

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



DƯƠNG THANH TÙNG

**ĐÁNH GIÁ ĐỊNH LƯỢNG KHẢ NĂNG CHỐNG LẠI
MẤT MÁT GÓI TIN CỦA THUẬT TOÁN MÃ HÓA
ILBC TRONG CÁC HỆ THỐNG THÔNG TIN THOẠI**

LUẬN VĂN THẠC SĨ NGÀNH CÔNG NGHỆ THÔNG TIN

Hà Nội - 2016

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

DƯƠNG THANH TÙNG

**ĐÁNH GIÁ ĐỊNH LƯỢNG KHẢ NĂNG CHỐNG LẠI
MẤT MẮT GÓI TIN CỦA THUẬT TOÁN MÃ HÓA ILBC
TRONG CÁC HỆ THỐNG THÔNG TIN THOẠI**

Ngành: Công nghệ Thông tin

Chuyên ngành: Truyền dữ liệu và Mạng máy tính

Mã số:

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. HOÀNG XUÂN TÙNG

Hà Nội - 2016

LỜI CAM ĐOAN

Tôi xin cam đoan nội dung của luận văn “***Đánh giá định lượng khả năng chống lại mất mát gói tin của thuật toán mã hóa iLBC trong các hệ thống thông tin thoại***” là sản phẩm do tôi thực hiện dưới sự hướng dẫn của Tiến sĩ Hoàng Xuân Tùng. Trong toàn bộ nội dung của luận văn, những điều được trình bày là do tôi nghiên cứu được từ các tài liệu tham khảo. Tất cả các tài liệu tham khảo đều có xuất xứ rõ ràng và được trích dẫn hợp pháp.

Tôi xin chịu trách nhiệm cho lời cam đoan của mình.

Hà Nội, ngày tháng 6 năm 2016

Người cam đoan

Dương Thanh Tùng

LỜI CẢM ƠN

Tôi xin bày tỏ lòng biết ơn sâu sắc đến thầy hướng dẫn của tôi, Tiến sĩ Hoàng Xuân Tùng. Thầy đã giúp tôi có những cơ hội để có thể theo đuổi nghiên cứu lĩnh vực mình yêu thích. Trong suốt quá trình thực hiện luận văn, thầy đã tận tình hướng dẫn cho tôi, góp ý cho tôi về đường lối, đồng thời đưa ra những lời khuyên bổ ích để tôi có thể hoàn thành luận văn của mình.

Tiếp đến, tôi xin chân thành cảm ơn các thầy cô giáo trong Khoa Công nghệ Thông tin, Đại học Công nghệ - Đại học Quốc gia Hà Nội đã truyền đạt cho tôi những kiến thức và kinh nghiệm vô cùng quý báu trong quá trình học tập và nghiên cứu.

Tôi cũng muốn cảm ơn các bạn cùng lớp và các đồng nghiệp đã cho tôi những lời động viên, những hỗ trợ và góp ý về mặt chuyên môn.

Cuối cùng, tôi xin cảm ơn gia đình, bạn bè, những người đã luôn bên cạnh ủng hộ và động viên tôi.

MỤC LỤC

LỜI CAM ĐOAN

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

LỜI MỞ ĐẦU	1
CHƯƠNG 1 – TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN THOẠI VÀ CÁC VẤN ĐỀ MÃ HÓA, GIẢI MÃ ÂM THOẠI.....	2
1. 1. Âm thanh thoại và quá trình số hóa tín hiệu âm thanh.....	2
1.1.1. Âm thanh thoại.....	2
1.1.2. Số hóa âm thanh thoại.....	2
1. 2. Tổng quan về hệ thống thông tin thoại.....	3
1.2.1. Giới thiệu về hệ thống thông tin thoại.....	3
1.2.2. Các yếu tố ảnh hưởng đến chất lượng trong hệ thống thông tin thoại...4	
1. 3. Mã hóa – giải mã tín hiệu âm thanh trong hệ thống thông tin thoại.....	5
1.3.1. Chức năng của bộ mã hóa – giải mã trong hệ thống thoại.....	5
1.3.2. Các phương pháp mã hóa tín hiệu thoại.....	7
1.3.2.1. Phương pháp mã hóa tín hiệu dạng sóng (Waveform coding).....	7
1.3.2.2. Phương pháp mã hóa tiếng nói Vocoder.....	8
1.3.2.3. Phương pháp mã hóa lai (Hybrid coding).....	9
1. 4. Đánh giá chất lượng âm thanh thoại.....	10
1.4.1. Các yêu cầu đối với một bộ mã hóa âm thoại.....	10
1.4.2. Các tham số liên quan đến chất lượng thoại.....	11
1.4.3. Các phương pháp đánh giá chất lượng thoại phổ biến.....	11
CHƯƠNG 2 – ILBC CODEC	16
2. 1. Giới thiệu về iLBC Codec và kỹ thuật xử lý tiếng nói dựa trên mã hóa dự đoán tuyến tính.....	16
2.1.1. Giới thiệu iLBC Codec.....	16
2.1.2. Kỹ thuật xử lý tiếng nói dựa trên mã hóa dự đoán tuyến tính.....	18
2. 2. Quá trình mã hóa iLBC Codec (Encoder).....	25
2.2.1. Tổng quan về quá trình mã hóa iLBC Codec.....	25
2.2.2. Các nguyên tắc mã hóa.....	27
2. 3. Quá trình giải mã iLBC Codec (Decoder).....	29
2.3.1. Tổng quan về quá trình giải mã iLBC Codec.....	29
2.3.2. Các nguyên tắc giải mã.....	31
CHƯƠNG 3 – ĐÁNH GIÁ KHẢ NĂNG CHỐNG LẠI MẤT MẮT GÓI TIN TRÊN ĐƯỜNG TRUYỀN THOẠI.....	33
3. 1. Khái niệm chống mất mát gói tin trên đường truyền thoại.....	33
3. 2. Phân loại kỹ thuật chống mất mát gói tin.....	33

3.2.1.	<i>Kỹ thuật chống mất gói từ phía gửi</i>	34
3.2.2.	<i>Kỹ thuật bù mất gói từ phía nhận</i>	36
3.3.	Đánh giá khả năng chống lại mất mát gói tin của iLBC Codec	41
3.3.1.	<i>Phân tích khả năng chống mất mát gói tin của iLBC Codec</i>	41
3.3.2.	<i>Phương pháp đánh giá khả năng chống mất mát gói tin của iLBC Codec</i> 44	
	CHƯƠNG 4 – ĐÁNH GIÁ BẰNG THỰC NGHIỆM	47
4.1.	Quá trình thực hiện	47
4.2.	Kết quả của quá trình thực nghiệm	50
	KẾT LUẬN	53
	TÀI LIỆU THAM KHẢO	54

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

STT	KÍ HIỆU	Ý NGHĨA
1	ACELP	Algebraic Code Excited Linear Prediction
2	AMR-WB	Adaptive Multi-Rate Wideband
3	AMDF	Average Magnitude Difference Function
4	CELP	Code Excited Linear Predictive
5	EMBSD	Enhanced Modified Bark Spectral Distortion
6	FEC	Forward Error Correction
7	IP	Internet Protocol
8	LAN	Local Area Network
9	LPC	Linear Predictive Coding
10	LTP	Long-Term Predictive
11	LSP	Line Spectrum Pair
12	LSF	Line Spectral Frequency
13	MOS	Mean Opinion Score
14	MMSE	Minimum of Mean Squared Error
15	MSE	Mean Squared Error
16	MNB	Measuring Normalizing Blocks
17	PEAQ	Perceptual Evaluation of Audio Quality
18	PESQ	Perceptual Evaluation of Speech Quality
19	PSQM	Perceptual Speech Quality Measure
20	PAMS	Perceptual Assesment of Speech Quality
21	RELP	Residual-Excited Linear Predictive
22	RTP	Real-Time Protocol
23	RMSE	Root Mean Square Energy
24	SNR	Signal-to-Noise Ratio
25	STP	Short-Term Predictive
26	TCP	Transmission Control Protocol
27	VoIP	Voice Over Internet Protocol

LỜI MỞ ĐẦU

Ngày nay, nhu cầu liên lạc của con người càng trở nên phổ biến và rộng khắp, các yêu cầu về loại hình dịch vụ thông tin ngày càng phong phú. Điều này đòi hỏi các thiết bị cũng như các dịch vụ xử lý thông tin phải phát triển để đáp ứng được sự nhanh nhạy, chính xác của thông tin. Tuy nhiên các dịch vụ này lại chiếm rất nhiều băng thông đường truyền và đôi khi chất lượng thông tin không được tốt do nhiều các yếu tố khách quan tác động đến.

Để sử dụng một cách hiệu quả nhất cơ sở hạ tầng viễn thông, kỹ thuật chuyên mạch gói đã ra đời. Kỹ thuật này chia dữ liệu cần vận chuyển thành các gói (hay các khung) có kích thước và định dạng xác định. Mỗi gói như vậy sẽ được vận chuyển riêng rẽ và đến nơi nhận bằng các đường truyền khác nhau. Khi toàn bộ các gói dữ liệu đã đến nơi nhận thì chúng sẽ được hợp lại thành dữ liệu ban đầu. Tuy nhiên, sự hiệu quả của kỹ thuật chuyên mạch gói cũng đi kèm với các nhược điểm. Trong các hệ thống thông tin thoại, yếu tố mất mát gói tin có ảnh hưởng rất lớn đến chất lượng dịch vụ. Mất gói xảy ra khi các gói gửi từ nguồn đến đích vượt quá khoảng thời gian cho phép chờ nhận hoặc không đến được đích. Có rất nhiều giải thuật với mục đích giải quyết vấn đề mất gói được áp dụng ở bộ phát để tạo dư thừa cho việc mất gói, hoặc ở bộ thu để che giấu các gói bị mất. Trong hệ thống thông tin trên nền giao thức IP, các kỹ thuật này được các nhà phát triển, các tổ chức, thống nhất và đưa ra trên các bộ xử lý tín hiệu thoại, nhằm khắc phục các nhược điểm nêu trên.

Các bộ xử lý này được gọi chung là Codec, và từ khi ra đời cho đến nay đã có rất nhiều chuẩn Codec được áp dụng rộng rãi. Mỗi Codec có một đặc điểm riêng, bù đắp cho nhau nhưng tựu chung lại là cân bằng được giữa yếu tố băng thông yêu cầu và chất lượng gói tin sau quá trình giải mã. Để làm rõ một khía cạnh của vấn đề này, tôi đã lựa chọn việc tìm hiểu iLBC Codec, đánh giá yếu tố chống mất mát gói tin của nó và so sánh với một Codec khác có tính chất tương tự. Việc tìm hiểu được đặc tính của Codec sẽ giúp lựa chọn và có những giải pháp tốt hơn khi xây dựng một hệ thống thông tin thoại.

Bố cục bài luận văn được chia thành 4 chương, với nội dung cốt lõi tập trung vào 3 vấn đề chính:

- **Phần 1:** Giới thiệu tổng quan về hệ thống thông tin thoại và các vấn đề liên quan đến mã hóa, giải mã tín hiệu trong hệ thống.
- **Phần 2:** Trình bày khái niệm về bộ mã hóa và giải mã tín hiệu thoại nói chung và iLBC Codec nói riêng. Các thuật toán xử lý tín hiệu thoại dựa trên mã hóa dự đoán tuyến tính và quá trình thực hiện việc mã hóa, giải mã tín hiệu của iLBC Codec.
- **Phần 3:** Đánh giá khả năng chống lại mất mát gói tin của các Codec, cụ thể là phân tích các đặc trưng của iLBC Codec về kỹ thuật mã hóa, giải mã nhằm bù mất gói trên đường truyền thoại. Cuối cùng là đánh giá định lượng khả năng bù mất gói của các Codec bằng thực nghiệm, thực hiện trên phần mềm mô phỏng Matlab và Simulink.

CHƯƠNG 1 – TỔNG QUAN VỀ HỆ THỐNG THÔNG TIN THOẠI VÀ CÁC VẤN ĐỀ MÃ HÓA, GIẢI MÃ ÂM THOẠI

1. 1. Âm thanh thoại và quá trình số hóa tín hiệu âm thanh

1.1.1. Âm thanh thoại

Âm thanh (Sound) là các dao động cơ học của các phân tử, nguyên tử hay các hạt vật chất lan truyền trong không gian, được cảm nhận trực tiếp qua tai người bởi sự va đập vào màng nhĩ và kích thích bộ não. Sóng âm tần được đặc trưng bởi biên độ, tần số (bước sóng) và vận tốc lan truyền. Đối với tai người, âm thanh cảm nhận được bởi sóng có dao động trong dải tần từ 20Hz đến 20kHz. Tín hiệu âm thanh được chia thành 2 loại dựa trên dải tần:

- Âm thanh dải tần cơ sở (âm thanh tiếng nói thoại, gọi tắt là âm thanh thoại): có dải tần từ 300Hz đến 4kHz.

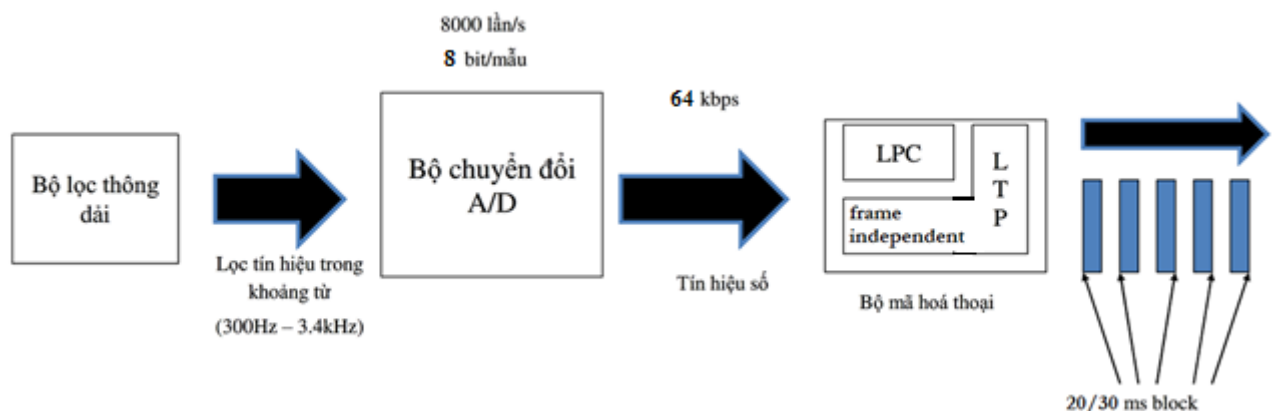
- Âm thanh dải rộng (tiếng nói trình diễn, âm nhạc...): có dải tần số từ 100Hz đến 20kHz.

Audio là âm thanh thoại thu nhận được, được xử lý và tái tạo bởi các thiết bị điện tử, các đối tượng truyền thông đa phương tiện. Trong bài luận văn chỉ đề cập đến âm thanh thoại. Âm thanh thoại có một số đặc điểm như sau:

- Giới hạn dải phổ tín hiệu ~ 4kHz.
- Tần số lấy mẫu $f_s = 8\text{kHz}$ tương đương với chu kỳ $T_e = 125\mu\text{s}$.
- Lượng tử hóa các giá trị với mã hóa 8 bit.
- Tốc độ bit tiêu chuẩn: $8\text{bit} \times 8\text{kHz} = 64\text{kbps}$.

1.1.2. Số hóa âm thanh thoại

Đầu tiên, tiếng nói được microphone biến đổi sang tín hiệu điện ở dạng tương tự. Microphone bao gồm một màng 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, tín hiệu được cho qua bộ lọc thông dải trong khoảng tần số từ 300Hz đến 3,4kHz. Sau đó, tín hiệu này được biến đổi sang tín hiệu số bằng bộ chuyển đổi tương tự - số (A/D Converter) dùng kỹ thuật điều chế xung mã PCM với tần số lấy mẫu là 8kHz và mã hoá mỗi mẫu bằng 8 bit. Do đó, luồng tín hiệu số sau khi được biến đổi có tốc độ 64kbps [1, tr.2-3].



Hình 1.1 – Số hóa và mã hóa tín hiệu thoại

Việc xử lý âm thanh tương tự và âm thanh số ban đầu cho chất lượng âm thanh ở mức khá tốt. Tuy nhiên để truyền qua hệ thống thông tin thoại đòi hỏi nhiều hơn nữa về tỉ lệ giữa băng thông sử dụng và chất lượng tín hiệu. Codec ra đời để giải quyết vấn đề này.

Tín hiệu số ở đầu ra của bộ chuyển đổi A/D có tốc độ 64kbps 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. Ví dụ, thay vì truyền đi luồng số từ tiếng nói thì sử dụng công nghệ 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. Tín hiệu số ở đầu ra của bộ chuyển đổi A/D có tốc độ 64kbps được chia thành từng đoạn có chiều dài 20ms, như vậy mỗi đoạn chứa 1280 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 13kbps. Tốc độ truyền này đảm bảo chất lượng tín hiệu thu được ở mức chấp nhận được và băng thông yêu cầu cho đường truyền sẽ được giảm đi rất nhiều.

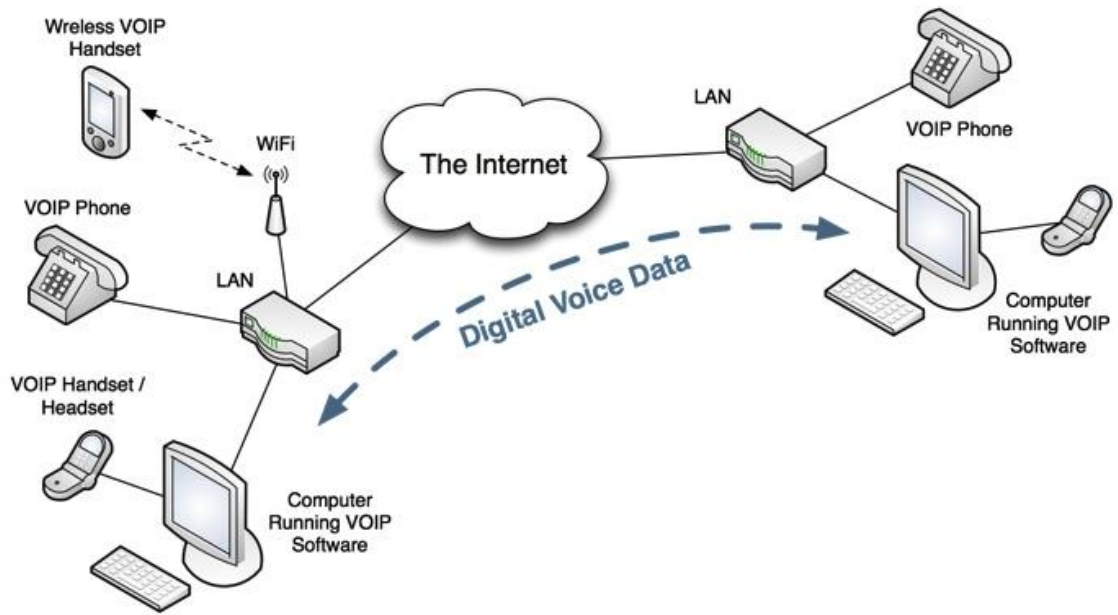
1. 2. Tổng quan về hệ thống thông tin thoại

1.2.1. Giới thiệu về hệ thống thông tin thoại

Hệ thống thông tin thoại là một tập hợp tất cả những thành phần tham gia hình thành nên một mô hình truyền – nhận các tín hiệu âm thoại. Có rất nhiều các hệ thống như vậy hiện nay như hệ thống vệ tinh, hệ thống mạng cục bộ LAN (Local Area Network), Internet,... Đặc điểm chung của các hệ thống này là đều tiếp nhận, xử lý tín hiệu thông tin thoại sau đó truyền tải từ bên gửi đến bên nhận để truyền đạt thông tin. Trong các hệ thống trên có một hệ thống rất phổ biến với mọi người hiện nay, đó là hệ thống truyền thông tin thoại qua kênh truyền sử dụng bộ giao thức TCP/IP, hay còn gọi là VoIP (Voice over Internet Protocol). Trong khuôn khổ bài luận văn sẽ chỉ đề cập đến việc mã hóa, giải mã và xử lý tiếng nói trên mô hình này.

VoIP, hay truyền giọng nói trên giao thức IP, là công nghệ truyền tiếng nói của con người (âm thanh thoại) qua mạng thông tin sử dụng bộ giao thức TCP/IP. Nó sử dụng các gói dữ liệu IP (trên mạng LAN, WAN, Internet) với thông tin được truyền tải là các gói tin âm thanh đã được mã hoá. Công nghệ này bản chất là dựa trên chuyển mạch gói, nhằm thay thế công nghệ truyền thoại cũ dùng chuyển mạch kênh. Nó nén (ghép) nhiều kênh thoại trên một đường truyền tín hiệu, và những tín hiệu này được truyền qua mạng Internet, do vậy có thể làm giảm giá thành so với chuyển mạch kênh.

Để thực hiện việc này, điện thoại IP, thường được tích hợp sẵn các giao thức báo hiệu chuẩn như giao thức SIP hay H.323, kết nối tới một tổng đài IP (IP PBX) của doanh nghiệp hay của nhà cung cấp dịch vụ. Điện thoại IP có thể là điện thoại thông thường (chỉ khác là thay vì nối với mạng điện thoại qua đường dây giao tiếp RJ11 thì điện thoại IP nối trực tiếp vào mạng LAN qua cáp Ethernet, giao tiếp RJ45) hoặc phần mềm thoại (soft-phone) cài trên các thiết bị đầu cuối (máy tính, điện thoại,...).



Hình 1.2 – Ví dụ về một hệ thống thông tin thoại

1.2.2. Các yếu tố ảnh hưởng đến chất lượng trong hệ thống thông tin thoại

Chất lượng của âm thanh được khôi phục qua hệ thống thoại là mục tiêu cơ bản của dịch vụ truyền thoại qua hệ thống, và luôn đòi hỏi đạt được ở mức độ tốt nhất mặc dù các tiêu chuẩn đã được Liên minh viễn thông quốc tế ITU (International Telecommunication Union) phát triển. Có 3 yếu tố chính ảnh hưởng lớn tới chất lượng của dịch vụ thoại [14, tr.44-46]:

a. Độ trễ (Delay):

Độ trễ là khoảng thời gian truyền một thông điệp từ nút này đến nút khác trong hệ thống mạng. Độ trễ trong quá trình truyền gói tin từ nguồn tới đích được phân thành:

- Trễ xử lý: là thời gian đóng gói hay xử lý các gói tin tại các nút. Trễ này phụ thuộc vào từng loại thiết bị khác nhau.
- Trễ lan truyền: là thời gian truyền một bit thông tin trên đường liên kết từ nguồn tới đích. Trễ lan truyền phụ thuộc vào khoảng cách truyền giữa hai nút mạng.
- Trễ truyền tin: là khoảng thời gian cần thiết để truyền đi một đơn vị dữ liệu. Ví dụ trong chuyển mạch gói, đó là khoảng thời gian để truyền hết tất cả các bit của một gói tin lên đường truyền. Loại trễ này phụ thuộc vào kích thước của gói tin và băng thông của đường truyền.
- Trễ hàng đợi: là thời gian xử lý tại hàng đợi trong các nút mạng. Trong mạng chuyển mạch gói, trễ hàng đợi được tính bằng khoảng thời gian gói chờ từ khi vào hàng đợi đến khi ra khỏi hàng đợi. Trễ hàng đợi phụ thuộc vào số lượng gói tin gửi đến một nút mạng.

Hai vấn đề gây ra bởi sự trễ đầu cuối trong một mạng thoại là tiếng vang và chòng tiếng. Tiếng vang trở thành vấn đề khi trễ vượt quá 50ms. Đây là một vấn đề

ảnh hưởng đáng kể đến chất lượng, nên các hệ thống thông tin thoại phải kiểm soát và cung cấp các phương tiện loại bỏ tiếng vang. Hiện tượng chùng tiếng (giọng người này gối lên giọng người kia) trở nên đáng kể nếu trễ một chiều (one-way delay) lớn hơn 250ms.

b. Sự biến thiên độ trễ (Jitter):

Jitter là sự biến thiên thời gian trễ gây nên bởi sự trễ đường truyền khác nhau trên mạng. Loại bỏ jitter đòi hỏi thu thập các gói tin và giữ chúng đủ lâu để cho phép các gói chậm nhất đến để được phát lại đúng thứ tự, làm cho độ trễ tăng lên.

c. Sự mất mát gói tin (Packet Losses)

Mạng IP không thể cung cấp một sự bảo đảm rằng các gói tin sẽ được chuyển tới đích hết. Các gói tin sẽ bị loại bỏ khi quá tải và trong thời gian tắc nghẽn, gây nên sự mất mát gói tin (Packet Loss) của hệ thống. Truyền thoại rất nhạy cảm với việc mất gói, tuy nhiên, việc truyền lại gói của TCP thường không phù hợp. Các cách tiếp cận được sử dụng để bù lại các gói mất là thêm vào cuộc nói chuyện bằng cách phát (play) lại gói cuối cùng, và gửi đi thông tin dư thừa. Tuy vậy, sự tổn thất gói trên 10% nói chung là không chấp nhận được.

Việc duy trì chất lượng thoại ở mức độ chấp nhận được bất chấp những thay đổi của mạng (như tắc nghẽn hay mất kết nối) đạt được nhờ những kỹ thuật như nén tiếng, triệt im lặng. Trong nhiều năm trước đây, có một số thành tựu trong việc xử lý tín hiệu số, chuyển mạch mạng chất lượng cao đã được phối hợp để hỗ trợ và khuyến khích công nghệ thoại trên mạng IP.

Quá trình tiền xử lý bằng phần mềm của cuộc đàm thoại cũng có thể được sử dụng để tối ưu hoá chất lượng âm thanh. Một kỹ thuật, được gọi là triệt im lặng, sẽ xác định mỗi khi có một khoảng trống trong lời thoại (talk spurt) và loại bỏ sự truyền các khoảng nghỉ, hơi thở, và các khoảng im lặng khác. Điều đó có thể lên tới 50-60% thời gian của một cuộc gọi, giúp tiết kiệm băng tần đáng kể. Bởi lẽ sự thiếu các gói được hiểu là sự im lặng hoàn toàn ở đầu ra, cho nên lúc này bên nhận sẽ thực hiện việc bổ sung các tiếng động khi nhận tín hiệu.

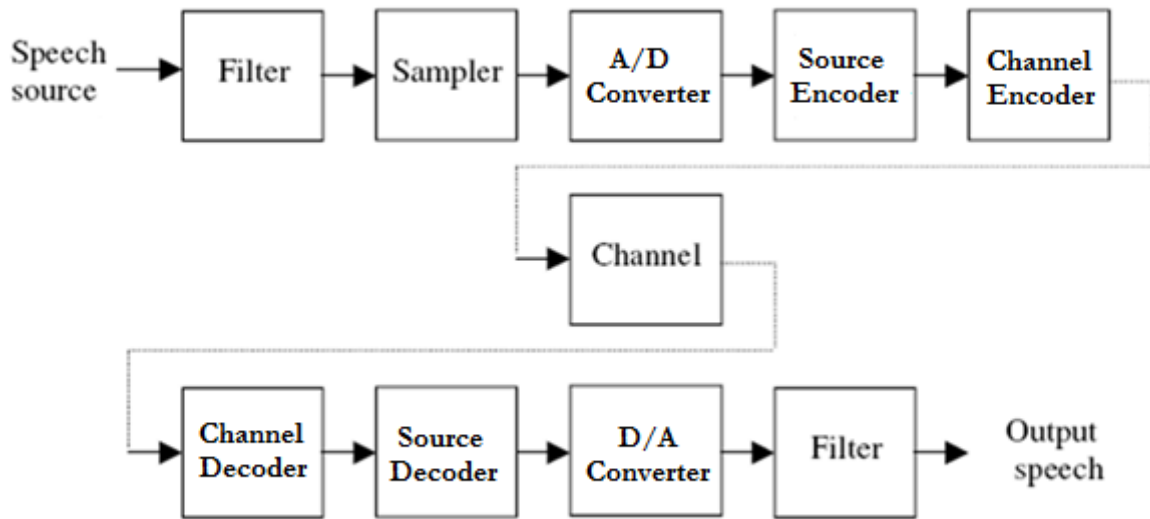
1. 3. Mã hóa – giải mã tín hiệu âm thanh trong hệ thống thông tin thoại

1.3.1. Chức năng của bộ mã hóa – giải mã trong hệ thống thoại

Trong các hệ thống truyền dẫn số, thông tin được chuyển đổi thành một chuỗi các tổ hợp xung, sau đó truyền trên đường truyền. Khi đó, thông tin tương tự (ví dụ tiếng nói của con người) phải được chuyển đổi sang dạng số nhờ các bộ chuyển đổi tương tự - số (A/D converter). Độ chính xác của chuyển đổi A/D sẽ quyết định chất lượng của hệ thống.

Trong hệ thống thông tin thoại, trước tiên tín hiệu âm thoại tương tự liên tục đầu vào (Speech source) sẽ được số hóa bởi bộ một bộ lọc chuẩn (Filter), bộ lấy mẫu (bộ chuyển đổi thời gian rời rạc, lượng tử hóa - Sampler), và bộ chuyển tín hiệu tương tự sang tín hiệu số (A/D Converter). Tín hiệu ngõ ra là tín hiệu âm thoại thời gian rời rạc

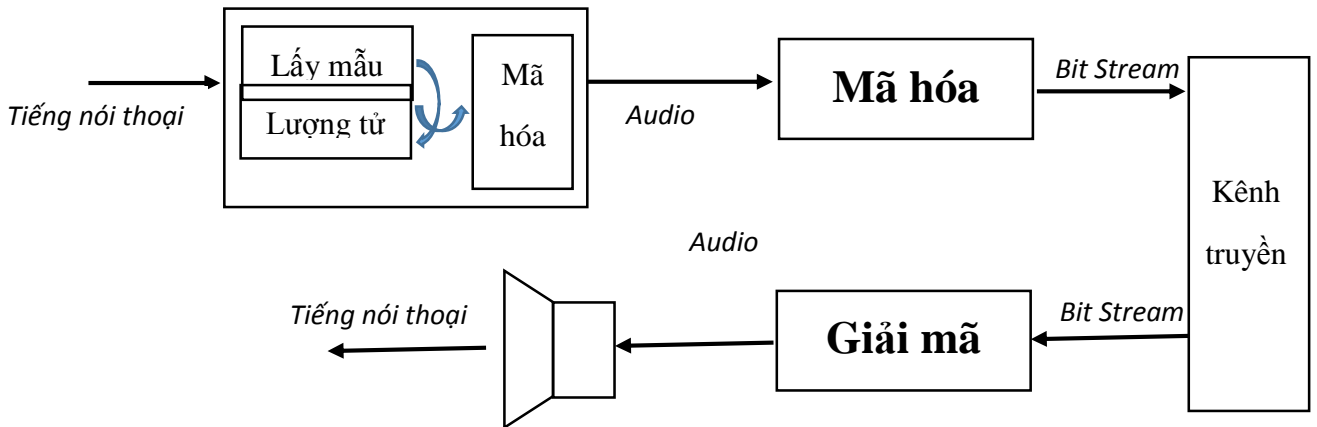
với các giá trị lấy mẫu cũng rời rạc hóa. Tín hiệu này được xem là tín hiệu số của âm thoại.



Hình 1.3 - Sơ đồ khối của hệ thống thông tin thoại

Thông thường, hầu hết các hệ thống mã hóa âm thoại được thiết kế để hỗ trợ các ứng dụng viễn thông, với tần số giới hạn giữa 300Hz và 3400Hz. Tần số lấy mẫu tối thiểu phải lớn hơn hai lần băng thông của tín hiệu liên tục thời gian. Giá trị 8kHz thường được lựa chọn là tần số lấy mẫu chuẩn cho tín hiệu thoại. Bộ mã hóa kênh (Channel Encoder) thực hiện việc mã hóa hiệu chỉnh lỗi của chuỗi bit truyền trước khi tín hiệu được truyền trên kênh truyền (Channel), nơi mà tín hiệu sẽ bị thay đổi do mất gói tin, trễ, nhiễu ... Bộ giải mã kênh (Channel Decoder) thực hiện việc hiệu chỉnh lỗi để có được tín hiệu đã mã hóa, sau đó tín hiệu được đưa vào bộ giải mã (Source Decoder) để có được tín hiệu âm thoại số có cùng tốc độ với tín hiệu ban đầu. Lúc này, tín hiệu số sẽ được chuyển sang dạng tương tự thời gian liên tục nhờ bộ chuyển đổi tín hiệu số về tín hiệu tương tự (D/A Converter). Như vậy, bộ phận thực hiện việc xử lý tín hiệu thoại chủ yếu của mô hình hệ thống xử lý thoại là bộ mã hóa và giải mã.

Liên hệ với các mô hình thực tế, các khối Filter, Sampler và A/D Converter nằm trên các thiết bị thu âm thanh như microphone, có nhiệm vụ thu và số hóa âm thanh rồi chuyển cho bộ mã hóa – giải mã Codec. Tương tự, ở phía nhận, các khối D/A Converter và Filter là một phần của hệ thống phát âm như tai nghe, loa. Bộ Codec chỉ có thể làm việc với tín hiệu số. Bộ Codec ở bên gửi có nhiệm vụ mã hóa, làm giảm dung lượng của âm thanh, sau đó đóng gói vào các đoạn (chunk) để đặt vào các gói dữ liệu chuyển đi trên mạng IP. Codec ở bên người nhận sẽ nhận, giải mã các gói tin, lấy ra các mẫu, tiếp tục đưa vào khối Source Decoder để chuyển đổi sang âm thanh số, chuyển cho bộ phát âm. Việc điều phối và quản lý các kênh được thực hiện thông qua một hệ thống máy chủ chuyển mạch. Hệ thống máy chủ có thể là một phần mềm cài trên máy chủ như Asterisk. Thông thường, khi xử lý các bài toán về truyền thoại, mô hình được đơn giản hóa như hình 1.4 dưới đây.



Hình 1.4: Sơ đồ khối đơn giản hóa của hệ thống thông tin thoại

Đối với bộ mã hóa, tín hiệu âm thoại đầu vào được phân tích và xử lý (Analysis and Processing) nhằm thu được các tham số đại diện cho một khung truyền. Các tham số này được mã hóa và lượng tử hóa với mã nhị phân và được gửi đi như là một chuỗi bit đã được nén. Các giá trị nhị phân đó được đóng gói và biểu diễn thành chuỗi bit, chúng được sắp xếp thứ tự truyền dựa vào các thông số đã quyết định trước và được truyền đến bộ giải mã.

Bộ giải mã thực hiện việc phân tích chuỗi bit nhận được, các giá trị nhị phân được phục hồi sau quá trình phân tích và dùng để kết hợp với các thông số tương ứng của bộ giải mã để có được các thông số đã được lượng tử. Các thông số giải mã này sẽ kết hợp với nhau và được xử lý để tạo lại tín hiệu âm thoại tổng hợp – Synthetic speech.

Mục tiêu chính của mã hóa thoại là tối đa hóa chất lượng nghe tại một tốc độ bit nào đó, hoặc tối thiểu hóa tốc độ bit ứng với một chất lượng đặc thù. Tốc độ bit tương ứng với âm thoại nào sẽ được truyền hoặc lưu trữ phụ thuộc vào chi phí của việc truyền hay lưu trữ, chi phí của mã hóa tín hiệu thoại số, và các yêu cầu về chất lượng của âm thoại đó. Trong hầu hết các bộ mã hóa âm thoại, tín hiệu được xây dựng lại sẽ khác với tín hiệu nguyên thủy. Tốc độ bit truyền bị giảm bởi việc biểu diễn tín hiệu âm thoại (hoặc các thông số trong mô hình tạo âm thoại) với độ chính xác bị giảm, và bởi quá trình loại bỏ các thông tin dư thừa của tín hiệu.

1.3.2. Các phương pháp mã hóa tín hiệu thoại

Mã hóa tín hiệu thoại được chia thành 3 phương pháp chính [1, tr.20-24]:

- Mã hóa tín hiệu dạng sóng (Waveform coding).
- Mã hóa tiếng nói Vocoder (Vocal coder).
- Mã hóa lai (Hybrid coding).

1.3.2.1. Phương pháp mã hóa tín hiệu dạng sóng (Waveform coding)

Mã hóa dạng sóng thực hiện thay đổi biên độ của tín hiệu tương tự (âm thanh thoại) được mô tả bằng một số của giá trị được đo. Sau đó các giá trị này được mã hóa thành các bit và truyền đi. Phương pháp này cho chất lượng thoại ở mức độ rất cao, do tín hiệu bên nhận được gần giống nhất với tín hiệu bên gửi.

a. Mã hóa trong miền thời gian:

Phương pháp này bao gồm các chuẩn từ G.710 đến G.719, được quy định bởi Liên minh Viễn thông quốc tế ITU.

➤ *Mã hóa điều chế xung mã PCM (Pulse Code Modulation)*: Lượng tử hóa đã được chuẩn hóa với chuẩn G.711, là phương pháp mã hóa cơ bản, mã hóa trực tiếp tín hiệu lấy mẫu tiếng nói dùng các luật lượng tử hóa μ -law, a-law.

Đầu vào của bộ lượng tử là tín hiệu tương tự đã được đưa qua bộ 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 quantization). Bộ lượng tử quyết định xem với mỗi giá trị đầu ra là giá trị lớn nhất của miền giá trị. Trong kiểu mã hóa 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 nhiều lượng tử, nhưng lại có thể đủ lớn để miền giá trị của cả bộ lượng tử có độ lớn thích hợp. Với một bộ lượng tử N bit có bước lượng tử là S , thì miền giá trị là $R=2^N \cdot S$.

➤ *Mã hóa dự đoán – điều chế xung mã sai phân DPCM (Differential Pulse Code Modulation)*: 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 $t+1$. 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.

➤ *Mã hóa dự đoán thích nghi – điều chế xung mã sai phân thích nghi ADPCM (Adaptive Differential Pulse Code Modulation)*: Là phương pháp mở rộng của DPCM (lượng tử hóa được chuẩn hóa với chuẩn G.726). Phương pháp này vẫn dùng một số bit nhất định để mã hóa sự sai khác giữa tín hiệu 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 việc điều chế.

Với mục tiêu làm giảm tốc độ bit hơn nữa mà chất lượng tín hiệu vẫn giữ ở mức tương đương, người ta sử dụng phương pháp thích nghi động giá trị của bước lượng tử trước những thay đổi của biên độ tín hiệu vào. Mục đích là duy trì miền giá trị lượng tử phù hợp với miền giá trị của tín hiệu vào. Đây được gọi là phương pháp thích nghi 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 một số thống kê về tín hiệu có liên quan đến biên độ của nó. Có nhiều bước thực hiện để 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 một khối (block) mẫu thu được trong một thời gian ngắn, về năng lượng, sự biến đổi và những yếu tố khác.

b. Mã hóa trong miền tần số

- Mã hóa các dải tần con SBC (Subband Coding), ví dụ G.722 Codec.
- Mã hóa dựa trên phép biến đổi Transform Coding.

1.3.2.2. Phương pháp mã hóa tiếng nói Vocoder

Mã hóa Vocoder là bộ mã hóa tham số. Thay cho việc truyền tín hiệu mô tả trực tiếp dạng của đường tín hiệu thoại, nguyên lý mã hóa Vocoder dựa trên bộ mô phỏng hệ thống nguồn phát âm của con người, tạo ra âm thanh tiếng nói từ tập các tham số, mô tả đường cong của tín hiệu được phát ra như thế nào. Vocoder làm việc với 2 kiểu nguồn kích thích là nguồn xung tạo ra âm hữu thanh và nguồn nhiễu trắng tạo ra âm vô thanh. Từ đó, nó mô phỏng hệ thống phát âm bằng hệ thống lọc dự đoán tuyến tính PLC được kích thích bằng hai trạng thái nguồn.

Ưu điểm của phương pháp này là tín hiệu có thể được truyền đi với tốc độ bit rất thấp, phân tích được các tham số nguồn kích thích, có thể sửa đổi nội dung tiếng nói theo ý muốn. Tuy nhiên, nhược điểm của nó là phụ thuộc nhiều vào mô hình thoại, tiếng nói nhận được là tiếng nói tổng hợp, không hoàn toàn giống với giọng nói thực và chất lượng ở mức trung bình.

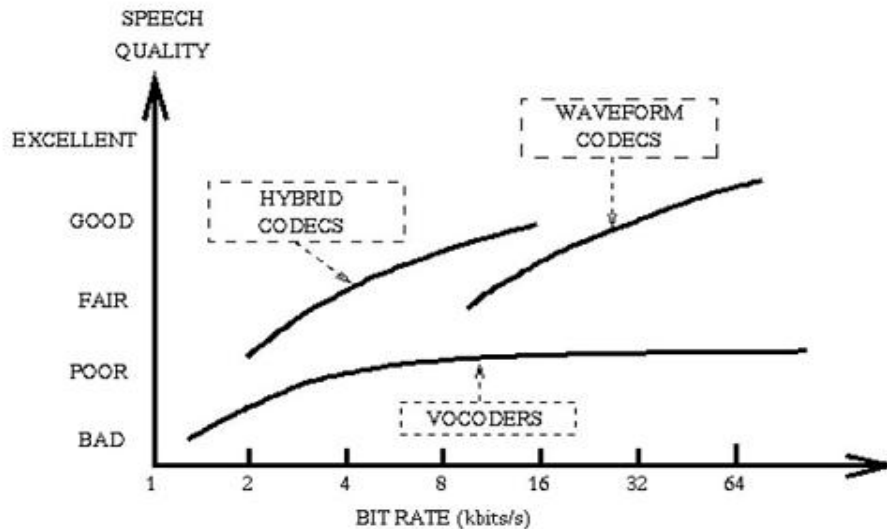
iLBC Codec được trình bày trong bài luận văn cũng sử dụng phương pháp mã hóa tiếng nói Vocoder.

1.3.2.3. Phương pháp mã hóa lai (Hybrid coding)

Kỹ thuật phổ biến của phương pháp mã hóa lai là mã hóa dựa trên kết hợp phân tích bằng cách tổng hợp AbS (Analysis by Synthesis).

Phương pháp này sử dụng mô hình phát âm của người tương tự như phương pháp mã hóa tiếng nói Vocoder, nhưng mặt khác, tín hiệu kích thích đạt được từ bộ phân tích tín hiệu tiếng nói của chính người nói và được chọn sao cho dạng sóng tiếng nói khi được tái tạo giống với dạng sóng tiếng nói ban đầu nhất. Thuật toán tìm ra sóng kích thích này quyết định độ phức tạp của bộ mã hóa. Có nhiều loại mã hóa lai khác nhau theo kỹ thuật phân tích tạo ra tín hiệu kích thích như mã hóa dự đoán kích thích mã CELP (Code Excited Linear Predictive), mã hóa dự đoán tuyến tính kích thích xung đều RPE-LTP (Regular Pulse Excited – Long Term Predictive), mã hóa dự đoán tuyến tính kích thích đa xung MPE-LTP (Multi Pulse Excited – Long Term Predictive),...

Đánh giá chung về 3 phương pháp: Mã hóa dạng sóng nhìn chung không cho phép đạt chất lượng tiếng nói tốt ở tốc độ bit dưới 16kbps. Mã hóa Vocoder có thể đạt được tốc độ bit rất thấp nhưng nhược điểm là rất khó nhận diện được người nói. Còn mã hóa lai thường được dùng theo chuẩn GSM. Hình dưới đây thể hiện sự so sánh giữa 3 phương pháp mã hóa về bit-rate và chất lượng giọng nói tương ứng:



Hình 1.5 – So sánh 3 phương pháp mã hóa âm thoại

1. 4. Đánh giá chất lượng âm thanh thoại

1.4.1. 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 ban đầu. Nguyên nhân là do khi cố gắng làm tăng chất lượng 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 độ 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 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 một 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 từng ứng dụng. Có rất nhiều khía cạnh về mặt chất lượng như 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 nhiều*: 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 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 về 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 nhiều vấn đề trong việc thực hiện trao đổi tiếng nói hai chiều trong thời gian thực.

- *Khả năng cắt bỏ khoảng lặng*: khi nói chuyện không phải âm thoại được phát ra liên tục mà có những khoảng lặng. Đó là những lúc dừng lại lấy hơi hay 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.

1.4.2. Các tham số liên quan đến chất lượng thoại

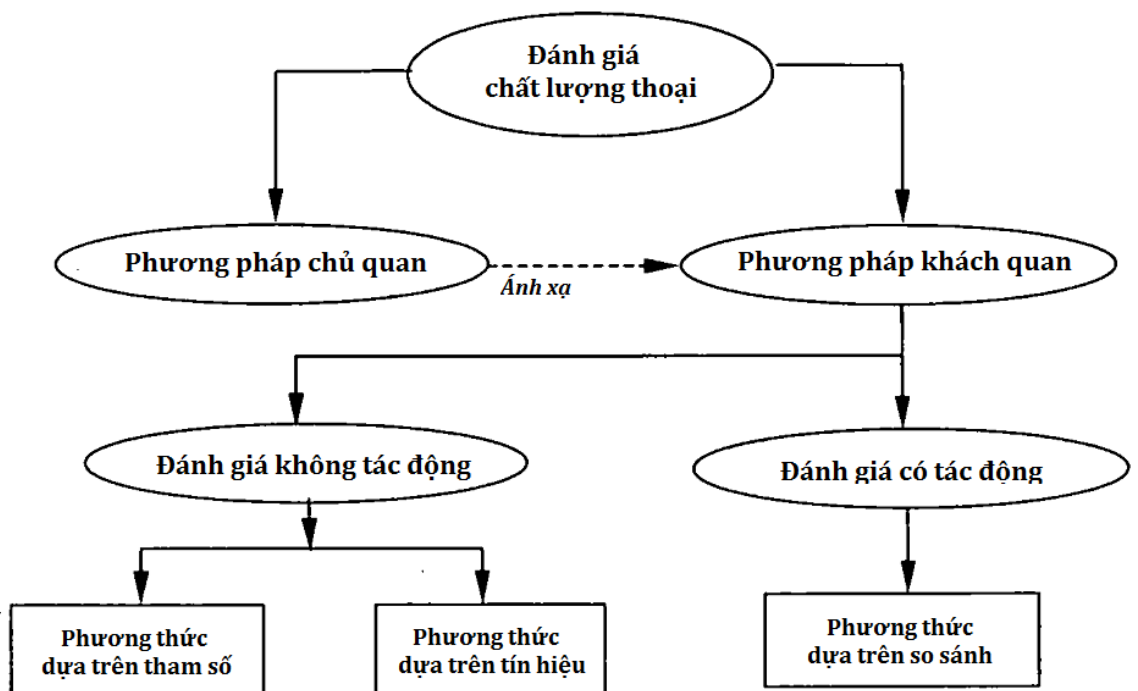
Các tham số truyền dẫn cơ bản liên quan đến chất lượng thoại là:

- Tham số đánh giá cường độ âm lượng/tổn hao tổng thể OLR (Overall Loudness Rating).
- Độ trễ: Thời gian truyền dẫn tín hiệu giữa hai đầu cuối gây ra những khó khăn trong việc hội thoại. Yếu tố 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.
- Tiếng vọng (echo).
- Cắt ngưỡng (clipping): là hiện tượng mất phần đầu hoặc phần cuối của cụm tín hiệu thoại, do quá trình xử lý khoảng lặng bị sai.
- Các tính chất liên quan đến độ nhạy tần số.
- Nhiễu xuyên âm, nhiễu nền.

1.4.3. Các phương pháp đánh giá chất lượng thoại phổ biến

Việc đánh giá chất lượng thoại trong mạng có thể được thực hiện bằng cách đánh giá các tham số truyền dẫn có ảnh hưởng đến chất lượng thoại và xác định tác động 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).

Sơ đồ phân loại đánh giá chất lượng thoại [10, tr.30]:



Hình 1.6 - Phân loại các phương pháp đánh giá chất lượng thoại

Những phương pháp sử dụng MOS đều mang tính chất chủ quan do chúng phụ thuộc 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:

a. 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. Phương pháp này được quy định trong khuyến nghị ITU-T P.800. Người đánh giá chất lượng sẽ trực tiếp tham gia cuộc hội thoại. Đánh giá MOS bao gồm 2 bước thực hiện: Đánh giá phân loại tuyệt đối ACR (Absolute Category Rating) và đánh giá phân loại độ suy giảm DCR (Degradation Category Rating).

Việc đánh giá MOS thông thường gồm 12-24 người tham gia lắng nghe riêng biệt, và đưa ra thang điểm từ 1-5. Sau đó, trung bình cộng điểm của từng người chính là giá trị MOS. Việc chọn các Codec sẽ cho các chất lượng hội thoại khác nhau trên cùng một đường truyền, do đó MOS có thể đánh giá chất lượng Codec nhưng không hoàn toàn chính xác.

Bảng đánh giá chất lượng dịch vụ VoIP bằng chỉ số MOS:

Điểm	Chất lượng tương ứng (ACR)	Mức độ suy giảm tín hiệu (DCR)
5	Xuất sắc	Không thể nhận ra.
4	Tốt	Có thể nhận ra nhưng không gây khó chịu.
3	Trung bình	Hơi khó nghe một chút.
2	Yếu	Khó để nghe.
1	Kém	Rất khó để nghe được.

Bảng tham chiếu chỉ số chất lượng hội thoại:

Khoảng điểm	Chất lượng cuộc gọi
4.0 - 5.0	Chất lượng ở mức mong muốn.
3.6 – 4.0	Chất lượng ở mức chấp nhận được.
1.0 – 3.6	Chất lượng kém, không nên sử dụng dịch vụ.

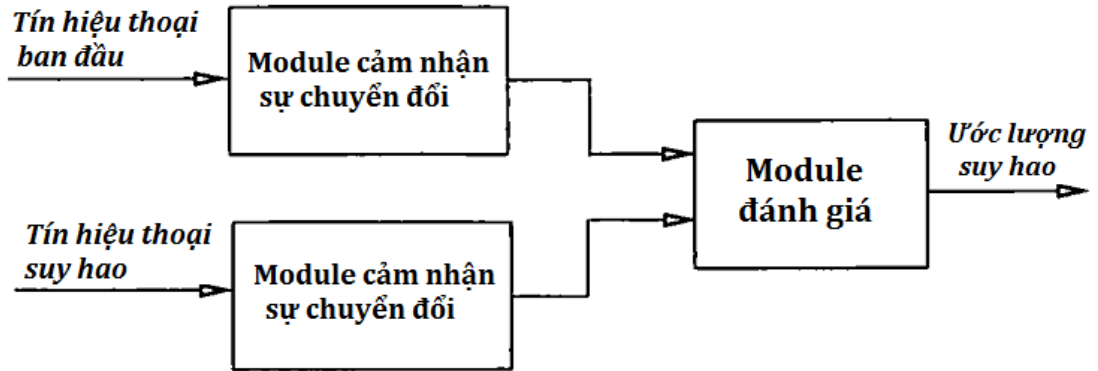
b. Các phương pháp đánh giá khách quan:

Sử dụng một số mô hình để ước lượng mức chất lượng theo thang điểm MOS. Phương pháp đánh giá khách quan có thể được phân thành [10, tr.32-34]:

- **Phương pháp đánh giá có tác động (Intrusive method):** dựa trên việc so sánh tín hiệu thoại truyền dẫn với một tín hiệu chuẩn đã biết. Quá trình đánh giá sẽ tác động vào quá trình thực hiện trao đổi tín hiệu của hệ thống. Một số phương pháp cụ thể có thể kể đến đó là PSQM (Perceptual Speech Quality

Measure), PAMS (Perceptual Assessment of Speech Quality), MNB (Measuring Normalizing Blocks), EMBSD (Enhanced Modified Bark Spectral Distortion), PESQ (Perceptual Evaluation of Speech Quality).

Mô hình chính của phương pháp này như sau:



Hình 1.7 – Mô hình đánh giá chất lượng thoại có tác động

Tương tự như phương pháp MOS, phương pháp này yêu cầu chi phí cao và tốn thời gian. Nó được tự động ánh xạ thang điểm với thang điểm của MOS. Về cơ bản, bộ đánh giá gồm 2 tín hiệu: tín hiệu tham khảo và tín hiệu bị suy hao, được gửi cùng một lúc tới bộ đánh giá và kết quả cho ra thang điểm đánh giá.

- **Phương pháp đánh giá không tác động (Non-Intrusive method):** bao gồm 2 phương thức là dựa trên tham số và dựa trên tín hiệu [10, tr.38-41].

Về cơ bản, phương pháp này đánh giá chất lượng tín hiệu thoại mà không có sự tác động vào quá trình thực hiện việc chuyển đổi tín hiệu. Đặc trưng của phương pháp này có thể kể đến đó là phương pháp đánh giá sử dụng mô hình E (E-model).

E-model khắc phục được các nhược điểm của phương pháp đánh giá có tác động, bằng cách đánh giá tất cả các yếu tố ảnh hưởng tới chất lượng hội thoại (Packet Loss, Delay, Jitter...). Bộ đánh giá của E-model bao gồm nhiều yếu tố ảnh hưởng tới chất lượng giọng nói, sau đó tính toán ra giá trị R (R-factor), nằm trong khoảng (0-100). R-factor có thể được chuyển đổi sang thang điểm MOS theo một phép ánh xạ quy đổi.

Công thức tính R: $R_{obj} = R_0 - I_s - I_d - I_e + A$, với:

R_0 : Tỷ số S/N (tín hiệu/nhiều), bao gồm nhiễu mạch và nhiễu phòng.

I_s : Các yếu tố làm suy giảm chất lượng (như nhiễu lượng tử, nội âm,...).

I_d : Giá trị làm suy giảm tín hiệu do trễ (Delay).

I_e : Giá trị làm suy giảm tín hiệu do thiết bị (như Codec, jitter, mất gói).

A : Hệ số tích cực, để bù vào các suy hao khác khi có những tiến bộ của thiết bị truy nhập của người sử dụng ($A=0$ với đường truyền có dây và $A=5$ với truyền không dây).

Chuẩn ITU G.107 cũng cho phép ánh xạ từ giá trị R-factor về thang điểm MOS như sau:

- $R < 0$: $MOS = 1$
- $0 < R < 100$: $MOS = 1 + 0.035R + R(R-60)/(100-R)7.10^{-6}$
- $R > 100$: $MOS = 4.5$

Ánh xạ giữa R-factor và MOS score:

R	Mức độ hài lòng của người sử dụng	MOS
100		
94	Rất hài lòng	4.4
90		4.3
80	Hài lòng	4.0
70	Một số người không hài lòng	3.6
60	Nhiều người không hài lòng	3.1
50	Hầu hết không hài lòng	2.6
0	Không khuyến nghị	1.0

Giá trị mặc định →

Hình 1.8 – Ánh xạ thang điểm giữa R-factor và MOS

Theo chuẩn ITU G.107, R-factor có thể tính đơn giản:

$$R = 93.2 - I_d - I_e - A, \text{ với:}$$

$$I_e = a + b \ln(1 + cP/1000)$$

P : phần trăm mất gói tin.

a, b, c : tham số tùy thuộc từng codec.

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3)$$

d : độ trễ một chiều (tính theo ms).

H là hàm truyền: $H(x) = 0$ với $x < 0$ và $H(x) = 1$ với mọi giá trị x còn lại.

Bảng giá trị a, b, c của một số Codec (do tổ chức ITU đưa ra):

Tham số	G.711	G.729 (10ms)	G.729 (20ms)	G.723.1	iLBC
bitrate(kb/s)/framesize(ms)	64/20	8/10	8/20	6.3/30	15.2/20
a	0	10	10	15	10
b	30	25.21	25.21	36.59	19.8
c	15	15	20.2	6	29.7

Tuy vậy E-model cũng có những nhược điểm như sau:

- Giá trị tham chiếu từ E-model Score (R-factor) sang MOS Score (thang điểm MOS) không quá chính xác so với thang điểm MOS chủ quan.
- E-model không tính toán được độ mất gói và độ trễ khi cuộc hội thoại chuyển tiếp giữa các vùng mạng (Handoff).

Ngoài phương pháp MOS được kể ra ở trên, có thể sử dụng phương pháp tổng quát, áp dụng cho nhiều trường hợp đánh giá sai số, đó là phương pháp tính sai số trung bình bình phương MSE (Mean Squared Error) để đánh giá chất lượng tín hiệu. MSE có một điểm giống với phương pháp đánh giá khách quan không tác động (Intrusive method), đó là đều so sánh giữa tín hiệu ban đầu và tín hiệu sau khi suy hao bởi quá trình truyền trong hệ thống. Tuy nhiên, trong khi Intrusive method thiên về đánh giá tổng quan chất lượng tín hiệu thoại toàn hệ thống thì phương pháp MSE lại hiệu quả hơn trong việc đánh giá chi tiết về khả năng chống mất mát gói tin thoại sau khi nhận được của hệ thống. Phương pháp này sẽ được sử dụng trong bài luận văn để đánh giá định lượng khả năng chống mất mát gói tin của thuật toán mã hóa – giải mã trong iLBC Codec và sẽ được nói cụ thể hơn ở chương 3.

CHƯƠNG 2 – ILBC CODEC

2. 1. Giới thiệu về iLBC Codec và kỹ thuật xử lý tiếng nói dựa trên mã hóa dự đoán tuyến tính

2.1.1. Giới thiệu iLBC Codec

Kể từ khi hệ thống VoIP được triển khai trong thực tế, đã có rất nhiều Codec được sử dụng trong các môi trường thoại băng tần rộng (Wide band) và hẹp (Narrow band). Sự phát triển của công nghệ đòi hỏi các yếu tố cấu thành nên một hệ thống thông tin thoại cũng phải phát triển theo, trong đó có các giải thuật mã hóa – giải mã tín hiệu thoại hay còn gọi là Codec. Để một hệ thống hoạt động trơn tru, ổn định, đòi hỏi các Codec phải đáp ứng được yêu cầu về việc tiêu tốn băng thông thấp nhưng chất lượng dịch vụ phải cao.

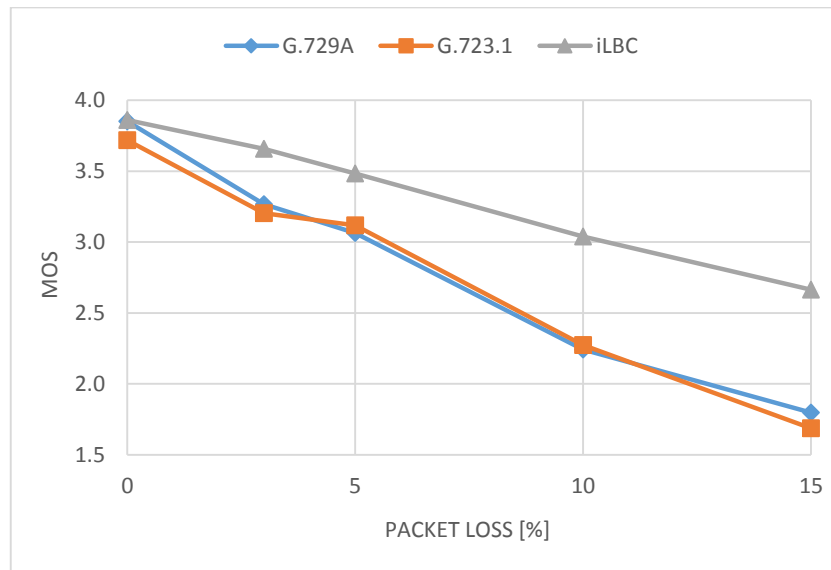
Để giải quyết các yêu cầu này, xuất hiện các Codec tốc độ bit thấp như Speex, G.723.1, G.729A,... Hầu hết các Codec đều dựa trên kỹ thuật dự đoán tuyến tính kích thích mã CELP (Code Excited Linear Prediction), ví dụ theo chuẩn ITU-T, có các Codec như ITU G.729A, G.723.1, GSM-EFR, 3GPP-AMR. Kỹ thuật CELP cho kết quả mã hóa ở tốc độ bit thấp và chất lượng rất tốt dành cho các mạng chuyển mạch. Có được tính hiệu quả này là do kỹ thuật CELP sử dụng phương pháp phụ thuộc frame giữa các đoạn tín hiệu thoại gần kề nhau. Hiệu năng của kỹ thuật CELP cũng do đó mà phụ thuộc nhiều vào việc mã hóa các frame đã nhận trước đó. Ngoài ra, kỹ thuật này còn phải dựa trên bộ ghi nhớ trong quá trình mã hóa, và bị lan truyền lỗi khi gói tin bị mất hoặc bị trễ trên đường truyền. Đây chính là một nhược điểm lớn của các hệ thống truyền tin chuyển mạch gói, vì rõ ràng chỉ cần một gói tin đơn lẻ bị mất là ảnh hưởng đến chất lượng của rất nhiều các gói tin về sau.

Để khắc phục nhược điểm vừa nêu trên của các Codec sử dụng kỹ thuật CELP, tổ chức Global IP Sound (GIPS) đã nghiên cứu và cho ra đời iLBC Codec. iLBC được tổ chức IETF (Internet Engineering Task Force) thông qua vào tháng 3/2002.

Một số cột mốc quan trọng trong quá trình phát triển của iLBC Codec:

- Bắt đầu được phát triển từ năm 2000.
- Đóng vai trò là bản nháp internet của tổ chức IETF vào tháng 2 năm 2002.
- Có khả năng hỗ trợ các frame với kích cỡ 20ms năm 2003.
- Vào tháng 4/2004, iLBC chính thức được sử dụng hợp lệ, lâu dài trong CableLabs PacketCableTM 1.1 Audio/Video Codec Specification dành cho các thiết bị đa phương tiện đầu cuối và các media gateway.
- Tháng 12 năm 2004 được IETF hoàn thiện (trở thành bản tài liệu RFC 3951 và 3952).
- Tổ chức Cable Television Laboratories (CableLabs®) cũng đã thông qua chính sách coi iLBC là một chuẩn mã hóa và giải mã (Codec) dành cho hệ thống VoIP thông qua các ứng dụng sử dụng đường Cab.

Hình dưới đây mô tả kết quả của quá trình đánh giá bởi tổ chức Dynastat, so sánh iLBC Codec ở frame 30ms với hai codec tiêu chuẩn khác là G.729A và G.723.1:



Hình 2.1 – So sánh chất lượng thoại của iLBC Codec và G.729A, G.723.1 Codec trong điều kiện mất mát gói tin

Kết quả cho thấy rõ sự vượt trội của iLBC Codec khi sử dụng trong thực tế, trong điều kiện mất mát gói tin càng lớn thì chất lượng iLBC mang lại càng cao hơn so với các Codec khác. Không những vậy, trong điều kiện không bị mất gói (điều kiện gần như lý tưởng) thì iLBC cũng bằng, thậm chí là tốt hơn so với các Codec kia.

Điều kiện: Kênh truyền sạch (lý tưởng)				
iLBC	> G.723.1	>= G.729	>= G.729A	<= G.729E
Điều kiện: Mất gói tin				
iLBC	>> G.723.1	>> G.729	>> G.729A	> G.729E

Hình 2.2 – Kết quả đánh giá iLBC và các Codec thoại khác trong từng điều kiện

iLBC (internet Low Bitrate Codec) là một bộ mã hóa – giải mã tín hiệu âm thanh thoại, thích hợp sử dụng trong các dịch vụ thoại trên nền IP. Codec giúp cải thiện chất lượng thoại tốt hơn trong mạng IP khi xảy ra các vấn đề về trễ hoặc mất gói tin, bằng tính năng chống mất mát gói tin trên đường truyền. iLBC được thiết kế cho môi trường thoại băng tần hẹp, với tần số lấy mẫu là 8kHz. Nó cũng sử dụng phương pháp mã hóa dự đoán tuyến tính độc lập theo khối (Block-independent Linear Predictive Coding) và hỗ trợ hai loại frame của chuỗi tín hiệu với kích cỡ: 20ms ở tốc độ 15,2kbps và 30ms ở tốc độ 13,33kbps.

- Khi Codec hoạt động ở frame 20ms, mỗi frame sẽ chứa 304 bit (38 bytes).
- Khi Codec hoạt động ở frame 30ms, mỗi frame sẽ chứa 400 bit (50 bytes).

Hai kiểu frame tuy khác nhau về kích cỡ nhưng cơ chế hoạt động tương tự nhau. Trong trường hợp có sự khác nhau, để phân biệt, nó sẽ được ký hiệu x/y, với x đại diện cho frame 20ms và y đại diện cho frame 30ms.

Khác với phương pháp dự đoán dài hạn LTP (Long-Term Predictive) và kỹ thuật CELP sử dụng trong các Codec thoại hiện tại, iLBC Codec thực hiện mã hóa độc lập phần tín hiệu lỗi (residual signal) của mã dự đoán tuyến tính LPC theo từng frame một. Codec kiểm soát được sự suy giảm chất lượng tín hiệu thoại đối với sự gia tăng về độ trễ hoặc mất gói tin trên đường truyền. Điều này khác với hoạt động của các Codec dựa trên kỹ thuật CELP, khi mà các Codec này chỉ được thiết kế cho việc chống chịu lỗi mất bit thay vì mất mát gói tin.

Một phương pháp đo đặc thích hợp cho Codec nói chung, được sử dụng trong trường hợp mất gói tin là số lượng frame/packet sử dụng cho việc khôi phục gói tin bị mất. Nhưng trong iLBC Codec, giá trị này bằng 0. Gói tin gần nhất với gói tin bị mất luôn luôn được giải mã, mà không phải phụ thuộc vào bất cứ gói tin nào xung quanh khác.

Trong khi iLBC là một Codec thoại sử dụng ở băng tần hẹp, sử dụng toàn bộ dải băng tần sẵn có 4kHz, thì các Codec tốc độ bit thấp khác chỉ sử dụng dải băng tần từ 300Hz đến 3400Hz. Điều này cho thấy iLBC tận dụng được băng tần tốt hơn các Codec khác. Thêm vào đó, tính chất tần số của của tín hiệu thoại khi được mã hóa bởi iLBC được bắt chước một cách chính xác tính chất của tín hiệu ban đầu, do đó kết quả thu được là âm thanh tự nhiên và chính xác hơn so với các Codec thoại khác.

Trong hệ thống mã hóa thoại, thuật toán của Codec sử dụng một bộ đáp ứng để kiểm soát sự mất mát gói tin tương tự như thuật toán mã hóa PCM với tính năng bù gói tin bị mất (theo như chuẩn ITU-T G.711, hoạt động ở bit rate cố định 64 kbps). Thuật toán cho phép mã hóa ở tốc độ bit cố định, với sự thống nhất về mặt chất lượng tín hiệu và tốc độ bit gần như tốt nhất. Với những ưu điểm này, các ứng dụng cho thuật toán mã hóa iLBC thường là các ứng dụng về giao tiếp thời gian thực như điện thoại, hội thảo truyền hình, truyền thanh trực tiếp, lưu trữ.

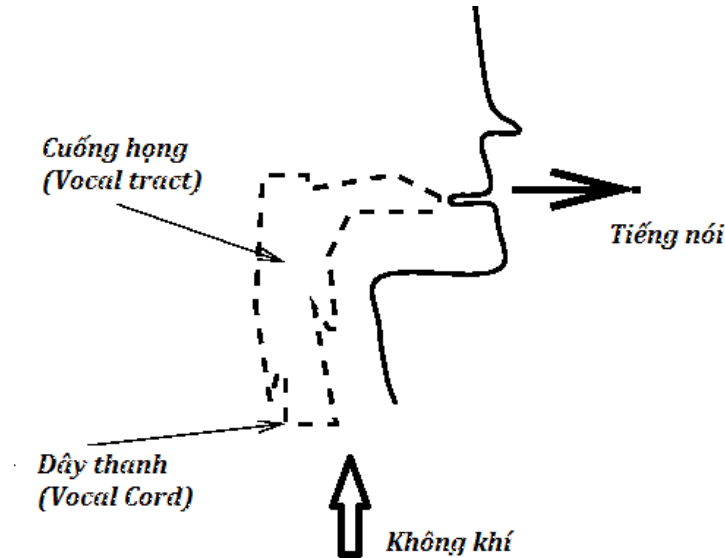
2.1.2. Kỹ thuật xử lý tiếng nói dựa trên mã hóa dự đoán tuyến tính

a. Mô hình dự đoán tuyến tính

Quá trình tạo nên tiếng nói của con người khi phát âm:

- Không khí được đẩy từ phổi xuyên qua tuyến âm và ra khỏi miệng từ đó hình thành nên tiếng nói.
- Với những âm hữu thanh, các dây thanh dao động (đóng và mở). Tốc độ dao động của dây thanh xác định pitch (âm sắc) của giọng nói. Phụ nữ và trẻ em thường có pitch cao (dao động dây thanh nhanh), trong khi đó đàn ông trưởng thành thường có pitch thấp (dao động dây thanh chậm).
- Với những âm vô thanh (phụ âm xát và âm bật) thì dây thanh không dao động nhưng những phần còn lại thì luôn luôn mở.

- Hình dạng của tuyến âm sẽ xác định âm thanh được tạo ra. Khi phát âm, tuyến âm sẽ thay đổi hình dạng của nó để tạo ra các âm thanh khác nhau.
- Hình dạng của tuyến âm thay đổi tương đối chậm (thay đổi từ 10ms đến 100ms).
- Thông lượng không khí đến từ phổi sẽ xác định độ to nhỏ của âm thanh.



Hình 2.3 - Mô hình vật lý của cơ quan phát âm người

Từ nguyên lý tạo tiếng nói của con người như trên, một mô hình toán học với kỹ thuật mã hóa dự đoán tuyến tính được dùng để mô phỏng việc tạo tiếng nói.

Mối quan hệ giữa mô hình vật lý và mô hình toán học:

Bộ máy phát âm (Cường họng)	<=>	$H(z)$ (Bộ lọc LPC)
Không khí	<=>	$u(n)$ (Kích thích)
Dao động của dây thanh âm	<=>	V (Âm hữu thanh)
Chu kỳ dao động của dây thanh âm	<=>	T (Chu kỳ cao độ)
Phụ âm sát và phụ âm bật	<=>	UV (Âm vô thanh)
Độ lớn không khí	<=>	G (Độ lợi)

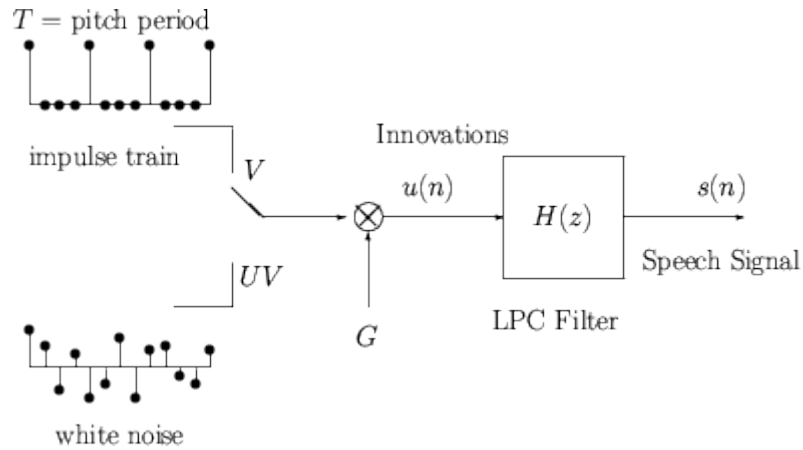
Bộ lọc LPC được xác định bởi hàm truyền đạt:

$$H(z) = \frac{1}{1 - a_1z^{-1} - a_2z^{-2} - \dots - a_{10}z^{-10}}$$

Khi đó, mối quan hệ giữa đầu vào và đầu ra của bộ lọc được xác định bởi phương trình sai phân tuyến tính:

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n)$$

Như vậy, các tham số cần thiết cho một bộ tổng hợp LPC là các hệ số LPC (đặc trưng cho bộ lọc LPC và tín hiệu nguồn kích thích).

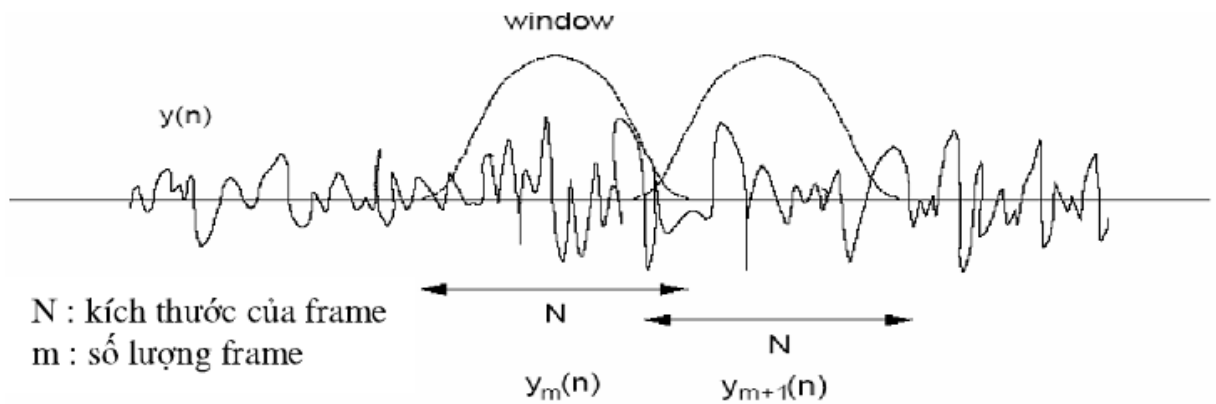


Hình 2.4 - Mô hình tổng hợp tiếng nói bằng phương pháp LPC

b. Kỹ thuật phân khung Framing, Overlap-Adding trong xử lý tín hiệu tiếng nói

Kỹ thuật này thực hiện chức năng phân khung, tức là chia chuỗi tín hiệu đầu vào thành các khung (frame) có kích thước cố định N (thường là 160, 180, 256, 512 hoặc 1024 mẫu tín hiệu).

Để cho tín hiệu được xử lý liên tục không gián đoạn và tín hiệu ra mượt mà hơn thì các khung thường gối chồng lên nhau, các khung kề nhau chồng lên nhau ở M mẫu tín hiệu, tức là 2 khung liên tiếp thì trùng nhau M mẫu tín hiệu. Nếu M càng lớn thì tín hiệu đầu ra sẽ càng trơn, mịn hơn. Thường thì chọn M bằng $1/2$, $2/3$ hoặc $3/4$ chiều dài của khung N .



Hình 2.5 - Phân tích tín hiệu tiếng nói thành các frame

c. Phương pháp phân tích mã hóa dự đoán tuyến tính

Tín hiệu tiếng nói thay đổi theo thời gian, ứng với một đoạn tiếng nói ngắn (gọi là segment hay frame), tiếng nói được xem là tín hiệu dừng. Nói cách khác, mô hình bộ máy phát âm là không đổi trên mỗi segment. Thông thường, mỗi segment có chiều dài 20ms, nếu tiếng nói được lấy mẫu tại tần số 8kHz thì số mẫu trong một segment là 160 mẫu [2, tr.10].

Xét một frame tiếng nói: Đối với mô hình mô phỏng tiếng nói dựa trên kỹ thuật dự đoán tuyến tính ở trên, bộ máy phát âm được xem như bộ lọc toàn cực với đầu vào bộ lọc là một chuỗi nhiễu trắng hay là một dãy xung tựa tuần hoàn; đầu ra bộ lọc là tiếng nói số. Bộ lọc này là bộ lọc đệ quy nhưng chỉ lấy đầu vào là mẫu âm kích thích

hiện tại $u(n)$ để tính thay vì lấy m mẫu quá khứ của $u(n)$. Tiếng nói đầu ra được mô tả bằng công thức [2, tr.10]:

$$s(n) = Gu(n) - \sum_{k=1}^M a_k s(n-k)$$

Với a_k là hệ số dự đoán tuyến tính. M là bậc dự đoán.

Gọi $\bar{s}(n)$ là mẫu hiện tại của tiếng nói được dự đoán tuyến tính từ M mẫu quá khứ của tiếng nói.

$$\bar{s}(n) = \sum_{k=1}^M a_k s(n-k)$$

Như vậy, độ sai lệch $e(n)$ giữa mẫu tiếng nói thực và mẫu dự đoán sẽ là:

$$e(n) = s(n) - \bar{s}(n) = s(n) - \sum_{k=1}^M a_k s(n-k)$$

Do đó, tổng của sai lệch dự đoán bình phương của cả frame được tính như sau:

$$E = \sum_n e^2(n) = \sum_n \left(s(n) - \sum_{k=1}^M a_k s(n-k) \right)^2$$

- Xác định các thông số bộ lọc và độ lợi:

Để xác định các hệ số bộ lọc dự đoán tuyến tính, ta phải tối thiểu hóa tổng sai lệch bình phương E bằng cách thiết lập đạo hàm của E đối với a_k (hệ số dự đoán tuyến tính) bằng phương trình:

$$\frac{\partial E}{\partial a_k} = 0 \text{ với } 1 \leq k \leq M$$

Hệ số bộ lọc a_k được giải bằng thuật toán **Levinson-Durbin**. Giải thuật này thực hiện việc tìm bộ dự đoán bậc thứ M từ bậc dự đoán thứ $M-1$. Đây là quá trình lặp lại đệ quy cho đến khi tìm được lần đầu tiên bộ dự đoán bậc zero, sau đó sẽ dùng bộ bậc zero để tính bộ dự đoán bậc 1 và quá trình tiếp tục cho đến khi tính toán được bộ dự đoán có bậc cần tìm.

Độ lợi G (Gain) được xác định bởi năng lượng RMSE (Root Mean Square Energy) của từng khung tín hiệu nguồn, được tính theo công thức:

$$E_n = \left[\frac{1}{180} \sum_{i=0}^{179} u_n^2(i) \right]^{1/2}$$

Với E_n là năng lượng RMSE của frame thứ n .

$u_n(i)$ là mẫu tín hiệu nguồn thứ i trong frame thứ n .

- Xác định âm hữu thanh/vô thanh (Voiced/Unvoiced) và xác định chu kỳ cao độ (pitch period):

Âm hữu thanh/vô thanh (gọi chung là Voicing) và pitch được xác định dựa trên dạng sóng của tín hiệu nguồn.

Pitch sẽ được xác định cho từng frame bằng cách tối thiểu hoá hàm vi sai biên độ trung bình AMDF (Average Magnitude Difference Function). Ở đây hàm vi sai biên độ trung bình sẽ được tính theo tín hiệu nguồn trong một frame dữ liệu và chu kỳ τ nằm trong khoảng xác định của pitch:

$$AMDF(\tau) = \frac{1}{T} \int_0^T |u(t) - u(t - \tau)| dt$$

Với $u(t)$ là tín hiệu nguồn của tín hiệu tiếng nói ban đầu.

$\tau \in [20;156]$ là khoảng xác định của pitch.

Sau khi có tập giá trị hàm AMDF, pitch và voicing sẽ được xác định dựa trên các giá trị AMDF nhỏ nhất, khoảng cách giữa 2 giá trị AMDF nhỏ nhất sẽ cho chu kỳ pitch, và khi đó âm hữu thanh (voiced) là 1. Ngược lại nếu pitch không tồn tại thì âm vô thanh (unvoiced) là 0.

- Bộ mã hóa LPC -10:

Trong bộ mã hóa LPC-10, tín hiệu tiếng nói được chia thành khung có chiều dài 20 ms hay 160 mẫu với tần số lấy mẫu là 8kHz. Bậc của bộ lọc LPC là 10. Thông thường, 10 thông số bộ lọc dự đoán tuyến tính được chuyển sang hệ số cặp phổ vạch LSP (Line Spectrum Pair) tương đương vì LSP có độ ổn định cao hơn.

Quá trình mã hóa LPC sẽ tính toán và truyền đi các thông số bộ lọc, dấu hiệu xác định hữu thanh/vô thanh và chu kỳ pitch của frame tiếng nói đó.

Vì tốc độ của bộ mã hóa LPC-10 là 2.4kbps, do vậy, số bit cần dùng để mã hóa 1 frame là: $2400 * 0.02 = 48$ bit.

Tóm lại, quá trình phân tích và tổng hợp theo LPC-10e được chia làm 2 quá trình như sau:

- Phân tích LPC: Từ tín hiệu tiếng nói ban đầu S , tìm ra vector tham số A tốt nhất mà đặc trưng cho tín hiệu S .
- Tổng hợp LPC: Từ vector tham số A , tổng hợp thành tín hiệu tiếng nói S .

d. Phương pháp mã hóa CELP

Mã hoá dự đoán tuyến tính kích thích mã (hay bằng vector) CELP (Code-Excited Linear Predictive) là phương pháp mã hoá tiếng nói ở tốc độ bit thấp nhưng tiếng nói tổng hợp vẫn có chất lượng cao tương đương với các bộ mã hoá dạng sóng có tốc độ bit trung bình, tuy nhiên độ phức tạp của bộ mã hoá CELP sẽ cao hơn.

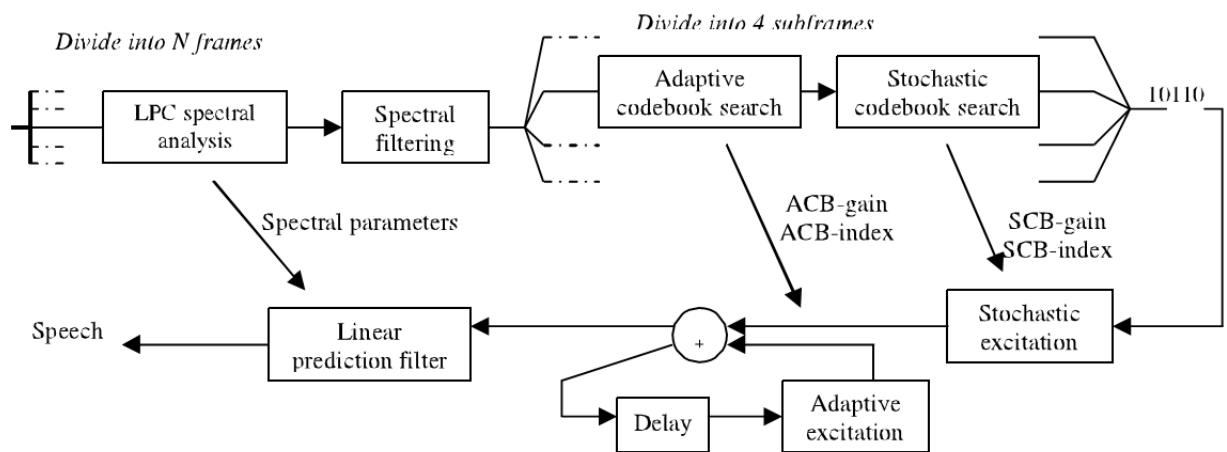
Trong thuật toán CELP sử dụng khái niệm codebook. Đây là nơi chứa các vector (hay từ mã) kích thích và mỗi vector đó có độ dài cố định. Các vector này được tạo thành từ các dãy nhiễu Gauss. Tín hiệu kích thích cho mỗi đoạn tiếng nói sẽ được tìm kiếm trong codebook. Do đó, mỗi vector kích thích trong codebook sẽ có số phần tử bằng với số mẫu trong một đoạn tiếng nói. Ngoài codebook chứa các vector tín hiệu kích thích thường được gọi là stochastic codebook, còn có một codebook khác liên quan đến bộ dự đoán dài hạn LTP gọi là codebook thích nghi (adaptive codebook). Codebook thích nghi chứa các giá trị độ trễ τ (có giá trị từ 20 - 147). Việc tìm giá trị τ

và độ lợi pitch β cho bộ lọc dài hạn LTP (Long-Term Predictive) là quá trình tìm một giá trị thích hợp sao cho tín hiệu kích thích tổng hợp đưa vào bộ lọc ngắn hạn STP (Short-Term Predictive) là tốt nhất, tức là làm cho sai số dự đoán là nhỏ nhất [8, tr.299-301].

Ở phương pháp này, bộ mã hóa và giải mã có chung một codebook chứa các tín hiệu kích thích ngẫu nhiên (white noise Gauss). Với mỗi tín hiệu tiếng nói, bộ phát sẽ tìm trong bảng mã tín hiệu kích thích lên bộ lọc LPC cho tín hiệu gần giống nhất, sau đó chỉ việc truyền chỉ số của tín hiệu kích thích này đến bộ thu. Bộ thu dựa vào chỉ số này để tìm ra tín hiệu kích thích thích hợp. CELP đòi hỏi hơn 500 triệu phép tính nhân và cộng trong 1s, cho chất lượng tiếng nói cao cùng với việc kích thích được mã hóa 0.25 bit/mẫu và đạt tốc độ 4.8kbps.

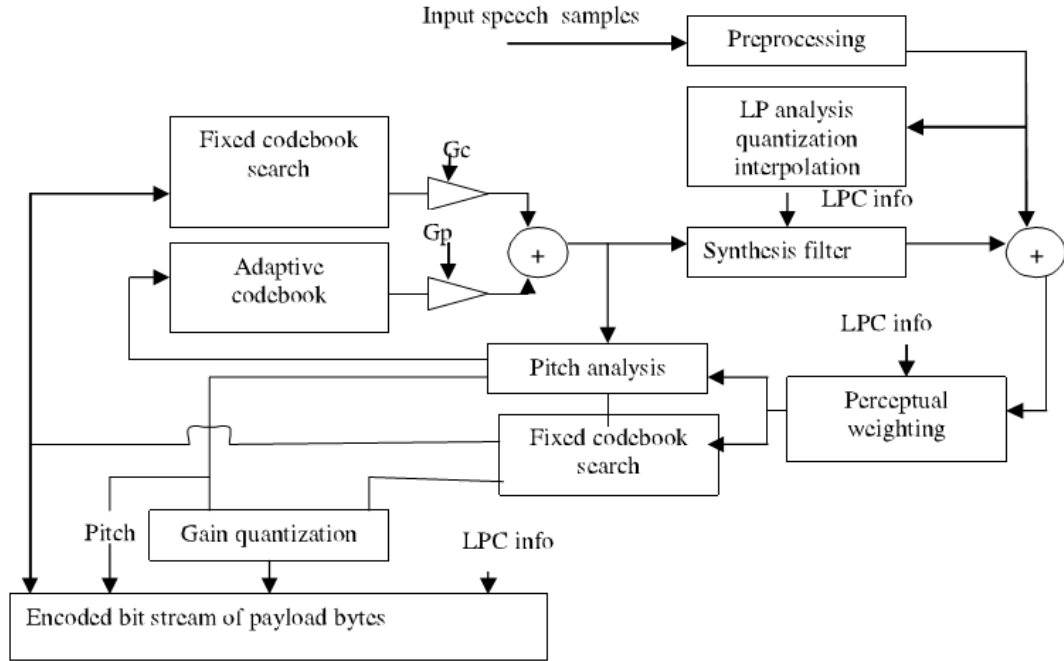
Nguyên tắc mã hóa CELP cũng tương tự như mã hóa LPC trong phương pháp mã hóa Vocoder, chỉ khác ở một số điểm sau: Kích thước frame của CELP là 30ms, tương ứng với 240 mẫu; CELP thực hiện mã hóa trực tiếp tín hiệu, đòi hỏi nhiều bit cho quá trình mã hóa đồng thời độ phức tạp của tính toán cũng cao hơn LPC; CELP có sử dụng thêm bộ lọc chu kỳ pitch và lượng tử hóa vector (Khác với phép lượng tử thông thường mỗi mẫu biên độ tương ứng với một giá trị lượng tử hay một tổ hợp bit để biểu diễn, trong lượng tử vector một nhóm mẫu biên độ ứng với một nhóm mẫu lượng tử mới được tương ứng với một tổ hợp bit được gọi là một vector lượng tử).

Thuật toán CELP được mô tả như *Hình 2.6* [2, tr.12]:



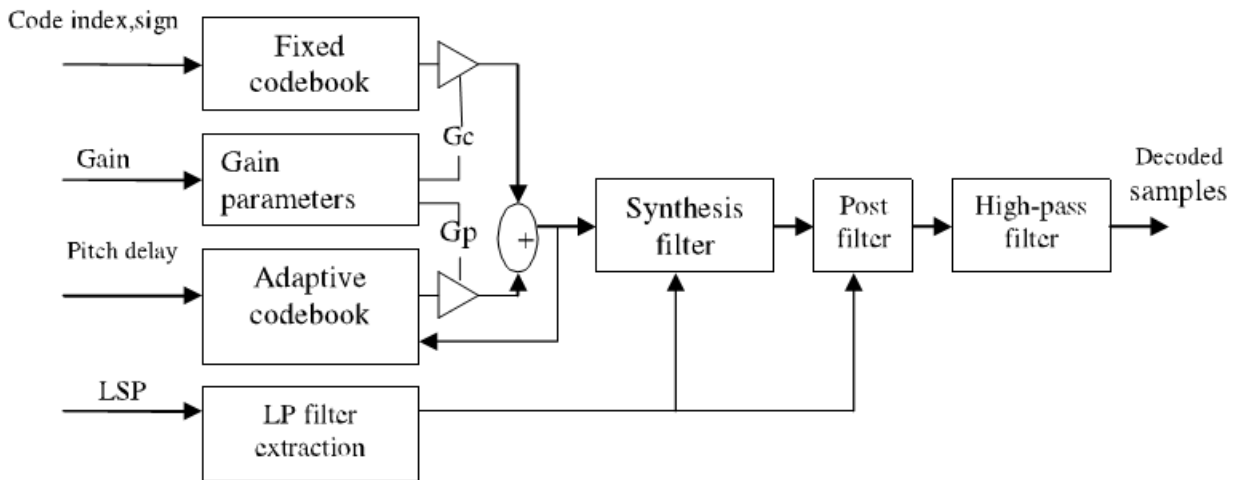
Hình 2.6 - Thuật toán CELP

- Mô hình mã hóa CELP [2]



Hình 2.7 - Bộ mã hóa CELP

- Mô hình giải mã CELP [2]

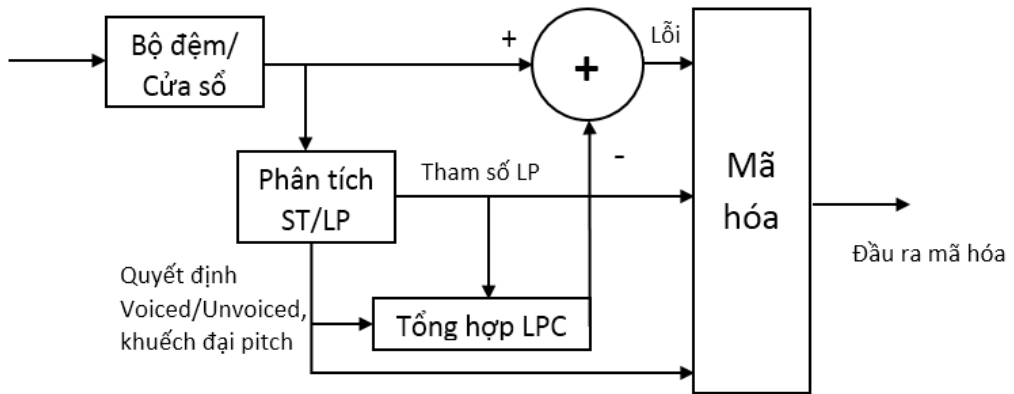


Hình 2.8 - Bộ giải mã CELP

e. Mã hóa dự đoán tuyến tính kích thích bằng lỗi RELP

Mã hóa dự đoán tuyến tính kích thích bằng lỗi RELP (Residual-Excited Linear Predictive) là phương pháp mã hóa thuộc kiểu LPC Vocoder, nhưng có liên quan tới kỹ thuật DPCM trong mã hóa dạng sóng. Trong kỹ thuật mã hóa này, sau khi ước lượng các thông số mô hình (hệ số dự đoán tuyến tính LP) và các thông số kích thích (quyết định voiced/unvoiced, pitch, hệ số khuếch đại) từ một khung tiếng nói, tiếng nói được tổng hợp tại bộ phát và trừ đi tín hiệu gốc để có được tín hiệu lỗi (residual signal). Tín hiệu lỗi sẽ được lượng tử, mã hóa và phát đến bộ thu cùng với các thông số LPC. Tại bộ thu, tín hiệu lỗi được cộng với tín hiệu tổng hợp. Chất lượng tiếng nói cũng nhờ vào việc cộng thêm lỗi đó mà được nâng cao hơn.

Sơ đồ thuật toán mã hóa RELP:



Hình 2.9 – Thuật toán RELP

2. 2. Quá trình mã hóa iLBC Codec (Encoder)

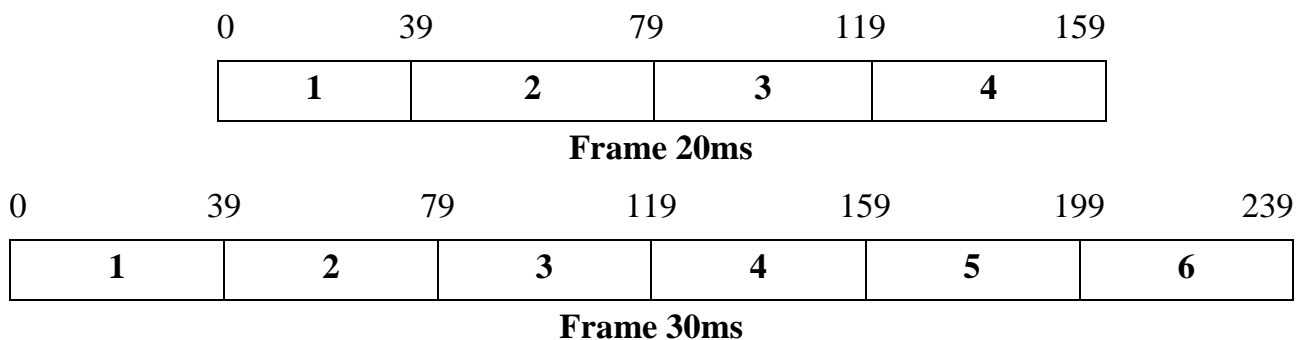
2.2.1. Tổng quan về quá trình mã hóa iLBC Codec

Bản chất của thuật toán mã hóa iLBC là mã hóa dự đoán tuyến tính LPC (Linear Predictive Coding) và mã hóa theo block của các tín hiệu lỗi (residual signal) sau khi đã thực hiện mã hóa dự đoán tuyến tính [7, tr.23-24]. Đối với mỗi block gồm 160/240 mẫu (tương ứng với frame kích cỡ 20ms/30ms), các bước sau đây sẽ được thực hiện:

- Một tập các bộ lọc LPC được tính toán để lọc tín hiệu thoại, sau đó đưa ra những residual signal. Các thành phần có năng lượng lớn của residual signal trong block được lượng tử.
- Trạng thái chủ yếu có chiều dài 57/58 mẫu (20ms/30ms) và tạo ra một trạng thái khởi đầu cho các từ mã động (dynamic codebook) được xây dựng từ các phần đã được mã hóa của residual signal. Những từ mã động này được sử dụng để mã hóa những phần còn lại của residual signal.
- Theo phương thức này, các block sẽ được mã hóa độc lập, giúp loại trừ được sự lan truyền của mất mát gói tin. Điều này sẽ giúp chống mất mát gói tin ở điều kiện cao nhất.

Đầu vào của bộ mã hóa theo tiêu chuẩn là tín hiệu PCM được lấy mẫu đều ở tần số 8kHz, mã hóa mỗi mẫu với 16 bit. Tín hiệu được chia thành các block chứa 160/240 mẫu tương ứng với các frame là 20ms/30ms.

Mỗi block lại được chia thành 4/6 block con (sub-block) liên tiếp nhau (tương ứng với frame 20ms/30ms), mỗi block con có 40 mẫu.



Hình 2.10 – Phân chia mẫu trong một frame của thuật toán iLBC

Đối với mỗi tín hiệu frame 30ms, bộ mã hóa thực hiện hai bước phân tích mã hóa dự đoán tuyến tính. Bước phân tích đầu tiên áp dụng trong phạm vi từ sub-block thứ hai kéo dài tới giữa sub-block thứ năm. Bước phân tích LPC thứ hai áp dụng từ sub-block thứ năm cho tới cuối sub-block thứ sáu.

Đối với tín hiệu frame size 20ms, một bước phân tích mã hóa dự đoán tuyến tính được thực hiện qua sub-frame thứ ba.

Đối với mỗi bước phân tích LPC, thu được một tập các đường phổ tần LSFs (Line Spectral Frequency), được lượng tử hóa, nội suy để thu được các hệ số LSF cho mỗi sub-block. Sau đó, việc mã hóa dự đoán tuyến tính phần residual signal được tính toán bằng cách sử dụng các bộ lọc phân tích được lượng tử và nội suy.

Hai sub-block kề nhau của residual signal biểu thị năng lượng với trọng số lớn nhất được chọn. Trong hai sub-block này, trạng thái khởi đầu được chọn từ 57/58 mẫu đầu tiên hoặc 57/58 mẫu cuối cùng của hai sub-block kề nhau. Đoạn được chọn là một trong những đoạn có năng lượng lớn hơn. Trạng thái khởi đầu được mã hóa cùng với lượng tử vô hướng.

Một thủ tục mã hóa bảng mã động (dynamic codebook) được dùng để mã hóa:

- 23/22 mẫu còn lại (tương ứng với 20/30ms) trong hai sub-block chứa trạng thái khởi đầu.
- Các sub-block sau trạng thái khởi đầu.
- Các sub-block trước trạng thái khởi đầu.

Do đó, phần mã hóa có thể là 23/22 mẫu còn lại của hai sub-block kề nhau chứa trạng thái khởi đầu, hoặc là một sub-block chứa 40 mẫu. Các mẫu này có thể được đánh chỉ mục trước hoặc sau, phụ thuộc vào vị trí của trạng thái khởi đầu.

Từ mã (codebook coding) sẽ dựa trên bảng mã thích ứng (adaptive codebook, được xây dựng từ bộ nhớ bảng mã) chứa các kích thích mẫu LPC được giải mã từ các phần đã mã hóa của block). Những mẫu này được đánh chỉ mục cùng một thời điểm với target vector (vector đích, mô tả frame của âm thoại), kết thúc tại mẫu đứng ngay trước mẫu đầu tiên được mô tả trong target vector. Codebook được sử dụng trong 3 giai đoạn theo phương pháp lọc liên tục, cho kết quả là 3 code vector gains (độ lợi dự đoán) được mã hóa lần lượt với 5 bit, 4 bit, và 3 bit lượng tử vô hướng.

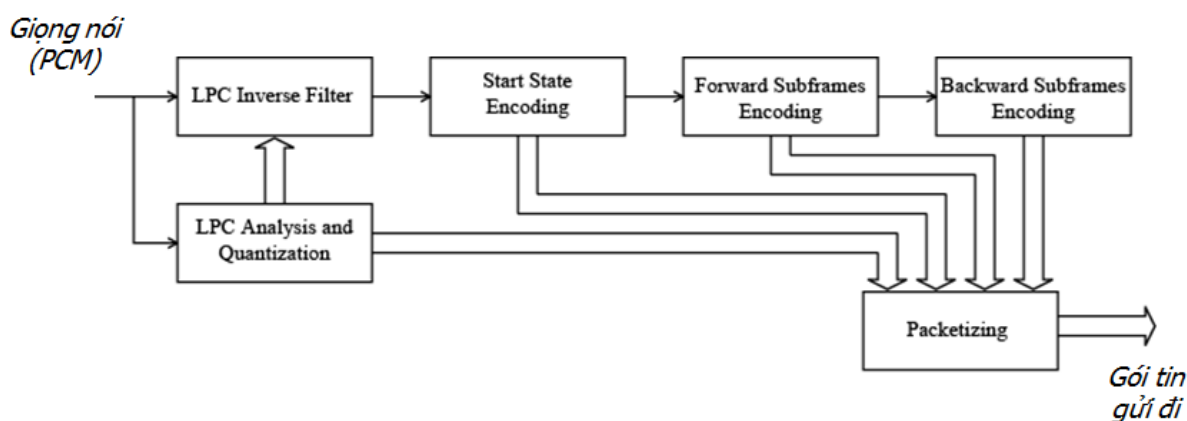
Phương thức tìm codebook mã sử dụng kỹ thuật “noise shaping” (kỹ thuật làm tăng dải động trong miền tần số phổ biến mà tai người có thể nghe) lấy từ bộ lọc LPC, và tiêu chí chủ yếu là giảm thiểu sai số trung bình bình phương MMSE (Minimum of Mean Squared Error) giữa target vector và các code vector. Mỗi code vector trong codebook này thuộc một trong 2 codebook. Phần đầu tiên gồm các residual vector đã được mã hóa. Các code vector của codebook thứ hai được xây dựng bằng sự kết hợp tuyến tính giữa các vector trong phần codebook đầu tiên.

Việc mã hóa codebook kết hợp với sai số trung bình bình phương nhằm tạo ra một tín hiệu được mã hóa ít năng lượng hơn tín hiệu được lượng tử hóa vô hướng ở trạng thái khởi đầu, phương thức để tăng khuếch đại đó được thực hiện bằng cách tìm

kiểm, lọc ra tập các codebook tăng về năng lượng kết hợp sau khi mã hóa. Điều này được thực hiện bằng cách tìm kiếm một gain factor (độ lợi) có giá trị lớn hơn trong codebook ở trạng thái đầu tiên, cũng như mở rộng các codebook gains ở bước tiếp theo bởi codebook gain ở trạng thái đầu tiên.

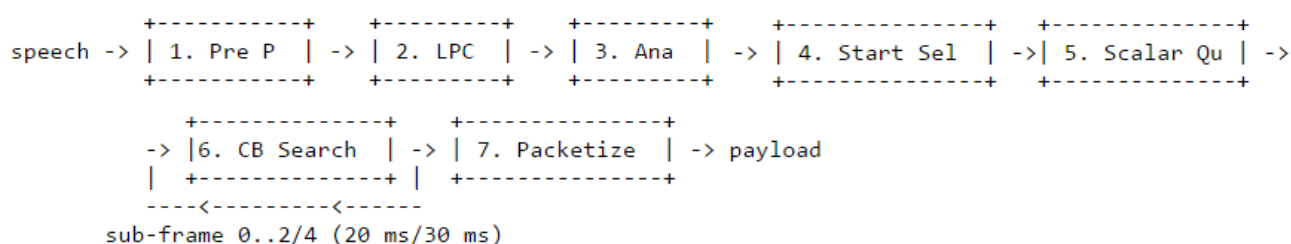
2.2.2. Các nguyên tắc mã hóa

iLBC Codec được mã hóa theo sơ đồ *Hình 2.11* [5]:



Hình 2.11 – Sơ đồ mã hóa của iLBC Codec

Cụ thể, quá trình mã hóa sẽ bao gồm các bước sau:



Hình 2.12 – Các bước mã hóa của iLBC Codec

Các bước thực hiện quá trình mã hóa sẽ theo thứ tự: Âm thoại được đưa vào (1) bộ tiền xử lý Pre Processing, được (2) mã hóa dự đoán tuyến tính LPC, (3) phân tích Analysis, tiếp đó là (4) lựa chọn trạng thái khởi đầu Start State Selection, (5) lượng tử vô hướng Scalar Quantization, sau đó được (6) tìm kiếm CodeBook Search, và cuối cùng (7) đóng gói Packetize để cho ra gói dữ liệu Payload.

Bước 1: Tiền xử lý tín hiệu với bộ lọc High-Pass (nếu cần):

Trong một số trường hợp, tín hiệu thoại thu được bao gồm cả tín hiệu điện một chiều hoặc nhiễu 50/60Hz. Nếu những tín hiệu này không được xóa đi trước khi tới bộ mã hóa, chúng sẽ được xóa bởi bộ lọc High-Pass.

Bước 2: Tính toán, phân tích các tham số LPC, lượng tử hóa, và nội suy:

Đầu vào bộ phân tích LPC là một tín hiệu thoại nằm trong bộ đệm, đã được lọc High-Pass, chứa 240/300 mẫu (mẫu loopback + mẫu của block = 80/60 + 160/240 = 240/300), với mẫu từ 0-79 (0-59) là của block trước đó và mẫu từ 80-239 (60-299) là của block hiện tại. Không sử dụng các mẫu phần look-ahead cho block tiếp theo. Khi block đầu tiên được xử lý, các mẫu trước đó được giả sử bằng 0.

Đối với mỗi block đầu vào, bộ phân tích LPC sẽ tính toán một/hai tập hệ số bộ lọc LPC sử dụng phương thức tự tương quan và giải thuật đệ quy Levinson-Durbin. Những hệ số này được chuyển đổi sang dạng các đường phổ tần LSFs (Line Spectrum Frequency). Trong trường hợp frame 20ms, một tập LSF duy nhất sẽ mô tả đặc tính của tần số được tính từ giữa sub-block thứ ba. Đối với frame 30ms, sẽ có hai tập, tập đầu tiên $lsf1$ mô tả thuộc tính tần số của tín hiệu đầu vào từ giữa sub-block thứ hai, và tập còn lại $lsf2$, mô tả đặc tính tần số được tính từ giữa sub-block thứ năm.

Sở dĩ có sự khác nhau giữa số tập hợp của frame 20ms và 30ms, đó là: một tập các hệ số tự tương quan được tính toán trên số lượng 80 (loopback) + 160 (block) = 240 mẫu. Những mẫu này vừa đủ để hình thành tín hiệu ban đầu. Các hệ số tự tương quan acf (autocorrelation coefficient) cũng được tính toán trên 240 mẫu đầu vào. Như vậy chỉ cần một tập lsf là đủ cho bộ phân tích LPC đối với frame 20ms.

Tóm lại, bước phân tích và lượng tử hóa này gồm có các bước chính sau:

- Tính toán các hệ số tự tương quan.
- Tính toán các hệ số LPC.
- Tính toán các hệ số LSF từ các hệ số LPC.
- Lượng tử các hệ số LSF.
- Kiểm tra tính ổn định của các hệ số LSF.
- Nội suy các hệ số LSF.

Bước 3: Sử dụng bộ lọc phân tích để tính toán các tín hiệu lỗi (residual signal):

Mỗi block chứa các mẫu tín hiệu thoại được lọc bởi bộ lọc phân tích lượng tử và nội suy LPC để thu được residual signal. Cụ thể, bộ lọc phân tích LPC tương với với mỗi sub-block 40 mẫu được sử dụng để lọc các mẫu tín hiệu của các sub-block tương tự nhau. Bộ ghi nhớ của bộ lọc ở cuối mỗi sub-block được áp dụng cho cả bộ lọc LPC của sub-block tiếp theo. Tín hiệu ở đầu ra của mỗi bộ lọc phân tích là các residual signal của các sub-block tương ứng.

Bước 4: Lựa chọn vị trí của 57/58 mẫu của trạng thái khởi đầu:

- Chọn lọc các trọng số cho từng sub-block trước khi lượng tử và mã hóa.

Bước 5: Lượng tử 57/58 mẫu trạng thái khởi đầu bằng phương pháp lượng tử vô hướng:

- Ước lượng trạng thái khởi đầu.
- Lọc All-Pass và lượng tử mở rộng (scale quantization).
- Lượng tử vô hướng.

Bước 6: Tìm codebook cho mỗi sub-frame và mã hóa các mẫu còn lại:

Bắt đầu với 23/22 mẫu, sau đó mã hóa các sub-block kế tiếp, và rồi mã hóa các sub-block trước đó. Bước này gồm một số thao tác chính như sau:

- Bảng ghi nhớ từ mã (codebook memory).
- Ước lượng trọng số của codebook memory và Target.
- Tạo Codebook: Tạo Base Codebook, mở rộng Codebook, tăng Codebook.

- Tìm kiếm Codebook: Tìm kiếm tại mỗi giai đoạn, lượng tử Gain ở mỗi giai đoạn, Chuẩn bị Target cho giai đoạn tiếp theo.

Bước 7: Mã hóa chính xác độ lợi Gain.

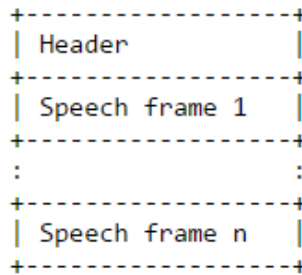
Bước 8. Định nghĩa luồng bit (Bit stream) và đóng gói (Packetizer).

Ở bước này, tín hiệu sau khi được mã hóa sẽ được đóng gói vào thành một chuỗi các gói tin, bao gồm nhiều frame ở dạng nhị phân. Phần đầu tiên của tệp tin là phần Header, chứa nội dung mang thông tin định danh file thuộc định dạng của iLBC. Header của iLBC tương ứng với chuỗi ký tự thuộc bảng mã ASCII như sau [6]:

- Đối với kiểu frame 30ms: "#!iLBC30\n", hoặc "0x23 0x21 0x69 0x4C 0x42 0x43 0x33 0x30 0x0A" theo kiểu Hexa.

- Đối với kiểu frame 20ms: "#!iLBC20\n", hoặc "0x23 0x21 0x69 0x4C 0x42 0x43 0x32 0x30 0x0A" theo kiểu Hexa.

Tiếp sau phần Header là các frame liên tiếp nhau. Thứ tự sắp xếp chuỗi bit stream iLBC như hình sau [6]:



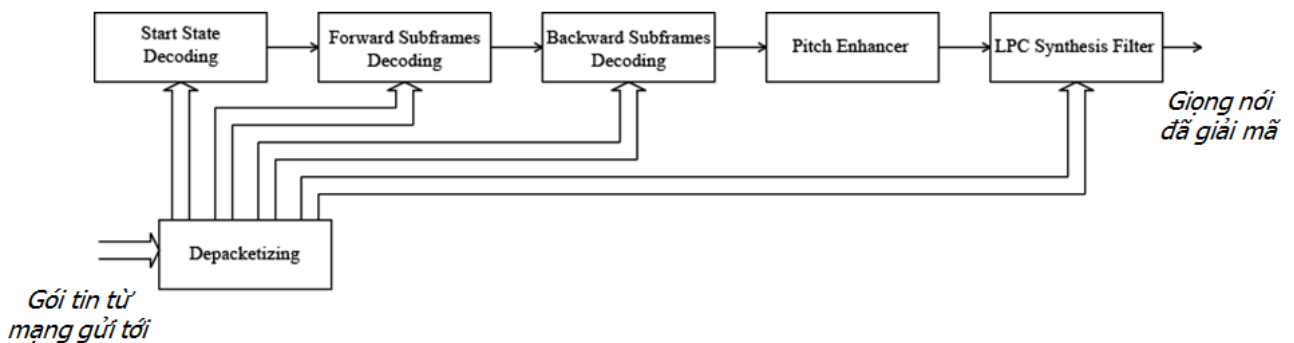
Hình 2.13 – Sắp xếp chuỗi bit đã mã hóa vào trong file nhị phân

Sau quá trình mã hóa, tín hiệu thoại ban đầu được đóng gói thành một chuỗi bit và gửi đi theo kênh truyền sang phía nhận. Quá trình gửi đi có thể sẽ bị mất mát gói tin do chất lượng đường truyền, và để khắc phục điều này, bên nhận sẽ có những cơ chế nhằm khôi phục lại gói tin bị mất, ví dụ như bù mất gói ở bước giải mã.

2. 3. Quá trình giải mã iLBC Codec (Decoder)

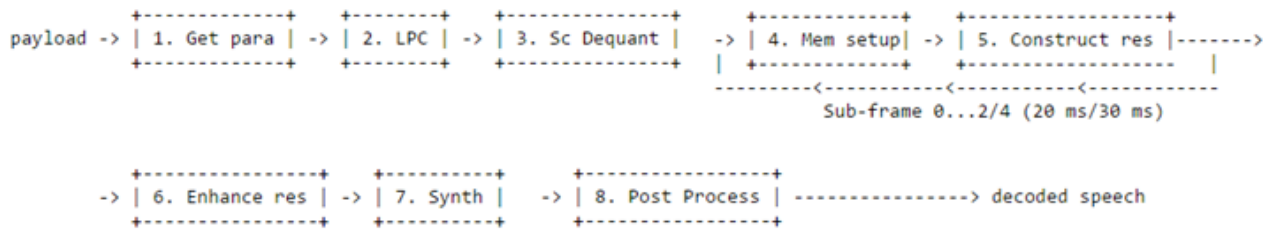
2.3.1. Tổng quan về quá trình giải mã iLBC Codec

Sơ đồ tổng quan quá trình giải mã iLBC Codec [5]:



Hình 2.14 - Sơ đồ giải mã của iLBC Codec

Cụ thể, quá trình giải mã sẽ bao gồm các bước ngược lại với tuần tự của quá trình mã hóa:



Hình 2.15 – Các bước giải mã của iLBC Codec

Các bước giải mã bao gồm:

- Trích xuất các tham số từ bit stream.
- Giải mã các mã dự đoán tuyến tính LPC và nội suy.
- Xây dựng 57/58 mẫu khởi đầu.
- Thiết lập bộ ghi nhớ (memory) bằng cách sử dụng dữ liệu từ residual signal đã giải mã. Bộ ghi nhớ này được sử dụng cho việc xây dựng codebook. Đối với các block đằng trước trạng thái khởi đầu, cả residual signal đã giải mã và target đều đảo ngược lại theo chiều thời gian. Các sub-frame được giải mã theo đúng thứ tự như khi mã hóa.
- Xây dựng các residual signal của sub-block này. Lặp lại hai bước này cho đến khi xây dựng xong residual signal của tất cả các sub-block.
- Tăng cường các giá trị residual với bộ lọc sau (post filter).
- Tổng hợp các residual signal.
- Xử lý sau (post process) với bộ lọc High-Pass (nếu cần).

Trong hệ thống truyền thông thoại bằng gói tin, thông thường sẽ sử dụng một bộ đệm đặt ở phía thiết bị nhận trong trường hợp các gói tin chứa các khối tín hiệu đã mã hóa bị mất. Tuy nhiên trong phần xử lý của iLBC Codec, sự logic này không được thực hiện. Đối với mỗi khối tín hiệu đã mã hóa nhận được, bộ giải mã sẽ thực hiện công việc giải mã. Trong quá trình giải mã, bộ giải mã sẽ thực hiện xử lý mã hóa dự đoán tuyến tính LPC cho mỗi khối tín hiệu bị mất.

Việc mã hóa cho mỗi khối tín hiệu được bắt đầu bằng cách giải mã và nội suy từ các hệ số LPC. Tiếp sau đó là quá trình giải mã trạng thái khởi đầu.

Đối với các đoạn codebook đã được mã hóa, mỗi đoạn sẽ được giải mã bằng cách sử dụng 3 code vector lấy từ codebook nhận được, tương tự như cách mà các code vector được xây dựng trong bộ mã hóa. 3 gain vector cũng được giải mã và kết quả thu được là tín hiệu đã giải mã từ tổng của 3 codebook vector.

Một giải thuật nâng cao được áp dụng cho các tín hiệu kích thích được tái tạo. Sự nâng cao này làm tăng tính tuần hoàn của vùng âm thoại, được tối ưu hóa dưới các ràng buộc mà tín hiệu thay đổi (là phần tín hiệu khác biệt giữa tín hiệu kích thích tăng cường và tín hiệu trước khi kích thích tăng cường) có năng lượng trong thời gian ngắn mà không vượt qua phần năng lượng sẵn có trong thời gian ngắn của tín hiệu kích thích trước khi tăng cường.

Hoạt động che giấu gói tin bị mất mát (Packet Loss Concealment) dễ dàng được đặt trong bộ giải mã. Hoạt động PLC có thể dựa trên việc lặp lại bộ lọc LPC và thu được residual signal bằng cách ước lượng dự đoán dài hạn từ các residual signal trước đó.

2.3.2. Các nguyên tắc giải mã

Quá trình giải mã thuật toán iLBC được thực hiện theo các bước sau [5]:

Bước 1: Xây dựng bộ lọc LPC.

Bước 2: Xây dựng trạng thái khởi đầu.

Bước 3: Thực hiện liên tục việc kích thích giải mã.

Bước 4: Giải mã bằng mã thích ứng nhiều giai đoạn: xây dựng tín hiệu kích thích được giải mã.

Bước 5: Che giấu gói tin mất mát (Packet Loss Concealment).

Khi xảy ra việc mất gói tin trên đường truyền, bộ mã hóa nhận được một tín hiệu thông báo thông tin tập hợp một block đã bị mất. Đối với các block, khuyến cáo rằng sử dụng một đơn vị Packet Loss Concealment để tạo một tín hiệu giải mã nhằm che giấu đi ảnh hưởng của gói tin bị mất đó. Đơn vị này là do bên giải mã thiết lập. Do đơn vị PLC chỉ được sử dụng ở bộ giải mã nên nó không ảnh hưởng đến toàn bộ quá trình hoạt động. Các quá trình thực hiện PLC khác có thể vẫn được sử dụng bình thường.

PLC hoạt động trong bộ lọc dự đoán tuyến tính LP, trong tín hiệu kích thích và dựa trên các nguyên tắc sau [5]:

- Đối với các Block nhận được đầy đủ và block trước đó cũng nhận được:

Nếu block được nhận đầy đủ, PLC chỉ ghi lại thông tin trạng thái của block hiện tại, để từ đó có thể sử dụng trong trường hợp block tiếp theo bị mất. Tất cả các hệ số của bộ lọc LP đối với mỗi sub-block và toàn bộ tín hiệu kích thích đã giải mã được lưu lại trong cấu trúc trạng thái của bộ giải mã. Tất cả các thông tin này sẽ được dùng tới khi block tiếp theo bị mất.

- Đối với các block không nhận được:

Nếu block không nhận được, sẽ thay thế bởi một block dựa trên sự lặp lại đồng bộ pitch (pitch-synchronous) của tín hiệu kích thích, tín hiệu này được lọc bởi bộ lọc LP cuối cùng của block trước đó. Thông tin của block trước đó được lưu trong cấu trúc trạng thái của bộ giải mã.

Quá trình phân tích sự tương quan được thực hiện trên tín hiệu kích thích của block trước đó để phát hiện chu kỳ tần số tuần hoàn (pitch periodicity) và giá trị của pitch. Biện pháp tương quan cũng được sử dụng để chọn ra mức âm lượng (mức độ mà kích thích của block trước đó được thể hiện hoặc tín hiệu tựa tuần hoàn). Sự kích thích trong block trước đó được sử dụng để tạo nên một kích thích cho block được thay thế, khi đó pitch của block trước đó sẽ được duy trì. Từ đó, sự kích thích mới này sẽ được xây dựng theo cách đồng bộ pitch. Để tránh việc block thay thế bị dội quá mức (buzzy-sounding), một kích thích ngẫu nhiên sẽ được kết hợp với kích thích tần số tuần hoàn mới (the new pitch

periodic excitation), và mối quan hệ giữa hai thành phần này được tính toán từ việc đo đạc sự tương quan.

Đối với block thay thế, tín hiệu kích thích vừa được xây dựng sẽ đi qua bộ lọc LP nhằm tạo ra các tín hiệu thay thế cho block bị mất.

Đối với trường hợp một số block liên tiếp bị mất, việc che giấu gói tin bị mất vẫn được thực hiện tương tự. Sự tương quan của block nhận được cuối cùng vẫn được sử dụng cùng với giá trị pitch. Bộ lọc LP của block cuối cùng nhận được cũng được sử dụng lại. Phần năng lượng của tín hiệu kích thích thay thế cho các block kề nhau bị mất cũng bị giảm đi, dẫn đến kích thích suy yếu đi, và do đó làm cho tín hiệu thoại cũng suy giảm.

➤ Đối với các block nhận được đầy đủ mà block trước đó không nhận được:

Trong trường hợp block nhận được hoàn chỉnh nhưng block trước đó lại không, thì tín hiệu thoại được giải mã trực tiếp từ block nhận được đó không được sử dụng cho đầu ra thực tế. Lý do là tín hiệu được giải mã trực tiếp đó không thể kết hợp một cách vừa vặn với tín hiệu thoại tổng hợp được tạo ra cho block bị mất trước đó. Nếu hai tín hiệu không được kết hợp với nhau một cách vừa vặn, âm thanh nghe được sẽ không liên tục, hay nói cách khác là bị vấp. Do vậy, sự phân tích tương quan giữa hai block về tín hiệu kích thích (sự kích thích của block được che giấu trước đó và block nhận được hiện tại) được thực hiện để tìm ra thời điểm kết hợp tốt nhất. Sau đó, một động tác xếp chồng lên một cách đơn giản được thực hiện để khớp kích thích trước đó vào kích thích của block hiện tại một cách vừa vặn.

Sự thực hiện chính xác việc che giấu gói tin không ảnh hưởng tới tính tương hợp của codec.

Bước 6: Tăng cường residual signal:

- Ước lượng pitch
- Chọn các chuỗi pitch đồng bộ
- Tính toán kích thích hợp lý.
- Kiểm tra residual signal xem có thỏa mãn điều kiện không.
- Sử dụng ràng buộc để tính toán các yếu tố hỗn hợp.
- Tổng hợp các tín hiệu thành phần.

Sau tất cả 6 bước trên, tín hiệu sẽ được khôi phục trở lại như lúc ban đầu ở phía bên gửi đi. Chất lượng tín hiệu lúc này phục thuộc vào mức độ lỗi trên đường truyền và khả năng khôi phục lại tín hiệu khi gặp lỗi của phía nhận. Việc đánh giá khả năng khôi phục lại lỗi gặp phải trên đường truyền của iLBC Codec sẽ được nói cụ thể hơn trong chương 3.

CHƯƠNG 3 – ĐÁNH GIÁ KHẢ NĂNG CHỐNG LẠI MẤT MÁT GÓI TIN TRÊN ĐƯỜNG TRUYỀN THOẠI

3. 1. Khái niệm chống mất mát gói tin trên đường truyền thoại

Trong một hệ thống truyền tin thoại, việc đường truyền bị lỗi hoặc không bị lỗi sẽ ảnh hưởng rất lớn đến chất lượng tín hiệu thu nhận được. Khả năng đường truyền không bị lỗi là trường hợp lý tưởng, vì trong thực tế thì hầu như các hệ thống truyền thông tin thoại đều gặp phải lỗi. Do mạng trên nền IP hoạt động theo cơ chế “best effort” (nỗ lực đến mức tối đa) nên độ tin cậy không cao, đôi khi xảy ra nghẽn mạng và các gói phải nằm lâu trong hàng đợi của các nút mạng, trường hợp xấu hơn là có thể rớt do tràn bộ đệm hoặc do độ ưu tiên của gói thấp, bị router loại bỏ. Hơn nữa, yếu tố mất mát gói tin ảnh hưởng rất lớn đến chất lượng của đường truyền thoại, do vậy cần phải có các biện pháp để khắc phục điều này.

Để truyền tin cậy dữ liệu trên mạng IP, người ta phải sử dụng cơ chế truyền lại các gói bị mất, được thực hiện ở tầng giao vận. Cơ chế gửi lại được kích hoạt khi đầu cuối nhận gửi bản tin thông báo mất gói. Tuy nhiên, việc gửi lại có nhược điểm là độ trễ lớn. Đối với các ứng dụng thoại thời gian thực thì cơ chế gửi lại không sử dụng được. Gói được xác định là mất khi độ trễ từ bên gửi đến bên nhận của gói vượt quá quy định. Như vậy, mất gói là không bao giờ đến đích (thời gian trễ bằng vô cùng), nhưng cũng có thể là gói mặc dù đến đích nhưng sau thời điểm quy định bên nhận nhận được (trường hợp này gọi là mất do trễ). Số lượng gói tin bị trễ sẽ giảm khi tăng bộ đệm, nghĩa là tăng thời gian trễ cho phép của gói. Nhưng tăng trễ bộ đệm đồng nghĩa với việc tăng trễ gửi – nhận, điều này ảnh hưởng đến chất lượng, độ cảm nhận của người nghe. Do đó, cần có sự cân nhắc thỏa đáng giữa hai tham số này. Mất gói làm gián đoạn quá trình hội thoại, gây nên mất tiếng, mất âm khiến cho đoạn hội thoại không hiểu được. Vì mất gói là điều không thể tránh khỏi, do vậy cần có các biện pháp để làm giảm ảnh hưởng của mất gói đến chất lượng thoại. Điều này đặt ra vấn đề chống mất mát gói tin ở các hệ thống này, bằng cách sử dụng các kỹ thuật kiểm soát lỗi hoặc sửa lỗi. Đó là kỹ thuật khôi phục mất gói, hay che giấu mất gói, nghĩa là làm cho người nghe không cảm thấy khó chịu do tác động của mất gói tin gây nên.

3. 2. Phân loại kỹ thuật chống mất mát gói tin

Về cơ bản, việc chống lại mất mát gói tin trong hệ thống thông tin thoại (Robustness to Packet Loss) được phân loại thành hai phương pháp chính như sau [16]:

- *Phương pháp dựa vào bên gửi và nhận gói tin (sender-receiver-based algorithms)*: sử dụng lược đồ tái gửi hoặc thêm vào các thông tin phụ (side information). Phương pháp này còn được gọi là phương pháp hướng codec, tức là căn cứ và đặc tính mã hóa của codec, ngăn chặn việc mất đồng bộ khi xảy ra mất gói. Ví dụ: G.722 Codec chống mất gói bằng cách thêm vào các thông tin phụ để

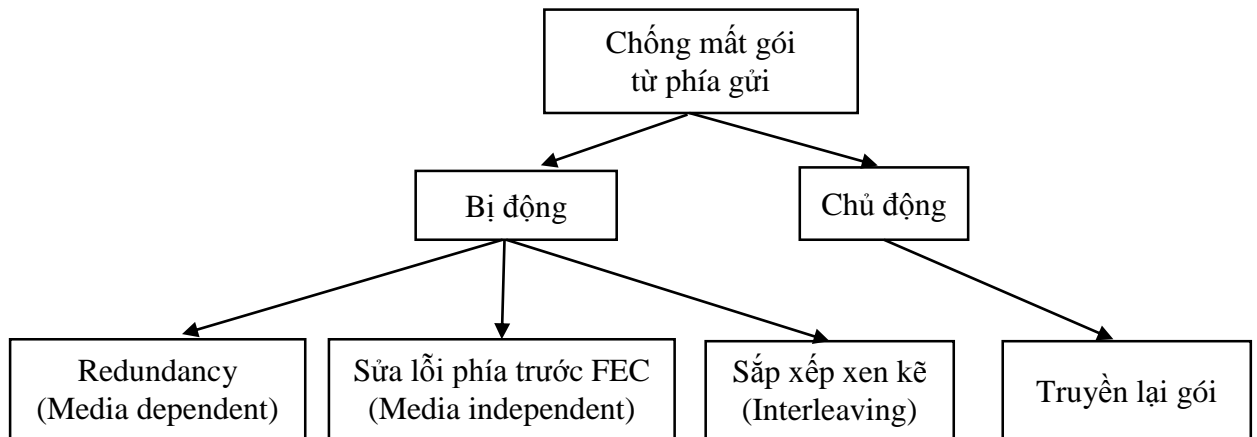
cho phép tạo lại các gói tin đã bị mất trên đường truyền và giúp bộ giải mã đồng bộ lại gói tin sau khi bị mất gói.

- *Phương pháp dựa vào bên nhận gói tin (receiver-based techniques)*: thực hiện việc tái tạo lại các gói tin bị mất sử dụng các tham số (như các hệ số dự đoán tuyến tính - Linear Prediction Coefficients, tham số kích thích - Excitations) của gói tin cuối cùng nhận được một cách chính xác, hoặc tùy vào tính chất mất gói của đường truyền mà áp dụng các kỹ thuật che giấu các gói bị mất bằng cách bù mất gói.

Hai phương pháp này đều cố gắng phục hồi gói bị mất để cho tiếng nói có chất lượng tốt nhất.

3.2.1. Kỹ thuật chống mất gói từ phía gửi

Kỹ thuật này có thể phân chia thành các phương pháp cụ thể như hình dưới đây [16]:



Hình 3.1 - Phân loại kỹ thuật chống mất gói từ phía gửi

a. Truyền gói lại

Cơ chế này sẽ thực hiện phát lại các gói lỗi hoặc mất, nó là phương pháp phổ biến nhất để khôi phục dựa vào phía gửi và thường sử dụng trong TCP. Tại phía gửi sẽ bổ sung các số thứ tự vào các gói thoại trước khi phát đi. Phía thu sẽ dựa vào số thứ tự này để biết các gói thoại nào đã bị mất và yêu cầu phía phát phát lại. Phương pháp này đảm bảo độ tin cậy cho dữ liệu nhưng lại có nhược điểm là độ trễ quá lớn và không phù hợp với các cuộc hội thoại thời gian thực. Nó chỉ phát huy tác dụng trong trường hợp phát thanh, quảng bá khi mà độ trễ không đòi hỏi cao bằng độ chính xác trong thông tin phát. Phương pháp này có hiệu quả không cao và chất lượng cuộc thoại không tốt [14].

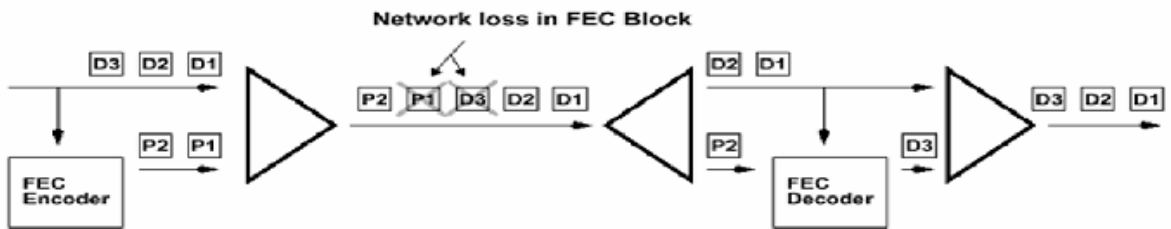
b. Sửa lỗi phía trước FEC (Forward Error Corection)

Đây là một phương pháp cơ bản và khá phổ biến, được sử dụng để bù lại gói tin đã mất mát trong quá trình truyền tải. Phương thức này sử dụng kỹ thuật chèn thêm các thông tin của gói tin vào các gói tin trước đó. Ý tưởng ở đây là gói tin thoại khi gửi đi sẽ kèm theo cả thông tin dư thừa của các gói tin khác. Trường hợp đơn giản nhất đó là gói tin truyền đi mang theo toàn bộ gói tin bên cạnh nó. Các kỹ thuật phức tạp hơn

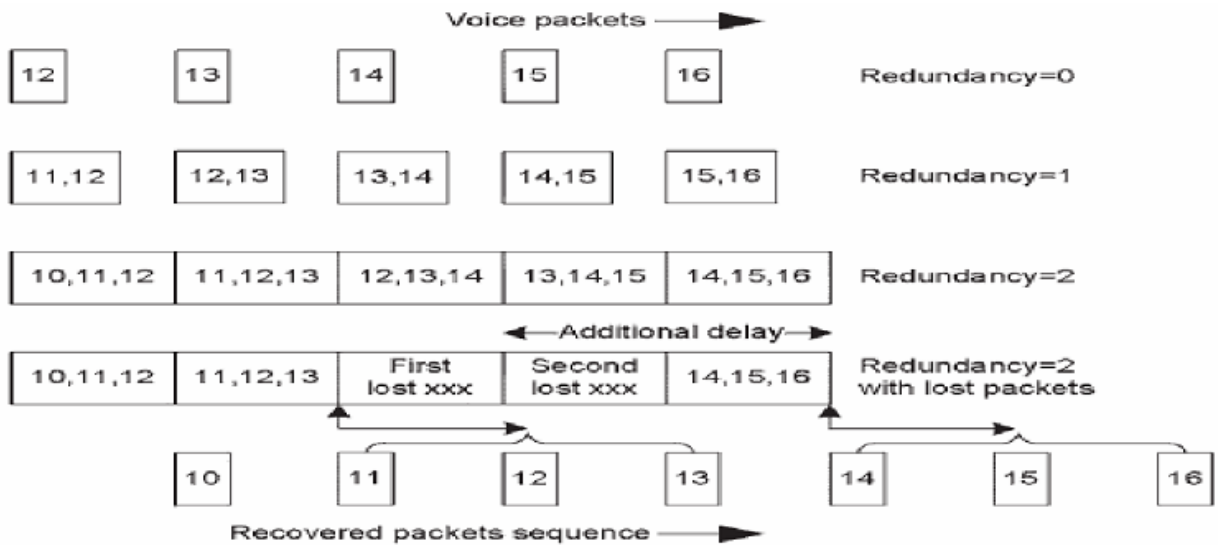
bao gồm việc truyền kèm theo toàn bộ một gói tin bất kỳ khác, nhưng được mã hóa với một codec khác yêu cầu ít băng thông hơn. Ví dụ như phương thức kết hợp giữa G.711 Codec, ADPCM, GSM 06.10. Trong trường hợp này, khi một gói tin bị mất, việc khôi phục lại nó chỉ đơn giản là khôi phục các thông tin liên quan đến gói tin từ các gói tin nhận được trước đó, và khả năng khôi phục một cách chính xác.

Nguyên lý của FEC là truyền đi k gói dữ liệu gốc kèm theo h gói chẵn lẻ dư thừa. Ví dụ $k=4, h=2$, như vậy các gói truyền đi có 4 gói mang thông tin thoại và 2 gói chẵn lẻ. Giả sử một gói thoại và 1 gói chẵn lẻ dư thừa bị mất thì tại phía thu vẫn có thể khôi phục lại toàn bộ các gói thoại dựa vào 3 gói thoại còn lại và 1 gói chẵn lẻ dư thừa. FEC làm việc hiệu quả ngay cả khi tỷ lệ h/k khá nhỏ. Ở bộ giải mã FEC, mất gói liên tiếp có thể sửa được với k khá lớn vì chúng sử dụng cả các bit dữ liệu và các bit chẵn lẻ cho việc sửa lỗi [12, tr.1].

Phương thức này được sử dụng trong rất nhiều ứng dụng khi mà việc mất mát thông tin chỉ xảy ra ở từng gói tin riêng lẻ, khá hiệu quả trong việc cải thiện chất lượng hội thoại, tuy nhiên lại có nhược điểm là tiêu tốn băng thông nhiều hơn.



Hình 3.2 - Cơ chế phục hồi FEC

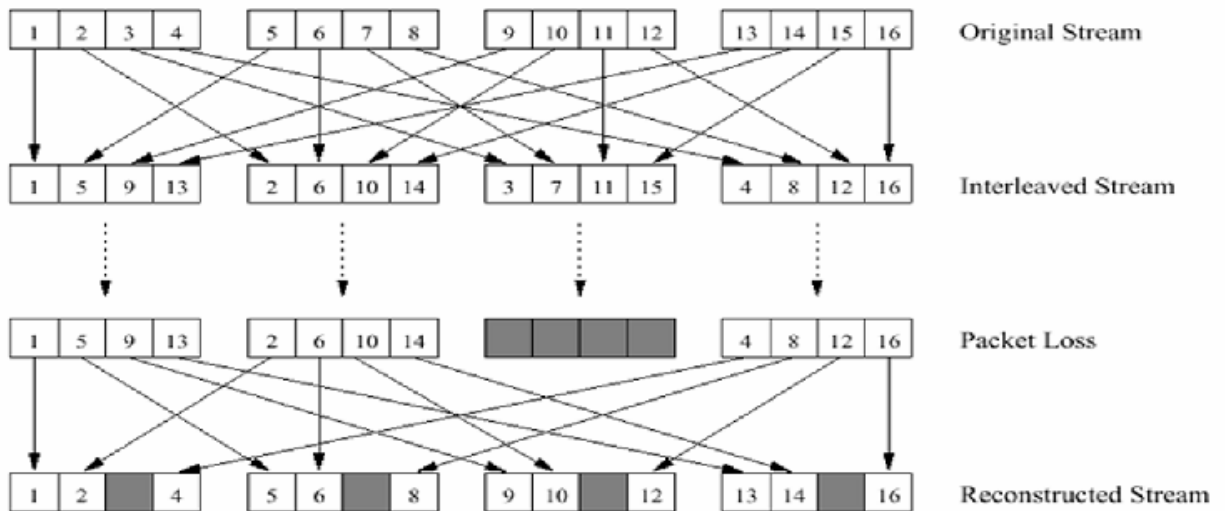


Hình 3.3 – Ví dụ cơ chế phục hồi Redundancy

c. Kỹ thuật sắp xếp xen kẽ (Interleaving)

Sắp xếp xen kẽ, hay còn gọi là ghép xen, giúp làm giảm ảnh hưởng của lỗi bit chùm khi độ trễ không phải là vấn đề chủ yếu. Trong phương pháp này, một gói tin thoại được chia thành các đơn vị nhỏ hơn và sắp xếp xen kẽ nhau trong các gói khác

nhau. Tại bộ thu, các khối dữ liệu được tập hợp lại và sắp xếp theo thứ tự rồi đưa đến bộ giải mã. Đối với ghép xen, ảnh hưởng của mất gói được chia đều trên các khoảng thời gian nhỏ tương ứng. Khi đó các khoảng thời gian này nhỏ hơn chiều dài một âm vị (phoneme). Do đó có thể dễ dàng nội suy lại khoảng đã mất và âm thoại vẫn nghe tốt như bình thường. Trong khi đó nếu không có ghép xen thì có thể có trường hợp khối dữ liệu bị mất lớn hơn so với khoảng thời gian 1 âm vị, khi đó sẽ làm giảm khả năng hiểu được của thoại. Ưu điểm của phương pháp sắp xếp xen kẽ là không làm tăng tải cho mạng, giúp giảm ảnh hưởng của mất gói chòm, tuy nhiên nhược điểm của nó là làm tăng độ trễ [14].



Hình 3.4 - Ví dụ cơ chế sắp xếp xen kẽ

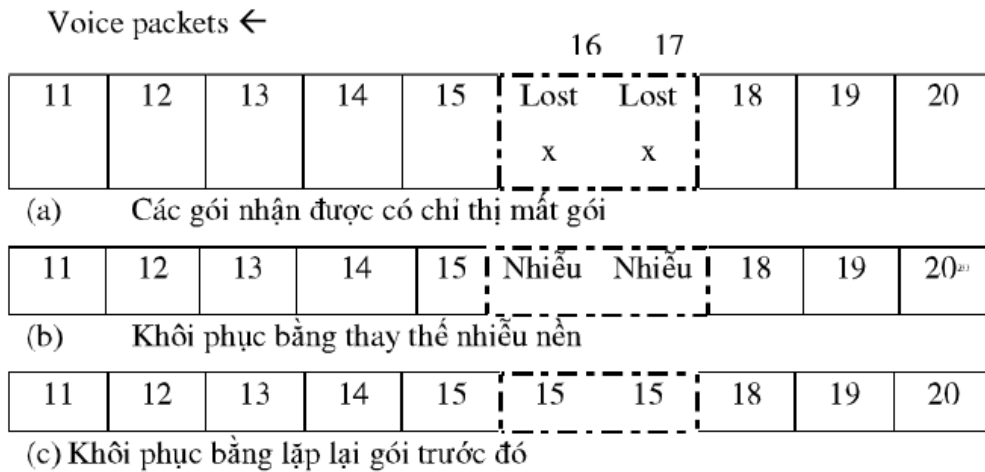
3.2.2. Kỹ thuật bù mất gói từ phía nhận

Kỹ thuật này thực hiện thay thế gói tin bị mất bởi gói tương tự nguyên bản. Như đã nói ở trên, trong một số hệ thống mã hóa thoại sử dụng kỹ thuật FEC (như SILK của Skype) yêu cầu tốc độ bit lớn trong quá trình truyền tải. Đây là nhược điểm lớn bởi vì tốc độ bit cao đòi hỏi băng thông phải lớn, và do đó dẫn đến tốn kém tài nguyên hệ thống. Nhiều bộ mã hóa thoại khác, ví dụ như G.723.1, G.729A, iLBC, sử dụng tốc độ bit thấp nhất có thể bằng cách sử dụng kỹ thuật PLC (Packet Loss Concealment) hay EC (Error Concealment) để che giấu việc mất mất gói tin. Theo phương pháp này, thay vì gửi kèm các thông tin của gói tin từ phía gửi, thì các thông tin quan trọng dùng để phục hồi lại các gói tin bị mất được đặt ở bộ giải mã ở phía nhận. Trong trường hợp bị mất gói tin, thông tin này sẽ được ngoại suy để che giấu việc mất mất thông tin, và chuyển sang gói tin thu được gần đó một cách mượt mà, không để lại dấu vết gì. Điều này có thể thực hiện được vì hầu hết các phần của tín hiệu thoại hoặc âm thanh thay đổi chậm theo thời gian, do vậy những đoạn tín hiệu thoại gần nhau thường tương tự như nhau. Kỹ thuật này làm việc tốt khi tỷ lệ mất gói nhỏ hơn 15% và khoảng thời gian mất gói nhỏ hơn 40ms. Khi chiều dài mất gói liên tiếp gần bằng chiều dài âm vị thì kỹ thuật này làm việc không hiệu quả.

Một số phương pháp bù mất gói ở phía thu có thể kể đến như: Bù mất gói bằng kỹ thuật chèn, bằng mô hình dự đoán tuyến tính LP (Linear Predictive), dùng kỹ thuật dự đoán tuyến tính đệ quy, tái tạo gói bị mất theo mô hình huấn luyện.

a. Bù mất gói bằng kỹ thuật chèn (Insertion)

Trong kỹ thuật chèn, bên nhận bù mất gói bằng cách thay thế nhiều nền vào vị trí các gói bị mất, hoặc lặp lại gói cuối cùng nhận được đối với tất cả các gói tin nằm trong khoảng thời gian bị mất gói. Cơ chế khôi phục mất gói này được minh họa như sau [2]:



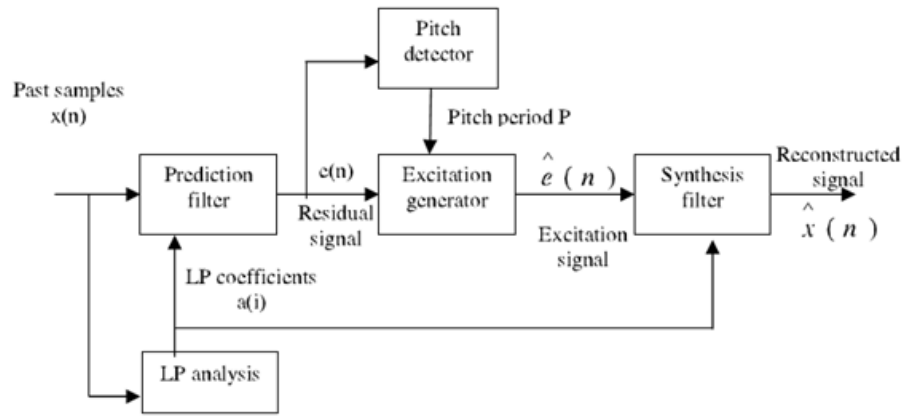
Hình 3.5 - Kỹ thuật phục hồi mất gói từ phía thu theo cơ chế chèn

b. Bù mất gói dựa vào mô hình LP

Phương pháp này hoạt động dựa trên kỹ thuật dự đoán tuyến tính (Linear Prediction). Các phần thông tin sẽ được duy trì ở phía nhận đó là tần số hài (pitch lag), hệ số dự đoán dài hạn LTP (Long-Term Predictive), hệ số dự đoán tuyến tính LPC (Linear Predictive Coefficient), và tín hiệu sử dụng để kích thích bộ lọc LPC. Khi một gói tin bị mất, hệ số LPC sẽ nói rộng về băng thông, nhằm làm giảm năng lượng đi một chút. Việc này giúp che giấu gói tin bị mất bằng cách làm giảm năng lượng giữa hai gói tin có năng lượng cao, nhằm giúp người nghe không bị ảnh hưởng. Khi đó, tín hiệu kích thích của frame trước đó hoạt động thông qua bộ lọc LPC và LTP giúp tạo ra tín hiệu cho các gói tin bị mất.

Cuối cùng, khi nhận được gói tin, sẽ có một công đoạn chuyển đổi giúp tín hiệu được khôi phục trở nên mượt mà so với tín hiệu bị mất, khiến cho phía thu khó có thể phát hiện ra phần tín hiệu khôi phục đó. Điều này được thực hiện bằng cách đo đạc phần năng lượng của gói tin khôi phục gần nhất, và gói tin nhận được tiếp theo. Khi đó, tín hiệu của gói tin nhận được nhân rộng bởi các hệ số tăng tuyến tính để đạt được sự chuyển đổi mượt mà giữa mức năng lượng của frame che giấu đi với mức năng lượng thay thế vào.

Kỹ thuật này dùng mô hình dự đoán tuyến tính LP (Linear Prediction) của quá trình tạo tiếng nói [2].



Hình 3.6 – Sơ đồ khối thuật toán bù mất gói dựa vào mô hình LP

Theo mô hình LP, tín hiệu tiếng nói $x(n)$ gồm hai thành phần:

- Thông số dự đoán chứa thông tin bộ máy phát âm.
- Tín hiệu dư thừa (sai lệch) chứa thông tin về tín hiệu kích thích.

$$x(n) = \sum_{i=1}^N a(i)x(n-i) + e(n)$$

Tiếng nói được tạo ra bằng cách cho tín hiệu kích thích đi qua một bộ lọc tổng hợp (bộ máy phát âm).

Phương pháp bù này được thực hiện hoàn toàn từ bên thu đối với tín hiệu PCM. Nguyên tắc cơ bản của thuật toán này là ước tính hệ số LP $\{a(i)\}$ và tín hiệu kích thích $\{e(n)\}$ của frame tiếng nói bị mất dựa vào thông tin được trích ra từ frame tiếng nói nhận tốt trước đó. Hai thông số này được kết hợp để tạo ra xấp xỉ cho gói bị mất.

- Bộ phân tích LP:

Khối này dùng để tìm các hệ số LP $\{a(i)\}$, mô phỏng hình dạng bộ máy phát âm của một frame tiếng nói.

Frame tiếng nói nhận tốt trước đó được đưa vào bộ phân tích LP bậc 10 để được tạo ra 10 hệ số dự đoán tuyến tính nhờ thuật toán Levinson-Durbin. 10 hệ số này được dùng làm hệ số của bộ lọc dự đoán và bộ lọc tổng hợp.

- Bộ lọc dự đoán:

Các mẫu thuộc frame trước đó được lọc bằng bộ lọc dự đoán và chỉ còn lại tín hiệu dư thừa. Tín hiệu này được dùng để dự đoán pitch trong frame tiếng nói và tạo tín hiệu kích thích cho frame bị mất.

$$e(n) = x(n) - \sum_{i=1}^{10} a(i)x(n-i)$$

- Bộ dự đoán pitch:

Khối này thực hiện ước đoán pitch của frame nhận tốt trước đó từ tín hiệu dư thừa. Thông tin pitch này được dùng để xây dựng tín hiệu kích thích của frame bị mất.

Đối với frame vô thanh, tín hiệu dư thừa không có chu kỳ pitch.

Đối với frame hữu thanh, vì tần số cơ bản của tiếng nói nằm trong khoảng 50Hz - 500Hz, chu kỳ pitch được xác định bằng cách tìm đỉnh của dãy chuẩn hóa $r_e(n)/r_e(0)$ trong khoảng thời gian tương ứng 3 đến 15ms trong frame tiếng nói 20ms theo công thức sau:

$$r_e(n) = \sum_{i=1}^{10} r_a(i)r(n-i)$$

Với $r_a(i)$, $r(n-i)$ lần lượt là chuỗi tự tương quan của các hệ số dự đoán và của các mẫu trong một frame.

$$r_a(n) = \sum_{i=1}^{10} a_k(i)a_k(i+n)$$

$$r(i) = \sum_{n=0}^{N-1} s(n)s(n+i)$$

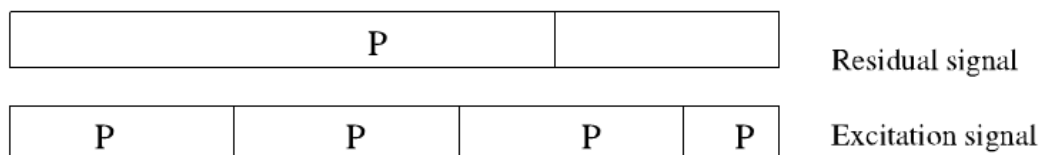
Chu kỳ pitch bằng giá trị $n=N_p$, với $r_e(N_p)/r_e(0)$ là lớn nhất.

- Bộ tạo tín hiệu kích thích:

Tín hiệu dư thừa và chu kỳ pitch của frame nhận được trước đó được dùng để tạo ra tín hiệu kích thích cho frame bị mất.

Nếu frame nhận được trước đó là vô thanh thì tín hiệu kích thích của frame bị mất là nhiễu.

Nếu frame nhận được trước đó là hữu thanh thì tín hiệu kích thích của frame bị mất được tạo ra bằng cách: lặp P (pitch) mẫu cuối cùng trong tín hiệu dư thừa nhiều lần cho tới khi điền đầy số mẫu trong tín hiệu kích thích.



Hình 3.7 - Cách tạo tín hiệu kích thích của frame bị mất từ frame trước đó

- Bộ lọc tổng hợp:

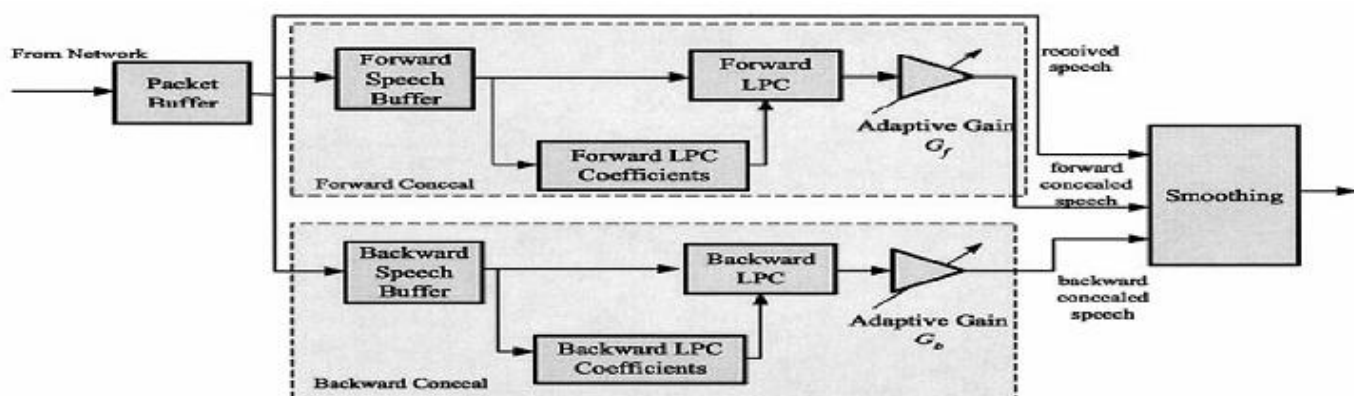
Tín hiệu tiếng nói của frame bị mất được tái tạo bằng cách cho tín hiệu kích thích khi đi qua bộ lọc tổng hợp để được thêm vào thông tin bộ máy phát. Các hệ số LP trong bộ phân tích LP được dùng làm hệ số trong bộ lọc tổng hợp. Bộ lọc tổng hợp là bộ lọc toàn cực bậc 10.

$$\hat{x}(n) = \sum_{i=1}^{10} a(i)\hat{x}(n-i) + \hat{e}(n)$$

Phương pháp bù mất gói dựa theo mô hình LP là phương pháp ngoại suy, nó chỉ dùng các frame được nhận trước đó, không dùng các frame sau vì như thế sẽ tăng thêm thời gian trễ, điều này không tốt cho các ứng dụng thời gian thực.

c. Bù mất gói dùng dự đoán tuyến tính đệ quy

Phương pháp này dùng dự đoán tuyến tính một cách đệ quy để ước lượng frame bị mất từ các mẫu (sample) tiếng nói thuộc các frame kế cận nhận được [2].



Hình 3.8 - Sơ đồ khối thuật toán bù mất gói dùng dự đoán tuyến tính đệ quy

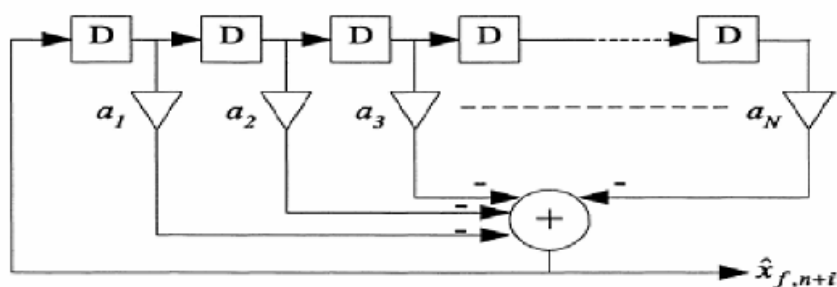
Bù mất gói đơn dùng cả dự đoán trước và dự đoán sau.

Công thức dự đoán trước (forward prediction) một mẫu từ các mẫu nhận được trước đó:

$$\hat{x}_{f,n} = -1 * \sum_{i=1}^N a_i x_{n-i}$$

Với $x_{f,n}$ là mẫu được dự đoán thuộc frame bị mất, x_{n-i} là các mẫu tiếng nói thuộc frame được nhận trước đó. N là bậc dự đoán; là hệ số LPC được tính từ M mẫu thuộc frame nhận được trước frame bị mất (x_{n-i} với $i = 1, 2, \dots, M$ và M là chiều dài của frame tiếng nói). Hệ số LPC được tính bằng công thức Levinson-Durbin.

Các mẫu sau đó, $x_{f,n+i}$ được dự đoán một cách đệ quy bởi các mẫu đã được dự đoán và các mẫu nhận được trước đó. Ví dụ, $x_{f,n}$ và x_i với $i = n-1, n-2, \dots, n-N+1$ được dùng để dự đoán mẫu $x_{f,n+1}$.



Hình 3.9 - Sơ đồ khối thuật toán dự đoán các mẫu LPC

Trong quá trình dự đoán các mẫu của gói bị mất, hệ số LPC của frame bị mất vẫn dùng nguyên lại hệ số LPC của frame nhận tốt trước đó. Quá trình dự đoán mẫu được lặp lại cho toàn bộ frame bị mất.

$$\hat{x}_{f,n+i} = -1 * \sum_{j=1}^N a_j \hat{x}_{f,n+i-j}$$

Khi dự đoán, độ lợi của tiếng nói dự đoán bị giảm dần. Vì vậy, cần có một độ lợi tăng tuyến tính G_f thêm vào. G_f bắt đầu bằng 1 tại đầu frame bị mất và bằng 1.8 tại cuối frame bị mất. Độ lợi được nhân vào các mẫu tiếng nói dự đoán.

Dự đoán sau (backward prediction) là dự đoán bằng cách dựa vào các mẫu phía sau. Công thức sau là dự đoán sau của một mẫu từ các mẫu nhận được sau đó:

$$\hat{x}_{b,n} = -1 * \sum_{i=1}^N b_i x_{n-i+N+1}$$

Với $x_{b,n}$ là mẫu được dự đoán sau thuộc frame bị mất, $x_{n-1+N+1}$ là các mẫu tiếng nói thuộc frame nhận được sau đó; b_i là hệ số LPC được tính từ M mẫu thuộc frame nhận được sau frame bị mất. Hệ số LPC cũng được tính bằng công thức Levinson-Durbin. Vì thực hiện dự đoán sau nên cần nhận được frame phía sau frame bị mất, do vậy, thời gian trễ xử lý tăng lên. Quá trình dự đoán mẫu được lặp lại cho toàn bộ frame bị mất với trật tự thời gian ngược.

Độ lợi thích ứng G_b được áp dụng theo cách tương tự. G_b bắt đầu bằng 1 tại cuối frame bị mất và bằng 1.8 tại đầu frame bị mất. Độ lợi được nhân vào các mẫu tiếng nói dự đoán.

Có hai ước đoán cho frame bị mất, đó là $x_{f,n+i}$ và $x_{b,n+i}$. Dự đoán trước $x_{f,n+i}$ dự đoán phần đầu của frame bị mất tốt hơn; $x_{b,n+i}$ dự đoán tốt hơn phần sau của frame bị mất.

Do vậy, cần kết hợp hai dự đoán với trọng số tuyến tính để đạt được dự đoán của một mẫu đơn giản:

$$X_{n+i} = (1-\alpha)x_{f,n+i} + \alpha x_{b,n+i}$$

Với α là trọng số tăng tuyến tính, $\alpha=0$ tại đầu frame bị mất và bằng 1 tại cuối frame bị mất.

Ở phương pháp dự đoán đệ quy, việc bù mất gói được thực hiện nội suy từ gói trước và gói sau của gói bị mất. Nội suy gói bị mất từ các gói ở hai chiều giúp cải thiện tốt hơn chất lượng tiếng nói, tuy nhiên độ trễ xử lý lại tăng lên vì cần thời gian cho việc đợi thêm gói thoại.

3. 3. Đánh giá khả năng chống lại mất mát gói tin của iLBC Codec

Các yếu tố chính ảnh hưởng tới khả năng chống lại mất mát gói tin của iLBC Codec bao gồm mã hóa dự đoán tuyến tính theo từng khung độc lập, sử dụng tính tương quan dài hạn, và mã hóa trạng thái khởi đầu (start state). Chi tiết cụ thể từng ảnh hưởng sẽ được phân tích theo các phần ở dưới đây.

3.3.1. Phân tích khả năng chống mất mát gói tin của iLBC Codec

a. Mã hóa dự đoán tuyến tính dài hạn theo từng khung độc lập

Bản chất của Codec là mã hóa dự đoán tuyến tính (LPC) và mã hóa tín hiệu lỗi (residual signal) dựa trên block sử dụng một bảng mã thích ứng (codebook) [7]. Thực tế, trong quá trình mã hóa phương pháp CELP cũng sử dụng bảng mã thích ứng với sự kích thích tín hiệu theo thời gian. Cách thức này cũng có nhiều vấn đề, và sau đây là một trong số những vấn đề đó:

- Nếu tín hiệu bị mất hoặc bị lỗi trong quá trình truyền, bảng mã thích ứng trong bộ giải mã sẽ khác với bảng mã trong bộ mã hóa.
- Ở phần đầu đoạn tín hiệu có âm thanh, bảng mã thích ứng không đủ để mô tả chính xác chu kỳ cơ bản (pitch cycle). Điều này dẫn đến việc giảm tín hiệu lời thoại trong tín hiệu được giải mã.

Trong kỹ thuật CELP, việc mong muốn tái tạo một vùng âm thoại có chu kỳ lớn, và không mong muốn nhận được nhiều có sự mâu thuẫn lẫn nhau. Thực tế, kết quả của mã hóa CELP là một tín hiệu bao gồm nhiều nền trong vùng thoại.

iLBC Codec sử dụng kỹ thuật mã hóa dự đoán tuyến tính LPC, đồng thời cho phép sử dụng các codebook (khái niệm có trong CELP) ở cả trước và sau khi mã hóa, bắt đầu từ một đoạn trong frame thoại, gọi là vector trạng thái khởi đầu. Trạng thái khởi đầu có thể được xác định bằng một đoạn mẫu có năng lượng phần residual cao nhất. Đoạn mẫu này sẽ chứa ít nhất một phần tần số chủ yếu khi trong nó có chứa âm thoại. Ta coi tín hiệu này là trạng thái khởi đầu cho việc mã hóa dự đoán tuyến tính dài hạn. Vị trí và hình dạng sóng của trạng thái này được mã hóa và truyền đi theo từng frame [7].

Bảng mã thích ứng được gắn với trạng thái khởi đầu các đoạn đã được giải mã. Sau đó bảng mã này được sử dụng cho mã hóa dự đoán dài hạn theo thời gian một chiều,..., tiếp tục trở về sau cho các khung tín hiệu còn lại từ lúc kết thúc trạng thái khởi động cho đến cuối frame tín hiệu. Trong suốt quá trình mã hóa này, bảng mã thích ứng được điều chỉnh liên tục với tín hiệu giải mã gần nhất. Tiếp theo, bảng mã thích ứng được gắn với đoạn trạng thái khởi động đã giải mã và đoạn tín hiệu mã hóa đầu tiên. Sau đó, bảng mã thích ứng được sử dụng cho mã hóa dự đoán dài hạn theo thời gian một chiều, ví dụ, ngược trở lại đầu frame tín hiệu, từ điểm bắt đầu của trạng thái khởi động tới điểm bắt đầu của frame. Theo cách này, gói dữ liệu có thể chứa đựng thông tin trạng thái khởi đầu, và thông tin về lag, gain đủ cho việc mã hóa chính xác và hoàn chỉnh frame [7].

Phương thức này nhấn mạnh rằng bảng mã thích ứng trong bộ giải mã sẽ độc lập với việc tiếp nhận hoặc việc mất gói tin trước đó. Hơn nữa, nếu một tín hiệu khởi đầu xuất hiện trong frame thì ít nhất một xung tần quan trọng (significant pitch pulse) sẽ nằm trong bảng mã thích ứng như một điểm khởi đầu (starting point) và do đó, pitch cycle trong frame có thể được mã hóa chính xác. Sẽ không còn xảy ra việc tái tạo một vùng âm thoại có chu kỳ lớn kéo theo việc mã hóa nhiều tới bảng mã thích ứng nữa, bởi vì trạng thái khởi đầu (start state) cơ bản đã chứa đầy đủ xung tần (pitch pulse). Có thể nhận thấy rằng nếu giải quyết được vấn đề xây dựng bảng mã codebook thì bảng mã thích ứng có thể được sử dụng trong tất cả các giai đoạn của quá trình mã hóa, hơn là chỉ sử dụng nó trong một giai đoạn duy nhất khi dùng bảng mã lớn cố định, thường phổ biến trong phương pháp CELP.

Khi việc tái tạo tần số cao được giải quyết (đối với ứng dụng sử dụng bảng mã thích ứng), nó có thể được áp dụng đối với các vấn đề liên quan tới tiếng nói thoại khác. Ví dụ, việc mã hóa vùng âm thoại hiện nay sử dụng việc tái tạo tần số cao khá linh hoạt đối với mã hóa vùng nhiều trong quá trình mã hóa trạng thái khởi đầu. Trong iLBC Codec những vấn đề về nhiễu được giải quyết bằng phương pháp lọc sau (Post Filter).

b. Chọn và mã hóa trạng thái khởi đầu

Đối với phương pháp mã hóa tuyến tính dài hạn độc lập giữa các frame thì việc mã hóa trạng thái khởi đầu khá là quan trọng. Về cơ bản, cần có sự chuẩn hóa toàn bộ

trạng thái khởi đầu bởi bộ lượng tử vô hướng cùng với nhiễu thu được để tạo ra các nhiễu lượng tử [7].

Đối với âm thanh thoại, trạng thái khởi đầu sẽ chứa đựng ít nhất một tần số pitch pulse quan trọng. Pitch pulse này được mô tả trong trạng thái khởi đầu sau giải mã. Việc lượng tử vô hướng tín hiệu dạng xung sẽ dẫn tới sự đánh đổi không tốt giữa quá tải (overload) và nhiễu hạt cho bộ lượng tử. Do đó, cần xử lý trước trạng thái khởi đầu với một bộ lọc All-Pass nhằm phân bố đồng đều năng lượng tín hiệu giữa các mẫu trong nó. Trong khi bộ lọc này có thể tối ưu cho tín hiệu cho trước, lựa chọn thay vào một thiết kế đơn giản để tránh phải truyền thêm các thông tin phụ nhằm xác định bộ lọc All-Pass này cho bộ giải mã. Nhận thấy bộ lọc All-Pass này là sự tổng hợp của bộ lọc tổng hợp LPC đã giải mã trước và sau theo thời gian. Kết quả là ngoại trừ ảnh hưởng của lặp lại trong khi tổng hợp, residual signal trước khi xử lý sẽ giống hệt với residual signal dự đoán, nếu bộ lọc tổng hợp LPC được thực hiện theo hướng ngược trở lại phía trước đó. Do đó, thay vì cho kết quả với một vùng peak tập trung ở đầu pitch cycle, phương pháp này cho một vùng phân bố dự đoán lỗi rộng hơn trên pitch cycle. Sau khi giải mã trạng thái khởi đầu, một bộ lọc All-Pass đảo ngược được thực hiện để bù lại ảnh hưởng của việc tiền xử lý này.

c. Mã hóa

Codec hoạt động theo từng frame với 160/240 mẫu/frame, khi đó sẽ có hai bộ phân tích LPC thực hiện: sử dụng một cửa sổ từ vị trí lựa chọn kéo dài tới đầu frame, và một cửa sổ còn lại kéo dài tới cuối frame. Hai bộ lọc LPC cùng được mã hóa sử dụng đường dải tần (Line Spectral Frequencies - LSF). Sau khi lọc phân tích với một bộ lọc phân tích LPC được nội suy chuẩn, các frame lỗi (residual frame) được chia thành 6 sub-frame, mỗi frame đó có gồm 40 mẫu. Hai sub-frame kế nhau có phần năng lượng cao nhất sẽ được chọn ra. Sau đó, 57 mẫu đầu hoặc cuối của hai sub-frame kế nhau được chọn ra làm trạng thái khởi đầu và được mã hóa. Việc lựa chọn đoạn 57/58 mẫu ở đầu và ở cuối được lặp lại nhằm tối đa hóa năng lượng tín hiệu. Trạng thái khởi đầu được mã hóa sử dụng 6 bit cho độ lớn và 3/4 bit cho vị trí lựa chọn [6].

240 mẫu, được mã hóa với 399 bits = 13.3 kbit/s (50 oct) 160 mẫu, được mã hóa với 303 bits = 15.2 kbit/s (38 oct)

Parameter	Bits
LPC	40
Start state position	4
Start state scale	6
Start state samples	174
Shapes	115
Gains	60
Total	399

Parameter	Bits
LPC	20
Start state position	3
Start state scale	6
Start state samples	171
Shapes	67
Gains	36
Total	303

Dựa trên trạng thái khởi đầu đã giải mã, bảng mã thích ứng được khởi tạo để mã hóa tiếp 23/24 mẫu của hai sub-frame có năng lượng cao đó. Đối với các sub-frame còn lại ở phía trước (theo trục thời gian), tất cả 80 mẫu kích thích LPC đã giải mã vừa rồi sẽ được sử dụng để khởi tạo tiếp. Tương tự như vậy, tất cả các mẫu đã giải mã trước đó, có thể là cả 127 mẫu, sẽ được sử dụng để tạo ra các bảng mã thích ứng cho các mẫu phía sau đó (theo trục thời gian). Trong cả ba tình huống, bảng mã thích ứng chứa đựng các đoạn kích thích LPC đã giải mã được dịch chuyển theo thời gian bằng một mẫu. Thêm vào đó, bảng mã thích ứng được nội suy cùng với các vector, các vector này có được bằng cách kết hợp tuyến tính giữa các vector kích thích LPC dịch chuyển theo thời gian. Các hệ số kết hợp tuyến tính được thay đổi theo thời gian và được huấn luyện nhằm tối thiểu hóa lỗi sai số trung bình bình phương khi sử dụng bảng mã thích ứng. Bảng mã thích ứng được chấp nhận với một cấu hình gồm nhiều giai đoạn gain-shape: việc mô tả mỗi sub-frame được lọc trong 3 giai đoạn.

d. Giải mã

Đầu tiên, bộ giải mã giải mã trạng thái khởi đầu, tiếp đó là các sub-frame trước (theo thời gian) và cuối cùng là các sub-frame sau (theo thời gian). Trước khi lọc tổng hợp, bộ mã hóa thực hiện lọc sau tần số cơ bản (pitch postfiltering) trong vùng residual signal. Khi một gói tin không nhận đúng thời điểm đã định, quá trình che giấu mất gói tin (Packet Loss Concealment) được thực hiện. Việc chống mất mát này cũng được thực hiện trong miền kích thích LPC.

3.3.2. Phương pháp đánh giá khả năng chống mất mát gói tin của iLBC Codec

a. Phương pháp đánh giá

Như đã nói ở chương 1, chất lượng tín hiệu âm thanh trong hệ thống truyền tin thoại bao gồm rất nhiều các yếu tố cấu thành như độ trễ, độ mất gói tin, jitter,... Để đánh giá yếu tố này người ta sử dụng chỉ số MOS. Mặc dù sự mất gói đóng vai trò quan trọng nhất trong chất lượng tín hiệu thu được, nhưng để đánh giá mất gói tin thì phương pháp MOS không hoàn toàn chính xác. Để thực hiện việc này, bài luận vẫn sử dụng phương pháp tính sai số trung bình bình phương MSE (Mean Squared Error) [15, tr.20].

MSE là một khái niệm trong thống kê học, nghĩa là sai số trung bình bình phương, của một phép ước lượng là trung bình của bình phương các sai số, tức là sự khác biệt giữa các ước lượng và những gì được đánh giá. MSE là một hàm rủi ro, tương ứng với giá trị kỳ vọng của sự mất mát sai số bình phương hoặc mất mát bậc hai. Sự khác biệt xảy ra do ngẫu nhiên, hoặc vì các ước lượng không tính đến thông tin có thể cho ra một ước tính chính xác hơn.

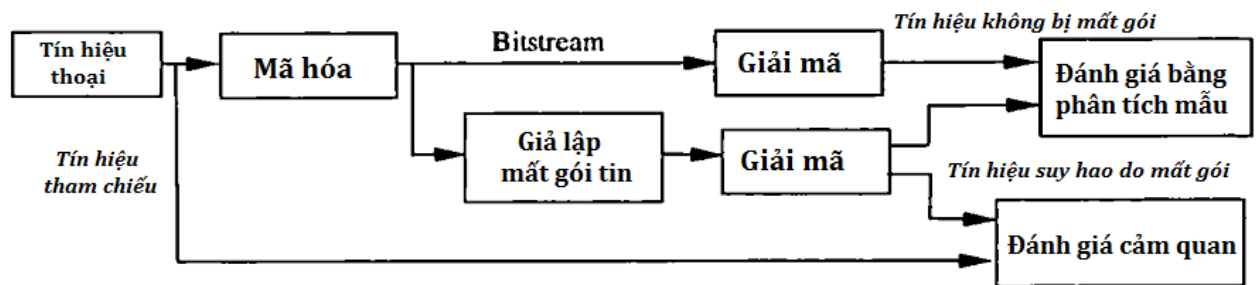
MSE là moment bậc hai của sai số, và do đó kết hợp cả hai phương sai của ước lượng và thiên vị của nó. Đối với một ước lượng không có thiên vị, MSE là phương sai của ước lượng. Cũng giống như các phương sai, MSE có cùng một đơn vị đo lường theo bình phương của số lượng được ước tính. Trong một trường hợp tương tự với độ lệch chuẩn, lấy căn bậc hai của MSE cho ra sai số root-mean-square (RMSE), hoặc độ

lệch root-mean-square (RMSD), trong đó có các đơn vị tương tự như đại lượng được ước tính. Đối với đại lượng không có thiên vị, các RMSE là căn bậc hai của phương sai, và được gọi là độ lệch chuẩn.

MSE là trung bình $\frac{1}{n} \sum_{i=1}^n$ của bình phương các sai số $(\hat{Y}_i - Y_i)^2$. Đây là định lượng dễ dàng tính được cho một mẫu cụ thể.

Như vậy, với những định nghĩa đơn giản về MSE, có thể nói đơn giản rằng việc đánh giá khả năng chống mất mát gói tin của iLBC Codec bằng phương pháp MSE chính là dựa vào sự so sánh giữa file ban đầu và file sau khi truyền qua hệ thống truyền tin thoại. Dễ dàng nhận thấy là nếu trên đường truyền xảy ra lỗi, các gói tin bị mất mát hoặc không đến được bên nhận, thì tín hiệu thoại ban đầu và tín hiệu sau khi giải mã sẽ có những sai khác nhau về mặt thuộc tính. Công việc đánh giá bây giờ bằng phương pháp MSE là so sánh các thuộc tính này của các tín hiệu trước và sau để đưa ra nhận định. Độ lệch về thuộc tính giữa tín hiệu ban đầu và tín hiệu sau giải mã cùng với tỉ lệ mất gói tương ứng sẽ nói lên được khả năng bù mất gói của Codec. Để có cái nhìn rõ hơn về iLBC Codec trong vấn đề này, quá trình đánh giá iLBC Codec sẽ thực hiện cùng với G.711 Codec.

Sơ đồ thực hiện việc đánh giá như sau [10, tr.48]:



Hình 3.10 – Sơ đồ đánh giá khả năng chống mất gói của Codec theo phương pháp MSE

b. Các bước mô phỏng và đánh giá

Bước 1: Xây dựng thuật toán.

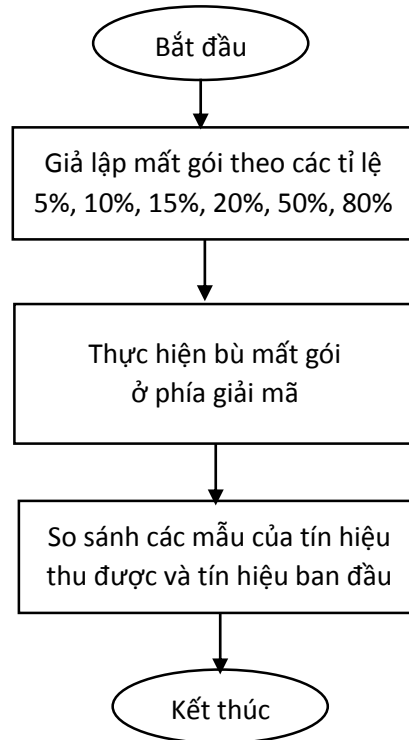
Bước 2: Mô phỏng cơ sở dữ liệu bị mất gói.

Bước 3: Thực hiện các phương pháp bù mất gói ở phía giải mã.

Bước 4: Thực hiện so sánh các mẫu của tín hiệu thu được và tín hiệu ban đầu.

Bước 5: Nhận xét & đánh giá.

c. Sơ đồ thực hiện mô phỏng



Hình 3.11 – Mô phỏng quá trình chống mất gói tin

d. Nhận xét

Việc đánh giá định lượng khả năng mất mát gói tin với phương pháp đánh giá cảm quan không cho kết quả lượng hóa chính xác. Do vậy quá trình thực nghiệm sẽ chỉ dựa trên phương pháp đánh giá định lượng bằng cách phân tích mẫu, đồng thời giúp kiểm tra lại phương pháp đánh giá cảm quan.

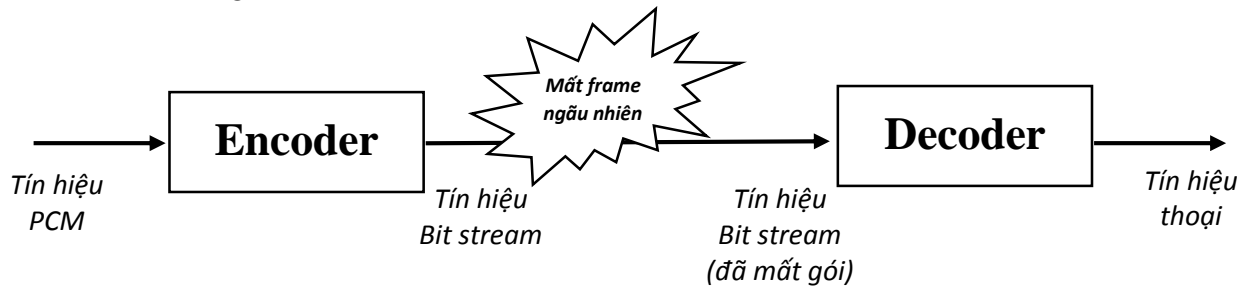
Quá trình thực nghiệm được thực hiện trên iLBC Codec và G.711 Codec. Mặc dù G.711 Codec hoạt động với tốc độ bit 64kbps, lớn hơn nhiều so với iLBC Codec, nhưng giải thuật mã hóa, giải mã G.711 cũng được tích hợp sẵn tính năng che giấu mất mát gói tin (Packet Loss Concealment). Từ đó, quá trình đánh giá khả năng chống mất gói của iLBC Codec sẽ khách quan hơn [17, tr.1].

CHƯƠNG 4 – ĐÁNH GIÁ BẰNG THỰC NGHIỆM

Quá trình đánh giá bằng thực nghiệm sử dụng các công cụ sau:

- Ứng dụng Simulink để mô phỏng quá trình mã hóa, mất ngẫu nhiên gói tin và giải mã gói tin sau khi thu nhận được đối với hai Codec là G.711 và iLBC.
- Phần mềm Matlab để tính toán, thống kê các giá trị MSE của tín hiệu sau khi bù mất gói so với tín hiệu ban đầu.

Sơ đồ thực nghiệm như sau:



Hình 4.1 - Sơ đồ khối thực nghiệm việc mất gói tin

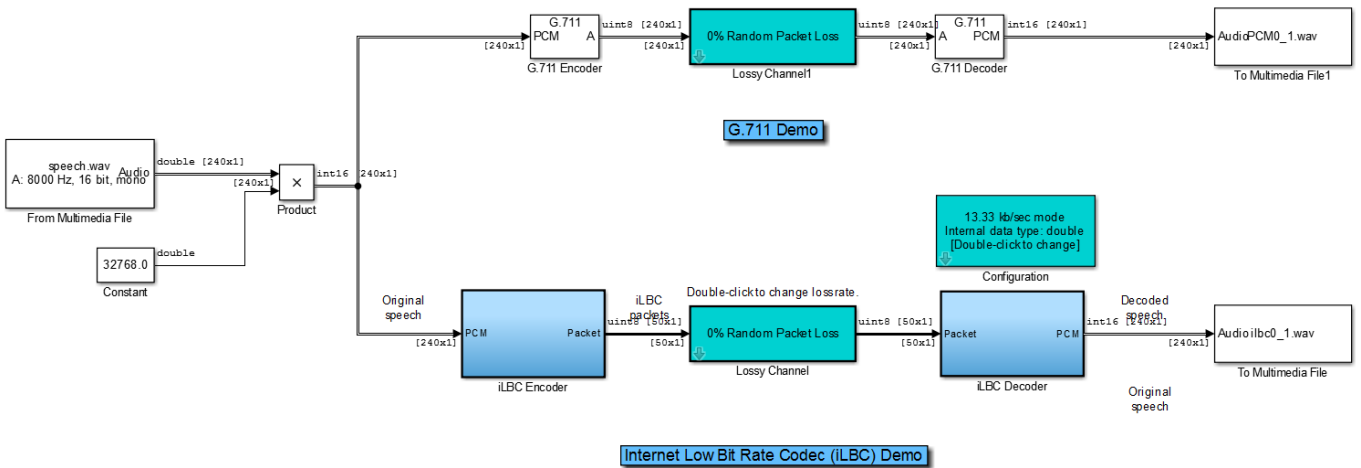
Quá trình thực nghiệm được làm theo các bước sau:

- Sử dụng ứng dụng Simulink mô phỏng một hệ thống mã hóa và giải mã file âm thanh sau khi đã làm mất gói tin một cách ngẫu nhiên. Một điểm cần lưu ý là các iLBC Codec có thể lựa chọn tốc độ mã hóa và giải mã ở kiểu frame có chiều dài 30ms, với tốc độ bit tương ứng là 13,33kbps hoặc kiểu frame 20ms, với tốc độ bit tương ứng 15,2kbps.
- Các giá trị gói tin bị mất được tính theo phần trăm số mẫu trong gói tin ban đầu, lần lượt là 5%, 10%, 15%, 20%, 50%, 80%.
- Sau các bước trên, tất cả các giá trị đo đạc được từ số lượng gói tin bị mất ảnh hưởng tới số mẫu/frame của âm thanh ban đầu và âm thanh sau khi được khôi phục sẽ được tính toán giá trị sai số, dựa theo công thức sai số trung bình bình phương MSE (Mean Squared Error).
- Đánh giá và vẽ biểu đồ so sánh dựa trên các giá trị MSE thu được ở trên.

4.1. Quá trình thực hiện

Thực nghiệm được thực hiện trên file âm thanh thoại ban đầu là file PCM, có định dạng .wav, được mã hóa đều 16 bit với tần số lấy mẫu là 8kHz, tên là *Speech.wav*.

File thực nghiệm sẽ được mã hóa ở dạng frame 30ms đối với iLBC Codec (Do iLBC hỗ trợ hai loại kích cỡ frame là 30ms và 20ms). Sơ đồ mô phỏng mã hóa và giả lập mất gói được thực hiện trên Simulink như sơ đồ sau:

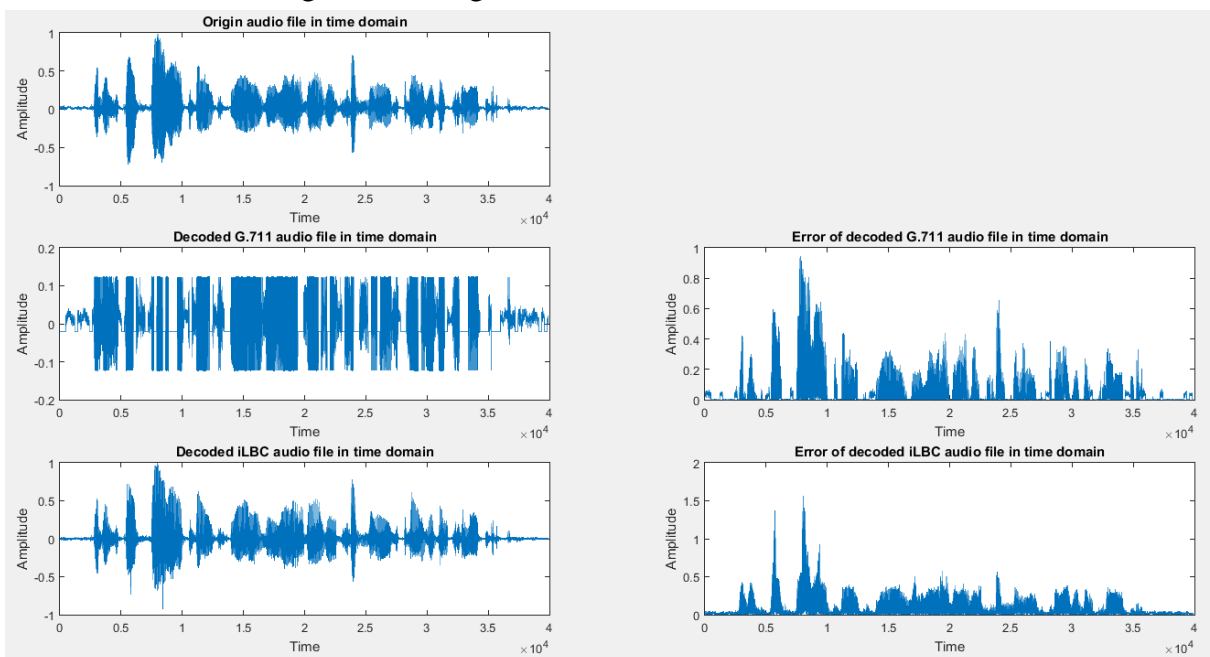


Hình 4.2 – Sơ đồ mô phỏng mã hóa và giả lập mất gói tin cho hai Codec

File âm thanh ban đầu sẽ được đưa đồng thời qua 2 bộ mã hóa tương ứng với G.711 Codec và iLBC Codec. Đầu ra của 2 bộ mã hóa này là các chuỗi bit stream, được đưa qua module giả lập mất gói (Loss Channel) với tỉ lệ mất gói có thể tùy chỉnh. Ứng với từng tỉ lệ mất gói tin, bộ giải mã của 2 Codec sẽ cho ra các file âm thanh đã được bù mất gói tương ứng.

Thực hiện nhập các tỉ lệ phần trăm mất gói vào cửa sổ Lossy Channel để thu được các file âm thanh được giải mã tương ứng. Mỗi file thu được sẽ được đặt tên theo một nguyên tắc chung. Ví dụ, ứng với tỉ lệ mất gói là 20% thì file giải mã của iLBC Codec thu được sẽ là *ilbc20_1.wav*, của G.711 Codec sẽ là *pcm20_1.wav*. Phần “_1” trong tên của file giải mã là thứ tự thực nghiệm giải mã (lần một là “_1”, lần hai là “_2”).

So sánh tín hiệu của 2 file lúc trước và sau khi mô phỏng mất gói tin 20% của hai Codec, biểu diễn trong miền thời gian:



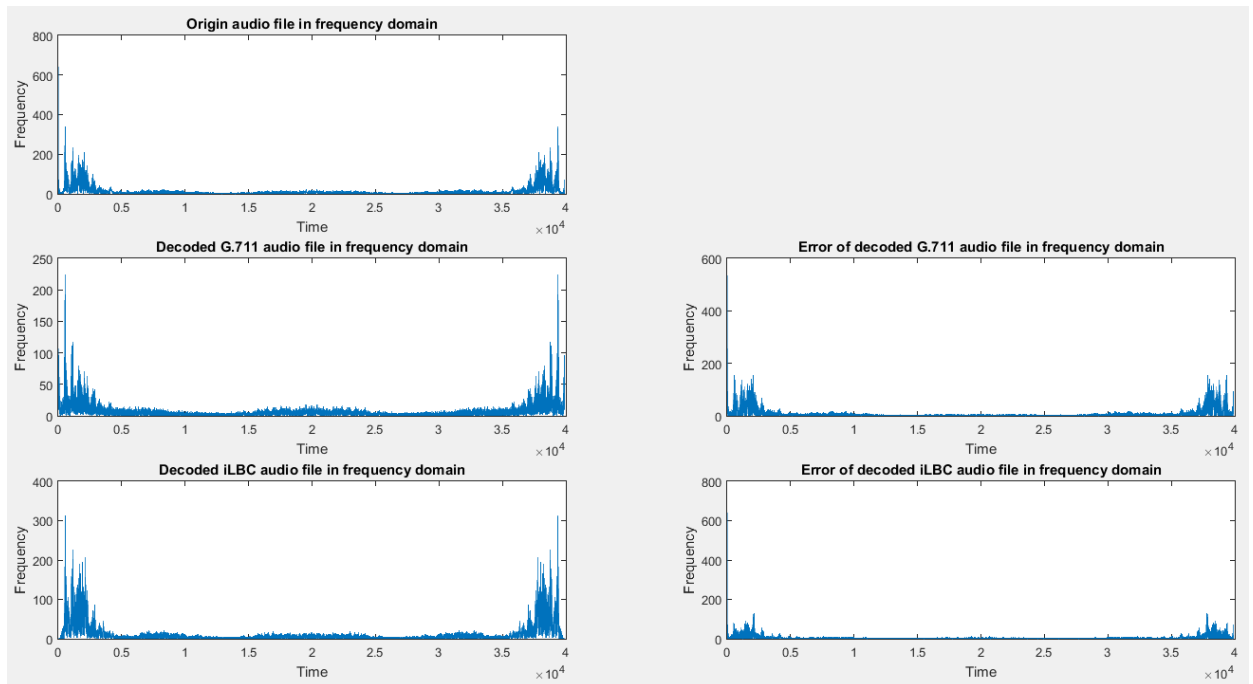
Hình 4.3 – Dạng sóng của các file trong miền thời gian trước và sau khi bù mất gói ở tỉ lệ 20%

Nhận thấy trong miền thời gian, các mẫu tham gia đánh giá để tính toán độ mất mát gói tin khi gói tin bị mất trên đường truyền không được chính xác. Do vậy thực nghiệm được làm trên miền tần số của các tín hiệu thu được. Quá trình sẽ so sánh các file sau giải mã với file ban đầu trong miền tần số, rồi từ đó tính giá trị MSE của từng mẫu.

So sánh tín hiệu của 2 file lúc trước và sau khi mô phỏng mất gói tin 20% của hai Codec trong miền tần số, sử dụng công thức biến đổi Fourier nhanh. Câu lệnh để biến đổi Fourier nhanh cho file âm thanh ban đầu trên Matlab như sau:

```
origin = audioread('speech.wav');
origin = abs(fft(origin));
```

Làm tương tự đối với các file âm thanh đã giải mã, ta cũng thu được tập hợp các biến đổi Fourier của các tín hiệu. Ví dụ đối với tỉ lệ mất gói là 20%, miền tần số của các mẫu sẽ được biểu diễn như sau:



Hình 4.4 – Biến đổi FFT của các file trước và sau khi bù mất gói ở tỉ lệ 20%

Để đánh giá chủ quan các file âm thanh này, thực hiện việc nghe trực tiếp từng file và đưa ra đánh giá chủ quan về chất lượng âm thanh. Nhận thấy có sự khác biệt về chất lượng giữa các file nhưng không nhiều, file sau khi khôi phục mất gói có chất lượng kém hơn một chút, một số vị trí bị rè hoặc nhảy từ nhưng không rõ lắm. Nhìn trên hình 4.3 ta cũng thấy sự khác nhau giữa phần tín hiệu lỗi giữa hai Codec đối với tín hiệu ban đầu. Tuy nhiên đây chỉ là ước lượng cảm quan. Đánh giá khách quan hơn bằng cách sử dụng phương pháp tính sai số trung bình bình phương (MSE) của các file đó so với file ban đầu, theo giá trị tần số.

Sử dụng phần mềm Matlab để phân tích các mẫu trên hai file đầu vào theo giá trị tần số và tính giá trị MSE giữa chúng:



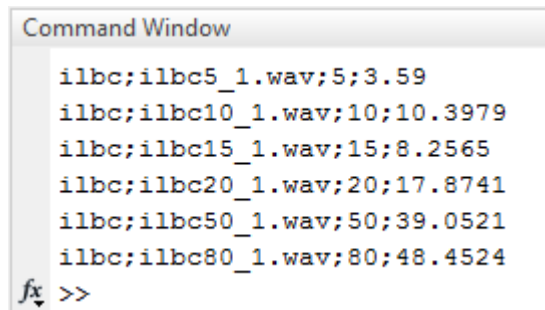
```

Editor - D:\Dropbox\LUANVAN\demo\msecalculate.m
fftcompare.m x msecalculate.m x +
1 - clear
2 - clc
3 - run=1;
4 - codec='ilbc';
5 - addpath('D:\Dropbox\LUANVAN\demo');
6 - x = audioread('speech.wav');
7 - x=abs(fft(x));
8 - maxx=max(x);
9 - length = length(x);
10
11 - packetLoss=[5, 10, 15, 20, 50, 80];
12 - for i = 1 : 6
13 -     loss=packetLoss(i);
14 -     filename=strcat(codec,num2str(loss),'_',num2str(run), '.wav');
15 -     y = audioread(filename);
16 -     y(length + 1:end) = [];
17 -     y=abs(fft(y));
18 -     MSE = sum((y - x).^2)/length;
19 -     MSE = MSE/maxx * 100;
20 -     disp(strcat(codec,';', filename,';', num2str(loss), '; ', num2str(MSE)));
21
22 - end
23

```

Hình 4.5 – Tính giá trị MSE của các file âm thanh

Sau khi chạy đoạn Script trên để tính MSE, kết quả thu được như sau:



```

Command Window
ilbc;ilbc5_1.wav;5;3.59
ilbc;ilbc10_1.wav;10;10.3979
ilbc;ilbc15_1.wav;15;8.2565
ilbc;ilbc20_1.wav;20;17.8741
ilbc;ilbc50_1.wav;50;39.0521
ilbc;ilbc80_1.wav;80;48.4524
fx >>

```

Làm tương tự đối với G.711 Codec, kết quả thu được các giá trị MSE của Codec này so với tín hiệu ban đầu theo từng tỉ lệ mất gói tin. Các kết quả này được lưu trữ lại phục vụ cho việc phân tích và thống kê để đưa ra kết quả cuối cùng.

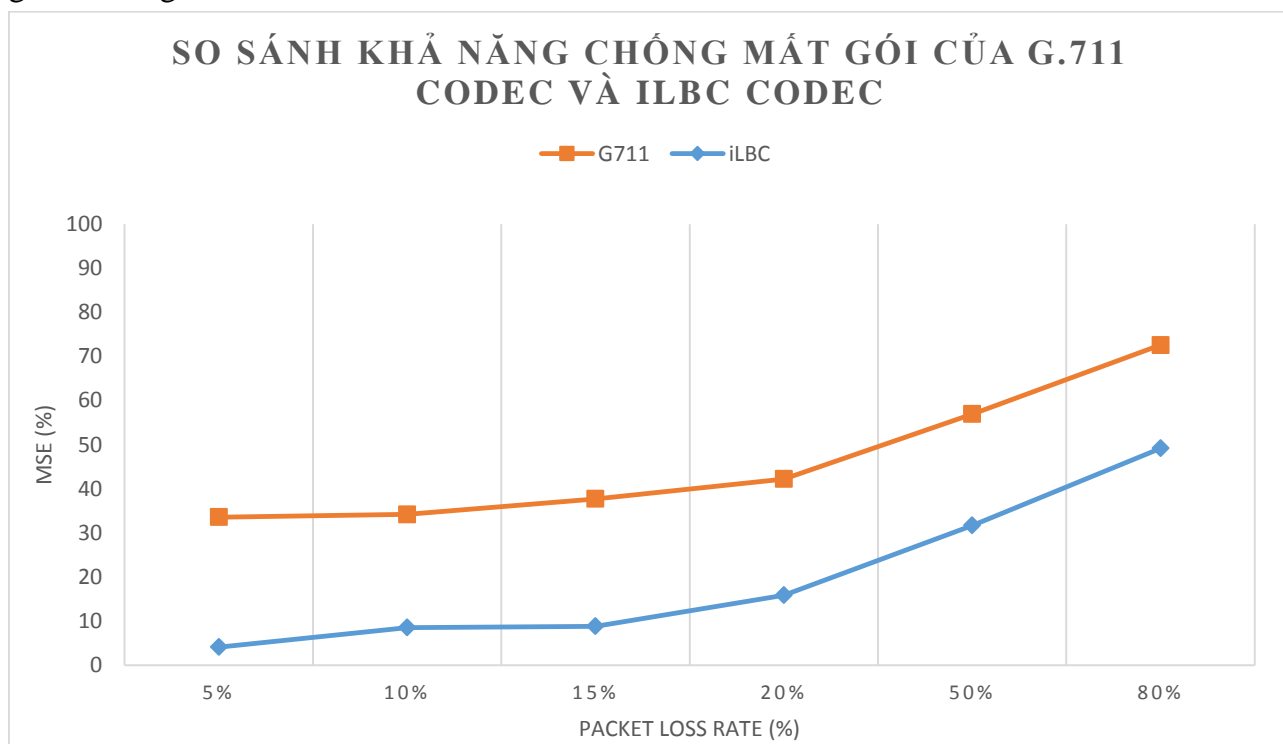
4. 2. Kết quả của quá trình thực nghiệm

Như vậy, đối với iLBC Codec, ứng với mỗi tỉ lệ gói tin bị mất (5%, 10%, 15%, 20%, 50%, 80%), ta thu được giá trị MSE tương ứng. Chú ý là giá trị MSE càng nhỏ thì sai số giữa các mẫu càng thấp, và file âm thanh được khôi phục càng giống với file ban đầu trước khi bị mất gói.

Làm tương tự đối với G.711 Codec, ứng với từng giá trị mất gói ta cũng thu được các giá trị MSE của Codec này. Sau 3 lần thực hiện thực nghiệm, các giá trị MSE thực nghiệm thu được ở bảng dưới đây (Cột trung bình là giá trị trung bình sau 3 lần đo):

	LẦN 1		LẦN 2		LẦN 3		TRUNG BÌNH	
PLR	G711	iLBC	G711	iLBC	G711	iLBC	G.711	iLBC
5%	33.178	3.59	34.0962	3.963	33.4435	4.782	33.5726	4.1117
10%	34.4731	10.3979	35.4731	8.6858	32.6413	6.5793	34.1958	8.5543
15%	37.0358	8.2565	39.9381	9.0378	36.1626	9.2011	37.7122	8.8318
20%	42.8885	17.8741	41.2652	10.1482	42.4233	19.5354	42.1923	15.8526
50%	57.6702	39.0521	54.3549	23.03	58.7473	32.9715	56.9241	31.6845
80%	71.1384	48.4524	72.9959	47.7667	73.5138	51.2692	72.5494	49.1628

Kết quả cuối cùng thu được được mô tả ở đồ thị dưới đây. Các thành phần phần trăm mất gói được biểu diễn ở trục hoành, trục tung là các giá trị MSE tương ứng. Các giá trị mất gói lần lượt là 5%, 10%, 15%, 20%, 50% và 80%.



Hình 4.6 – Kết quả thực nghiệm khả năng chống mất gói của G.711 Codec và iLBC Codec sau 3 lần thực nghiệm

Nhìn vào đồ thị ta có một số kết luận sau:

- Trong điều kiện mất gói tin càng nhiều (từ 5% đến 80%, tính từ trái qua phải), thì đồ thị có xu hướng đi lên trên, hay giá trị sai số trung bình bình phương MSE của các Codec cũng tăng lên, tức là khả năng khôi phục các frame bị mất của các Codec sẽ giảm dần theo phần trăm mất gói.
- Trong điều kiện mất gói dưới 15%, giá trị MSE của từng Codec chênh lệch nhau không nhiều (đồ thị không đi lên quá nhanh), điều đó có nghĩa là khi lượng gói tin của từng Codec bị mất dưới 15%, khả năng khôi phục của Codec

hoạt động tốt, các gói tin được phục hồi gần giống với gói tin ban đầu nhất. Khi tỉ lệ mất gói trên 15%, đồ thị có xu hướng đi lên nhanh hơn.

- Xét về tổng thể thì giá trị MSE của iLBC Codec thấp hơn của G.711 Codec, dù ở điều kiện mất gói như thế nào. Ví dụ, ở lần thực nghiệm thứ nhất:

- + Tỉ lệ mất gói 5%: $MSE_{G.711}(33.178) > MSE_{iLBC}(3.59)$

- + Tỉ lệ mất gói 10%: $MSE_{G.711}(34.4731) > MSE_{iLBC}(10.3979)$

Do đó có thể nói ngoài ưu điểm tốc độ bit sử dụng thấp hơn, khả năng bù mất gói của iLBC Codec cũng tốt hơn G.711 Codec.

KẾT LUẬN

Việc tiết kiệm băng thông đường truyền và nâng cao chất lượng hệ thống thoại là điều quan trọng mà các nhà sản xuất thiết bị, nhà cung cấp dịch vụ cần phải quan tâm. Tuy nhiên, khi truyền thoại trên mạng chuyển mạch gói, do thoại là dịch vụ thời gian thực nên nó yêu cầu thời gian trễ và tỷ lệ mất gói rất thấp. Mất gói lớn xảy ra làm chất lượng hội thoại kém đi, gây khó chịu cho người sử dụng. Vì vậy, việc nghiên cứu và áp dụng các biện pháp để chống mất mát gói tin, nâng cao chất lượng hệ thống thông tin thoại trong mạng IP là điều rất cần thiết.

Qua quá trình tìm hiểu, luận văn đã thực hiện được một số công việc sau:

- Trình bày lý thuyết về hệ thống thông tin thoại, các yếu tố ảnh hưởng đến chất lượng truyền dẫn thoại, cùng các phương pháp đánh giá chất lượng tiếng nói chủ quan và khách quan.

- Trình bày về hai kỹ thuật mã hóa tiếng nói dựa trên dự đoán tuyến tính thường được dùng trong mạng IP: LPC và CELP.

- Tìm hiểu được thuật toán mã hóa, giải mã và phân tích được khả năng chống mất gói của iLBC Codec.

- Đánh giá định lượng khả năng chống mất gói của iLBC Codec so với G.711 Codec, bằng ngôn ngữ Matlab và ứng dụng mô phỏng Simulink; thực hiện đánh giá định lượng bằng phương pháp đánh giá khách quan và kiểm nghiệm lại bằng phương pháp đánh giá chủ quan.

Tuy nhiên, việc thực hiện bù mất gói mới chỉ đánh giá được khả năng chống mất mát gói tin do mất ngẫu nhiên các gói (packet loss rate). Do vậy, hướng phát triển của đề tài trong tương lai sẽ là:

- Thực hiện các phương pháp đánh giá chống mất mát gói tin theo vị trí mất gói (packet loss location), kích cỡ mất gói (packet loss size) và mẫu gói tin bị mất (packet loss pattern) [13, tr.1].

- Thực hiện các phương pháp đánh giá do mất gói tin theo chùm (bursty loss).

TÀI LIỆU THAM KHẢO

TÀI LIỆU TIẾNG VIỆT:

1. Nguyễn Đại Hòa (2013), *Nghiên cứu kỹ thuật mã hóa tiếng nói trong di động*, Đồ án tốt nghiệp Đại học, Học viện Công nghệ Bưu chính Viễn thông, Khoa Viễn thông 2.
2. Trương Lê Phương Anh (2011), *Nâng cao chất lượng thoại trên mạng IP bằng kỹ thuật bù mất gói*, Luận văn Thạc sĩ kỹ thuật, ĐH Đà Nẵng.
3. *Các biện pháp đảm bảo chất lượng dịch vụ*, Đại học Sư phạm Kỹ thuật Hưng Yên, Khoa Công nghệ thông tin.
4. Nguyễn Thị Ngọc Diệp, *Thuật toán xử lý tiếng nói trong Speech Enhancement và đánh giá tính hiệu quả của thuật toán*, ĐH Bách khoa Đà Nẵng, Khoa Điện tử viễn thông.

TÀI LIỆU TIẾNG ANH:

5. S. Andersen - Aalborg University; A. Duric, Telio, H. Astrom, R. Hagen, W. Kleijn, J. Linden (December 2004), *Internet Low Bit Rate Codec (iLBC)*, Request for Comments: 3951, Global IP Sound.
6. A. Duric, Telio, S. Andersen - Aalborg University (December 2004), *Real-time Transport Protocol (RTP) Payload Format for internet Low Bit Rate Codec (iLBC) Speech*, Request for Comments: 3952.
7. S. V. Andersen, W. B. Kleijn, R. Hagen, J. Linden, M. N. Murthi, J. Skoglund, *iLBC - A linear predictive coder with robustness to packet losses*, Department of Communication Technology - Aalborg University, Denmark; Global IP Sound - Stockholm, Sweden and San Francisco, USA.
8. Wai C. Chu, *Speech Coding Algorithms – Foundation and Evolution of Standardized Coders*, Mobile Media Laboratory, DoCoMo USA Labs, San Jose, California.
9. A. M. Kondozi, *Digital Speech - Coding for Low Bit Rate Communication Systems*, University of Surrey, UK.
10. Lingfen Sun (January 2004), *Speech quality prediction for voice over internet protocol networks*, Doctor of Philosophy, School of Computing, Communications and Electronics Faculty of Technology.
11. Lijing Ding and Rafik A. Goubran (2003), *Assessment of Effects of Packet Loss on Speech Quality in VoIP*, Department of Systems and Computer Engineering, Carleton University.
12. Wenyu Jiang, Henning Schulzrinne (June 2002), *Comparisons of FEC and Codec Robustness on VoIP quality and bandwidth efficiency*, Columbia University, Department of Computer Science.

- 13.** Lingfen Sun; Emmanuel Ifeachor (2012), *Impact of Packet Loss Location on Perceived Speech Quality*, University of Plymouth.
- 14.** Alexander F. Ribadeneira (2007), *An analysis of the MOS under conditions of Delay, Jitter and Packet Loss and an Analysis of the Impact of Introducing Piggybacking and Reed Solomon FEC for VoIP*, Master of Science, College of Arts and Sciences, Georgia State University.
- 15.** Muhammad Azam (October 2011), *Methods for Recovery of Missing Speech Packets*, Master of Science, Blekinge Institute of Technology.
- 16.** Nam In Park, Hong Kook Kim, Min A Jung, Seong Ro Lee, Seung Ho Choi (2011), *Burst Packet Loss Concealment Using Multiple Codebooks and Comfort Noise for CELP-Type Speech Coders in Wireless Sensor Networks*, Paper, Gwangju Institute of Science and Technology (GIST), Mokpo National University, Seoul National University of Science and Technology.
- 17.** Artur Janicki, Bartłomiej Ksiezak, *Packet Loss Concealment Algorithm for VoIP Transmission in Unreliable Networks*, Institute of Telecommunications, Warsaw University of Technology.
- 18.** Mohamed CHIBANI (2007), *Increasing the Robustness of CELP Speech Codecs against Packet Losses*, Université de Sherbrooke, Québec, Canada.