



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0027276

(51)⁷ G01L 19/20

(13) B

(21) 1-2017-00383

(22) 23/07/2015

(86) PCT/CN2015/084931 23/07/2015

(87) WO2016/015591 04/02/2016

(30) 62/029,437 26/07/2014 US; 14/511,943 10/10/2014 US

(45) 25/02/2021 395

(43) 25/04/2017 349A

(73) HUAWEI TECHNOLOGIES CO., LTD. (CN)

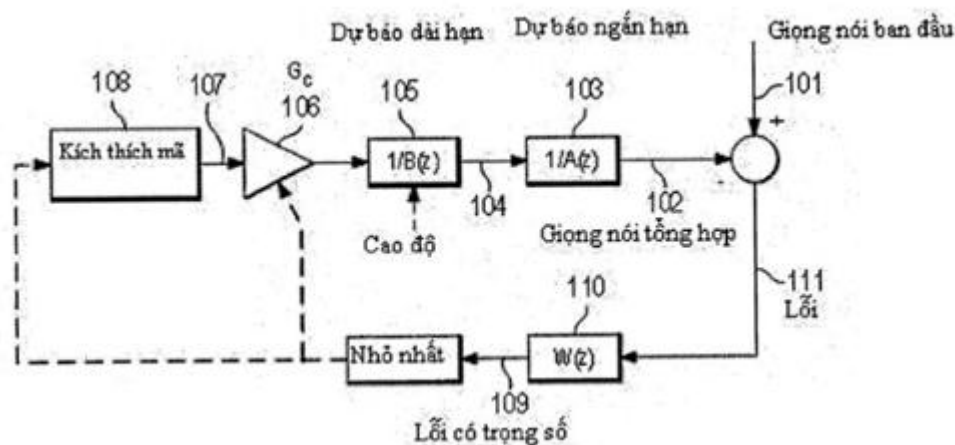
Huawei Administration Building, Bantian, Longgang, Shenzhen, Guangdong 518129, China

(72) GAO, Yang (US).

(74) Công ty Luật TNHH Phạm và Liên danh (PHAM & ASSOCIATES)

(54) PHƯƠNG PHÁP VÀ THIẾT BỊ XỬ LÝ TÍN HIỆU GIỌNG NÓI

(57) Sáng chế đề xuất phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh gồm lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số.



Lĩnh vực kỹ thuật được đề cập

Sáng chế liên quan đến lĩnh vực mã hóa tín hiệu. Cụ thể là, sáng chế thuộc lĩnh vực cải thiện phân loại giữa mã hóa miền thời gian và mã hóa miền tần số.

Tình trạng kỹ thuật của sáng chế

Mã hóa giọng nói liên quan đến quá trình giảm tốc độ bit của tệp tin giọng nói. Mã hóa giọng nói là ứng dụng nén dữ liệu của các tín hiệu âm thanh số chứa giọng nói. Mã hóa giọng nói sử dụng ước tính tham số giọng nói cụ thể sử dụng các kỹ thuật xử lý tín hiệu âm thanh để tạo mô hình tín hiệu giọng nói, kết hợp với các thuật toán nén dữ liệu tổng quát để biểu diễn các tham số tạo mô hình cuối cùng ở dòng bit thu nhỏ. Mục đích của việc mã hóa giọng nói là tiết kiệm không gian lưu trữ bộ nhớ cần thiết, băng thông truyền và công suất truyền nhờ giảm số lượng bit trên mẫu sao cho giọng nói được mã hóa (được giải nén) không khác biệt về mặt trực giác với giọng nói gốc.

Tuy nhiên, các bộ mã hóa giọng nói là các bộ mã hóa tổn hao, tức là, tín hiệu được giải mã khác với tín hiệu gốc. Do vậy, một trong các mục đích khi mã hóa giọng nói là nhằm giảm tối đa biến dạng (hoặc tổn hao nhận thấy được) ở tốc độ bit cụ thể, hoặc giảm tối đa tốc độ bit để có biến dạng cụ thể.

Mã hóa giọng nói khác các dạng mã hóa âm thanh khác ở chỗ giọng nói là tín hiệu đơn giản hơn nhiều so với phần lớn các tín hiệu âm thanh khác, và nhiều thông tin thống kê hơn khả dụng về các thuộc tính của giọng nói. Kết quả là, một số thông tin thính giác liên quan đến việc mã hóa âm thanh có thể là không cần thiết trong ngữ cảnh mã hóa giọng nói.

Khi mã hóa giọng nói, tiêu chí quan trọng nhất là việc giữ gìn độ rõ và “dễ nghe” của giọng nói, với lượng dữ liệu được truyền bị giới hạn.

Độ rõ của giọng nói còn gồm, bên cạnh nội dung nguyên văn thực, định danh người nói, các cảm xúc, âm điệu, âm sắc, v.v. đều quan trọng cho độ rõ hoàn chỉnh. Khái niệm trừu tượng hơn về độ dễ nghe của giọng nói bị giảm cấp là thuộc tính khác so với độ rõ, do có khả năng giọng nói giảm cấp hoàn toàn dễ hiểu, nhưng một cách khách quan gây khó chịu cho người nghe.

Một cách truyền thống, tất cả các phương pháp mã hóa giọng nói tham số tận dụng sự dư thừa có trong tín hiệu giọng nói để giảm lượng thông tin phải được gửi và để ước tính các tham số của các mẫu giọng nói của tín hiệu ở các khoảng ngắn. Sự dư thừa này chủ yếu phát sinh từ việc lặp lại của các dạng sóng giọng nói ở tốc độ một phần tư chu kỳ, và đường bao phổ thay đổi chậm của tín hiệu giọng nói.

Sự dư thừa của các dạng sóng giọng nói có thể được xem xét so với một số loại tín hiệu giọng nói khác nhau, như các tín hiệu giọng nói có thoại và không có thoại. Các âm có thoại, chẳng hạn, ‘a’, ‘b’, chủ yếu do các rung động của dây các thanh âm, và dao động. Do vậy, trên các khoảng thời gian ngắn, chúng được tạo mô hình chuẩn nhờ các tổng của các tín hiệu tuần hoàn như các đường hình sin. Nói theo cách khác, đối với giọng nói có thoại, tín hiệu giọng nói chủ yếu là tuần hoàn. Tuy nhiên, tính tuần hoàn này có thể biến thiên trên khoảng thời gian của đoạn giọng nói và dạng sóng tuần hoàn thường thay đổi từ đoạn này sang đoạn khác. Mã hóa giọng nói tốc độ bit thấp có thể hưởng lợi nhiều từ việc khai thác tính tuần hoàn này. Mã hóa giọng nói miền thời gian có thể hưởng lợi lớn từ việc khai thác tính tuần hoàn này. Chu kỳ giọng nói có thoại còn được gọi là cao độ, và việc dự báo cao độ thường được gọi là dự báo thời hạn dài (Long-Term Prediction, LTP). Ngược lại, các âm thanh không có thoại như ‘s’, ‘sh’, giống nhiều hơn. Đây là do tín hiệu

giọng nói không có thoại giống nhiều ngẫu nhiên nhiều hơn và có lượng dự báo được nhỏ hơn.

Trong trường hợp bất kỳ, mã hóa tham số có thể được sử dụng để giảm sự dư thừa của các đoạn giọng nói nhờ tách riêng thành phần kích thích của tín hiệu giọng nói với thành phần đường bao phổ, mà thay đổi ở tốc độ chậm hơn. Thành phần đường bao phổ thay đổi chậm có thể được biểu diễn bằng mã hóa dự báo tuyến tính (Linear Prediction Coding, LPC) còn được gọi là dự báo thời hạn ngắn (Short-Term Prediction, STP). Mã hóa giọng nói tốc độ bit thấp có thể còn hưởng lợi nhiều từ việc khai thác STP này. Ưu điểm mã hóa phát sinh từ tốc độ thấp mà ở đó các tham số thay đổi. Tuy nhiên, đối với các tham số này hiếm khi khác đáng kể với các giá trị được giữ trong vài mili giây.

Ở các chuẩn đã biết gần đây hơn như G.723.1, G.729, G.718, tốc độ hoàn chỉnh tăng cường (Enhanced Full Rate, EFR), bộ mã hóa tiếng nói kiểu lựa chọn được (Selectable Mode Vocoder, SMV), đa tốc độ thích ứng (Adaptive Multi-Rate, AMR), băng rộng đa chế độ tốc độ biến thiên (Variable-Rate Multimode Wideband, VMR-WB), hoặc băng rộng đa tốc độ thích ứng (Adaptive Multi-Rate Wideband, AMR-WB), kỹ thuật dự báo tuyến tính kích thích bằng mã (Code Excited Linear Prediction, CELP) đã được đưa vào sử dụng. CELP thường được hiểu như là tổ hợp kỹ thuật của kích thích bằng mã, LTP và STP. CELP chủ yếu được sử dụng để mã hóa tín hiệu giọng nói bằng cách hưởng lợi từ các đặc tính giọng nói con người cụ thể hoặc mô hình tạo giọng nói thanh âm con người. Mã hóa giọng nói CELP là nguyên lý thuật toán rất phổ biến trong khu vực nén giọng nói mặc dù các chi tiết của CELP để mã hóa/giải mã khác nhau có thể khác nhau đáng kể. Do tính phổ biến của nó, thuật toán CELP đã được sử dụng trong các chuẩn ITU-T, MPEG, 3GPP, và 3GPP2 khác nhau. Các biến thể của CELP gồm CELP đại số, CELP không căng, CELP độ trễ thấp và dự báo tuyến tính kích thích bằng tổng vector, và các

biến thể khác. CELP là thuật ngữ chung cho lớp các thuật toán và không dành cho mã hóa/giải mã cụ thể.

Thuật toán CELP dựa trên bốn ý tưởng chính. Trước hết, mô hình lọc nguồn của việc tạo giọng nói thông qua LP (linear prediction, dự báo tuyến tính) được sử dụng. Mô hình lọc nguồn tạo giọng nói tạo mô hình giọng nói dưới dạng tổ hợp nguồn âm thanh, như các dây thanh âm, và bộ lọc âm tuyến tính, đường âm thanh (và đặc tính bức xạ). Khi triển khai mô hình lọc nguồn tạo giọng nói, nguồn âm thanh, hoặc tín hiệu kích thích, thường được tạo mô hình dưới dạng chuỗi xung tuần hoàn, đối với giọng nói có thoại, hoặc nhiễu trắng cho giọng nói không có thoại. Thứ hai, từ điển mã cố định và thích ứng được sử dụng làm đầu vào (kích thích) mô hình LP. Thứ ba, thực hiện tìm kiếm trong vòng kín trong “miền có trọng số về mặt nhận thức”. Thứ tư, áp dụng lượng tử hóa vector (vector quantization, VQ).

Bản chất kỹ thuật của sáng chế

Theo phương án thực hiện sáng chế, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh gồm lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số.

Theo phương án thực hiện khác của sáng chế, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh bao gồm lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên. Theo cách khác, phương pháp lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới. Tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ.

Theo phương án thực hiện khác của sáng chế, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa bao gồm lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số bao gồm dữ liệu âm thanh khi tín hiệu số không bao gồm tín hiệu cao độ ngắn và tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường. Phương pháp còn bao gồm lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên. Tín hiệu số bao gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm là thấp. Phương pháp còn gồm lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian và tín hiệu số bao gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm rất mạnh.

Theo phương án thực hiện khác của sáng chế, thiết bị xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh bao gồm bộ lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số.

Mô tả vắn tắt các hình vẽ

Để hiểu rõ hơn sáng chế, và các ưu điểm của nó, thực hiện tham khảo các phần mô tả dưới đây có dựa vào các hình vẽ đi kèm, trong đó:

Fig.1 minh họa các hoạt động được thực hiện trong quá trình mã hóa giọng nói gốc sử dụng bộ mã hóa CELP đã biết;

Fig.2 minh họa các hoạt động được thực hiện trong quá trình giải mã giọng nói gốc sử dụng bộ giải mã CELP;

Fig.3 minh họa bộ mã hóa CELP đã biết;

Fig.4 minh họa bộ giải mã CELP cơ bản tương ứng với bộ mã hóa trên Fig.3;

Fig.5 và Fig.6 minh họa các ví dụ về các tín hiệu giọng nói dạng lược đồ và quản hệ của nó với kích thước khung và kích thước khung phụ trong miền thời gian;

Fig.7 minh họa ví dụ của phổ băng rộng có thanh âm gốc;

Fig.8 minh họa phổ băng rộng có thanh âm được mã hóa của phổ băng rộng có thanh âm gốc được minh họa trên Fig.7 bằng cách sử dụng gấp đôi mã hóa độ trễ cao độ;

Fig.9A và Fig.9B minh họa lược đồ mã hóa/giải mã trực giác miền tần số đặc trưng, trong đó Fig.9A minh họa bộ mã hóa miền tần số trong đó Fig.9B minh họa bộ giải mã miền tần số;

Fig.10 minh họa lược đồ của các hoạt động ở bộ mã hóa trước khi mã hóa tín hiệu giọng nói bao gồm dữ liệu âm thanh theo các phương án thực hiện sáng chế;

Fig.11 minh họa hệ thống truyền thông 10 theo phương án thực hiện sáng chế;

Fig.12 minh họa sơ đồ khối của hệ thống xử lý có thể được sử dụng để triển khai các thiết bị và các phương pháp được bộc lộ ở đây;

Fig.13 minh họa sơ đồ khối của thiết bị xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số; và

Fig.14 minh họa sơ đồ khối của thiết bị xử lý các tín hiệu giọng nói khác trước khi mã hóa tín hiệu số.

Mô tả chi tiết sáng chế

Trong hệ thống truyền thông tín hiệu số âm thanh/giọng nói hiện đại, tín hiệu số được nén ở bộ mã hóa, và thông tin nén hoặc dòng bit có thể được đóng gói và được gửi đến bộ giải mã theo từng khung qua kênh truyền thông. Bộ giải mã nhận và giải mã thông tin nén để có tín hiệu số âm thanh/giọng nói.

Ở hệ thống truyền thông tín hiệu số âm thanh/giọng nói hiện đại, tín hiệu số được nén ở bộ mã hóa, và thông tin nén hoặc dòng bit có thể được đóng gói và gửi đến bộ giải mã theo từng khung qua kênh truyền thông. Hệ thống của cả bộ mã hóa và bộ giải mã đều được gọi là codec. Việc nén âm thanh/giọng nói có thể được sử dụng để giảm số lượng bit biểu diễn tín hiệu âm thanh/giọng nói nhờ đó giảm băng thông và/hoặc tốc độ bit cần để truyền. Nói chung, tốc độ bit cao hơn sẽ cho chất lượng âm thanh cao hơn, trong khi tốc độ bit thấp hơn sẽ có chất lượng âm thanh thấp hơn.

Fig.1 minh họa các hoạt động được thực hiện trong quá trình mã hóa giọng nói gốc sử dụng bộ mã hóa CELP đã biết.

Fig.1 minh họa bộ mã hóa CELP ban đầu đã biết trong đó lỗi có trọng số 109 giữa giọng nói được tổng hợp 102 và giọng nói gốc 101 được giảm tối đa thường bằng cách sử dụng cách phân tích nhờ tổng hợp, nghĩa là việc mã hóa (phân tích) được thực hiện bằng cách tối ưu hóa trực giác tín hiệu giải mã (tổng hợp) trong vòng kín.

Nguyên lý cơ bản mà tất cả các bộ mã hóa giọng nói triển khai chính là việc các tín hiệu giọng nói là các dạng sóng có độ tương quan cao. Như là minh họa, giọng nói có thể được biểu diễn bằng cách sử dụng mô hình tự hồi quy (autoregressive, AR) như ở phương trình (1) dưới đây.

$$X_n = \sum_{i=1}^P a_i X_{n-i} + e_n \quad (1)$$

Ở phương trình (1), mỗi mẫu được biểu diễn dưới dạng tổ hợp tuyến tính của P mẫu trước đó cộng nhiễu trắng. Các hệ số trọng số $a_1, a_2, \dots, a_P$, được gọi là hệ số dự báo tuyến tính (Linear Prediction Coefficient, LPC). Đối với mỗi khung, các hệ số trọng số $a_1, a_2, \dots, a_P$, được chọn sao cho phổ $\{X_1, X_2, \dots, X_N\}$, được tạo nhờ sử dụng mô hình trên đây, gần như so khớp phổ của khung thoại đầu vào.

Theo cách khác, các tín hiệu giọng nói có thể còn được biểu thị bằng tổ hợp của mô hình điều hòa và mô hình nhiễu. Phần điều hòa của mô hình một cách hiệu quả là biểu diễn chuỗi Fourier của thành phần điều hòa của tín hiệu. Nói chung, đối với các tín hiệu có thanh âm, mô hình nhiễu cộng điều hòa của giọng nói gồm hỗn hợp của cả điều hòa và nhiễu. Phần điều hòa và nhiễu trong giọng nói có thoai phụ thuộc nhiều yếu tố gồm các đặc tính của người nói (chẳng hạn, ở mức độ nhất định giọng nói của người nói là bình thường hoặc nghe rõ cả hơi thở); ký tự đoạn giọng nói (chẳng hạn, ở mức độ nào đó đoạn giọng nói là tuần hoàn) và trên tần số đó. Các tần số cao hơn của giọng nói có thoai có tỷ lệ các thành phần dạng nhiễu cao hơn.

Mô hình LP và mô hình nhiễu điều hòa là hai phương pháp chính để tạo mô hình và mã hóa các tín hiệu giọng nói. Mô hình LP cụ thể là tốt ở việc tạo mô hình đường bao phổ của giọng nói trong khi mô hình nhiễu điều hòa tốt ở việc tạo mô hình cấu trúc tinh chỉnh của giọng nói. Hai phương pháp có thể được kết hợp để tận dụng các điểm mạnh tương đối của chúng.

Như nêu trước đó, trước khi mã hóa CELP, tín hiệu đầu vào micrô của máy thu phát cầm tay được lọc và lấy mẫu, chẳng hạn, ở tốc độ 8000 mẫu trên giây. Sau đó, mỗi mẫu được lượng tử hóa, chẳng hạn, với 13 bit trên mẫu. Giọng nói được lấy mẫu được phân đoạn thành nhiều đoạn hoặc khung 20ms (chẳng hạn, trong trường hợp 160 mẫu).

Tín hiệu giọng nói được phân tích và mô hình LP của nó, các tín hiệu kích thích và cao độ được trích rút. Mô hình LP biểu diễn đường bao phổ của giọng nói. Nó được biến đổi thành tập hợp của các hệ số LSF (line spectral frequencies, tần số phổ đường thẳng), là biểu diễn thay thế của các tham số LP, do các hệ số LSF có các thuộc tính lượng tử hóa tốt. Các hệ số LSF có thể được lượng tử hóa vô hướng hoặc hiệu quả hơn chúng

có thể được lượng tử hóa vector nhờ sử dụng các từ điển mã vector LSF được đào tạo trước đó.

Việc kích thích mã gồm từ điển mã bao gồm các vector mã, có các thành phần được lựa chọn độc lập sao cho mỗi vector mã có thể có phổ gần như ‘trắng’. Đối với mỗi khung phụ của thoại đầu vào, mỗi vector mã được lọc qua bộ lọc dự báo tuyến tính khoảng ngắn 103 và bộ lọc LTP 105, và đầu ra được so sánh với các mẫu giọng nói. Ở mỗi khung phụ, vector mã có đầu ra so khớp tốt nhất với giọng nói đầu vào (lỗi được giảm nhỏ nhất) được chọn để biểu diễn khung phụ đó.

Kích thích được mã hóa 108 thông thường bao gồm tín hiệu dạng xung hoặc tín hiệu dạng nhiễu, được xây dựng về mặt toán học hoặc lưu lại trong từ điển mã. Từ điển mã khả dụng cho cả bộ mã hóa lẫn bộ giải mã nhận. Kích thích được mã hóa 108, có thể là từ điển mã cố định hoặc ngẫu nhiên, có thể là từ điển lượng tử hóa vector được mã hóa cứng (ngầm hoặc tường minh) vào codec. Từ điển mã cố định này có thể là LP được kích thích bằng mã đại số hoặc được lưu trữ tường minh.

Vector mã từ từ điển mã được tạo tỷ lệ bằng độ khuếch đại phù hợp để khiến năng lượng bằng năng lượng của giọng nói đầu vào. Do đó, đầu ra kích thích được mã hóa 108 được tạo tỷ lệ bằng độ khuếch đại G_c 107 trước khi đưa qua các bộ lọc tuyến tính.

Bộ lọc dự báo tuyến tính khoảng ngắn 103 tạo dạng phổ ‘trắng’ của vector mã để giống phổ của giọng nói đầu vào. Tương đương, trong miền thời gian, bộ lọc dự báo tuyến tính khoảng ngắn 103 đưa vào các mối tương quan thời gian ngắn (tương quan với các mẫu trước đó) trong chuỗi trắng. Bộ lọc tạo dạng kích thích có mô hình toàn cực có dạng $1/A(z)$ (bộ lọc dự báo tuyến tính khoảng ngắn 103), trong đó $A(z)$ được gọi là bộ lọc dự báo và có thể thu được nhờ sử dụng LP (chẳng hạn, thuật toán Levinson–Durbin). Theo một hoặc nhiều các phương án thực hiện, bộ lọc

toàn cục có thể được sử dụng do nó là biểu diễn tốt của dây thanh âm con người và do nó dễ tính.

Bộ lọc dự báo tuyến tính khoảng ngắn 103 thu được nhờ phân tích tín hiệu gốc 101 và được biểu diễn bởi nhóm hệ số:

$$A(z) = \sum_{i=1}^P 1 + a_i \cdot z^{-i}, i=1,2,\dots,P \quad (2)$$

Như nêu trước đó, các vùng của giọng nói có thoại biểu diễn tính tuần hoàn thời hạn dài. Chu kỳ này, được biết là cao độ, được đưa vào phổ tổng hợp bởi bộ lọc cao độ $1/(B(z))$. Đầu ra của bộ lọc LTP 105 tùy thuộc vào cao độ và độ khuếch đại cao độ. Theo một hoặc nhiều các phương án thực hiện, cao độ có thể được ước tính từ tín hiệu gốc, tín hiệu dư, hoặc tín hiệu gốc có trọng số. Theo một phương án thực hiện, hàm LTP ($B(z)$) có thể được biểu diễn nhờ sử dụng phương trình (3) như sau.

$$B(z) = 1 - G_p \cdot z^{-Pitch} \quad (3)$$

Bộ lọc trọng số 110 liên quan đến bộ lọc STP nêu trên. Một trong các bộ lọc trọng số đặc trưng có thể được biểu diễn dưới dạng được mô tả ở phương trình (4).

$$W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \quad (4)$$

trong đó $\beta < \alpha$, $0 < \beta < 1$, $0 < \alpha \leq 1$.

Theo phương án thực hiện khác, bộ lọc trọng số $W(z)$ có thể được suy ra từ bộ lọc LPC nhờ sử dụng mở rộng băng thông như được minh họa theo một phương án thực hiện ở phương trình (5) dưới đây.

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (5),$$

Ở phương trình (5), $\gamma_1 > \gamma_2$, vốn là các hệ số mà các cực di chuyển về gốc.

Do đó, đối với mỗi khung của giọng nói, các LPC và cao độ được tính toán và các bộ lọc được cập nhật. Đối với mỗi khung phụ của giọng nói,

vector mã tạo đầu ra lọc ‘tốt nhất’ được chọn để biểu diễn khung phụ. Giá trị lượng tử hóa tương ứng của độ khuếch đại phải được truyền đến bộ giải mã để giải mã thích hợp. Các LPC và các giá trị cao độ phải được lượng tử hóa và gửi mỗi khung để tái tạo các bộ lọc ở bộ giải mã. Do đó, chỉ mục kích thích được mã hóa, chỉ mục độ khuếch đại được lượng tử hóa, chỉ mục tham số dự báo dài hạn được lượng tử hóa, và chỉ mục tham số dự báo khoảng ngắn được lượng tử hóa được truyền đến bộ giải mã.

Fig.2 minh họa các hoạt động được thực hiện trong quá trình giải mã giọng nói gốc sử dụng bộ giải mã CELP.

Tín hiệu giọng nói được tái tạo ở bộ giải mã bằng cách đưa các vector mã nhận được qua các bộ lọc tương ứng. Kết quả là, mỗi khối ngoại trừ hậu xử lý có định nghĩa tương tự như được mô tả trong bộ mã hóa trên Fig.1.

Dòng bit CELP được mã hóa được nhận và dỡ gói 80 ở thiết bị nhận. Đối với mỗi khung phụ nhận được, chỉ mục kích thước được mã hóa nhận được, chỉ mục khuếch đại được lượng tử hóa, chỉ mục tham số dự báo dài hạn được lượng tử hóa, và chỉ mục tham số dự báo khoảng ngắn được lượng tử hóa, được sử dụng để tìm thấy các tham số tương ứng nhờ sử dụng các bộ giải mã tương ứng, chẳng hạn, bộ giải mã độ khuếch đại 81, bộ giải mã LTP 82, và bộ giải mã STP 83. Chẳng hạn, các vị trí các dấu hiệu biên độ của các xung kích thích và vector mã đại số của kích thích mã 402 có thể được xác định từ chỉ mục kích thích được mã hóa nhận được.

Như được thể hiện trên Fig.2, bộ giải mã là tổ hợp của vài khối gồm kích thích được mã hóa 201, LTP 203, STP 205. Bộ mã hóa ban đầu còn gồm khối hậu xử lý 207 sau khi giọng nói được tổng hợp 206. Việc hậu xử lý có thể còn gồm hậu xử lý ngắn hạn và hậu xử lý dài hạn.

Fig.3 minh họa bộ mã hóa CELP đã biết.

Fig.3 minh họa bộ mã hóa CELP cơ bản sử dụng từ điển mã thích ứng bổ sung để cải thiện dự báo tuyến tính dài hạn. Tạo kích thích bằng cách tính tổng các phần góp từ từ điển mã thích ứng 307 và kích thích mã 308, có thể là từ điển mã cố định hoặc ngẫu nhiên như nêu trên. Các mục trong từ điển mã thích ứng bao gồm các phiên bản trễ kích thích. Điều này có thể mã hóa hiệu quả các tín hiệu tuần hoàn như các âm có thanh âm.

Như được thể hiện trên Fig.3, từ điển mã thích ứng 307 bao gồm kích thích tổng hợp trước 304 hoặc chu kỳ cao độ kích thích trước lặp lại ở chu kỳ cao độ. Độ trễ cao độ có thể được mã hóa ở giá trị nguyên khi nó lớn hoặc dài. Độ trễ cao độ thường được mã hóa ở giá trị phân số chính xác hơn khi nó nhỏ hoặc ngắn. Thông tin chu kỳ của cao độ được sử dụng để tạo thành phần thích ứng của kích thích. Sau đó, thành phần kích thích được tạo tỷ lệ bằng độ khuếch đại G_p 305 (còn được gọi là độ khuếch đại cao độ).

LTP đóng vai trò quan trọng trong việc mã hóa giọng nói có thoại do giọng nói có thoại có tính tuần hoàn mạnh. Các chu kỳ cao độ liên kế của giọng nói có thoại giống nhau, nghĩa là về mặt toán học độ khuếch đại cao độ G_p trong biểu thức kích thích dưới đây cao hoặc gần bằng 1. Kích thích thu được có thể được biểu diễn dưới dạng ở phương trình (6) dưới dạng tổ hợp của các kích thích riêng rẽ.

$$e(n) = G_p \cdot e_p(n) + G_c \cdot e_c(n) \quad (6)$$

trong đó, $e_p(n)$ là một khung phụ của các chuỗi mẫu được đánh chỉ mục n , đến từ từ điển mã thích ứng 307 bao gồm kích thích trước 304 thông qua vòng lặp phản hồi (Fig.3). $e_p(n)$ có thể được lọc thông thấp một cách thích ứng khi khu vực tần số thấp thông thường tuần hoàn hơn hoặc điều hòa cao hơn khu vực tần số cao. $e_c(n)$ từ từ điển mã kích thích được mã hóa 308 (còn được gọi là từ điển mã cố định) vốn là phần góp kích thích dòng điện. Ngoài ra, $e_c(n)$ có thể còn được tăng cường như bằng cách sử

dụng tăng cường lọc thông cao, tăng cường cao độ, tăng cường phân tán, tăng cường foc-măng, và các tăng cường khác.

Đối với giọng nói có thoại, phần gộp $e_p(n)$ từ từ điển mã thích ứng 307 có thể là nổi bật và độ khuếch đại cao độ G_p 305 xấp xỉ giá trị 1. Kích thích thường xuyên được cập nhật cho mỗi khung phụ. Kích thước khung phụ đặc trưng bằng 20 miligiây và kích thước khung phụ đặc trưng bằng 5 miligiây.

Như được mô tả trên Fig.1, kích thích được mã hóa cố định 308 được tạo tỷ lệ bằng độ khuếch đại G_c 306 trước khi đi qua các bộ lọc tuyến tính. Hai thành phần kích thích được tạo tỷ lệ từ kích thích được mã hóa cố định 108 và từ điển mã thích ứng 307 được thêm vào cùng nhau trước khi lọc qua bộ lọc dự báo tuyến tính khoảng ngắn 303. Hai độ khuếch đại (G_p và G_c) được lượng tử hóa và được truyền đến bộ giải mã. Do đó, chỉ mục kích thích được mã hóa, chỉ mục từ điển mã thích ứng, các chỉ mục khuếch đại được lượng tử hóa, và chỉ mục tham số dự báo khoảng ngắn được lượng tử hóa được truyền đến thiết bị âm thanh nhận.

Dòng bit CELP được mã hóa nhờ sử dụng thiết bị được minh họa trên Fig.3 được nhận ở thiết bị nhận. Fig.4 minh họa bộ giải mã tương ứng của thiết bị nhận.

Fig.4 minh họa bộ giải mã CELP cơ bản tương ứng với bộ mã hóa trên Fig.3. Fig.4 gồm khối hậu xử lý 408 nhận giọng nói tổng hợp 407 từ bộ giải mã chính. Bộ giải mã này giống Fig.3 ngoại trừ từ điển mã thích ứng 307.

Đối với mỗi khung phụ nhận được, chỉ mục kích thích được mã hóa nhận được, chỉ mục khuếch đại kích thích được mã hóa lượng tử hóa, chỉ mục cao độ lượng tử hóa, chỉ mục khuếch đại từ điển mã thích ứng được lượng tử hóa, và chỉ mục tham số dự báo khoảng ngắn được lượng tử hóa, được sử dụng để tìm các tham số tương ứng bằng cách sử dụng các bộ giải mã tương ứng, chẳng hạn, bộ giải mã độ khuếch đại 81, bộ giải mã

cao độ 84, bộ giải mã độ khuếch đại từ điển mã thích ứng 85, và bộ giải mã STP 83.

Theo các phương án thực hiện khác nhau, bộ giải mã CELP là tổ hợp của vài khối và bao gồm kích thích được giải mã 402, từ điển mã thích ứng 401, STP 406, và hậu xử lý 408. Mỗi khối ngoại trừ hậu xử lý có định nghĩa tương tự như được mô tả trong bộ mã hóa trên Fig.3. Hậu xử lý có thể còn gồm hậu xử lý ngắn hạn và hậu xử lý dài hạn.

Khối kích thích mã (dựa vào nhãn 308 trên Fig.3 và 402 trên Fig.4) minh họa vị trí của từ điển mã cố định (FCB) để mã hóa CELP tổng quát. Vector mã được chọn từ FCB được tạo tỷ lệ bằng độ khuếch đại thường được lưu ý là G_c 306.

Fig.5 và Fig.6 minh họa các ví dụ của các tín hiệu giọng nói dạng lược đồ và quan hệ của nó với kích thước khung và kích thước khung phụ trong miền thời gian. Fig.5 và Fig.6 minh họa khung gồm các khung phụ.

Các mẫu của giọng nói đầu vào được phân chia thành các khối của mỗi mẫu, được gọi là các khung, chẳng hạn, từ 80 đến 240 mẫu hoặc khung. Mỗi khung được phân chia thành các khối nhỏ hơn của mỗi mẫu, được gọi là các khung phụ. Ở tốc độ lấy mẫu 8kHz, 12,8kHz, hoặc 16kHz, thuật toán mã hóa giọng nói là sao cho khoảng thời gian khung danh định trong khoảng từ mười đến ba mươi mili giây, và đặc trưng hai mươi mili giây. Trên Fig.5 được minh họa, khung có kích thước khung 1 và kích thước khung phụ 2, trong đó mỗi khung được phân chia thành 4 khung phụ.

Dựa vào các phần dưới hoặc đáy trên Fig.5 và Fig.6, các vùng có thanh âm ở giọng nói giống như tín hiệu gần như tuần hoàn trong biểu diễn miền thời gian. Việc đóng và mở tuần hoàn của các nếp gấp thanh âm của người nói dẫn đến cấu trúc điều hòa trong các tín hiệu giọng nói có thoại. Do vậy, trên chu kỳ thời gian ngắn, các đoạn giọng nói có thoại có thể được xem là tuần hoàn cho toàn việc phân tích và xử lý thực

tế. Tính tuần hoàn liên kết với các đoạn này được định nghĩa là “Chu kỳ cao độ” hoặc đơn giản là “cao độ” trong miền thời gian và “Tần số cao độ hoặc tần số cơ bản f_0 ” trong miền tần số. Đảo của chu kỳ cao độ là tần số cơ bản của giọng nói. Các thuật ngữ cao độ và tần số cơ bản của giọng nói thường được sử dụng qua lại.

Đối với hầu hết giọng nói có thoại, một khung chứa nhiều hơn hai chu kỳ cao độ. Fig.5 còn minh họa ví dụ rằng chu kỳ cao độ 3 nhỏ hơn kích thước khung phụ 2. Ngược lại, Fig.6 minh họa ví dụ trong đó chu kỳ cao độ 4 lớn hơn kích thước khung phụ 2 và nhỏ hơn kích thước nửa khung.

Để mã hóa hiệu quả hơn tín hiệu giọng nói, tín hiệu giọng nói có thể được phân loại thành các lớp khác nhau và mỗi lớp được mã hóa theo cách khác nhau. Chẳng hạn, ở một số chuẩn như G.718, VMR-WB, hoặc AMR-WB, tín hiệu giọng nói được phân loại thành UNVOICED, TRANSITION, GENERIC, VOICED, và NOISE.

Đối với mỗi lớp, bộ lọc LPC hoặc STP luôn được sử dụng để biểu diễn đường bao phổ. Tuy nhiên, kích thích cho bộ lọc LPC có thể khác. Các lớp UNVOICED và NOISE có thể được mã hóa bằng kích thích nhiễu và một số tăng cường kích thích. Lớp TRANSITION có thể được mã hóa bằng kích thích xung và một số tăng cường kích thích mà không sử dụng từ điển mã thích ứng hoặc LTP.

GENERIC có thể được mã hóa bằng cách tiếp cận CELP đã biết như CELP đại số được sử dụng trong G.729 hoặc AMR-WB, trong đó một khung 20ms chứa bốn khung phụ 5ms. Cả thành phần kích thích từ điển mã thích ứng lẫn thành phần kích thích từ điển mã cố định được tạo với một số tăng cường kích thích cho mỗi khung phụ. Các độ trễ cao độ cho từ điển mã thích ứng ở các khung phụ thứ nhất và thứ ba được mã hóa ở khoảng đủ từ giới hạn cao độ nhỏ nhất PIT_MIN đến giới hạn cao độ lớn nhất PIT_MAX . Các độ trễ cao độ cho từ điển mã thích ứng ở các khung

phụ thứ hai và thứ tư được mã hóa vi sai với độ trễ cao độ được mã hóa trước đó.

Các lớp VOICED có thể được mã hóa theo cách sao cho chúng hơi khác với lớp GENERIC. Chẳng hạn, độ trễ cao độ ở khung phụ thứ nhất có thể được mã hóa trong khoảng đủ từ giới hạn cao độ nhỏ nhất PIT_MIN đến giới hạn cao độ lớn nhất PIT_MAX . Các độ trễ cao độ ở các khung phụ còn lại có thể được mã hóa vi sai với độ trễ cao độ được mã hóa trước đó. Như minh họa, giả sử tốc độ lấy mẫu kích thước là 12,8kHz, sau đó giá trị PIT_MIN ví dụ có thể là 34 và PIT_MAX có thể là 231.

Các phương án thực hiện sáng chế để cải thiện phân loại mã hóa miền thời gian và mã hóa miền tần số sẽ được mô tả.

Nói chung, tốt hơn sử dụng mã hóa miền thời gian cho tín hiệu giọng nói và mã hóa miền tần số cho tín hiệu âm nhạc để đạt chất lượng tốt nhất ở tốc độ bit khá cao (chẳng hạn, $24\text{kbps} \leq \text{tốc độ bit} \leq 64\text{kbps}$). Tuy nhiên, đối với một số tín hiệu giọng nói cụ thể như tín hiệu cao độ ngắn, tín hiệu giọng hát, hoặc tín hiệu giọng nói rất âm ỉ, có thể tốt hơn là sử dụng mã hóa miền tần số. Đối với một số tín hiệu âm nhạc cụ thể như tín hiệu rất tuần hoàn, tốt hơn có thể sử dụng mã hóa miền thời gian bằng cách hưởng lợi từ độ khuếch đại LTP rất cao. Tốc độ bit là tham số quan trọng để phân loại. Thông thường, mã hóa miền thời gian ưu tiên tốc độ bit thấp và mã hóa miền tần số ưu tiên tốc độ bit cao. Phân loại hoặc lựa chọn tốt nhất giữa mã hóa miền thời gian và mã hóa miền tần số cần được quyết định cẩn thận, còn xem xét khoảng tốc độ bit và đặc tính của các thuật toán mã hóa.

Ở các phần tiếp theo, việc dò giọng nói thường và tín hiệu cao độ ngắn sẽ được mô tả.

Giọng nói thường là tín hiệu giọng nói loại bỏ tín hiệu giọng hát, tín hiệu giọng nói cao độ ngắn, hoặc tín hiệu kết hợp thoại/âm nhạc. Giọng

nói thường có thể còn thay đổi nhanh tín hiệu giọng nói, phổ và/hoặc năng lượng của nó thay đổi nhanh hơn phần lớn các tín hiệu âm nhạc. Thông thường, thuật toán mã hóa miền thời gian tốt hơn thuật toán mã hóa miền tần số để mã hóa tín hiệu giọng nói thông thường. Phần sau là thuật toán lấy làm ví dụ để dò tín hiệu giọng nói thông thường.

Đối với ứng viên cao độ P , mỗi tương quan cao độ chuẩn hóa thường được định nghĩa ở dạng toán học như ở phương trình (8).

$$R(P) = \frac{\sum_n s_w(n) \cdot s_w(n-P)}{\sqrt{\sum_n \|s_w(n)\|^2 \cdot \sum_n \|s_w(n-P)\|^2}} \quad (8)$$

Ở phương trình (8), $s_w(n)$ là tín hiệu giọng nói có trọng số, tử số là tương quan, và mẫu số là hệ số chuẩn hóa năng lượng. Giả sử *Voicing* ký hiệu giá trị mỗi tương quan cao độ chuẩn hóa trung bình của bốn khung phụ trong khung thoại hiện tại, *Voicing* có thể được tính toán như ở phương trình (9) dưới đây.

$$Voicing = [R_1(P_1) + R_2(P_2) + R_3(P_3) + R_4(P_4)] / 4 \quad (9)$$

$R_1(P_1)$, $R_2(P_2)$, $R_3(P_3)$, và $R_4(P_4)$ là bốn mối tương quan cao độ chuẩn hóa được tính toán cho mỗi khung phụ; P_1 , P_2 , P_3 , và P_4 đối với mỗi khung phụ là các ứng viên cao độ tốt nhất được tìm thấy trong quãng cao độ từ $P=PIT_MIN$ đến $P=PIT_MAX$. Mỗi tương quan cao độ được làm tròn từ khung trước đến khung hiện tại có thể được tính toán như ở phương trình (10).

$$\begin{aligned} & \text{if } (Voicing > Voicing_sm) \text{ and } (speech_class \neq UNVOICED) \\ & \quad Voicing_sm \Leftarrow (3 \cdot Voicing_sm + Voicing) / 4 \\ & \text{else if } (VAD=1) \\ & \quad Voicing_sm \Leftarrow (31 \cdot Voicing_sm + Voicing) / 32 \end{aligned}$$

(10)

Ở phương trình (10), VAD là dò hoạt động thoại (Voice Activity Detection) và VAD=1 đề cập việc tín hiệu giọng nói thoát ra. Giả sử F_s là tốc độ lấy mẫu, năng lượng lớn nhất ở vùng tần số rất thấp $[0, F_{MIN}=F_s$

$/PIT_MIN]$ (Hz) là $Energy0$ (dB), năng lượng lớn nhất ở vùng tần số thấp $[F_{MIN}, 900]$ (Hz) là $Energy1$ (dB), và năng lượng lớn nhất ở vùng tần số cao $[5000, 5800]$ (Hz) là $Energy3$ (dB), tham số độ nghiêng phổ $Tilt$ được định nghĩa như sau.

$$Tilt = energy3 - \max\{energy0, energy1\} \quad (11)$$

Tham số độ nghiêng phổ làm tròn được lưu ý như ở phương trình (12).

$$Tilt_sm \leftarrow (7 \cdot Tilt_sm + Tilt)/8 \quad (12)$$

Độ nghiêng phổ chênh lệch của khung hiện tại và khung trước có thể được nêu như ở phương trình (13).

$$Diff_tilt = | tilt - old_tilt | \quad (13)$$

Độ nghiêng phổ chênh lệch được làm tròn được nêu như ở phương trình (14).

$$\begin{aligned} & \text{if } ((Diff_tilt > Diff_tilt_sm) \text{ and } (speech_class \neq UNVOICED)) \\ & \quad Diff_tilt_sm \leftarrow (3 \cdot Diff_tilt_sm + Diff_tilt)/4 \\ & \text{else if } (VAD=1) \\ & \quad Diff_tilt_sm \leftarrow (31 \cdot Diff_tilt_sm + Diff_tilt)/32 \end{aligned} \quad (14)$$

Năng lượng tần số thấp chênh lệch của khung hiện tại và khung trước đó là

$$Diff_energy1 = | energy1 - old_energy1 | \quad (15)$$

Năng lượng chênh lệch được làm tròn được nêu nhờ phương trình (16).

$$\begin{aligned} & \text{if } ((Diff_energy1 > Diff_energy1_sm) \text{ and } (speech_class \neq UNVOICED)) \\ & \quad Diff_energy1_sm \leftarrow (3 \cdot Diff_energy1_sm + Diff_energy1)/4 \\ & \text{else if } (VAD = 1) \\ & \quad Diff_energy1_sm \leftarrow (31 \cdot Diff_energy1_sm + Diff_energy1)/32 \end{aligned} \quad (16)$$

Ngoài ra, cờ giọng nói thường được chú thích như là $Speech_flag$ được quyết định và thay đổi trong suốt các khu vực có thoại nhờ xem xét biến thiên năng lượng $Diff_energy1_sm$, biến thiên thoại $Voicing_sm$, và biến thiên độ nghiêng phổ $Diff_tilt_sm$ như được nêu ở phương trình (17).

```

if (speech_class  $\neq$  UNVOICED ) {
    Diff_Sp = Diff_energy1_sm · Voicing_sm · Diff_tilt_sm
    if (Diff_Sp > 800) Speech_flag = 1 // switch to normal speech
    if (Diff_Sp < 100) Speech_flag = 0 // switch to non normal speech
}
(17)

```

Các phương án thực hiện sáng chế để dò tín hiệu cao độ ngắn sẽ được mô tả.

Phần lớn các codec CELP vận hành tốt cho các tín hiệu giọng nói thông thường. Tuy nhiên, các codec CELP tốc độ bit thấp thường thất bại đối với các tín hiệu âm nhạc và/hoặc các tín hiệu giọng hát. Nếu khoảng mã hóa cao độ từ PIT_MIN đến PIT_MAX và độ trễ cao độ thực nhỏ hơn PIT_MIN , hiệu năng mã hóa CELP có thể xấu về mặt trực giác do cao độ gấp đôi hoặc cao độ gấp ba. Chẳng hạn, quãng cao độ từ $PIT_MIN=34$ đến $PIT_MAX=231$ cho tần số lấy mẫu $F_s=12,8$ kHz thích ứng hầu hết các thoại của con người. Tuy nhiên, độ trễ cao độ thực của âm nhạc thông thường hoặc tín hiệu giọng hát có thể ngắn hơn nhiều giới hạn nhỏ nhất $PIT_MIN=34$ được định nghĩa trong thuật toán CELP ví dụ nêu trên.

Khi độ trễ cao độ thực là P , tần số cơ bản được chuẩn hóa tương ứng (hoặc điều hòa thứ nhất) là $f_0=F_s/P$, trong đó F_s là tần số lấy mẫu và f_0 là vị trí của đỉnh điều hòa thứ nhất trong phổ. Do vậy, đối với tần số lấy mẫu cho trước, giới hạn cao độ nhỏ nhất PIT_MIN thực sự định nghĩa giới hạn tần số điều hòa cơ bản lớn nhất $F_M=F_s/PIT_MIN$ cho thuật toán CELP.

Fig.7 minh họa ví dụ của phổ băng rộng có thanh âm gốc. Fig.8 minh họa phổ băng rộng có thanh âm được mã hóa của phổ băng rộng có thanh âm gốc được minh họa trên Fig.7 bằng cách sử dụng gấp đôi mã hóa độ trễ cao độ. Nói cách khác, Fig.7 minh họa phổ trước khi mã hóa và Fig.8 minh họa phổ sau khi mã hóa.

Ở ví dụ được thể hiện trên Fig.7, phổ được tạo bởi các đỉnh điều hòa 701 và đường bao phổ 702. Tần số điều hòa cơ bản thực (vị trí của đỉnh điều hòa thứ nhất) đã vượt qua giới hạn tần số điều hòa cơ bản lớn nhất F_M sao cho độ trễ cao độ được truyền cho thuật toán CELP không thể bằng độ trễ cao độ thực và có thể gấp đôi hoặc nhiều lần độ trễ cao độ thực.

Độ trễ cao độ sai được truyền với nhiều độ trễ cao độ thực có thể gây thoái hóa chất lượng rõ ràng. Theo cách khác, khi độ trễ cao độ thực cho tín hiệu âm nhạc điều hòa hoặc tín hiệu giọng hát nhỏ hơn giới hạn độ trễ nhỏ nhất PIT_MIN được định nghĩa trong thuật toán CELP, độ trễ được truyền có thể gấp đôi, gấp ba hoặc nhiều lần độ trễ cao độ thực.

Kết quả là, phổ của tín hiệu được mã hóa với độ trễ cao độ được truyền có thể như được thể hiện trên Fig.8. Như được minh họa trên Fig.8, bên cạnh gồm các đỉnh điều hòa 801 và đường bao phổ 802, các đỉnh nhỏ không mong muốn 803 giữa các đỉnh điều hòa thực có thể được thấy trong khi phổ đúng giống như phổ trên Fig.7. Các đỉnh phổ nhỏ kia trên Fig.8 có thể gây biến dạng trực giác không dễ chịu.

Theo các phương án thực hiện sáng chế, một giải pháp để giải quyết vấn đề này khi CELP thất bại đối với một số tín hiệu cụ thể chính là việc mã hóa miền tần số được sử dụng thay cho mã hóa miền thời gian.

Thông thường, các tín hiệu điều hòa âm nhạc hoặc các tín hiệu giọng hát cố định hơn các tín hiệu giọng nói thông thường. Độ trễ cao độ (hoặc tần số cơ bản) của tín hiệu giọng nói thông thường luôn thay đổi. Tuy nhiên, độ trễ cao độ (hoặc tần số cơ bản) của tín hiệu âm nhạc hoặc tín hiệu giọng hát thường thay đổi tương đối chậm trong khoảng thời gian tương đối dài. Khoảng cao độ rất ngắn được định nghĩa từ PIT_MIN0 đến PIT_MIN . Ở tần số lấy mẫu $F_s=12,8\text{kHz}$, định nghĩa ví dụ của khoảng cao độ rất ngắn có thể từ $PIT_MIN0 \leq 17$ đến $PIT_MIN=34$. Do ứng viên cao độ rất ngắn, năng lượng từ 0Hz đến $F_{MIN}=F_s / PIT_MIN\text{Hz}$ phải

tương đối đủ thấp. Các điều kiện khác như VAD và phân loại có thoại có thể được thêm vào trong quá trình dò sự tồn tại tín hiệu cao độ ngắn.

Hai tham số sau có thể giúp dò sự tồn tại khả thi của tín hiệu cao độ rất ngắn. Một tham số bao gồm “*Thiếu năng lượng tần số thấp*” và một tham số khác bao gồm “*Độ rõ phổ*”. Như đã nêu trên, giả sử năng lượng lớn nhất trong vùng tần số $[0, F_{MIN}]$ (Hz) là $Energy0$ (dB), năng lượng lớn nhất trong vùng tần số $[F_{MIN}, 900]$ (Hz) là $Energy1$ (dB), tỷ lệ năng lượng tương đối giữa $Energy0$ và $Energy1$ được nêu ở phương trình (18) dưới đây.

$$Ratio = Energy1 - Energy0 \quad (18)$$

Tỷ lệ năng lượng này có thể được đánh trọng số bằng cách nhân giá trị mới tương quan cao độ chuẩn hóa trung bình *Voicing*, được thể hiện dưới đây ở phương trình (19).

$$Ratio \leftarrow Ratio \cdot \max\{Voicing, 0.5\} \quad (19)$$

Nguyên nhân đánh trọng số ở phương trình (19) bằng cách sử dụng hệ số *Voicing* chính là việc dò cao độ ngắn có nghĩa đối với giọng nói có thoại hoặc âm nhạc điều hòa, và không có nghĩa đối với giọng nói không có thoại hoặc âm nhạc không điều hòa. Trước khi sử dụng tham số *Ratio* để dò việc thiếu năng lượng tần số thấp, tốt hơn là được làm trơn để giảm sự không chắc chắn như ở phương trình (20).

$$\begin{aligned} & \text{if } (VAD = 1) \{ \\ & \quad LF_EnergyRatio_sm \leftarrow (15 \cdot LF_EnergyRatio_sm + Ratio) / 16 \quad (20) \\ & \} \end{aligned}$$

Nếu $LF_lack_flag=1$ nghĩa là thiếu năng lượng tần số thấp được dò thấy (ngược lại $LF_lack_flag=0$), LF_lack_flag có thể được xác định bởi thủ tục sau.

$$\begin{aligned} & \text{if } ((LF_EnergyRatio_sm > 30) \text{ or } (Ratio > 48) \text{ or} \\ & \quad (LF_EnergyRatio_sm > 22 \text{ and } Ratio > 38)) \{ \\ & \quad LF_lack_flag = 1 ; \end{aligned}$$

```

}
else if (LF_EnergyRatio_sm < 13) {
    LF_lack_flag = 0;
}
else {
    LF_lack_flag keeps unchanged.
}

```

Các tham số liên quan *Độ rõ nét phổ* được xác định theo cách sau. Giải sử $Energy1$ (dB) là năng lượng lớn nhất trong vùng tần số thấp $[F_{MIN}, 900]$ (Hz), i_{peak} là vị trí đỉnh điều hòa năng lượng lớn nhất trong vùng tần số $[F_{MIN}, 900]$ (Hz) và $Energy2$ (dB) là năng lượng trung bình trong vùng tần số $[i_{peak}, i_{peak} + 400]$ (Hz). Một tham số độ rõ nét phổ được định nghĩa như ở phương trình (21).

$$SpecSharp = \max\{Energy1 - Energy2, 0\} \quad (21)$$

Tham số độ rõ nét phổ được làm tròn được tạo như sau.

```

if (VAD = 1) {
    SpecSharp_sm = (7 * SpecSharp_sm + SpecSharp) / 8
}

```

Một cờ độ rõ nét phổ biểu thị sự tồn tại khả thi của tín hiệu cao độ ngắn được đánh giá nhờ phương trình sau.

```

if ( SpecSharp_sm > 50 or SpecSharp > 80 ) {
    SpecSharp_flag = 1; //possible short pitch or tones
}
if ( SpecSharp_sm < 8 ) {
    SpecSharp_flag = 0;
}

```

if không điều kiện nào ở trên thỏa mãn, $SpecSharp_flag$ không đổi.

Theo các phương án thực hiện khác nhau, các tham số ước tính trên có thể được sử dụng để cải thiện phân loại hoặc lựa chọn mã hóa miền thời gian và mã hóa miền tần số. Giả sử $Sp_Aud_Deci=1$ ký hiệu rằng mã hóa miền tần số được lựa chọn và $Sp_Aud_Deci=0$ ký hiệu rằng mã hóa miền thời gian được lựa chọn. Thủ tục dưới đây lấy thuật toán lấy làm ví dụ để cải thiện phân loại mã hóa miền thời gian và mã hóa miền tần số cho các tốc độ bit mã hóa khác nhau.

Các phương án thực hiện sáng chế có thể được sử dụng để cải thiện các tốc độ bit cao, chẳng hạn, tốc độ bit mã hóa lớn hơn hoặc bằng 46200bps. Khi tốc độ bit mã hóa rất cao và tín hiệu cao độ ngắn có thể tồn tại, mã hóa miền tần số được lựa chọn do mã hóa miền tần số có thể tạo chất lượng đáng tin cậy và chắc chắn trong khi mã hóa miền thời gian có nguy cơ gây ảnh hưởng xấu từ việc dò cao độ sai. Ngược lại, khi tín hiệu cao độ ngắn không tồn tại và tín hiệu là giọng nói không có thoại hoặc giọng nói thông thường, mã hóa miền thời gian được lựa chọn do mã hóa miền thời gian có thể cho chất lượng tốt hơn mã hóa miền tần số cho tín hiệu giọng nói thông thường.

/ đối với tín hiệu cao độ ngắn khả thi, lựa chọn mã hóa miền tần số */*

if (LF_lack_flag=1 or SpecSharp_flag=1) {

Sp_Aud_Deci = 1; // lựa chọn mã hóa miền tần số

}

/ đối với giọng nói không có thoại hoặc giọng nói thông thường, lựa chọn mã hóa miền thời gian */*

if (LF_lack_flag=0 and SpecSharp_flag=0) {

if ((Tilt>40) and (Voicing<0.5) and (speech_class=UNVOICED) and (VAD=1)) {

Sp_Aud_Deci = 0; // lựa chọn mã hóa miền thời gian

```

    }
    if (Speech_flag=1) {
        Sp_Aud_Deci = 0; // lựa chọn mã hóa miền thời gian
    }
}

```

Các phương án thực hiện sáng chế có thể được sử dụng để cải thiện mã hóa tốc độ bit trung gian, chẳng hạn, khi tốc độ bit mã hóa giữa 24,4kbps và 46200bps. Khi tín hiệu cao độ ngắn có thể tồn tại và tính chu kỳ thanh âm là thấp, mã hóa miền tần số được lựa chọn do mã hóa miền tần số có thể cho chất lượng đáng tin cậy và chắc chắn trong khi mã hóa miền thời gian có nguy cơ gây ảnh hưởng xấu từ tính chu kỳ thanh âm thấp. Khi tín hiệu cao độ ngắn không tồn tại và tín hiệu là giọng nói không có thoại hoặc giọng nói thông thường, mã hóa miền thời gian được lựa chọn do mã hóa miền thời gian có thể cho chất lượng tốt hơn mã hóa miền tần số đối với tín hiệu giọng nói thông thường. Khi tính chu kỳ thanh âm rất mạnh, mã hóa miền thời gian được lựa chọn do mã hóa miền thời gian có thể hưởng lợi nhiều từ độ khuếch đại LTP cao với tính chu kỳ thanh âm rất mạnh.

Các phương án thực hiện sáng chế có thể còn được sử dụng để cải thiện các tốc độ bit cao, chẳng hạn, tốc độ bit mã hóa nhỏ hơn 24,4kbps. Khi có tín hiệu cao độ ngắn và tính chu kỳ thanh âm không thấp với việc dò thấy trễ cao độ ngắn thấp, mã hóa miền tần số không được lựa chọn do mã hóa miền tần số không thể cho chất lượng đáng tin cậy và chắc chắn ở tốc độ thấp trong khi mã hóa miền thời gian có thể hưởng lợi tốt từ chức năng LTP.

Thuật toán sau minh họa phương án cụ thể trong các phương án thực hiện nêu trên như minh họa. Tất cả các tham số có thể được tính toán như mô tả trước đó theo một hoặc nhiều các phương án thực hiện.

```

/* chuẩn bị các tham số hoặc các ngưỡng */
if (khung trước là mã hóa miền thời gian) {
    DPIT=0.4;
    TH1=0.92;
    TH2=0.8;
}
else {
    DPIT=0.9;
    TH1=0.9;
    TH2=0.7;
}

Stab_Pitch_Flag = ( $|P_0 - P_1| < DPIT$ ) and ( $|P_1 - P_2| < DPIT$ ) and ( $|P_2 - P_3| < DPIT$ );
High_Voicing = (Voicing_sm > TH1) and (Voicing > TH2);

/* đối với tín hiệu cao độ ngắn khả thi với tính tuần hoàn thấp (thanh
âm thấp), lựa chọn mã hóa miền tần số */
if ( (LF_lack_flag=1) or (SpecSharp_flag=1) ) {

    if ( ( (Stab_Pitch_Flag=0 or High_Voicing=0) and (
Tilt_sm <=-50) )
        or (Tilt_sm <=-60) )
    {
        Sp_Aud_Deci = 1; // lựa chọn mã hóa miền tần số
    }
}

/* đối với tín hiệu không có thanh âm hoặc tín hiệu giọng nói thông
thường, lựa chọn mã hóa miền thời gian */
if ( LF_lack_flag=0 and SpecSharp_flag=0 )
{

```

```

    if ( Tilt>40 and Voicing<0.5 and speech_class=UNVOICED
and Vad=1)
    {
        Sp_Aud_Deci = 0; // lựa chọn mã hóa miền thời gian
    }
    if ( Speech_flag=1)
    {
        Sp_Aud_Deci = 0; // lựa chọn mã hóa miền thời gian
    }
}
/* đối với tín hiệu có thanh âm mạnh, lựa chọn mã hóa miền thời gian
*/
if ( Ttilt_sm>-60 and ( speech_class is not UNVOICED ) )
{
    if ( High_Voicing=1 and
        (Stab_Pitch_Flag=1 or (LF_lack_flag=0 and
SpecSharp_flag=0) ) )
    {
        Sp_Aud_Deci = 0; // lựa chọn mã hóa miền thời gian
    }
}

```

Theo các phương án thực hiện khác nhau, việc phân loại hoặc lựa chọn mã hóa miền thời gian và mã hóa miền tần số có thể được sử dụng để cải thiện đáng kể chất lượng trực giác của một số tín hiệu giọng nói cụ thể hoặc tín hiệu âm nhạc.

Mã hóa âm thanh dựa trên công nghệ dàn lọc được sử dụng rộng rãi trong việc mã hóa miền tần số. Khi xử lý tín hiệu, dàn lọc là mảng các bộ lọc thông dải tách riêng tín hiệu đầu vào thành nhiều thành phần, mỗi thành phần mang một băng phụ tần số của tín hiệu đầu vào ban đầu. Quá

trình phân rã được thực hiện bởi dàn lọc được gọi là phân tích, và đầu ra phân tích dàn lọc được gọi là tín hiệu băng phụ có nhiều băng phụ như có các bộ lọc trong dàn lọc. Quá trình tái tạo được gọi là tổng hợp dàn lọc. Khi xử lý tín hiệu số, thuật ngữ dàn lọc thường được áp dụng cho dàn bộ nhận, có thể còn biến đổi xuống các băng phụ thành tần số trung tâm thấp có thể được lấy mẫu lại ở tốc độ được giảm. Kết quả tổng hợp tương tự đôi khi có thể còn đạt được nhờ lấy mẫu dưới các băng phụ thông dải. Đầu ra của phân tích dàn lọc có thể ở dạng các hệ số phức tạp. Mỗi hệ số phức có phần thực và phần ảo lần lượt biểu diễn số hạng cosin và số hạng sin cho mỗi băng phụ của dàn lọc.

Phân tích dàn lọc và tổng hợp dàn lọc là một loại cặp biến đổi biến đổi tín hiệu miền thời gian thành các hệ số miền tần số và biến đổi ngược các hệ số miền tần số thành tín hiệu miền thời gian. Các cặp biến đổi phổ biến khác, như (FFT và $iFFT$), (DFT và $iDFT$), và ($MDCT$ và $iMDCT$), có thể còn được sử dụng trong việc mã hóa âm thanh/giọng nói.

Khi ứng dụng các dàn lọc để nén tín hiệu, một số tần số quan trọng về mặt trực giác hơn các tần số khác. Sau khi phân rã, các tần số quan trọng về mặt trực giác có thể được mã hóa với độ phân giải tốt, do các độ chênh lệch nhỏ ở các tần số này có thể nhận thấy được về mặt trực giác để đảm bảo việc sử dụng phương tiện mã hóa bảo tồn các khác biệt này. Mặt khác, các tần số ít quan trọng về mặt trực giác không được sao chép chính xác. Do vậy, phương tiện mã hóa thô hơn có thể được sử dụng, mặc dù một số chi tiết tinh chỉnh hơn sẽ bị mất trong quá trình mã hóa. Phương tiện mã hóa thô hơn điển hình có thể dựa trên khái niệm mở rộng băng thông (Bandwidth Extension, BWE), còn được biết là mở rộng băng cao (High Band Extension, HBE). Một cách tiếp cận BWE hoặc HBE cụ thể phổ biến hơn được biết đến là bản sao băng phụ (Sub Band Replica, SBR) hoặc sao chép băng phổ (Spectral Band Replication, SBR). Các kỹ thuật này tương tự trong việc chúng mã hóa hoặc giải mã một số băng

phụ tần số (thông thường các băng cao) không tốn chi phí tốc độ bit, nhờ đó cho tốc độ bit thấp hơn đáng kể so với cách mã hóa/giải mã thông thường. Với công nghệ SBR, cấu trúc tinh chỉnh phổ trong băng tần số cao được sao chép từ băng tần số thấp, và nhiều ngẫu nhiên có thể được thêm vào. Tiếp theo, đường bao phổ của băng tần số cao được tạo dạng bằng cách sử dụng thông tin phụ được truyền từ bộ mã hóa đến bộ giải mã.

Việc sử dụng nguyên lý tâm thính học hoặc hiệu ứng làm mất nạ trực giác để thiết kế nén âm thanh hợp lý. Thiết bị âm thanh/giọng nói được nhằm tương tác với con người, với các khả năng và giới hạn nhận thức. Thiết bị âm thanh truyền thống cần tái tạo các tín hiệu với độ trung thực cao nhất so với bản gốc. Mục đích hiệu quả hơn và có định hướng thích hợp hơn là đạt độ trung thực mà người nhận thức được. Đây là mục đích của các bộ mã hóa trực giác.

Mặc dù một mục đích chính của các bộ mã trực giác âm thanh số là giảm dữ liệu, mã hóa trực giác có thể còn được sử dụng để cải thiện biểu diễn âm thanh số thông qua phân phối bit cải tiến. Một trong các ví dụ của các bộ mã hóa trực giác có thể là các hệ thống đa băng, phân chia lên phổ theo kiểu bắt chước các băng tới hạn âm thính học. Bằng cách tạo mô hình nhận thức con người, các bộ mã hóa trực giác có thể xử lý các tín hiệu theo cách người làm, và tận dụng các hiện tượng như làm mất nạ. Trong khi đây là mục đích của chúng, quá trình phụ thuộc vào thuật toán chính xác. Do thực tế là khó có mô hình trực giác chính xác bao gồm hành vi nghe của con người thông thường, độ chính xác biểu thức toán học bất kỳ của mô hình trực giác vẫn bị giới hạn. Tuy nhiên, với độ chính xác giới hạn, khái niệm nhận thức đã giúp thiết kế các codec âm thanh. Nhiều phương tiện mã hóa âm thanh MPEC đã hưởng lợi từ việc khai thác hiệu ứng mất nạ trực giác. Một số codec chuẩn ITU còn sử dụng khái niệm trực giác. Chẳng hạn, ITU G.729.1 thực hiện phân phối bit động

dựa trên khái niệm mặt nạ trực giác. Khái niệm phân phối bit động dựa trên tầm quan trọng trực giác còn được sử dụng trong codec 3GPP EVS gần đây.

Fig.9A và Fig.9B minh họa sơ đồ của codec trực giác miền tần số thông thường. Fig.9A minh họa bộ mã hóa miền tần số trong khi Fig.9B minh họa bộ giải mã miền tần số.

Trước hết tín hiệu gốc 901 được biến đổi thành miền tần số để có các hệ số miền tần số không được lượng tử hóa 902. Trước khi lượng tử hóa các hệ số, chức năng mặt nạ (độ quan trọng trực giác) phân chia phổ tần số thành nhiều băng phụ (thường được tạo khoảng cách đều nhau để đơn giản). Mỗi băng phụ phân phối động số lượng bit cần thiết trong khi duy trì tổng số bit được phân phối đều cho tất cả các băng phụ không vượt quá giới hạn trên. Một số băng phụ có thể được phân phối 0 bit nếu được đánh giá dưới ngưỡng mặt nạ. Khi thực hiện xác định cái nào có thể được bỏ qua, phần dư được phân phối số lượng bit khả dụng. Do các bit không bị lãng phí trên phổ mặt nạ, chúng có thể được phân phối theo số lượng lớn hơn cho phần còn lại của tín hiệu.

Theo các bit được phân phối, các hệ số được lượng tử hóa và dòng bit 703 được gửi đến bộ giải mã. Mặc dù khái niệm mặt nạ trực giác đã giúp nhiều trong quá trình thiết kế codec, song vẫn không hoàn hảo do các nguyên nhân khác nhau và các giới hạn.

Dựa vào Fig.9B, hậu xử lý phía bộ giải mã có thể còn cải thiện chất lượng trực giác của tín hiệu được giải mã được tạo với các tốc độ bit bị giới hạn. Trước hết, bộ giải mã sử dụng các bit nhận được 904 để tái tạo các hệ số được lượng tử hóa 905. Sau đó, được hậu xử lý bởi môđun được thiết kế thích hợp 906 để có các hệ số tăng cường 907. Biến đổi đảo được thực hiện trên các hệ số tăng cường để có đầu ra miền thời gian cuối cùng 908.

Fig.10 minh họa sơ đồ của các hoạt động ở bộ mã hóa trước khi mã hóa tín hiệu giọng nói bao gồm dữ liệu âm thanh theo các phương án thực hiện sáng chế.

Dựa vào Fig.10, phương pháp bao gồm lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian (hộp 1000) dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và độ trễ cao độ của tín hiệu số.

Lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước xác định liệu tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ (hộp 1010). Ngoài ra, xác định liệu tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên (hộp 1020). Nếu tín hiệu số bao gồm tín hiệu cao độ ngắn và tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên, mã hóa miền tần số được lựa chọn để mã hóa tín hiệu số.

Ngược lại, xác định liệu tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới (hộp 1030). Nếu tín hiệu số bao gồm tín hiệu cao độ ngắn và tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới, mã hóa miền thời gian được lựa chọn để mã hóa tín hiệu số.

Ngược lại, xác định liệu tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên (hộp 1040). Tính chu kỳ thanh âm được xác định tiếp theo (hộp 1050). Nếu tín hiệu số bao gồm tín hiệu cao độ ngắn và tốc độ bit mã hóa là trung gian và tính chu kỳ thanh âm là thấp, mã hóa miền tần số được lựa chọn để mã hóa tín hiệu số. Theo cách khác, nếu tín hiệu số bao gồm tín hiệu cao độ ngắn và tốc độ bit mã hóa là trung gian và tính chu kỳ thanh âm rất mạnh, mã hóa miền thời gian được lựa chọn để mã hóa tín hiệu số.

Theo cách khác, dựa vào hộp 1010, tín hiệu số không bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ. Xác định liệu tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường (hộp 1070). Nếu tín hiệu số không bao gồm

tín hiệu cao độ ngắn và nếu tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường, mã hóa miền thời gian được lựa chọn để mã hóa tín hiệu số.

Do đó, theo các phương án thực hiện khác nhau, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh gồm lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số. Tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ. Theo các phương án thực hiện khác nhau, phương pháp lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên, và lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới. Tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên khi tốc độ bit mã hóa lớn hơn hoặc bằng 46200bps. Tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới khi tốc độ bit mã hóa nhỏ hơn 24,4kbps.

Tương tự, theo phương án thực hiện khác, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh bao gồm lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên. Theo cách khác, phương pháp lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới. Tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ. Tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên khi tốc độ bit mã hóa lớn hơn hoặc bằng 46200bps. Tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới khi tốc độ bit mã hóa nhỏ hơn 24,4kbps.

Tương tự, theo phương án thực hiện khác, phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa bao gồm lựa chọn mã hóa miền thời gian

để mã hóa tín hiệu số bao gồm dữ liệu âm thanh khi tín hiệu số không bao gồm tín hiệu cao độ ngắn và tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường. Phương pháp còn bao gồm lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên. Tín hiệu số bao gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm là thấp. Phương pháp còn gồm lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian và tín hiệu số bao gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm rất mạnh. Giới hạn tốc độ bit dưới bằng 24,4kbps và giới hạn tốc độ bit trên bằng 46,2kbps.

Fig.11 minh họa hệ thống truyền thông 10 theo phương án thực hiện sáng chế.

Hệ thống truyền thông 10 có các thiết bị truy nhập âm thanh 7 và 8 được ghép nối với mạng 36 qua các liên kết truyền thông 38 và 40. Theo một phương án thực hiện, thiết bị truy nhập âm thanh 7 và 8 là các thiết bị VOIP (voice over internet protocol, giao thức thoại trên giao internet) và mạng 36 là WAN (wide area network, mạng diện rộng), PTSN (public switched telephone network, mạng điện thoại chuyển mạch công cộng) và/hoặc mạng internet. Theo phương án thực hiện khác, các liên kết truyền thông 38 và 40 là có dây và/hoặc các kết nối băng rộng không dây. Theo phương án thực hiện khác, các thiết bị truy nhập âm thanh 7 và 8 là các điện thoại di động hoặc tế bào, các liên kết 38 và 40 là các kênh điện thoại di động không dây và mạng 36 biểu thị mạng điện thoại di động.

Thiết bị truy nhập âm thanh 7 sử dụng micrô 12 để biến đổi âm thanh, như âm nhạc hoặc thoại của người thành tín hiệu đầu vào âm thanh tương tự 28. Giao diện micrô 16 biến đổi tín hiệu đầu vào âm thanh tương tự 28 thành tín hiệu âm thanh số 33 cho đầu vào bộ giải mã 22 của CODEC 20. Bộ mã hóa 22 tạo tín hiệu âm thanh được mã hóa TX để truyền đến mạng 26 qua giao diện mạng 26 theo các phương án thực hiện sáng chế. Bộ giải

mã 24 trong CODEC 20 nhận tín hiệu âm thanh được mã hóa RX từ mạng 36 qua giao diện mạng 26, và biến đổi tín hiệu âm thanh được mã hóa RX thành tín hiệu âm thanh số 34. Giao diện loa 18 biến đổi tín hiệu âm thanh số 34 thành tín hiệu âm thanh 30 thích hợp để điều khiển loa 14.

Theo các phương án thực hiện sáng chế, trong đó thiết bị truy nhập âm thanh 7 là thiết bị VOIP, một số hoặc tất cả các thành phần trong thiết bị truy nhập âm thanh 7 được triển khai trong máy thu phát cầm tay. Theo một số phương án thực hiện, tuy nhiên, micrô 12 và loa 14 là các khối riêng rẽ, và giao diện micrô 16, giao diện loa 18, CODEC 20 và giao diện mạng 26 được triển khai trong máy tính cá nhân. CODEC 20 có thể được triển khai khi chạy phần mềm trên máy tính hoặc bộ xử lý riêng, hoặc bởi phần cứng dành riêng, chẳng hạn, trên mạch tích hợp ứng dụng cụ thể (application specific integrated circuit, ASIC). Giao diện micrô 16 được triển khai bởi bộ biến đổi tương tự sang số (analog-to-digital, A/D), cũng như hệ mạch giao diện khác được đặt trong máy cầm tay và/hoặc trong máy tính. Tương tự, giao diện loa 18 được triển khai bởi bộ biến đổi số sang tương tự (digital-to-analog, D/A) và hệ mạch giao diện khác được đặt trong máy cầm tay và/hoặc trong máy tính. Theo các phương án thực hiện khác, thiết bị truy nhập âm thanh 7 có thể được triển khai và phân vùng theo cách khác đã biết theo giải pháp kỹ thuật.

Theo các phương án thực hiện sáng chế trong đó thiết bị truy nhập âm thanh 7 là điện thoại di động hoặc tế bào, các phần tử trong thiết bị truy nhập âm thanh 7 được triển khai trong máy tay cầm tế bào. CODEC 20 được triển khai bởi phần mềm chạy trên bộ xử lý trong máy tay cầm hoặc bởi phần cứng chuyên dụng. Theo các phương án thực hiện khác của sáng chế, thiết bị truy nhập âm thanh có thể được triển khai ở các thiết bị khác như các hệ thống truyền thông đường dây ngang hàng và số không dây, như truyền thông nội bộ (intercom), và các máy tay cầm vô tuyến. Ở các ứng dụng như các thiết bị vô tuyến người tiêu dùng, thiết bị truy nhập âm

thanh có thể chứa CODEC với chỉ bộ mã hóa 22 hoặc bộ giải mã 24, chẳng hạn, trong hệ thống micro số hoặc thiết bị phát lại âm nhạc. Theo các phương án thực hiện khác của sáng chế, CODEC 20 có thể được sử dụng mà không có micro 12 và loa 14, chẳng hạn, ở các trạm gốc tế bào truy nhập PTSN.

Việc xử lý giọng nói để cải thiện phân loại có thoại/không thoại được mô tả theo các phương án thực hiện khác của sáng chế có thể được triển khai ở bộ mã hóa 22 hoặc bộ giải mã 24, chẳng hạn. Việc xử lý giọng nói để cải thiện phân loại có thoại/không thoại có thể được triển khai ở phần cứng hoặc phần mềm theo các phương án thực hiện khác nhau. Chẳng hạn, bộ mã hóa 22 hoặc bộ giải mã 24 có thể là một phần của vi mạch xử lý tín hiệu số (DSP).

Fig.12 minh họa sơ đồ khối của hệ thống xử lý có thể được sử dụng để triển khai các thiết bị và các phương pháp được bộc lộ ở đây. Các thiết bị cụ thể có thể sử dụng tất cả các thành phần được thể hiện, hoặc chỉ tập con của các thành phần, và các mức tích hợp có thể thay đổi từ thiết bị này sang thiết bị khác. Ngoài ra, thiết bị có thể chứa nhiều thực thể của thành phần, như các khối xử lý, các bộ xử lý, các bộ nhớ, các bộ truyền, các bộ nhận, v.v.. Hệ thống xử lý có thể bao gồm khối xử lý có một hoặc nhiều thiết bị nhập/xuất, như loa, micro, chuột, màn hình cảm ứng, vùng phím bấm, bàn phím, máy in, phần hiển thị, và tương tự. Khối xử lý có thể gồm khối xử lý trung tâm (central processing unit, CPU), bộ nhớ, thiết bị lưu trữ khối, bộ điều hợp video, và giao diện I/O được nối với đường truyền.

Đường truyền có thể là một hoặc nhiều loại bất kỳ trong các kiến trúc đường truyền gồm đường truyền bộ nhớ hoặc bộ điều khiển bộ nhớ, đường truyền ngoại vi, đường truyền video, hoặc tương tự. CPU có thể bao gồm loại bộ xử lý dữ liệu điện tử bất kỳ. Bộ nhớ có thể bao gồm loại bộ nhớ hệ thống bất kỳ như bộ nhớ truy nhập ngẫu nhiên tĩnh (static

random access memory, SRAM), bộ nhớ truy nhập ngẫu nhiên động (dynamic random access memory, DRAM), DRAM đồng bộ (SDRAM), bộ nhớ chỉ đọc (read-only memory, ROM), tổ hợp của chúng, hoặc tương tự. Theo phương án thực hiện, bộ nhớ có thể gồm ROM để sử dụng lúc khởi động, và DRAM cho chương trình và lưu trữ dữ liệu để sử dụng trong khi thực thi các chương trình.

Thiết bị lưu trữ khối có thể bao gồm loại thiết bị lưu trữ bất kỳ được tạo cấu hình để lưu trữ dữ liệu, các chương trình, và loại thông tin khác và khiếm dữ liệu, các chương trình, và thông tin khác truy nhập được qua đường truyền. Thiết bị lưu trữ khối có thể bao gồm, chẳng hạn, một hoặc nhiều ổ trạng thái rắn, ổ đĩa cứng, ổ đĩa từ, ổ đĩa quang, hoặc tương tự.

Bộ điều hợp video và giao diện I/O cung cấp các giao diện để ghép nối các thiết bị I/O với khối xử lý. Như được minh họa, các ví dụ của các thiết bị I/O gồm phần hiển thị được ghép nối với bộ điều hợp video và chuột/bàn phím/máy in được ghép nối với giao diện I/O. Các thiết bị khác có thể được ghép nối với khối xử lý, và một vài thẻ giao diện có thể được tận dụng. Chẳng hạn, giao diện nối tiếp như đường truyền nối tiếp đa năng (Universal Serial Bus, USB) (không được thể hiện trên hình vẽ) có thể được sử dụng để tạo giao diện cho máy in.

Khối xử lý còn gồm một hoặc nhiều giao diện mạng, có thể bao gồm các liên kết có dây, như cáp Ethernet hoặc tương tự, và/hoặc các liên kết không dây để truy nhập các nút hoặc các mạng khác nhau. Giao diện mạng cho phép khối xử lý để truyền thông với các khối từ xa qua các mạng. Chẳng hạn, giao diện mạng có thể cung cấp truyền thông không dây qua một hoặc nhiều bộ truyền/anten truyền và một hoặc nhiều bộ nhận/anten nhận. Theo phương án thực hiện, khối xử lý được ghép nối với LAN hoặc WAN để xử lý dữ liệu và truyền thông với các thiết bị mạng từ xa, như các khối xử lý khác, mạng Internet, các phương tiện lưu trữ từ xa, hoặc tương tự.

Trong khi sáng chế đã được mô tả có dựa vào các phương án thực hiện minh họa, song phần mô tả không được nhằm hiểu theo nghĩa giới hạn. Các cải biến khác và các kết hợp của các phương án thực hiện minh họa, cũng như các phương án thực hiện khác của sáng chế, sẽ rõ ràng với chuyên gia trong lĩnh vực khi tham khảo phần mô tả. Chẳng hạn, các phương án thực hiện nêu trên có thể được kết hợp với nhau.

Dựa vào Fig.13, thiết bị 130 theo phương án thực hiện để xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số được mô tả. Thiết bị gồm:

bộ lựa chọn mã hóa 131 được tạo cấu hình để lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số.

Trong đó khi tín hiệu số gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên, và

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới.

Trong đó khi tín hiệu số gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên, và trong đó tính chu kỳ thanh âm là thấp.

Trong đó khi tín hiệu số không gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường.

Trong đó khi tín hiệu số gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên và tính chu kỳ thanh âm rất mạnh.

Thiết bị còn gồm khối mã hóa 132, khối mã hóa được tạo cấu hình để mã hóa tín hiệu số sử dụng mã hóa miền tần số được lựa chọn bởi bộ chọn 131 hoặc mã hóa miền thời gian được lựa chọn bởi bộ chọn 131.

Bộ lựa chọn mã hóa và khối mã hóa có thể được triển khai bởi CPU hoặc bởi các mạch phần cứng như FPGA, ASIC.

Dựa vào Fig.14, theo phương án thực hiện, thiết bị 140 để xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số được mô tả. Thiết bị gồm:

khối lựa chọn mã hóa 141, khối lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số bao gồm dữ liệu âm thanh khi tín hiệu số không gồm tín hiệu cao độ ngắn và tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường;

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên, và tín hiệu số gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm là thấp; và

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian và tín hiệu số gồm tín hiệu cao độ ngắn và tính chu kỳ thanh âm rất mạnh.

Thiết bị còn gồm khối mã hóa thứ hai 142, khối mã hóa thứ hai được tạo cấu hình để mã hóa tín hiệu số sử dụng mã hóa miền tần số được lựa chọn bởi khối lựa chọn mã hóa 141 hoặc mã hóa miền thời gian được lựa chọn bởi khối lựa chọn mã hóa 141.

Khối lựa chọn mã hóa và khối mã hóa có thể được triển khai bởi CPU hoặc bởi một số mạch phần cứng như FPGA, ASIC.

Mặc dù sáng chế và các ưu điểm của nó đã được mô tả chi tiết, song nên hiểu rằng các thay đổi, thay thế có thể được thực hiện ở đây mà không xa rời phạm vi của sáng chế như được định nghĩa bởi các điểm yêu cầu bảo hộ đi kèm. Chẳng hạn, nhiều dấu hiệu và chức năng nêu trên có thể được triển khai trong phần mềm, phần cứng, hoặc phần sụn, hoặc kết hợp của chúng. Ngoài ra, phạm vi sáng chế không được nhằm giới hạn ở quá trình, máy móc, sản xuất, thành phần chất, phương tiện, phương pháp và các bước theo các phương án thực hiện sáng chế. Khi người có kiến thức trung bình trong lĩnh vực sẽ hiểu rõ từ phần bộc lộ sáng chế, các quá trình, các máy móc, sản xuất, các thành phần chất, các phương tiện, các phương pháp, hoặc các bước, hiện tại có hoặc sau này được phát triển, thực hiện chức năng gần như tương tự hoặc đạt kết quả gần như tương tự với các phương án thực hiện tương ứng được mô tả ở đây có thể được sử dụng theo sáng chế. Do đó, các điểm yêu cầu bảo hộ đi kèm được nhằm gồm trong phạm vi của chúng các quá trình, máy móc, sản xuất, các thành phần chất, các phương tiện, các phương pháp, hoặc các bước.

YÊU CẦU BẢO HỘ

1. Phương pháp xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh, phương pháp bao gồm các bước:

lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên:

tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và

dò thấy trễ cao độ ngắn của tín hiệu số; trong đó việc dò thấy trễ cao độ ngắn bao gồm dò xem liệu tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, trong đó giới hạn độ trễ cao độ là cao độ cho phép nhỏ nhất cho thuật toán kỹ thuật dự báo tuyến tính kích thích bằng mã (Code Excited Linear Prediction Technique, CELP) để mã hóa tín hiệu số.

2. Phương pháp theo điểm 1, trong đó tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới.

3. Phương pháp theo điểm 2, trong đó tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới khi tốc độ bit mã hóa nhỏ hơn 24,4kbps.

4. Phương pháp theo điểm 1, trong đó tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên.

5. Phương pháp theo điểm 4, trong đó tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên khi tốc độ bit mã hóa lớn hơn hoặc bằng 46200bps.

6. Phương pháp theo điểm 1, trong đó tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi mã hóa tốc độ bit là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên, và trong khi tính chu kỳ thoại thấp.

7. Phương pháp theo điểm 1, trong đó tín hiệu số không bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường.

8. Phương pháp theo điểm 1, trong đó tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên và tính chu kỳ thanh âm rất mạnh.

9. Phương pháp theo điểm 1, còn bao gồm mã hóa tín hiệu số sử dụng mã hóa miền tần số được chọn hoặc mã hóa miền thời gian được chọn.

10. Phương pháp theo điểm 1, trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên độ trễ cao độ của tín hiệu số bao gồm việc dò tín hiệu cao độ ngắn dựa trên xác định tham số để dò thiếu năng lượng tần số thấp hoặc tham số cho độ rõ nét phổ.

11. Thiết bị xử lý các tín hiệu giọng nói trước khi mã hóa tín hiệu số bao gồm dữ liệu âm thanh, thiết bị bao gồm bộ lựa chọn mã hóa được tạo cấu hình để lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian dựa trên tốc độ bit mã hóa sẽ được sử dụng để mã hóa tín hiệu số và dò thấy trễ cao độ ngắn của tín hiệu số, trong đó việc dò độ trễ cao độ ngắn bao gồm dò xem liệu tín hiệu số có bao gồm tín hiệu cao độ ngắn mà độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, trong đó giới hạn độ trễ cao độ là cao độ cho phép nhỏ nhất đối với thuật toán CELP để mã hóa tín hiệu số.

12. Thiết bị theo điểm 11, trong đó khi tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để:

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới.

13. Thiết bị theo điểm 12, trong đó tốc độ bit mã hóa thấp hơn giới hạn tốc độ bit dưới khi tốc độ bit mã hóa nhỏ hơn 24,4kbps.

14. Thiết bị theo điểm 11, trong đó tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, và trong đó lựa chọn mã hóa miền tần số hoặc mã hóa miền thời gian bao gồm bước:

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên.

15. Thiết bị theo điểm 4, trong đó tốc độ bit mã hóa cao hơn giới hạn tốc độ bit trên khi tốc độ bit mã hóa lớn hơn hoặc bằng 46200bps.

16. Thiết bị theo điểm 11, trong đó khi tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để

lựa chọn mã hóa miền tần số để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên, và trong đó tính chu kỳ thanh âm là thấp.

17. Thiết bị theo điểm 11, trong đó khi tín hiệu số không bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tín hiệu số được phân loại thành giọng nói không có thoại hoặc giọng nói thông thường.

18. Thiết bị theo điểm 11, trong đó khi tín hiệu số bao gồm tín hiệu cao độ ngắn mà với nó độ trễ cao độ ngắn hơn giới hạn độ trễ cao độ, bộ lựa chọn mã hóa được tạo cấu hình để

lựa chọn mã hóa miền thời gian để mã hóa tín hiệu số khi tốc độ bit mã hóa là trung gian giữa giới hạn tốc độ bit dưới và giới hạn tốc độ bit trên và tính chu kỳ thanh âm rất mạnh.

19. Thiết bị theo điểm 11, trong đó thiết bị còn bao gồm khối mã hóa được tạo cấu hình để mã hóa tín hiệu số sử dụng mã hóa miền tần số được lựa chọn bởi bộ chọn hoặc mã hóa miền thời gian được lựa chọn bởi bộ chọn.

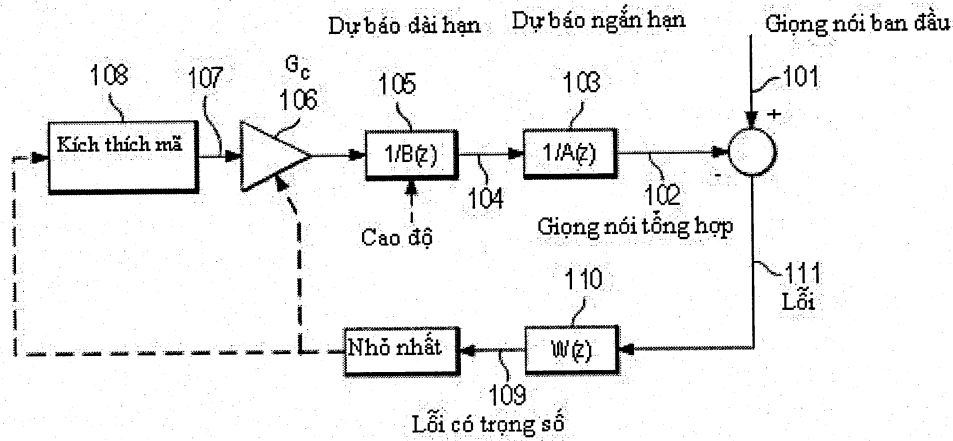


Fig. 1

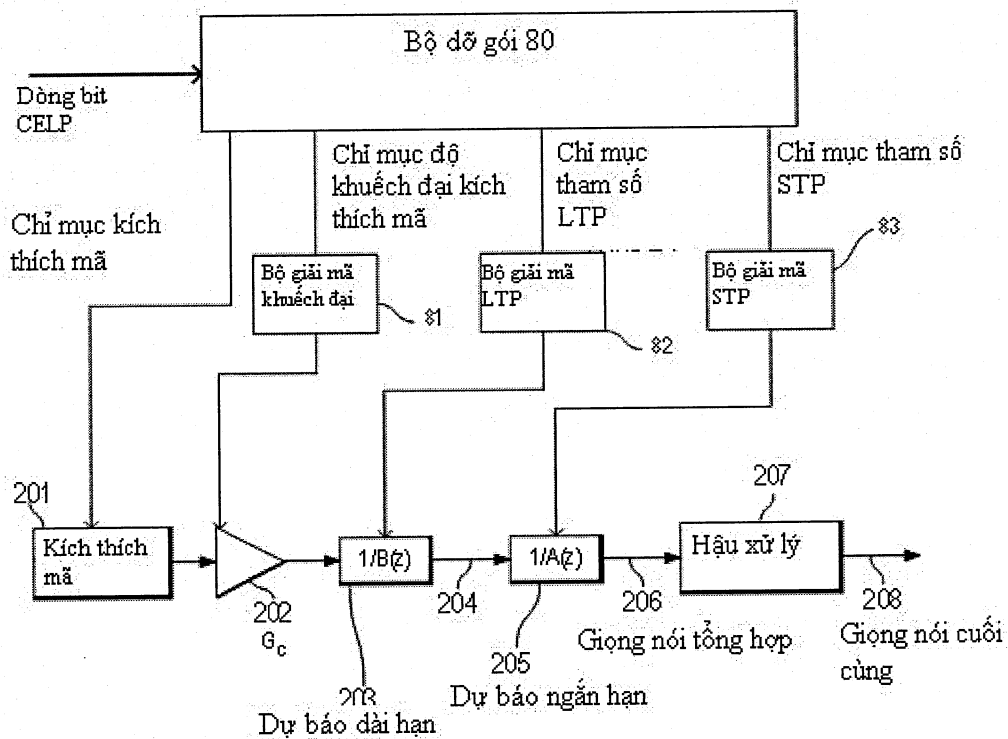


Fig. 2

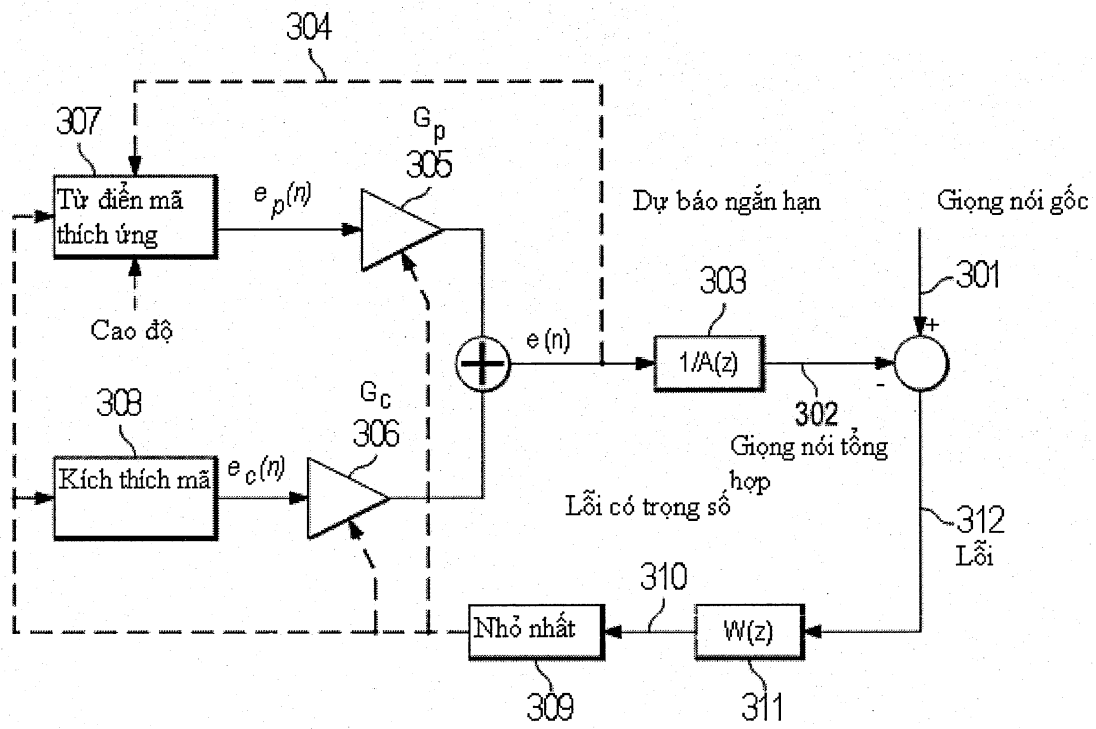
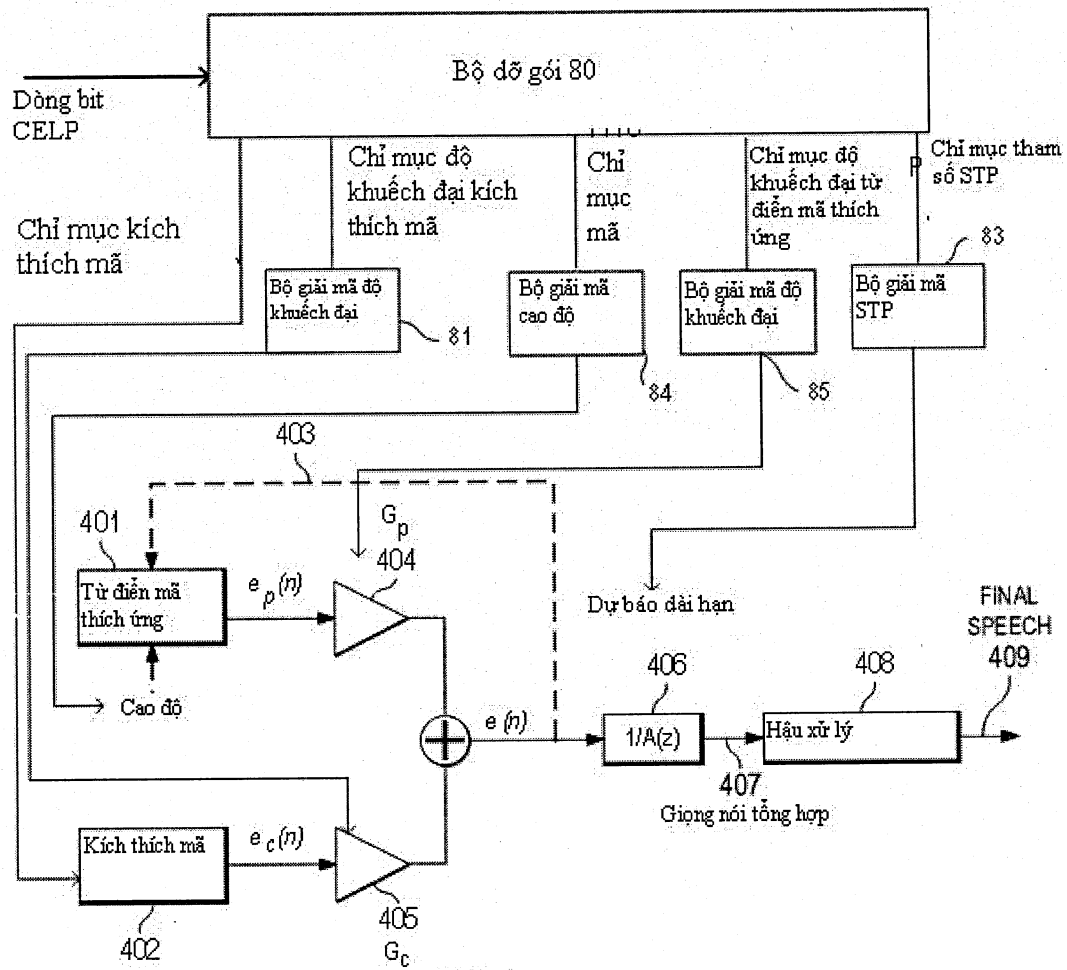


Fig. 3



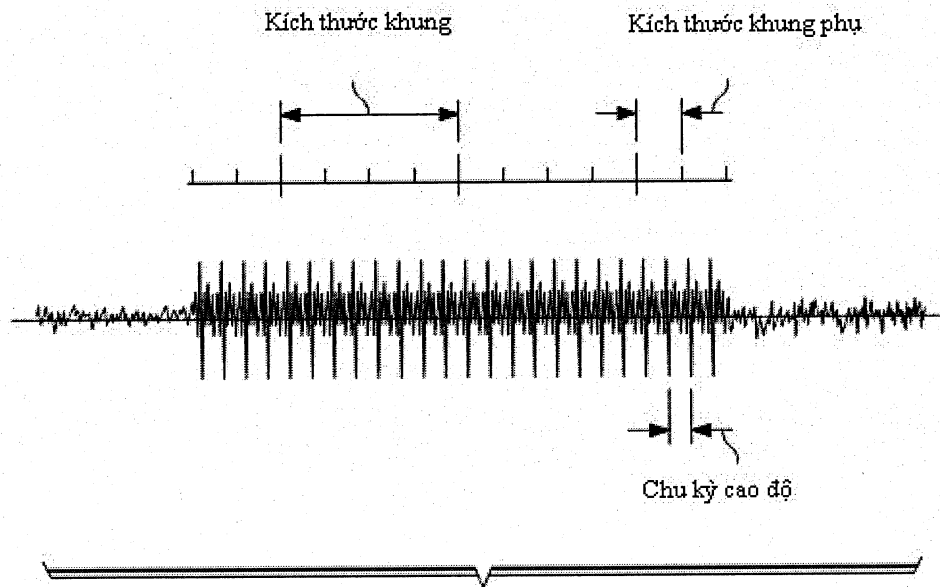


Fig.5

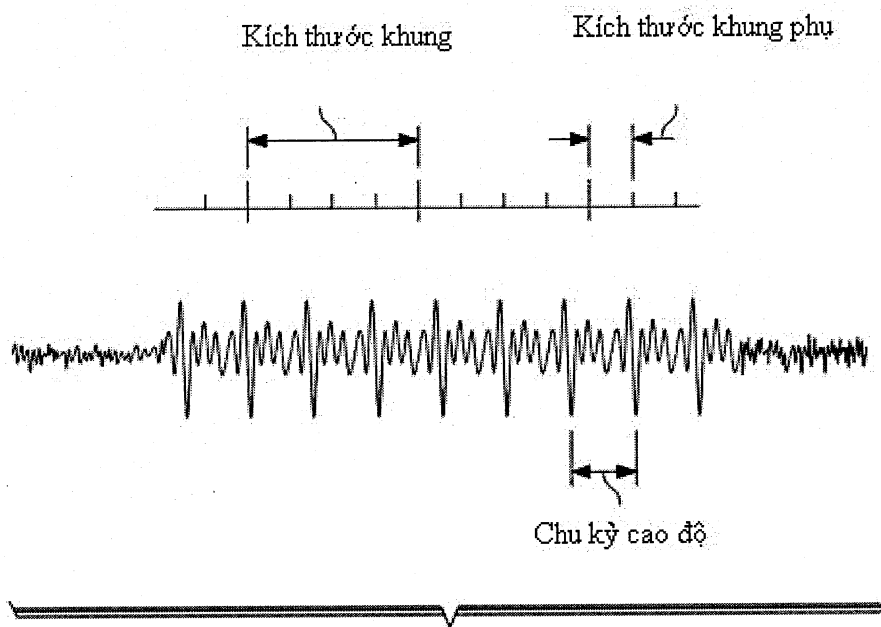


Fig.6

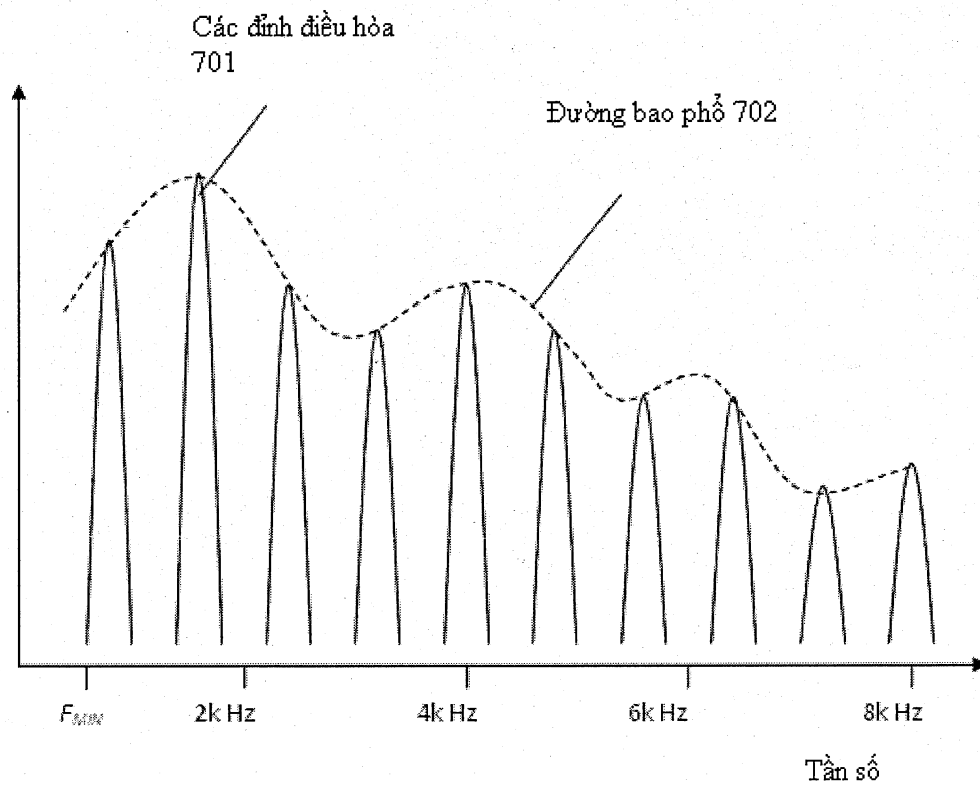


Fig.7

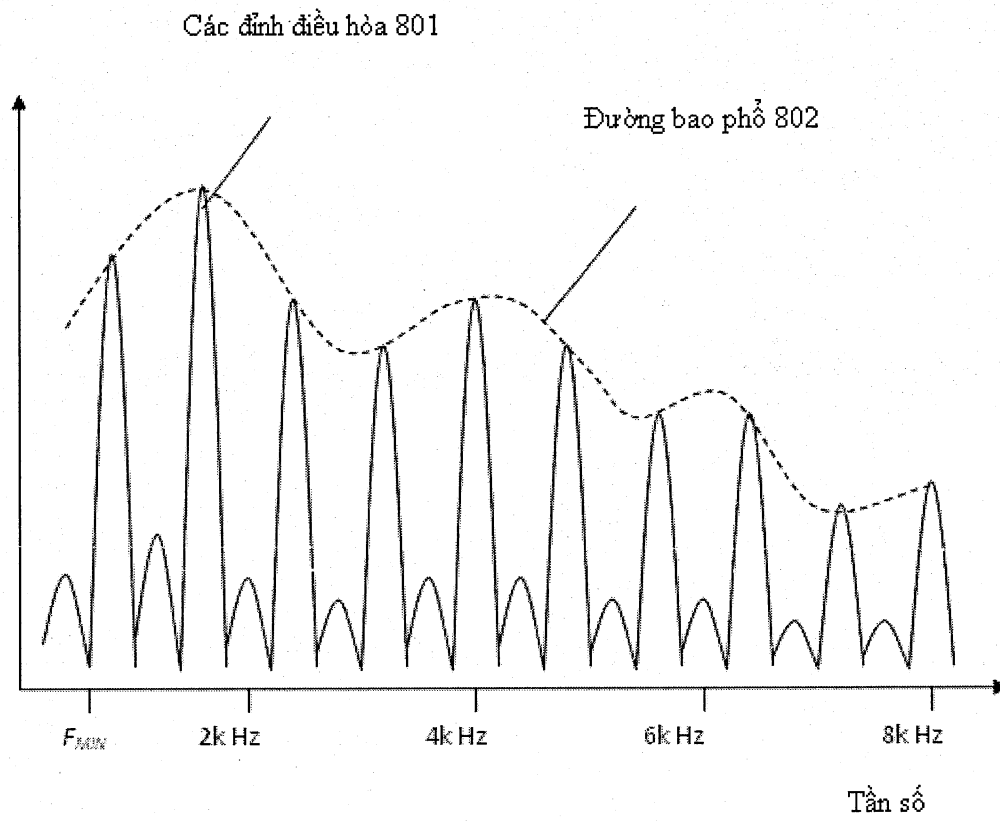


Fig.8

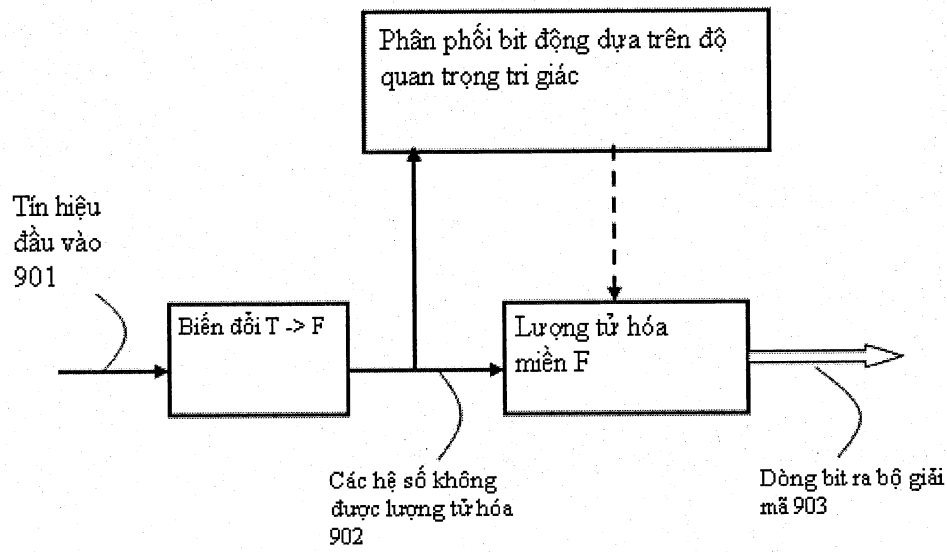


Fig. 9A

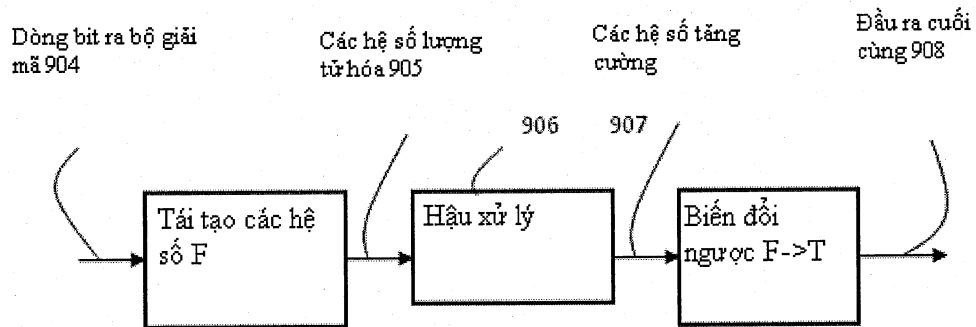


Fig. 9B

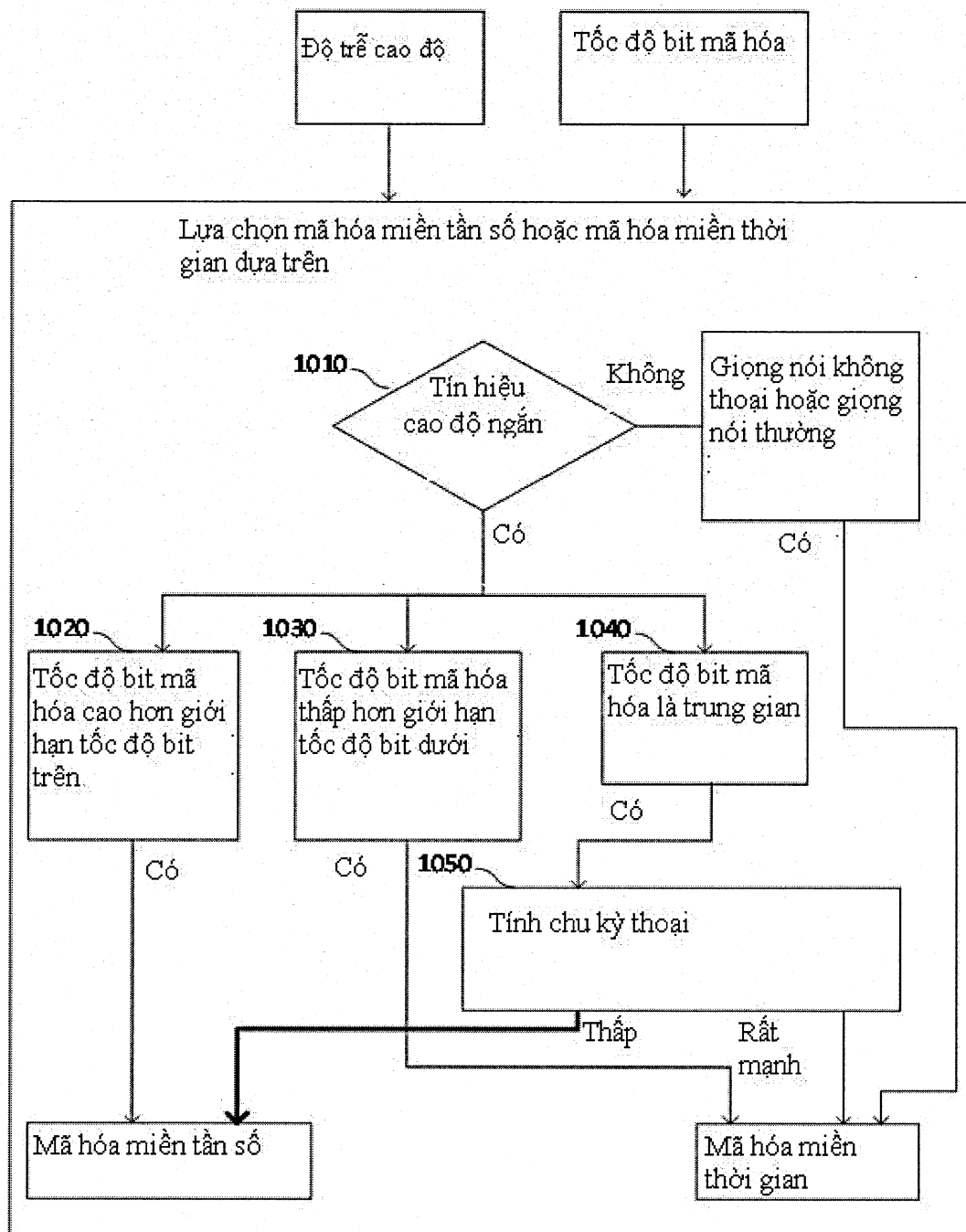


Fig.10

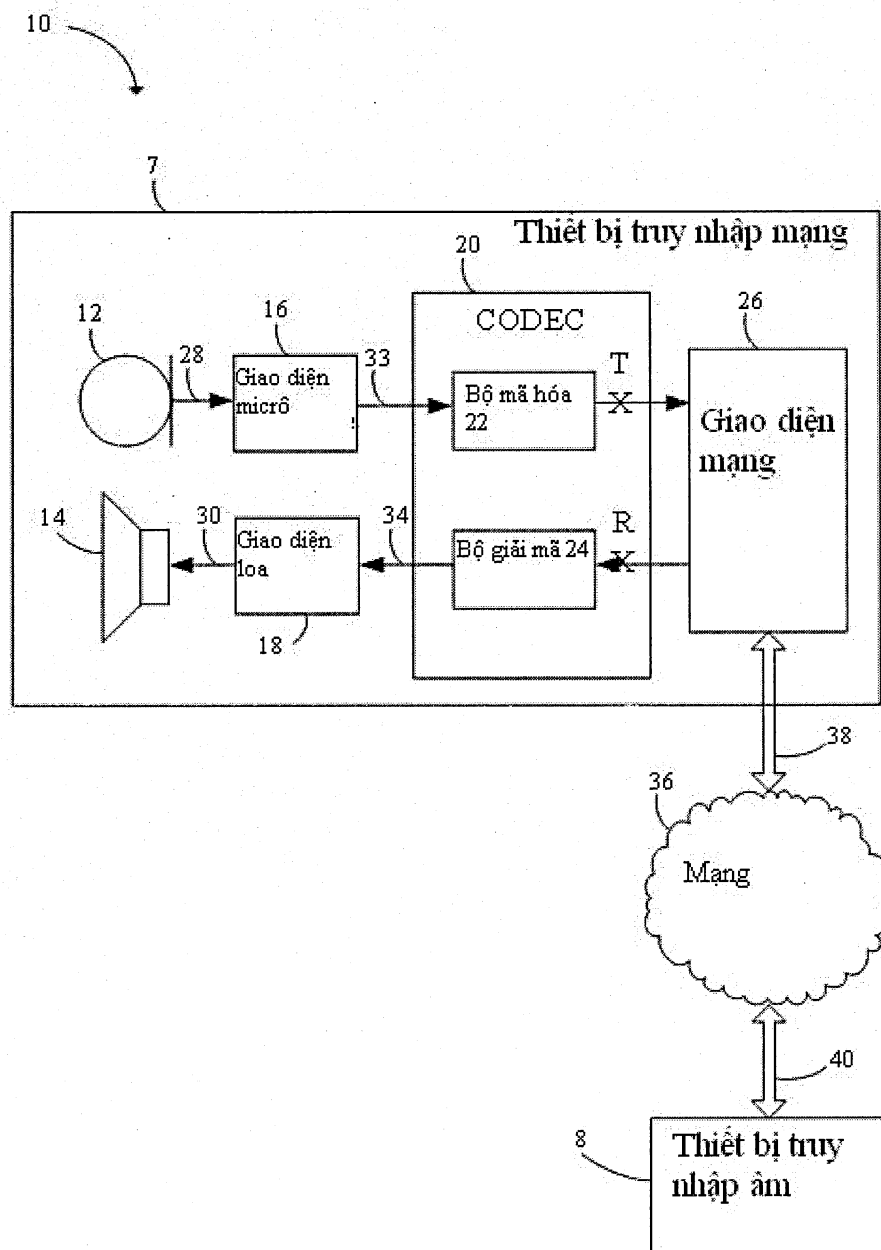


Fig.11

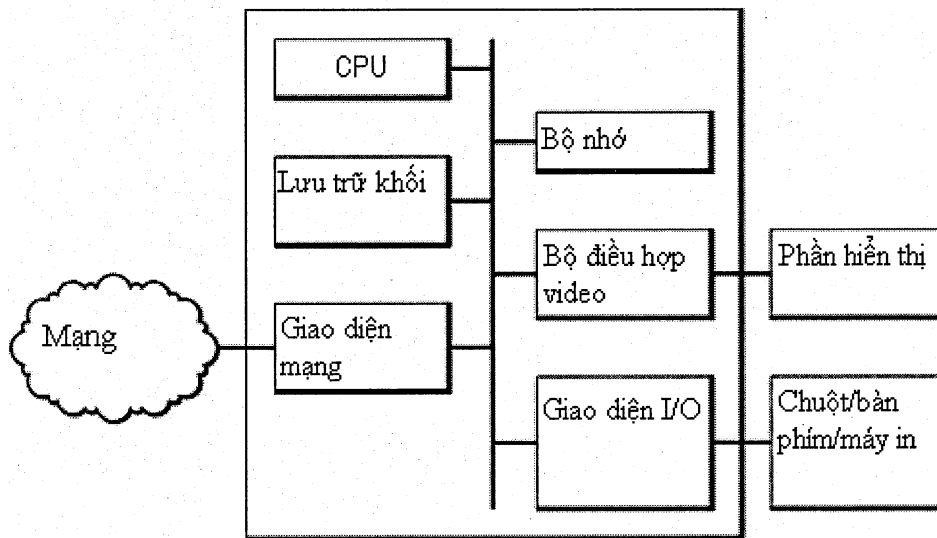


Fig.12

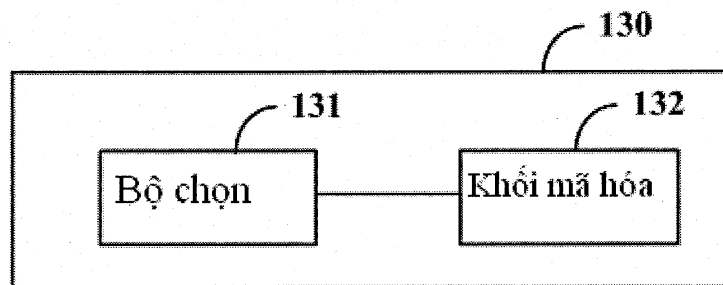


Fig.13

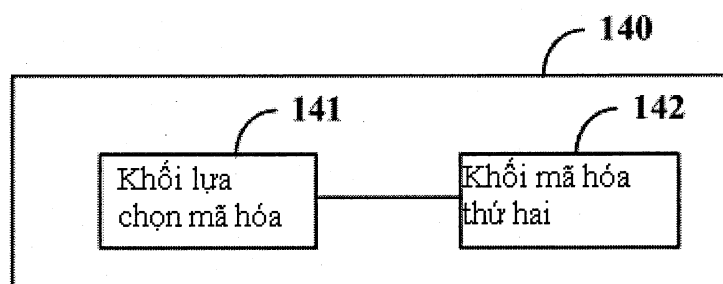


Fig.14