏng được kĩ thuật mã hoá tiếng nói chạy được trên PC. Trước tiên, chương trình sẽ thực hiện nén tín hiệu tiếng nói ở file mẫu có sẵn dưới định dạng .wav bằng codec RPE-LTP. Sau đó, sẽ tổng hợp các thông số lại để tạo thành tín hiệu tiếng nói ở ngõ ra. Với chương trình mô phỏng này, em hy vọng chương trình này phần nào giúp ta có thể hình dung được kĩ thuật mã hoá này.

Em xin cám ơn sự giúp đỡ tận tình của thầy Phạm Thanh Đàm đã hướng dẫn em thực hiện bài báo cáo này. Do thời gian và kiến thức có hạn nên báo cáo thực hiện vẫn còn nhiều thiếu sót, em rất mong sự nhận xét, đánh giá, đóng góp từ thầy cô và bạn bè. Em sẽ cố gắng tìm hiểu thêm. Một lần nữa, em xin chân thành cám ơn.

Tài liệu tham khảo

[1].A. M. Kondoz, "*Digital Speech – Coding for Low Bit Rate Communication Systems, 2nd*", John Wiley & Sons, Ltd, 2004.

[2]. Raymond Steele and Lajos Hanzo, "*Mobile Radio Communication 2nd*", John Wiley & Sons, Ltd, 1992.

[3]. "*GSM 06.10*", ETSI, 1997.

[4]. Randy Goldberg and Lance Riek, "*A Practical Handbook of Speech Coders"*, CRC Press LLC, 2000.

[5]. Wai C. Chu, "*Speech coding algorithms"*, John Wiley & Sons, Ltd, 2003.

[6]. Phạm Thanh Đàm, "Thông tin di động", Học viện Công nghệ Bưu chính Viễn thông Tp.HCM, 2010.

Chữ viết tắt

A/D	Analog to Digital	
AB	Access Burst	Cụm truy xuất
AbS	Analysis by Synthesis	Phân tích bằng tổng hợp
ADPCM	Adaptive Differently PCM	Điều chế mã xung vi sai thích ứng
DB	Dummy Burst	Cụm giả
DM	Delta Modulation	Điều chế Delta
DPCM	Differential PCM	Điều chế mã xung vi sai
FC	Frequency Correction Burst	Cụm điều chỉnh tần số
FEC	Forward Error Correction	Mã sửa lỗi hướng đi
GMSK	Gaussian Minimum Shift Keying	Điều chế khoá chuyển pha cực tiểu
GSM	Global System For Mobile Communications	Hệ thống thông tin di động toàn cầu
LAR	Logarithm Area Ratio	Tỉ số vùng logarith
LP	Linear Prediction	Dự đoán tuyến tính
LPC	Linear Prediction Coding	Mã hoá dự đoán tuyến tính
LTP	Long Term Predictor	Dự đoán dài hạn
MOS	Mean Opinion Score	Điểm số ý kiến trung bình
MPE-LTP	Multi-Pulse Excited LPC Codec with Long term Predictor	Dự đoán tuyến tính kích thích đa xung với bộ dự đoán dài hạn
NB	Normal Burst	Cụm thường
PCM	Pulse Code Modulation	Điều chế xung mã
PDF	Probability Density Function	Hàm mật độ xác suất
QMF	Quadrature Mirror Filter	Bộ lọc gương cầu phương
QoS	Quality of Service	Chất lượng dịch vụ
RELP	Residual Excited Linear Prediction	Dự đoán tuyến tính kích thích bằng tín hiệu sau dự đoán
RPE	Regular Pulse Excitation	Kích thích xung đều
RPE-LTP	Regular Pulse Excited - Long Term Prediction	Kích thích xung đều - Dự đoán dài hạn
SB	Synchronization Burst	Cụm đồng bộ
SNR	Signal to Noise Ratio	Tỉ số tín hiệu trên nhiễu
STP	Short term Predictor	Dự đoán ngắn hạn