



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

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

(11)



1-0036807

(51)⁷ G10L 19/06; G10L 21/0208

(13) B

(21) 1-2018-01736

(22) 23/09/2016

(86) PCT/EP2016/072701 23/09/2016

(87) WO 2017/050972 30/03/2017

(30) 15186901.3 25/09/2015 EP; 16175469.2 21/06/2016 EP

(45) 25/09/2023 426

(43) 25/06/2018 363A

(73) FRAUNHOFER-GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (DE)

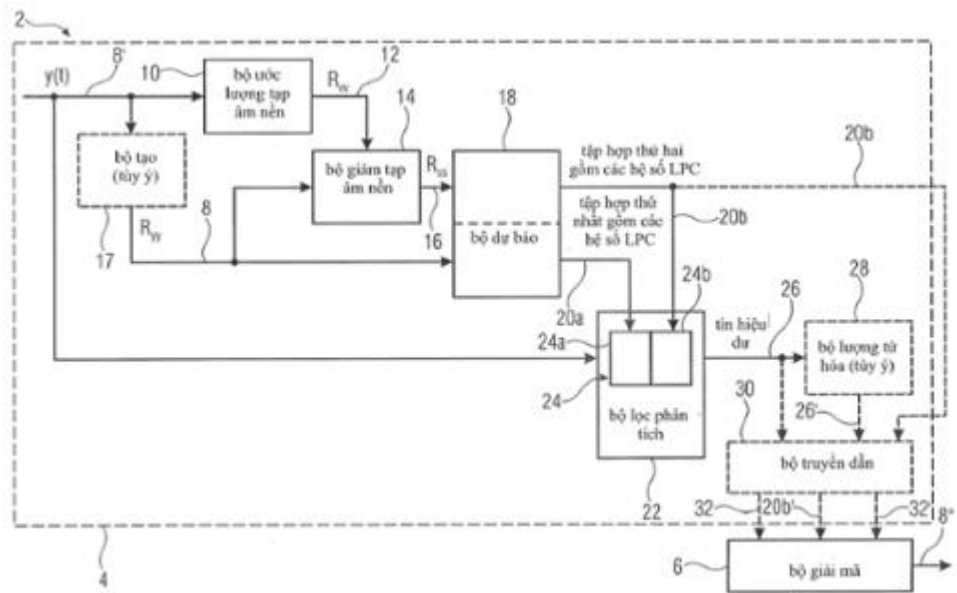
Hansastrasse 27c, 80686 Muenchen, Germany

(72) FISCHER, Johannes (DE); BAECKSTROEM, Tom (FI); JOKINEN, Emma (FI).

(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)

(54) BỘ MÃ HÓA, PHƯƠNG PHÁP MÃ HÓA TÍN HIỆU ÂM THANH VỚI TẬP ÂM
NỀN ĐƯỢC GIẢM SỬ DỤNG MÃ HÓA DỰ BÁO TUYẾN TÍNH VÀ HỆ
THỐNG MÃ HÓA VÀ GIẢI MÃ

(57) Sáng chế đề cập đến bộ mã hóa, phương pháp mã hóa tín hiệu âm thanh với tập âm nền được giảm sử dụng mã hóa dự báo tuyến tính và hệ thống mã hóa và giải mã. Bộ mã hóa để mã hóa tín hiệu âm thanh với tập âm nền được giảm sử dụng mã hóa dự báo tuyến tính. Bộ mã hóa bao gồm bộ ước lượng tập âm nền được tạo cấu hình để ước lượng tập âm nền của tín hiệu âm thanh, bộ giảm tập âm nền được tạo cấu hình để tạo tín hiệu âm thanh được giảm tập âm nền bằng việc trừ tập âm nền được ước lượng của tín hiệu âm thanh từ tín hiệu âm thanh, và bộ dự báo được tạo cấu hình để đưa tín hiệu âm thanh trải qua phân tích dự báo tuyến tính để thu tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) và để đưa tín hiệu âm thanh được giảm tạo âm nền trải qua phân tích dự báo tuyến tính để thu được tập hợp thứ hai gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC). Hơn nữa, bộ mã hóa bao gồm bộ lọc phân tích bao gồm tầng nối tiếp gồm các bộ lọc miền thời gian được điều khiển bởi tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến bộ mã hóa để mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính, phương pháp và hệ thống tương ứng bao gồm bộ mã hóa và bộ giải mã. Nói cách khác, sáng chế đề cập đến phương pháp nâng cao và/hoặc mã hóa tiếng nói liên kết, ví dụ nâng cao và mã hóa liên kết tiếng nói bằng việc kết hợp trong bộ mã hóa-giải mã CELP (dự báo tuyến tính kích thích theo bảng mã).

Tình trạng kỹ thuật của sáng chế

Các thiết bị liên lạc và giao tiếp đã trở nên phổ biến và có thể được sử dụng trong các điều kiện bất lợi, yêu cầu đối với các phương pháp nâng cao tiếng nói có thể giải quyết các môi trường bất lợi đã tăng lên. Kết quả là, ví dụ, bây giờ trong điện thoại di động sử dụng phổ biến các phương pháp giảm nhiễu âm như là một khối/bước xử lý trước cho tất cả các bước xử lý tiếng nói sau đó ví dụ như mã hóa tiếng nói. Có các phương pháp khác nhau mà kết hợp việc nâng cao tiếng nói vào các bộ mã hóa tiếng nói [1, 2, 3, 4]. Trong khi các thiết kế cải thiện chất lượng tiếng nói được truyền dẫn, xử lý nối tiếp không cho phép tối ưu hóa/tối thiểu hóa cảm quan liên kết của chất lượng hoặc tối thiểu hóa liên kết của nhiễu âm lượng tử hóa và nhiễu âm ít nhất là khó khăn.

Mục tiêu của bộ mã hóa-giải mã tiếng nói là cho phép truyền dẫn tiếng nói chất lượng cao với lượng tối thiểu dữ liệu được truyền. Để đạt được mục tiêu này, các phép biểu diễn hiệu quả của tín hiệu là cần thiết, ví dụ tạo mô hình đường bao quang phổ của tín hiệu tiếng nói bằng dự báo tuyến tính, tần số cơ sở bằng bộ dự báo dài hạn và phần còn lại với bảng mã. Phép biểu diễn là cơ sở của bộ mã hóa - giải mã tiếng nói sử dụng mô hình dự báo tuyến tính kích thích mã hóa (code excited linear prediction - CELP) mà được sử dụng trong các tiêu chuẩn mã hóa tiếng nói chính ví dụ đa tốc độ thích ứng (Adaptive Multi-Rate - AMR), AMR-băng rộng (AMR-Wide-Band - AMR-WB), Mã hóa âm thanh và tiếng nói hợp nhất (Unified Speech and Audio Coding - USAC) và Dịch vụ thoại nâng cao (Enhanced Voice Service - EVS) [5, 6, 7, 8, 9, 10, 11].

Đối với giao tiếp tiếng nói tự nhiên, người nói thường sử dụng các thiết bị ở chế độ rảnh tay. Trong các tình huống này, các ống phóng thành thường xa miệng mà ở đó tín hiệu tiếng nói có thể dễ dàng trở nên biến dạng do nhiễu âm ví dụ như tiếng vọng hoặc tạp âm nền. Sự thoái hóa này không chỉ ảnh hưởng đến chất lượng tiếng nói cảm quan mà còn ảnh hưởng đến độ rõ của tín hiệu tiếng nói và do đó có thể cản trở một cách nghiêm trọng độ tự nhiên của cuộc hội thoại. Để nâng cao trải nghiệm giao tiếp, có lợi khi áp dụng các phương pháp nâng cao tiếng nói để giảm nhiễu âm và giảm các hiệu ứng tiếng vọng. Lĩnh vực của kỹ thuật nâng cao tiếng nói là hoàn thiện và có nhiều phương pháp có sẵn [12]. Tuy nhiên, phần lớn các thuật toán hiện tại dựa trên các phương pháp chồng lấp-cộng, ví dụ như các phép biến đổi như biến đổi Fourier thời gian ngắn (short-time Fourier transform - STFT) mà áp dụng chồng lấp-cộng dựa trên các sơ đồ tạo cửa sổ, trong khi ngược lại, các bộ mã hóa-giải mã CELP tạo mô hình tín hiệu với bộ dự báo tuyến tính/bộ lọc dự báo tuyến tính và chỉ áp dụng việc tạo cửa sổ trên phần dư. Những khác biệt cơ bản như vậy khiến cho việc hợp nhất các phương pháp nâng cao và mã hóa khó khăn. Tuy nhiên, rõ ràng rằng việc tối ưu hoá kết hợp giữa nâng cao và mã hóa có thể có khả năng cải thiện chất lượng, giảm độ trễ và độ phức tạp tính toán.

Do đó, tồn tại nhu cầu về một phương pháp được cải thiện.

Tài liệu tham khảo

[1] M. Jeub and P. Vary, “Enhancement of reverberant speech using the CELP postfilter,” in Proc. ICASSP, April 2009, pp. 3993–3996.

[2] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, “Noise reduction for dual-microphone mobile phones exploiting power level differences,” in Proc. ICASSP, March 2012, pp. 1693–1696.

[3] R. Martin, I. Wittke, and P. Jax, “Optimized estimation of spectral parameters for the coding of noisy speech,” in Proc. ICASSP, vol. 3, 2000, pp. 1479–1482 vol.3.

[4] H. Taddei, C. Beaugeant, and M. de Meuleneire, “Noise reduction on speech codec parameters,” in Proc. ICASSP, vol. 1, May 2004, pp. I-497–500 vol.1.

[5] 3GPP, “Mandatory speech CODEC speech processing functions; AMR speech

Codec; General description,” 3rd Generation Partnership Project (3GPP), TS 26.071, 12 2009. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/26071.htm>

[6] —, “Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions,” 3rd Generation Partnership Project (3GPP), TS 26.190, 12 2009. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/26190.htm>

[7] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, “The adaptive multirate wideband speech codec (AMR-WB),” *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620–636, Nov 2002.

[8] ISO/IEC 23003–3:2012, “MPEG-D (MPEG audio technologies), Part 3: Unified speech and audio coding,” 2012.

[9] M. Neuendorf, P. Gournay, M. Multus, J. Lecomte, B. Bessette, R. Geiger, S. Bayer, G. Fuchs, J. Hilpert, N. Rettelbach, R. Salami, G. Schuller, R. Lefebvre, and B. Grill, “Unified speech and audio coding scheme for high quality at low bitrates,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 1–4.

[10] 3GPP, “TS 26.445, EVS Codec Detailed Algorithmic Description; 3GPP Technical Specification (Release 12),” 3rd Generation Partnership Project (3GPP), TS 26.445, 12 2014. [Online]. Available: <http://www.3gpp.org/ftp/Specs/html-info/26445.htm>

[11] M. Dietz, M. Multus, V. Eksler, V. Malenovsky, E. Norvell, H. Poblath, L. Miao, Z. Wang, L. Laaksonen, A. Vasilache, Y. Kamamoto, K. Kikuri, S. Ragot, J. Faure, H. Ehara, V. Rajendran, V. Atti, H. Sung, E. Oh, H. Yuan, and C. Zhu, “Overview of the EVS codec architecture,” in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, April 2015, pp. 5698–5702.

[12] J. Benesty, M. Sondhi, and Y. Huang, *Springer Handbook of Speech Processing*. Springer, 2008.

[13] T. Bäckström, “Computationally efficient objective function for algebraic

codebook optimization in ACELP,” in Proc. Interspeech, Aug. 2013.

[14] —, “Comparison of windowing in speech and audio coding,” in Proc. WASPAA, New Paltz, USA, Oct. 2013.

[15] J. Fischer and T. Bäckström, “Comparison of windowing schemes for speech coding,” in Proc EUSIPCO, 2015.

[16] M. Schroeder and B. Atal, “Code-excited linear prediction (CELP): High-quality speech at very low bit rates,” in Proc. ICASSP. IEEE, 1985, pp. 937–940.

[17] T. Bäckström and C. R. Helmrich, “Decorrelated innovative codebooks for ACELP using factorization of autocorrelation matrix,” in Proc. Interspeech, 2014, pp. 2794–2798.

[18] soundeffects.ch, “Civilisation soundscapes library,” accessed: 23.09.2015. [Online]. Available: <https://www.soundeffects.ch/de/geraeusch-archive/soundeffects.ch-produkte/civilisation-soundscapes-d.php>

[19] Method for the subjective assessment of intermediate quality levels of coding systems, ITU-R Recommendation BS.1534, 2003. [Online]. Available: <http://www.itu.int/rec/R-REC-BS.1534/en>.

[20] P. P. Vaidyanathan, “The theory of linear prediction,” in Synthesis Lectures on Signal Processing, vol. 2, pp. 1–184. Morgan & Claypool publishers, 2007.

[21] J. Allen, “Short-term spectral analysis, and modification by discrete Fourier transform,” IEEE Trans. Acoust., Speech, Signal Process., vol. 25, pp. 235–238, 1977.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất nguyên lý được cải thiện về xử lý tín hiệu âm thanh sử dụng mã hóa dự báo tuyến tính. Mục đích này được giải quyết bằng đối tượng bảo hộ trong các điểm yêu cầu bảo hộ độc lập.

Các phương án của sáng chế thể hiện bộ mã hóa để mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính. Bộ mã hóa bao gồm bộ ước lượng tạp âm nền được tạo cấu hình để ước lượng tạp âm nền của tín hiệu âm thanh, bộ giảm tạp âm nền được tạo cấu hình để tạo tín hiệu âm thanh được giảm tạp âm nền bằng

việc trừ tạp âm nền được ước lượng của tín hiệu âm thanh từ tín hiệu âm thanh, và bộ dự báo được tạo cấu hình để đưa tín hiệu âm thanh trải qua phân tích dự báo tuyến tính để thu tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) và để đưa tín hiệu âm thanh được giảm tạp âm nền trải qua phân tích dự báo tuyến tính để thu được tập hợp thứ hai gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC). Hơn nữa, bộ mã hóa bao gồm bộ lọc phân tích bao gồm tầng nối tiếp gồm các bộ lọc miền thời gian được biểu diễn bởi tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC.

Sáng chế dựa trên việc phát hiện rằng bộ lọc phân tích được cải thiện trong môi trường mã hóa dự báo tuyến tính tăng các đặc tính xử lý tín hiệu của bộ mã hóa. Cụ thể hơn, sử dụng tầng nối tiếp hoặc chuỗi gồm các bộ lọc miền thời gian được kết nối nối tiếp cải thiện tốc độ xử lý hoặc thời gian xử lý của tín hiệu âm thanh đầu vào nếu các bộ lọc đã nêu được áp dụng cho bộ lọc phân tích của môi trường mã hóa dự báo tuyến tính. Đây là lợi thế do việc chuyển đổi đổi thời gian - tần số được sử dụng điển hình và việc biến đổi tần số-thời gian nghịch đảo của tín hiệu âm thanh miền thời gian quay về để giảm tạp âm nền bằng việc lọc các băng tần mà bị chi phối bởi nhiễu âm được bỏ qua. Nói cách khác, việc thực hiện việc giảm hoặc xóa bỏ tạp âm nền như một phần của bộ lọc phân tích, việc giảm tạp âm nền có thể được thực hiện trong miền thời gian. Do đó, quy trình chồng lấp-và-cộng, ví dụ là phép biến đổi cosin rời rạc cải biên (nghịch đảo) (MDCT/IDMCT - [inverse] modified discrete cosine transform), mà có thể được sử dụng cho việc chuyển đổi thời gian/tần số/thời gian được bỏ qua. Phương pháp chồng lấp-và-cộng giới hạn đặc tính xử lý thời gian thực của bộ mã hóa, do việc giảm tạp âm nền không thể thực hiện trên khung đơn mà chỉ trên các khung liên tiếp.

Nói cách khác, bộ mã hóa như mô tả ở trên có khả năng thực hiện việc giảm tạp âm nền và do đó toàn bộ việc xử lý của bộ lọc phân tích trên khung âm thanh đơn, và do đó cho phép xử lý thời gian thực tín hiệu âm thanh. Xử lý thời gian thực có thể liên quan tới xử lý tín hiệu âm thanh mà không có độ trễ đáng kể đối với những người tham gia. Độ trễ đáng kể có thể xảy ra ví dụ trong hội nghị qua điện thoại nếu một người dùng phải chờ hồi đáp của người dùng khác do độ trễ xử lý của tín hiệu âm thanh. Độ trễ được cho phép cực đại có thể nhỏ hơn 1 giây, ưu tiên nhỏ hơn 0,75 giây và ưu tiên hơn nữa là nhỏ hơn 0,25 giây. Cần lưu ý rằng thời gian xử lý này liên quan đến việc xử lý toàn bộ

của tín hiệu âm thanh từ người gửi đến người nhận vào do đó gồm có, ngoài việc xử lý tín hiệu của bộ mã hóa còn thời gian truyền dẫn tín hiệu âm thanh và xử lý tín hiệu trong bộ giải mã tương ứng.

Theo các phương án, tầng nối tiếp gồm các bộ lọc miền thời gian và do đó bộ lọc phân tích bao gồm bộ lọc dự báo tuyến tính hai lần sử dụng tập hợp thứ nhất thu được gồm các hệ số LPC và nghịch đảo một lần của bộ lọc dự báo tuyến tính khác sử dụng tập hợp thứ hai thu được gồm các hệ số LPC. Việc xử lý tín hiệu có thể được đề cập đến như là lọc Wiener. Do đó, nói cách khác, tầng nối tiếp gồm các bộ lọc miền thời gian có thể bao gồm bộ lọc Wiener.

Theo các phương án khác, bộ ước lượng tap âm nền có thể ước lượng tương quan tự động của tap âm nền như là phép biểu diễn tap âm nền của tín hiệu âm thanh. Hơn nữa, bộ giảm tap âm nền có thể tạo phép biểu diễn tín hiệu âm thanh được giảm tap âm nền bằng việc trừ tương quan tự động của tap âm nền từ tương quan tự động được ước lượng của tín hiệu âm thanh, trong đó tương quan âm thanh được ước lượng của tín hiệu âm thanh là phép biểu diễn của tín hiệu âm thanh và trong đó phép biểu diễn tín hiệu âm thanh được giảm tap âm nền là tương quan tự động của tín hiệu âm thanh được giảm tap âm nền. Việc sử dụng ước lượng của các hàm tương quan tự động thay cho việc sử dụng tín hiệu âm thanh miền thời gian để tính toán các hệ số LPC và để thực hiện việc giảm tap âm nền cho phép xử lý tín hiệu hoàn toàn trong miền thời gian. Do đó, tương quan tự động của tín hiệu âm thanh và tương quan tự động của tap âm nền có thể được tính toán bằng cách nhân chập hoặc bằng việc sử dụng tích phân nhân chập của khung âm thanh hoặc phần con của khung âm thanh. Do đó, tương quan tự động của tap âm nền có thể được thực hiện trong khung hoặc thậm chí chỉ trong khung con mà có thể được xác định như là khung hoặc một phần của khung trong đó (hầu như) không tín hiệu âm thanh phía trước nào ví dụ như tiếng nói là có mặt. Hơn nữa, tương quan tự động của tín hiệu âm thanh được giảm tap âm nền có thể được tính toán bằng việc trừ tương quan tự động của tap âm nền và tương quan tự động của tín hiệu âm thanh (bao gồm tap âm nền). Việc sử dụng tương quan tự động của tín hiệu âm thanh được giảm tap âm nền và tín hiệu âm thanh (điển hình có tap âm nền) cho phép tính toán các hệ số LPC lần lượt cho tín hiệu âm thanh được giảm tap âm nền và tín hiệu âm thanh. Các hệ số LPC được giảm tap âm nền có thể được đề cập đến như là tập hợp thứ hai gồm các hệ số LPC,

trong đó các hệ số LPC của tín hiệu âm thanh có thể được đề cập đến như là tập hợp thứ nhất gồm các hệ số LPC. Do đó, tín hiệu âm thanh có thể được xử lý hoàn toàn trong miền thời gian, do việc ứng dụng tầng nối tiếp gồm các bộ lọc miền thời gian cũng thực hiện việc lọc của chúng trên tín hiệu âm thanh trong miền thời gian.

Trước khi các phương án được mô tả cụ thể sử dụng các hình vẽ đi kèm, cần phải chỉ ra rằng các phần tử tương tự hoặc giống nhau về chức năng có cùng một số chỉ dẫn trong các hình và việc mô tả lặp lại các phần tử được có cùng một số chỉ dẫn được bỏ qua. Do đó, các mô tả được cung cấp cho các phần tử có cùng số chỉ dẫn có thể thay đổi với nhau.

Mô tả vắn tắt các hình vẽ

Các phương án của sáng chế sẽ được thảo luận sau đây viện dẫn đến các hình vẽ đi kèm, trong đó:

Fig.1 thể hiện sơ đồ khối dạng giản lược của hệ thống bao gồm bộ mã hóa để mã hóa tín hiệu âm thanh và bộ giải mã;

Fig.2 thể hiện sơ đồ khối dạng giản lược của a) sơ đồ mã hóa nâng cao tầng nối tiếp, b) sơ đồ mã hóa tiếng nói CELP, và c) sơ đồ mã hóa nâng cao liên kết theo sáng chế;

Fig.3 thể hiện sơ đồ khối dạng giản lược của phương án của Fig.2 với ký hiệu khác;

Fig.4 thể hiện sơ đồ đường dạng giản lược của cường độ cảm quan tỉ số tín hiệu trên nhiễu (SNR - signal-to-noise ration), như được xác định trong phương trình 23 cho phương pháp liên kết được đề xuất (J) và phương pháp tầng nối tiếp (C), trong đó tín hiệu đầu vào được suy giảm bằng nhiễu âm ô tô không cố định và các kết quả được thể hiện cho hai tốc độ bit khác nhau (7,2 kbit/giây được biểu thị bằng ký hiệu 7 và 13,2 kbit/giây được biểu thị bằng ký hiệu 13);

Fig.5 thể hiện sơ đồ đường dạng giản lược của cường độ cảm quan SNR, như được xác định trong phương trình 23 cho phương pháp liên kết được đề xuất (J) và phương pháp tầng nối tiếp (C) trong đó tín hiệu đầu vào được suy giảm bằng nhiễu âm trắng cố định và kết quả được thể hiện cho hai tốc độ bit khác nhau (7,2 kbit/giây được biểu thị

bằng ký hiệu 7 và 13,2 kbit/giây được biểu thị bằng ký hiệu 13);

Fig.6 thể hiện sơ đồ dạng giản lượng thể hiện minh họa các điểm số MUSHRA đối với người nói Tiếng Anh khác nhau (nữ (F) và (nam (M)) đối với hai nhiều âm khác nhau (nhiều âm trắng (W) và nhiều âm ô tô (C)), đối với hai SNR đầu vào khác nhau (10 dB (1) và 20 dB (2)), trong đó tất cả các mục được mã hóa tại hai tốc độ bit (7,2 kbit/s (7) và 13,2 kbit/s (13)) đối với phương pháp liên kết được đề xuất (JE) và nâng cao tăng nối tiếp (CE), trong đó REF là tham chiếu ẩn, LP là neo thông thấp 3.5 kHz và Mix là hỗn hợp bị biến dạng;

Fig.7 thể hiện sơ đồ gồm các điểm số MUSHRA khác nhau, được mô phỏng trên hai tốc độ bit khác nhau, so sánh việc nâng cao liên kết mới (JE) với phương pháp tăng nối tiếp (CE); và

Fig.8 thể hiện lưu đồ của phương pháp mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính.

Mô tả chi tiết sáng chế

Sau đây, các phương án của sáng chế sẽ được mô tả cụ thể. Các phần tử thể hiện trong các hình vẽ tương ứng có cùng chức năng hoặc chức năng tương tự kết hợp cùng với các ký hiệu tham chiếu tương tự.

Phần sau sẽ mô tả phương pháp nâng cao và mã hóa liên kết, dựa trên việc lọc Wiener [12] và mã hóa CELP. Lợi ích của việc kết hợp này là 1) bao gồm việc lọc Wiener trong chuỗi xử lý không làm tăng độ trễ thuật toán thấp của bộ mã hóa-giải mã CELP và việc tối ưu hóa liên kết tối thiểu hóa một cách đồng thời độ biến dạng do lượng tử hóa và tạp âm nền. Hơn nữa, độ phức tạp tính toán của sơ đồ liên kết là thấp hơn so với độ phức tạp tính toán của phương pháp tăng nối tiếp. Việc thực thi dựa trên các công việc gần đây về việc tạo cửa sổ phần dư của các bộ mã hóa-giải mã kiểu CELP [13, 14, 15], mà cho phép kết hợp bộ lọc Wiener vào bộ lọc của bộ mã hóa-giải mã CELP theo cách mới. Với phương pháp này, nó có thể chứng minh rằng chất lượng khách quan và chủ quan được cải thiện khi so với hệ thống tăng nối tiếp.

Phương pháp được đề cập để nâng cao và mã hóa liên kết tiếng nói, từ đó tránh sự chồng chất các lỗi do xử lý tăng nối tiếp và ngoài ra nâng cao chất lượng đầu ra cảm

quan. Nói cách khác, phương pháp được đề cập tránh sự chong chát các lỗi do xử lý tăng nối tiếp, như tối thiểu hóa liên kết của nhiều âm và biến dạng lượng tử hóa được thực hiện bằng việc lọc Wiener tối ưu trong miền cảm quan.

Fig.1 thể hiện sơ đồ khối dạng giản lược của hệ thống 2 bao gồm bộ mã hóa 4 và bộ giải mã 6. Bộ mã hóa 4 được tạo cấu hình mã hóa tín hiệu âm thanh 8' với tập âm nền được giảm sử dụng mã hóa dự báo tuyến tính. Do đó, bộ mã hóa 4 có thể bao gồm bộ ước lượng tập âm nền 10 được tạo cấu hình để ước lượng phép biểu diễn tập âm nền 12 của tín hiệu âm thanh 8'. Bộ mã hóa có thể còn bao gồm bộ giảm tập âm nền 14 được tạo cấu hình để tạo phép biểu diễn của tín hiệu âm thanh được giảm tập âm nền 16 bằng việc trừ phép biểu diễn của tập âm nền được ước lượng 12 của tín hiệu âm thanh 8' từ phép biểu diễn của tín hiệu âm thanh 8. Do đó, bộ giảm tập âm nền 14 có thể nhận phép biểu diễn của tập âm nền 12 từ bộ ước lượng tập âm nền 10. Đầu vào khác của bộ giảm tập âm nền có thể là tín hiệu âm thanh 8' hoặc phép biểu diễn của tín hiệu âm thanh 8. Một cách tùy ý, bộ giảm tập âm nền và có thể bao gồm bộ tạo được tạo cấu hình để tạo nội tại phép biểu diễn của tín hiệu âm thanh 8, ví dụ tương quan tự động 8 của tín hiệu âm thanh 8'.

Hơn nữa, bộ mã hóa 4 bao gồm bộ dự báo 18 được tạo cấu hình để đưa phép biểu diễn tín hiệu âm thanh 8 trải qua phân tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) 20a và đưa phép biểu diễn của tín hiệu âm thanh được giảm tập âm nền 16 trải qua phân tích dự báo tuyến tính để thu được tập hợp thứ hai gồm các hệ số 20b của bộ lọc dự báo tuyến tính. Tương tự như bộ giảm tập âm nền 14, bộ dự báo 18 có thể bao gồm bộ tạo để tạo nội tại phép biểu diễn của tín hiệu âm thanh 8 từ tín hiệu âm thanh 8'. Tuy nhiên, nó có thể là có lợi khi sử dụng bộ tạo thông thường hoặc trung tâm 17 để tính toán phép biểu diễn 8 của tín hiệu âm thanh 8' một lần và để cung cấp phép biểu diễn tín hiệu âm thanh, ví dụ như tương quan tự động của tín hiệu âm thanh 8', đến bộ giảm tập âm nền 14 và bộ dự báo 18. Do đó, bộ dự báo có thể nhận phép biểu diễn của tín hiệu âm thanh 8 và phép biểu diễn của tín hiệu âm thanh được giảm tập âm nền 16, ví dụ lần lượt là tương quan tự động của tín hiệu âm thanh và tương quan tự động của tín hiệu âm thanh được giảm tập âm nền, và để xác định, dựa trên các tín hiệu quay lại, lần lượt tập hợp thứ nhất gồm các hệ số LPC và tập hợp thứ hai gồm các hệ số LPC.

Nói cách khác, tập hợp thứ nhất gồm các hệ số LPC có thể được xác định từ phép biểu diễn của tín hiệu âm thanh 8 và tập hợp thứ hai gồm các hệ số LPC có thể được xác định từ phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền 16. Bộ dự báo có thể thực hiện thuật toán Levison-Durbin để tính toán tập hợp thứ nhất và thứ hai gồm các hệ số LPC từ tương quan tự động tương ứng.

Hơn nữa, bộ mã hóa bao gồm bộ lọc phân tích 22 bao gồm tầng nối tiếp 24 gồm các bộ lọc miền thời gian 24a, 24b được điều khiển bởi tập hợp thứ nhất thu được gồm các hệ số LPC 20a và tập hợp thứ hai thu được gồm các hệ số LPC 20b. Bộ lọc phân tích có thể áp dụng tầng nối tiếp gồm các bộ lọc miền thời gian, trong đó các hệ số lọc của bộ lọc miền thời gian 24a là tập hợp thứ nhất gồm các hệ số LPC và các hệ số lọc của bộ lọc miền thời gian thứ hai 24b là tập hợp thứ hai gồm các hệ số LPC, vào tín hiệu âm thanh 8' để xác định tín hiệu dư 26. Tín hiệu dư có thể bao gồm các thành phần tín hiệu của tín hiệu âm thanh 8' mà có thể không được biểu diễn bằng bộ lọc tuyến tính có tập hợp thứ nhất và/hoặc tập hợp thứ hai gồm các hệ số LPC.

Theo các phương án, tín hiệu dư có thể được cung cấp tới bộ lượng tử hóa 28 được tạo cấu hình để lượng tử hóa và/hoặc mã hóa tín hiệu dư và/hoặc tập hợp thứ hai gồm các hệ số LPC 24b trước khi truyền dẫn. Ví dụ, bộ lượng tử hóa có thể thực hiện kích thích được mã hóa biến đổi (transform coded excitation - TCX), dự báo tuyến tính được kích thích mã hóa (code excited linear prediction - CELP), hoặc mã hóa không hao hụt ví dụ như mã hóa entropy.

Theo một phương án khác, việc mã hóa tín hiệu dư có thể được thực hiện trong bộ truyền dẫn 30 thay thế cho việc mã hóa trong bộ lượng tử hóa 28. Do đó, bộ truyền dẫn, ví dụ, thực hiện kích thích được mã hóa biến đổi (transform coded excitation - TCX), dự báo tuyến tính được kích thích mã hóa (code excited linear prediction - CELP), hoặc mã hóa không hao hụt ví dụ như mã hóa entropy để mã hóa tín hiệu dư. Hơn nữa, bộ truyền dẫn có thể được tạo cấu hình để truyền dẫn tập hợp thứ hai gồm các hệ số LPC. Bộ nhận tùy ý là bộ giải mã 6. Do đó, bộ truyền dẫn 30 có thể nhận tín hiệu dư 26 hoặc tín hiệu dư được lượng tử hóa 26'. Theo một phương án, bộ truyền dẫn có thể mã hóa tín hiệu dư hoặc tín hiệu dư được lượng tử hóa, ít nhất nếu tín hiệu dư được lượng tử hóa chưa được mã hóa trong bộ lượng tử hóa. Sau khi mã hóa tùy ý tín hiệu dư hoặc tín hiệu

được lượng tử hóa thay thế, tín hiệu tương ứng được cung cấp đến bộ truyền dẫn được truyền dẫn như là tín hiệu dư được mã hóa 32 hoặc như là tín hiệu dư được mã hóa và được lượng tử hóa 32'. Hơn nữa, bộ truyền dẫn có thể nhận tập hợp thứ hai gồm các hệ số LPC 20b', mã hóa tùy ý tương tự, ví dụ với phương pháp mã hóa tương tự như được sử dụng để mã hóa tín hiệu dư, và mã hóa thêm tập hợp thứ hai được mã hóa gồm các hệ số LPC 20b', ví dụ đến bộ giải mã 6, mà không truyền dẫn tập hợp thứ nhất gồm các hệ số LPC. Nói cách khác, tập hợp thứ nhất gồm các hệ số LPC 20a không cần được truyền dẫn.

Bộ giải mã 6 có thể nhận thêm tín hiệu dư được mã hóa 32 hoặc tín hiệu dư được lượng tử hóa được mã hóa 32' và ngoài một trong số các tín hiệu dư 32 hoặc 32' còn có tập hợp thứ hai được mã hóa gồm các hệ số LPC 20b'. Bộ giải mã có thể giải mã các tín hiệu được nhận đơn lẻ và cung cấp tín hiệu dư được giải mã 26 đến bộ lọc tổng hợp. Bộ lọc tổng hợp có thể là nghịch đảo của bộ lọc đáp ứng xung hữu hạn dự báo tuyến tính (linear predictive FIR - finite impulse response) có tập hợp thứ hai gồm các hệ số LPC như là các hệ số của bộ lọc. Nói cách khác, bộ lọc có tập hợp thứ hai gồm các hệ số LPC được nghịch đảo để tạo thành bộ lọc tổng hợp của bộ giải mã 6. Do đó, đầu ra của bộ lọc tổng hợp và đầu ra của bộ giải mã là tín hiệu âm thanh được giải mã 8".

Theo các phương án, bộ ước lượng tạp âm nền có thể ước lượng tương quan tự động 12 của tạp âm nền của tín hiệu âm thanh như là phép biểu diễn của tạp âm nền của tín hiệu âm thanh. Hơn nữa, bộ giảm tạp âm nền có thể tạo phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền 16 bằng việc trừ tương quan tự động của tạp âm nền 12 từ tương quan tự động của tín hiệu âm thanh 8, trong đó tương quan tự động được ước lượng 8 của tín hiệu âm thanh là phép biểu diễn của tín hiệu âm thanh và trong đó phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền 16 là tương quan tự động của tín hiệu âm thanh được giảm tạp âm nền.

Fig.2 và Fig.3 đều đề cập đến cùng một phương án, tuy nhiên sử dụng kí hiệu khác nhau. Do đó, Fig.2 minh họa các phương pháp tăng nối tiếp và nâng cao/mã hóa liên kết, trong đó W_N and W_C lần lượt thể hiện độ trắng của nhiễu âm và các tín hiệu sạch tương ứng và W_N^{-1} và W_C^{-1} là các nghịch đảo tương ứng của chúng. Tuy nhiên, Fig.3 minh họa các phương pháp tăng nối tiếp và nâng cao/mã hóa liên kết trong đó

A_y và A_s lần lượt thể hiện các bộ lọc trắng của nhiễu âm và các tín hiệu sạch và H_y và H_s là các bộ lọc khôi phục (hoặc tổng hợp), các nghịch đảo tương ứng của chúng.

Các Fig.2 và Fig.3 đều thể hiện phần nâng cao và phân mã hóa của chuỗi xử lý tín hiệu vì vậy thực hiện nâng cao và mã hóa tăng nối tiếp. Phần nâng cao 34 có thể vận hành trong miền tần số, trong đó các khối 36a và 36b có thể thực hiện biến đổi tần số thời gian sử dụng, ví dụ MDCT và biến đổi thời gian tần số ví dụ IMDCT hoặc bất kỳ chuyển đổi nào thích hợp để thực hiện chuyển đổi tần số thời gian hoặc thời gian tần số. Các bộ lọc 38 và 40 có thể thực hiện việc giảm tạp âm nền của tín hiệu âm thanh được biến đổi tần số 42. Tại đây, các phần tần số của tạp âm nền có thể được lọc bằng việc giảm tác động của chúng lên phổ tần số của tín hiệu âm thanh 8'. Do đó, bộ chuyển đổi thời gian tần số 36b có thể thực hiện biến đổi nghịch đảo từ miền tần số sang miền thời gian. Sau khi việc giảm tạp âm nền được thực hiện trong phần nâng cao 34, phân mã hóa 35 có thể được thực hiện mã hóa tín hiệu âm thanh với tạp âm nền được giảm. Do đó, bộ lọc phân tích 22' tính toán tín hiệu dư 26'' sử dụng các hệ số LPC thích hợp. Tín hiệu dư có thể được lượng tử hóa và được cung cấp tới bộ lọc tổng hợp 44, mà trong trường hợp Fig.2a và Fig.3a là nghịch đảo của bộ lọc phân tích 22'. Do bộ lọc tổng hợp 42 là nghịch đảo của bộ lọc phân tích 22', trong trường hợp của Fig.2a và Fig.3a, các hệ số LPC được sử dụng để xác định tín hiệu dư 26 được truyền dẫn đến bộ giải mã để xác định tín hiệu âm thanh được giải mã 8".

Fig.2b và Fig.3b thể hiện giai đoạn mã hóa 35 mà không có bước giảm tạp âm nền được thực hiện trước đó. Do bước mã hóa 35 đã được mô tả trên Fig.2a và Fig.3a, phần mô tả thêm được loại bỏ để tránh lặp lại phần đã mô tả.

Fig.2c và Fig.3c đề cập đến khái niệm chính của mã hóa nâng cao liên kết. Nó thể hiện bộ lọc phân tích 22 bao gồm tầng nối tiếp của các bộ lọc miền thời gian sử dụng các bộ lọc A_y và H_s . Cụ thể hơn, tầng nối tiếp của các bộ lọc miền thời gian bao gồm hai lần bộ lọc dự báo tuyến tính sử dụng tập hợp thứ nhất thu được gồm các hệ số LPC 20a (A_y^2) và một lần nghịch đảo của bộ lọc được bao tuyến tính khác sử dụng tập hợp thứ hai được thu gồm các hệ số LPC 20b (H_s). Việc bố trí của bộ lọc hoặc cấu trúc bộ lọc này có thể được gọi là bộ lọc Wiener. Tuy nhiên, lưu ý rằng một bộ lọc dự báo xóa bỏ

$\mathbf{H}_s$ với bộ lọc phân tích $\mathbf{A}_y$. Nói cách khác, có thể áp dụng hai lần bộ lọc $\mathbf{A}_y$ (ký hiệu bởi $\mathbf{A}_y^2$), hai lần bộ lọc $\mathbf{H}_s$ (ký hiệu bởi $\mathbf{H}_s^2$) và một lần bộ lọc $\mathbf{A}_s$.

Như đã được mô tả trên Fig.1, các hệ số LPC đối với các bộ lọc này được xác định, ví dụ, sử dụng tương quan tự động. Do tương quan tự động có thể được thực hiện trong miền thời gian, việc chuyển đổi thời gian - tần số không cần phải thực hiện để thực thi nâng cao và mã hóa liên kết. Hơn nữa, phương pháp này là có lợi do chuỗi xử lý khác của việc truyền dẫn lượng tử hóa việc lọc tổng hợp vẫn giữ nguyên khi so sánh với giai đoạn mã hóa 35 được mô tả trên các Fig.2a và 3a. Tuy nhiên, lưu ý rằng các hệ số của bộ lọc LPC dựa trên tín hiệu được giảm tạp âm nên cần được truyền dẫn đến bộ giải mã để lọc tổng hợp thích hợp. Tuy nhiên, theo phương án khác, thay vì truyền dẫn các hệ số LPC, các hệ số của bộ lọc đã được tính toán của bộ lọc 24b (được biểu diễn bằng nghịch đảo của các hệ số của bộ lọc 20b) có thể được truyền dẫn để tránh sự nghịch đảo khác của bộ lọc tuyến tính đã có các hệ số LPC để suy ra bộ lọc tổng hợp 42, do việc nghịch đảo này đã được thực hiện trong bộ mã hóa. Nói cách khác, thay vì truyền dẫn các hệ số của bộ lọc 20b, ma trận nghịch đảo của các hệ số của bộ lọc này có thể được truyền dẫn, do đó tránh thực hiện việc nghịch đảo hai lần. Hơn nữa, cần lưu ý rằng bộ lọc phía bộ mã hóa 24b và bộ lọc tổng hợp 42 có thể là cùng một bộ lọc, được áp dụng lần lượt trong bộ mã hóa và bộ giải mã.

Nói cách khác đối với Fig.2, các bộ mã hóa-giải mã tiếng nói dựa trên mô hình CELP được dựa trên mô hình tạo tiếng nói mà giả sử rằng tương quan của tín hiệu tiếng nói đầu vào s_n có thể được tạo mô hình bằng bộ lọc dự báo tuyến tính với các hệ số $\mathbf{a} = [\alpha_0, \alpha_1, \dots, \alpha_M]^T$ trong đó M là thứ tự mô hình [16]. Phần dư $r_n = a_n * s_n$, là một phần của tín hiệu tiếng nói mà không thể dự báo bằng bộ lọc dự báo tuyến tính sau đó được lượng tử hóa sử dụng phép lượng tử hóa vectơ.

Để $\mathbf{s}_k = [s_k, s_{k-1}, \dots, s_{k-M}]^T$ là vectơ của tín hiệu đầu vào trong đó chỉ số trên T biểu thị phép hoán vị. Phần dư sau đó có thể được biểu thị như là:

$$r_k = \mathbf{a}^T \mathbf{s}_k. \quad (1)$$

Cho ma trận tương quan tự động $\mathbf{R}_{ss}$ của vectơ tín hiệu tiếng nói $\mathbf{s}_k$

$$\mathbf{R}_{ss} = E\{\mathbf{s}_k \mathbf{s}_k^T\}, \quad (2)$$

ước lượng bộ lọc dự báo của thứ tự M có thể được đưa ra như [20]

$$\mathbf{a} = \sigma_e^2 \mathbf{R}_{ss}^{-1} \mathbf{u}, \quad (3)$$

trong đó $\mathbf{u} = [1, 0, 0, \dots, 0]^T$ là lỗi dự báo vô hướng σ_e^2 được lựa chọn sao cho $\alpha_0 = 1$. Quan sát thấy rằng bộ lọc dự báo tuyến tính α_n là bộ lọc làm trắng, trong đó r_k là nhiễu âm trắng không tương quan. Ngoài ra, tín hiệu ban đầu s_n có thể được khôi phục từ phần dư r_n qua việc lọc IIR với bộ dự báo α_n . Bước tiếp theo là lượng tử hóa các véc tơ của phần dư $\mathbf{r}_k = [r_{kN}, r_{kN-1}, \dots, r_{kN-N+1}]^T$ với bộ lượng tử hóa véc tơ bằng $\tilde{\mathbf{r}}_k$, sao cho biến dạng cảm quan được tối thiểu hóa. Cho véc tơ của tín hiệu đầu ra bằng $\mathbf{s}'_k = [s_{kN}, s_{kN-1}, \dots, s_{kN-N+1}]^T$ và $\tilde{\mathbf{s}}'_k$ là phần tương ứng của nó, và $\mathbf{W}$ là ma trận nhân chập mà áp dụng cho việc gán trọng số cảm quan lên đầu ra. Vấn đề tối ưu hóa cảm quan sau đó có thể được thể hiện như sau:

$$\min_{\mathbf{r}_k} \|\mathbf{W}(\mathbf{s}'_k - \tilde{\mathbf{s}}'_k)\|^2 = \min_{\mathbf{r}_k} \|\mathbf{W}\mathbf{H}(\mathbf{r}_k - \tilde{\mathbf{r}}_k)\|^2, \quad (4)$$

trong đó $\mathbf{H}$ là ma trận nhân chập tương ứng với đáp ứng xung của bộ dự báo α_n .

Việc xử lý của mã hóa tiếng nói kiểu CELP được mô tả trên Fig.2b. Tín hiệu đầu vào trước tiên được làm trắng với bộ lọc $A(z) = \sum_{m=0}^M \alpha_m z^{-m}$ để thu tín hiệu dư. Các véc tơ của phần dư sau đó được lượng tử hóa trong khối Q. Cuối cùng, kết cấu đường bao quang phổ sau đó được khôi phục bởi việc lọc IIR, $A^{-1}(z)$ để thu được tín hiệu đầu ra được lượng tử hóa $\tilde{s}_k$. Do tín hiệu được tổng hợp lại được đánh giá trong miền cảm quan, phương pháp này được biết đến như là phương pháp phân tích bằng tổng hợp.

Lọc Wiener

Trong việc nâng cao tiếng nói kênh đơn, giả sử rằng tín hiệu y_n thu được, là hỗn hợp bổ sung của tín hiệu tiếng nói sạch được mong muốn s_n và một số nhiễu âm không mong muốn v_n , tức là

$$y_n = s_n + v_n. \quad (5)$$

Mục đích của việc xử lý nâng cao là để ước lượng tín hiệu tiếng nói sạch s_n , trong khi chỉ có thể truy cập đến tín hiệu nhiễu âm y_n và các lượng của các ma trận tương quan

$$\mathbf{R}_{ss} = E\{\mathbf{s}_k \mathbf{s}_k^T\} \quad \text{và} \quad \mathbf{R}_{yy} = E\{\mathbf{y}_k \mathbf{y}_k^T\} \quad (6)$$

Trong đó $\mathbf{y}_k = [y_k, y_{k-1}, \dots, y_{k-M}]^T$. Sử dụng ma trận bộ lọc $\mathbf{H}$, việc ước lượng của tín hiệu tín hiệu sạch $\hat{s}_n$ được xác định bằng

$$\hat{s}_k = \mathbf{H}\mathbf{y}_k. \quad (7)$$

Bộ lọc tối ưu trong chiều lỗi bình phương trung bình tối thiểu (minimum mean square error - MMSE) được biết đến như là bộ lọc Wiener có thể được suy ra một cách dễ dàng như là [12]

$$\mathbf{H} = \mathbf{R}_{ss}\mathbf{R}_{yy}^{-1}. \quad (8)$$

Thông thường, bộ lọc Wiener được áp dụng lên các cửa sổ chồng lấp của tín hiệu đầu vào và được khôi phục sử dụng phương pháp chồng lấp-cộng [21,12]. Phương pháp này được minh họa trong khối-Nâng cao của Fig.2a. Tuy nhiên, điều này làm tăng độ trễ thuật toán, tương ứng với độ dài của phần chồng lấp giữa các cửa sổ. Để tránh độ trễ này, mục tiêu là kết hợp việc lọc Wiener với phương pháp dựa trên dự báo tuyến tính.

Để đạt được kết nối này, tín hiệu tiếng nói được ước lượng $\hat{s}_k$ được thay vào phương trình 1, trong đó

$$\begin{aligned} r_k &= \mathbf{a}^T \hat{s}_k = \mathbf{a}^T \mathbf{H}\mathbf{y}_k = \sigma_e^2 \mathbf{u}^T \mathbf{R}_{ss}^{-1} \mathbf{R}_{ss} \mathbf{R}_{yy}^{-1} \mathbf{y}_k \\ &= \sigma_e^2 \mathbf{u}^T \mathbf{R}_{yy}^{-1} \mathbf{y}_k = \gamma \mathbf{a}'^T \mathbf{y}_k \end{aligned} \quad (9)$$

trong đó là hệ số định tỷ lệ và

$$\mathbf{a}' = \sigma_e^2 \mathbf{R}_{yy}^{-1} \mathbf{u} \quad (10)$$

là bộ dự báo tối ưu cho tín hiệu nhiễu âm y_n . Nói cách khác bằng việc lọc tín hiệu nhiễu âm với phần dư $\mathbf{a}'$ (được định tỷ lệ) của tín hiệu sạch được ước lượng được thu được. Việc định tỷ lệ là tỷ lệ giữa tỷ lệ giữa các lỗi phần dư được ước tính lần lượt của các tín hiệu sạch và các tín hiệu nhiễu âm, σ_e^2 và $\hat{\sigma}_e^2$, tức là, $\gamma = \sigma_e^2 / \hat{\sigma}_e^2$. Do đó, việc suy ra này thể hiện rằng việc lọc Wiener và dự báo tuyến tính là các phương pháp liên quan chặt chẽ và trong những phần sau đây, sự liên quan này sẽ được sử dụng để phát triển phương pháp nâng cao và mã hóa.

Kết hợp lọc Wiener vào bộ mã hóa-giải mã CELP

Mục đích là để hợp nhất việc lọc Wiener và các bộ mã hóa-giải mã CELP (được

mô tả trong phần 3 và phần 2) thành thuật toán liên kết. Bằng việc hợp nhất các thuật toán này độ trễ của việc tạo cửa sổ chồng lắp-cộng được yêu cầu bằng việc thực thi thông thường việc lọc Wiener có thể được tránh, và giảm độ phức tạp tính toán.

Việc thực thi kết cấu liên kết sau đó là đơn giản. Phần dư của tín hiệu tiếng nói được nâng cao có thể được thu được bởi phương trình 9. Do đó, tín hiệu tiếng nói được nâng cao có thể được khôi phục bởi lọc IIR phần dư với mô hình dự báo tuyến tính của tín hiệu sạch.

Đối với việc lượng tử hóa phần dư, phương trình 4 có thể được cải biên bằng việc thay thế tín hiệu sạch $\tilde{s}'_k$ với tín hiệu được ước lượng để thu được

$$\min_{\mathbf{r}_k} \|\mathbf{W}(\tilde{s}'_k - \tilde{s}'_k)\|^2 = \min_{\mathbf{r}_k} \|\mathbf{W}\mathbf{H}(\mathbf{r}_k - \tilde{\mathbf{r}}_k)\|^2 \quad (11)$$

Nói cách khác, chức năng mục tiêu với tín hiệu đích được nâng cao $\tilde{s}'_k$ vẫn tương tự như nếu truy cập vào tín hiệu sạch s'_k .

Tóm lại, cải biên duy nhất đối với tiêu chuẩn CELP là thay thế bộ lọc phân tích của tín hiệu sạch $\mathbf{a}$ với bộ lọc phân tích của tín hiệu nhiễu âm $\mathbf{a}'$. Các phần còn lại của thuật toán CELP giữ nguyên không đổi. Phương pháp được đề cập được minh họa trên Fig.2(c).

Rõ ràng rằng phương pháp được đề cập có thể được áp dụng trong các bộ mã hóa-giải mã CELP với các thay đổi tối thiểu bất cứ khi nào việc suy yếu nhiễu âm được mong muốn và khi truy cập vào ước lượng của tương quan tự động của tín hiệu tiếng nói sạch $\mathbf{R}_{ss}$. Nếu ước lượng của tương quan tự động của tín hiệu tiếng nói sạch không có sẵn, có thể được ước lượng sử dụng việc ước lượng tương quan tự động tín hiệu nhiễu âm $\mathbf{R}_{vv}$, bằng $\mathbf{R}_{ss} \approx \mathbf{R}_{yy} - \mathbf{R}_{vv}$ hoặc các ước lượng thông thường khác.

Phương pháp có thể được mở rộng dễ dàng đến các trường hợp như các thuật toán đa kênh với kỹ thuật điều hướng chùm sóng, miễn là ước lượng của tín hiệu sạch có thể thu được sử dụng các bộ lọc miền thời gian.

Lợi thế trong độ phức tạp tính toán của phương pháp được đề cập có thể được mô tả như sau. Lưu ý rằng trong phương pháp thông thường, cần thiết xác định ma trận-bộ lọc $\mathbf{H}$, được đưa ra bởi Phương trình 8. Nghịch đảo ma trận được yêu cầu là độ phức tạp

$\mathcal{O}(M^3)$. Tuy nhiên, trong phương pháp được đề cập chỉ Phương trình 3 được giải đối với tín hiệu nhiễu âm, mà có thể được thực hiện với thuật toán Levinson-Durbin (hoặc tương tự) với độ phức tạp $\mathcal{O}(N^2)$.

Dự báo tuyến tính kích thích mã hóa

Nói cách khác đối với Fig.3, các bộ mã hóa-giải mã tiếng nói dựa trên mẫu CELP sử dụng mô hình tạo tiếng nói mà giả sử rằng tương quan được xác định bởi mô hình ống nằm dưới [16] và do đó đường bao quang phổ của tín hiệu tiếng nói đầu vào s_n có thể được tạo mô hình bằng bộ lọc dự báo tuyến tính với các hệ số $\mathbf{a} = [\alpha_0, \alpha_1, \dots, \alpha_M]^T$ trong đó M là thứ tự mô hình. Phần dư $r_n = a_n * s_n$, một phần của tín hiệu tiếng nói mà không thể dự báo bằng bộ lọc dự báo tuyến tính (cũng được biết đến như là bộ dự báo 18) sau đó được lượng tử hóa sử dụng phép lượng tử hóa véc tơ.

Bộ lọc dự báo tuyến tính $\mathbf{a}_s$ đối với một khung của tín hiệu đầu vào s có thể được thu được, tối thiểu hóa

$$\min_{\mathbf{a}_s} \mathcal{E} \{ \|\mathbf{s}^* \mathbf{a}_s\|^2 - 2\sigma_s^2 (\mathbf{u}^* \mathbf{a}_s - 1) \}, \quad (12)$$

trong đó $\mathbf{u} = [1 \ 0 \ 0 \ \dots \ 0]^T$. Nghiệm như sau:

$$\mathbf{a}_s = \sigma_e^2 \mathbf{R}_{ss}^{-1} \mathbf{u}. \quad (13)$$

Với định nghĩa ma trận nhân chập $\mathbf{A}_s$, bao gồm các hệ số của bộ lọc α của $\mathbf{a}_s$

$$\mathbf{A}_s = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \alpha_1 & \ddots & & \vdots \\ \alpha_2 & \ddots & 1 & \ddots \\ \vdots & \ddots & \alpha_1 & 1 & 0 \\ \alpha_M & \dots & \alpha_2 & \alpha_1 & 1 \end{bmatrix}, \quad (14)$$

tín hiệu dư có thể được thu được bằng việc nhân khung tiếng nói đầu vào với ma trận nhân chập $\mathbf{A}_s$

$$\mathbf{e}_s = \mathbf{A}_s \cdot \mathbf{s}. \quad (15)$$

Việc tạo cửa sổ ở đây được thực hiện như trong các bộ mã hóa-giải mã CELP bằng

việc đáp ứng đầu vào-số không từ tín hiệu đầu vào và đưa nó tổng hợp lại[15].

Phép nhân trong Phương trình 15 là giống với phép nhân chập tín hiệu đầu vào với bộ lọc dự báo, và do đó tương ứng với việc lọc FIR. Tín hiệu ban đầu có thể được khôi phục từ phần dư, bằng phép nhân với bộ lọc khôi phục $\mathbf{H}_s$

$$\mathbf{s} = \mathbf{H}_s \cdot \mathbf{e}_s. \quad (16)$$

trong đó $\mathbf{H}_s$, bao gồm hồi đáp xung $\eta = [1, \eta_1, \dots, \eta_{N-1}]$ của bộ lọc dự báo

$$\mathbf{H}_s = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \eta_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \eta_{N-1} & \dots & \eta_1 & 1 \\ \vdots & & \vdots & \vdots \end{bmatrix} \quad (17)$$

sao cho phép toán này tương ứng với bộ lọc IIR.

Véc tơ phần dư được lượng tử hóa áp dụng cho việc lượng tử hóa véc tơ. Do đó, véc tơ được lượng tử hóa $\hat{\mathbf{e}}_s$ được lựa chọn, tối thiểu hóa khoảng cách cảm quan, trong chiều norm-2, đến tín hiệu sạch được khôi phục được mong muốn:

$$\min_{\hat{\mathbf{e}}_s} \|\mathbf{W}\mathbf{H}(\hat{\mathbf{e}}_s - \mathbf{e}_s)\|^2, \quad (18)$$

trong đó $\mathbf{e}_s$ là phần dư được lượng tử hóa và $\mathbf{W}(z) = A(0,92z)$ là bộ lọc được gán trọng số cảm quan, như được sử dụng trong bộ mã hóa-giải mã tiếng nói AMR-WB [6].

Ứng dụng lọc Wiener vào bộ mã hóa-giải mã CELP

Đối với việc ứng dụng nâng cao tiếng nói kênh đơn, giả sử rằng tín hiệu của ống phóng thanh đạt được y_n , là hỗn hợp bổ sung của tín hiệu tiếng nói sạch được mong muốn s_n và nhiễu âm không mong muốn v_n , sao cho $y_n = s_n + v_n$. Trong miền Z, tương đương với $Y(z) = S(z) + V(z)$.

Bằng việc áp dụng bộ lọc Wiener $B(z)$ có thể khôi phục tín hiệu tiếng nói $S(z)$ từ việc quan sát nhiễu âm $Y(z)$ bằng việc lọc, sao cho tín hiệu tiếng nói được ước lượng bằng $\hat{S}(z) := B(z)Y(z) \approx S(z)$. Nghiệm bình phương trung bình tối thiểu đối với bộ lọc Wiener như sau [12]

$$B(z) = \frac{|S(z)|^2}{|S(z)|^2 + |V(z)|^2}, \quad (19)$$

giả sử rằng lần lượt các tín hiệu tiếng nói và nhiễu âm s_n và v_n là không tương quan.

Trong bộ mã hóa-giải mã tiếng nói, ước lượng của quang phổ năng lượng là có sẵn cho tín hiệu nhiễu âm y_n , dưới dạng hồi đáp xung của mô hình dự báo tuyến tính $|A_y(z)|^{-2}$. Nói cách khác, $|S(z)|^2 + |V(z)|^2 \approx \gamma |A_y(z)|^{-2}$ trong đó γ là hệ số định tỉ lệ. Bộ dự báo tuyến tính có thể được tính toán từ ma trận tương quan tự động $\mathbf{R}_{yy}$ của tín hiệu nhiễu âm như thông thường.

Hơn nữa, có thể ước lượng quang phổ năng lượng của tín hiệu tiếng nói sạch $|S(z)|^2$ hoặc tương đương, ma trận tương quan tự động $\mathbf{R}_{ss}$ của tín hiệu tiếng nói sạch. Các thuật toán nâng cao thường giả sử tín hiệu tiếng nói là cố định, trong đó tương quan tự động của tín hiệu tiếng nói như là $\mathbf{R}_{vv}$ có thể được ước lượng từ khung không có tiếng nói của tín hiệu đầu vào. Ma trận tương quan tự động của tín hiệu tiếng nói sạch $\mathbf{R}_{ss}$ có thể được ước lượng như sau $\hat{\mathbf{R}}_{ss} = \mathbf{R}_{yy} - \mathbf{R}_{vv}$. Ở đây việc phòng ngừa thông thường để đảm bảo rằng $\hat{\mathbf{R}}_{ss}$ là dương là có lợi.

Sử dụng ma trận tương quan tự động đối với tiếng nói sạch $\hat{\mathbf{R}}_{ss}$, bộ dự báo tuyến tính tương ứng có thể được xác định mà hồi đáp xung trong miền Z là $\hat{A}_s^{-1}(z)$. Do đó $|S(z)|^2 \approx |\hat{A}_s(z)|^{-2}$ và phương trình 19 có thể được thể hiện như là

$$B(z) \approx \frac{|\hat{A}_s(z)|^{-2}}{|A_y(z)|^{-2}} = \frac{|A_y(z)|^2}{|\hat{A}_s(z)|^2}. \quad (20)$$

Nói cách khác, bằng việc lọc hai lần với bộ dự báo của các tín hiệu nhiễu âm và sạch lần lượt trong chế độ FIR và IIR, ước lượng Wiener của tín hiệu sạch có thể được thu được.

Các ma trận nhân trận có thể được chỉ ra lần lượt bằng $\hat{A}_s(z)$ và $A_y(z)$ bởi $\mathbf{A}_s$ và $\mathbf{A}_y$ tương ứng với việc lọc FIR với các bộ dự báo. Tương tự, để $\mathbf{H}_s$ và $\mathbf{H}_y$ là ma trận nhân chập tương ứng với lọc dự báo (IIR). Sử dụng các ma trận này, việc mã hóa CELP thông thường có thể được minh họa bằng lưu đồ như trên Fig.3b. Ở đây, có thể lọc tín hiệu

đầu s_n với A_s để thu phần dư, lượng tử hóa nó và khôi phục tín hiệu được lượng tử hóa bằng việc lọc H_s .

Phương pháp thông thường để tổ hợp nâng cao với mã hóa được minh họa trên Fig.3a, trong đó việc lọc Wiener được áp dụng như là khối xử lý trước trước khi mã hóa.

Cuối cùng, trong phương pháp được đề cập việc lọc Wiener được tổ hợp với các bộ mã hóa-giải mã tiếng nói dạng CELP. So với phương pháp tăng nối tiếp từ Fig.3a đến phương pháp liên kết được minh họa trên Fig.3b, rõ ràng rằng sơ đồ cửa sổ tạo cửa sổ chồng lấp cộng (overlap add windowing - OLA) bổ sung có thể được bỏ qua. Hơn nữa, bộ lọc đầu vào A_s tại bộ mã hóa hủy bỏ với H_s . Do đó, như được thể hiện trên Fig.3c, tín hiệu dư sạch $\tilde{e} = A_y^2 H_s y$ được ước lượng theo sau bởi việc lọc tín hiệu đầu vào được giảm y với tổ hợp bộ lọc $A_y^2 H_s$. Do đó, việc tối thiểu hóa lỗi như sau

$$\min_{\tilde{e}} \|W H_s (\hat{e} - \tilde{e})\|^2 \quad (21)$$

Do đó, phương pháp này tối thiểu hóa một cách liên kết khoảng cách giữa ước lượng sạch và tín hiệu được lượng tử hóa, trong đó việc tối thiểu hóa liên kết nhiều âm và nhiều âm lượng tử hóa trong miền cảm quan là có thể thực hiện được.

Hiệu suất của phương pháp mã hóa và nâng cao tiếng nói liên kết được đánh giá bằng cả biện pháp khách quan và chủ quan. Để phân tách hiệu quả của phương pháp mới và bộ mã hóa-giải mã CELP đơn giản được sử dụng, trong đó chỉ tín hiệu dư được lượng tử hóa, nhưng độ trễ và độ khuếch đại của dự báo dài hạn (long term prediction – LTP), mã hóa dự báo tuyến tính (linear predictive coding - LPC) và các thừa số khuếch đại không được lượng tử hóa. Phần dư được lượng tử hóa sử dụng phương pháp lặp theo cặp, trong đó hai xung được bổ sung liên tục bằng việc thử chúng trên mọi vị trí, như được mô tả trong [17]. Hơn nữa, để tránh bất kỳ ảnh hưởng nào của các thuật toán ước lượng, ma trận tương quan của tín hiệu tiếng nói sạch R_{ss} được giả sử là đã biết trong tất cả các trường hợp được mô phỏng. Với việc giả sử rằng tín hiệu tiếng nói và nhiễu âm là không tương quan, nó có giá trị bằng $R_{ss} = R_{yy} - R_{vv}$. Trong bất kỳ ứng dụng thực tiễn nào, ma trận tương quan nhiễu âm R_{vv} hoặc ma trận tương quan tiếng nói sạch R_{ss} phải được ước lượng từ tín hiệu loa phóng thanh thu được. Phương pháp thông thường để ước lượng ma trận tương quan nhiễu âm trong các đoạn dừng tiếng nói, giả sử rằng nhiễu

âm là cố định.

Trường hợp được đánh giá bao gồm hỗn hợp tín hiệu tiếng nói sạch được mong muốn và nhiễu âm bổ sung. Hai loại nhiễu âm có thể được xem xét là nhiễu âm trắng cố định và đoạn gồm các bản ghi nhiễu âm ô tô từ Civilisation Soundscapes Library [18]. Việc lượng tử hóa vectơ của phần dư được thực hiện với tốc độ bit bằng 2,8kbit/giây và 7,2kbit/giây lần lượt tương ứng với toàn bộ tốc độ bit bằng 7,2kbit/giây và và 13,2kbit/giây đối với bộ mã hóa-giải mã AMR-WB [6]. Tốc độ lấy mẫu bằng 12,8kHz được sử dụng cho tất cả mô phỏng.

Các tín hiệu được mã hóa và được nâng cao được đánh giá bằng cả biện pháp khách quan và chủ quan, do đó bài kiểm tra nghe được tiến hành và âm cường độ cảm quan tỉ số tín hiệu trên nhiễu âm (signal-to-noise ratio - SNR) được tính toán, như được xác định trong phương trình 23 và phương trình 22. Độ lớn cảm quan SNR được sử dụng như là xử lý nâng cao liên kết không ảnh hưởng lên pha của các bộ lọc, như là cả bộ lọc tổng hợp và khôi phục đều bị giới hạn bởi hạn chế của các bộ lọc pha tối thiểu, như là mỗi thiết kế của các bộ lọc dự báo.

Với định nghĩa về biến đổi Fourier như là toán tử $\mathcal{F}(\cdot)$, các giá trị quang phổ tuyệt đối của tham chiếu sạch được khôi phục và tín hiệu sạch được ước lượng trong miền cảm quan như sau:

$$S = |\mathcal{F}(\mathbf{W}\mathbf{H}_s\mathbf{e}_k)| \quad \text{và} \quad \hat{S} = |\mathcal{F}(\mathbf{W}\mathbf{H}_s\hat{\mathbf{e}}_k)|. \quad (22)$$

Xác định tỉ số tín hiệu cảm quan được cải biên trên nhiễu âm (modified perceptual signal to noise ratio - PSNR) bằng

$$\text{PSNR}_{\text{ABS}} = 10 \log_{10} \frac{\|S\|^2}{\|\hat{S} - S\|^2}. \quad (23)$$

Đối với việc đánh giá chủ quan, các phần tử tiếng nói được sử dụng từ tập hợp kiểm tra được sử dụng cho việc chuẩn hóa USAC [8], được tạo lỗi bởi nhiễu âm trắng và nhiễu âm ô tô như mô tả ở trên. Bài kiểm tra nghe đa kích thích với tham chiếu ẩn và neo (Multiple Stimuli with Hidden Reference and Anchor - MUSHRA) [19] với 14 người tham gia, sử dụng ống nghe tĩnh điện STAX trong môi trường cách âm. Kết quả của bài kiểm tra nghe được minh họa trong Fig.6 và các điểm số MUSHRA chênh lệch

trên Fig.7 thể hiện trung bình và khoảng tin cậy 95%.

Các kết quả kiểm tra MUSHRA tuyệt đối trên Fig.6 thể hiện rằng tham chiếu ẩn luôn được gán giá trị chính xác bằng 100 điểm. Hỗn hợp nhiều âm ban đầu nhận điểm trung bình thấp nhất cho mỗi mục, biểu thị rằng tất cả các phương pháp nâng cao cải thiện chất lượng cảm quan. Các điểm số trung bình cho tốc độ bit thấp thể hiện sự cải thiện đáng kể về mặt thống kê của các điểm MUSHRA 6,4 đối với trung bình tổng các mục so với phương pháp tăng nối tiếp. Đối với tốc độ bit cao, trung bình tổng các mục thể hiện sự cải thiện, tuy nhiên không đáng kể về mặt thống kê.

Để thu được sự so sánh cụ thể hơn giữa các phương pháp liên kết và nâng cao trước, các điểm MUSHRA chênh lệch được thể hiện trên Fig.7, trong đó chênh lệch giữa các phương pháp nâng cao trước và liên kết được tính toán đối với mỗi người nghe và mỗi mục. Các kết quả chênh lệch xác minh các điểm số MUSHRA tuyệt đối, bằng việc thể hiện sự cải thiện đáng kể về mặt thống kê đối với tốc độ bit thấp hơn, trong đó sự cải thiện đối với tốc độ bit cao hơn không đáng kể về mặt thống kê.

Nói cách khác, phương pháp để nâng cao và mã hóa tiếng nói liên kết được thể hiện, mà cho phép tối thiểu hóa toàn bộ nhiễu âm và nhiễu âm được lượng tử hóa. Trái lại, các phương pháp thông thường áp dụng nâng cao và mã hóa theo các bước xử lý tăng nối tiếp. Việc kết hợp cả hai bước xử lý là hấp dẫn về mặt độ phức tạp tính toán, do việc tạo cửa sổ và vận hành bộ lọc có thể được bỏ qua.

Các bộ mã hóa-giải mã loại CELP nhằm mục đích cung cấp độ trễ rất thấp và do đó tránh được sự chồng lấp của các cửa sổ xử lý lên các cửa sổ xử lý sau này. Trái lại, các phương pháp nâng cao thông thường, được ứng dụng trong miền tần số dựa trên việc tạo cửa sổ chồng lấp cộng, mà đưa ra độ trễ bổ sung tương ứng với độ dài chồng lấp. Phương pháp liên kết không yêu cầu việc tạo cửa sổ chồng lấp cộng, mà sử dụng sơ đồ tạo cửa sổ như được áp dụng trong các bộ mã hóa-giải mã tiếng nói [15], trong đó tránh được sự tăng trong độ trễ thuật toán.

Vấn đề đã biết với phương pháp được đề cập là khác với việc lọc Wiener quang phổ thông thường trong đó pha tín hiệu còn nguyên, phương pháp được đề cập áp dụng các bộ lọc miền thời gian, mà cải biên pha này. Việc cải biên pha này có thể được xử lý dễ dàng bằng việc áp dụng các bộ lọc thông toàn bộ. Tuy nhiên, do không nhận thấy bất

kỳ sự suy giảm cảm quan nào đối với cải biên pha, các bộ lọc thông toàn bộ được bỏ qua để giữ độ phức tạp tính toán thấp. Tuy nhiên, lưu ý rằng trong việc đánh giá khách quan, biên độ cảm quan SNR được đo, để cho phép so sánh công bằng các phương pháp. Biện pháp khách quan thể hiện rằng phương pháp được đề cập trung bình tốt hơn 3dB so với xử lý tăng nối tiếp.

Lợi ích về hiệu suất của phương pháp được đề cập được xác nhận thêm bởi các kết quả của bài kiểm tra nghe MUSHRA, mà thể hiện sự cải thiện trung bình bằng 6,4 điểm. Các kết quả minh họa rằng ứng dụng của mã hóa và nâng cao liên kết là có lợi cho toàn bộ hệ thống về mặt chất lượng và phức tạp tính toán, trong khi vẫn duy trì độ trễ thuật toán thấp của các bộ mã hóa-giải mã tiếng nói CELP.

Fig.8 thể hiện lưu đồ của phương pháp mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính. Phương pháp 800 bao gồm bước S802 ước lượng phép biểu diễn của nhiễu âm nền của tín hiệu âm thanh, bước S804 tạo phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền bằng việc trừ phép biểu diễn của tạp âm nền được ước lượng của tín hiệu âm thanh từ phép biểu diễn của tín hiệu âm thanh, bước S806 đưa phép biểu diễn tín hiệu âm thanh trải qua phép tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính và đưa phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền trải qua phép tích dự báo tuyến tính để thu được tập hợp thứ hai gồm các hệ số của bộ lọc dự báo tuyến tính, và bước S808 điều khiển tăng nối tiếp gồm các bộ lọc miền thời gian bằng bước thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC để thu tín hiệu dư từ tín hiệu âm thanh.

Cần hiểu rằng trong bản mô tả này, các tín hiệu trên các đường thẳng thoải được đặt tên bằng các số chỉ dẫn đối với các đường và thẳng thoải được biểu thị bằng chính các số chỉ dẫn của chúng mà được chỉ định cho các đường. Do đó, ký hiệu sao cho đường có tín hiệu nhất định biểu thị tín hiệu chính nó. Đường có thể là đường vật lý trong thực thi dây cứng. Tuy nhiên, trong việc thực thi bằng máy tính, đường dây vật lý không tồn tại, nhưng tín hiệu được biểu diễn bởi đường được truyền dẫn từ môđun tính toán này đến môđun tính toán khác.

Mặc dù sáng chế được mô tả trong ngữ cảnh gồm các sơ đồ khối trong đó các khối

biểu diễn các thành phần phần cứng thực hoặc logic, sáng chế cũng được thực thi bởi phương pháp thực thi bằng máy tính. Trong các trường hợp gần đây, các khối thể hiện các bước của phương pháp tương ứng trong đó các bước thể hiện các chức năng được thực hiện bởi các khối phần cứng vật lý hoặc logic tương ứng.

Mặc dù một vài khía cạnh đã được mô tả trong ngữ cảnh của thiết bị, rõ ràng rằng các khía cạnh này cũng thể hiện mô tả của phương pháp tương ứng, trong đó khối hoặc thiết bị tương ứng với bước xử lý của phương pháp hoặc đặc điểm của bước xử lý của phương pháp. Tương tự, các khía cạnh được mô tả trong ngữ cảnh của bước phương pháp cũng biểu diễn sự mô tả của khối hoặc mục hoặc dấu hiệu tương ứng của thiết bị tương ứng. Một số hoặc tất cả các bước phương pháp có thể được thực hiện bởi (hoặc sử dụng) thiết bị phần cứng, như ví dụ, bộ vi xử lý, máy tính có thể lập trình hoặc mạch điện tử. Theo một số phương án, một số một hoặc nhiều bước phương pháp quan trọng nhất có thể được thực hiện bởi thiết bị như vậy.

Các tín hiệu âm thanh được mã hóa hoặc được truyền dẫn theo sáng chế có thể được lưu trữ trên vật ghi lưu trữ kỹ thuật số hoặc có thể được truyền trong môi trường truyền dẫn như môi trường truyền không dây hoặc môi trường truyền có dây như Liên mạng.

Phụ thuộc vào các yêu cầu thực hiện nhất định, các phương án của sáng chế có thể được thực hiện trong phần cứng hoặc trong phần mềm. Phương án có thể được thực hiện sử dụng vật ghi lưu trữ số, ví dụ, đĩa mềm, DVD, Blu-Ray, CD, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ FLASH, có các tín hiệu điều khiển có thể đọc được bằng điện tử được lưu trữ trên đó, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính có thể lập trình sao cho phương pháp tương ứng được thực hiện. Do đó, vật ghi lưu trữ số có thể có khả năng đọc được bằng máy tính.

Một số phương án theo sáng chế bao gồm vật mang dữ liệu có các tín hiệu điều khiển có thể đọc được bằng điện tử, mà có khả năng kết hợp với hệ thống máy tính có thể lập trình được, sao cho một trong số các phương pháp được mô tả ở đây được thực hiện.

Nói chung, các phương án của sáng chế có thể được thực hiện như sản phẩm chương trình máy tính với mã chương trình, mã chương trình có tác dụng để thực hiện

một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Mã chương trình có thể, ví dụ, được lưu trữ trên vật mang có thể đọc được bằng máy.

Các phương án khác bao gồm chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây, được lưu trữ trên vật mang có thể đọc được bằng máy.

Do đó, nói cách khác, phương án của phương pháp theo sáng chế là, vật ghi có thể đọc được bằng máy tính bao gồm chương trình máy tính có mã chương trình để thực hiện một trong các phương pháp được mô tả ở đây, khi chương trình máy tính chạy trên máy tính.

Do đó, phương án khác của các phương pháp theo sáng chế là, vật mang dữ liệu (hoặc vật ghi lưu trữ số, hoặc vật ghi có thể đọc được bằng máy tính) gồm có, đã được ghi lại trên đó, chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây. Vật mang dữ liệu, môi trường lưu trữ số hoặc môi trường được ghi thường là hữu hình và/hoặc không chuyển dịch.

Do đó, phương án nữa của các phương pháp theo sáng chế là dòng dữ liệu hoặc chuỗi tín hiệu biểu diễn chương trình máy tính để thực hiện một trong các phương pháp đã được mô tả ở đây. Ví dụ, có thể cấu hình dòng dữ liệu hoặc chuỗi các tín hiệu để được truyền thông qua sự kết nối truyền thông dữ liệu, ví dụ thông qua Liên mạng.

Phương án khác gồm có phương tiện xử lý, ví dụ máy tính, hoặc thiết bị logic lập trình được, được cấu hình để hoặc được làm thích ứng để thực hiện một trong số các phương pháp được mô tả ở đây.

Phương án khác gồm có máy tính có chương trình máy tính được lắp trên máy tính để thực hiện một trong các phương pháp được mô tả ở đây.

Phương án nữa theo sáng chế bao gồm thiết bị hoặc hệ thống được cấu hình để chuyển (ví dụ, bằng điện tử hoặc quang học) chương trình máy tính để thực hiện một trong các phương pháp được mô tả ở đây đến bộ nhận. Bộ nhận có thể là, ví dụ, máy tính, thiết bị di động, thiết bị nhớ hoặc tương tự. Thiết bị hoặc hệ thống có thể, ví dụ, gồm có máy chủ tập tin để chuyển chương trình máy tính đến bộ nhận.

Theo một số phương án, thiết bị logic lập trình được (ví dụ, mảng công lập trình

được dạng trường) có thể được sử dụng để thực hiện một số hoặc tất cả các chức năng của các phương pháp được mô tả ở đây. Trong một số phương án, mảng công lập trình được dạng trường có thể kết hợp với bộ vi xử lý để thực hiện một trong số các phương pháp được mô tả ở đây. Thông thường, các phương pháp tốt hơn là được thực hiện bởi thiết bị phần cứng bất kỳ.

Các phương án được mô tả bên trên chỉ mang tính minh họa cho các nguyên lý của sáng chế. Cần hiểu rằng các sự biến thể và biến đổi của các phương án và các chi tiết được mô tả ở đây sẽ là rõ ràng đối với người có trình độ trong lĩnh vực kỹ thuật tương ứng. Do đó, mục đích của sáng chế là chỉ được giới hạn bởi phạm vi của các điểm yêu cầu bảo hộ sắp đưa ra dưới đây và không bị giới hạn bởi các chi tiết cụ thể được biểu diễn bằng bản mô tả và giải thích của các phương án ở đây.

YÊU CẦU BẢO HỘ

1. Bộ mã hóa mã hóa tín hiệu âm thanh với tập âm nền được giảm sử dụng mã hóa dự báo tuyến tính, bộ mã hóa bao gồm:

bộ ước lượng tập âm nền được tạo cấu hình để ước lượng phép biểu diễn tập âm nền của tín hiệu âm thanh ;

bộ giảm tập âm nền được tạo cấu hình để tạo phép biểu diễn tín hiệu âm thanh được giảm tập âm nền bằng việc trừ phép biểu diễn tập âm nền được ước lượng của tín hiệu âm thanh từ phép biểu diễn tín hiệu âm thanh ;

bộ dự báo được tạo cấu hình để đưa phép biểu diễn tín hiệu âm thanh trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) và đưa phép biểu diễn của tín hiệu âm thanh được giảm tập âm nền trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ hai của các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) (20a); và

bộ lọc phân tích bao gồm tầng nối tiếp gồm các bộ lọc miền thời gian được điều khiển bởi tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC để thu được tín hiệu dư từ tín hiệu âm thanh;

trong đó bộ mã hóa được thực hiện bằng thiết bị phần cứng hoặc sử dụng máy tính hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

2. Bộ mã hóa theo điểm 1, trong đó tầng nối tiếp của các bộ lọc miền thời gian bao gồm bộ lọc dự báo tuyến tính thứ nhất và bộ lọc dự báo tuyến tính thứ nhất sử dụng tập hợp thứ nhất được thu gồm các hệ số LPC và nghịch đảo của bộ lọc dự báo tuyến tính thứ ba sử dụng tập hợp thứ hai được thu gồm các hệ số LPC, hoặc

bộ lọc dự báo tuyến tính sử dụng tập hợp thứ nhất được thu gồm các hệ số LPC và được bình phương, và và nghịch đảo của bộ lọc dự báo tuyến tính khác sử dụng tập hợp thứ hai được thu gồm các hệ số LPC (20b).

3. Bộ mã hóa theo điểm 1, trong đó tầng nối tiếp của các bộ lọc miền thời gian là bộ lọc Wiener.

4. Bộ mã hóa theo điểm 1,

trong đó bộ ước lượng tạp âm nền được tạo cấu hình để ước lượng tương quan tự động của tạp âm nền như phép biểu diễn tạp âm nền (được ước lượng;

trong đó bộ giảm tạp âm nền được tạo cấu hình để tạo phép biểu diễn tín hiệu âm thanh được giảm tạp âm nền bằng việc trừ tương quan tự động của tạp âm nền từ tương quan tự động của tín hiệu âm thanh, trong đó tương quan tự động của tín hiệu âm thanh là phép biểu diễn của tín hiệu âm thanh và trong đó phép biểu diễn tín hiệu âm thanh được giảm tạp âm nền là tương quan tự động của tín hiệu âm thanh được giảm tạp âm nền.

5. Bộ mã hóa theo điểm 1, trong đó phép biểu diễn là tương quan tự động.

6. Bộ mã hóa theo điểm 1, còn bao gồm bộ truyền dẫn được tạo cấu hình để truyền dẫn tập hợp thứ hai của các hệ số LPC (20b).

7. Bộ mã hóa theo điểm 1, còn bao gồm bộ truyền dẫn được tạo cấu hình để truyền dẫn tín hiệu dư.

8. Bộ mã hóa theo điểm 1, còn bao gồm bộ lượng tử hóa được tạo cấu hình để lượng tử hóa và/hoặc mã hóa tín hiệu dư trước khi truyền dẫn.

9. Bộ mã hóa theo điểm 1, còn bao gồm bộ lượng tử hóa được tạo cấu hình để lượng tử hóa và/hoặc mã hóa tập hợp thứ hai gồm các hệ số LPC trước khi truyền dẫn.

10. Bộ mã hóa theo điểm 8, trong đó bộ lượng tử hóa được tạo cấu hình để sử dụng dự báo tuyến tính kích thích mã hóa (code-excited linear prediction - CELP), mã hóa entropy, hoặc kích thích được mã hóa biến đổi (transform coded excitation - TCX).

11. Hệ thống mã hóa và giải mã bao gồm:

bộ mã hóa để mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính, bộ mã hóa này bao gồm:

bộ ước lượng tạp âm nền được tạo cấu hình để ước lượng phép biểu diễn tạp âm nền của tín hiệu âm thanh;

bộ giảm tạp âm nền được tạo cấu hình để tạo phép biểu diễn tín hiệu âm thanh được giảm tạp âm nền bằng việc trừ phép biểu diễn tạp âm nền được ước lượng của tín

hiệu âm thanh từ phép biểu diễn tín hiệu âm thanh;

bộ dự báo được tạo cấu hình để đưa phép biểu diễn tín hiệu âm thanh trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC) và đưa phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ hai của các hệ số của bộ lọc dự báo tuyến tính (linear prediction filter - LPC); và

bộ lọc phân tích bao gồm tầng nối tiếp gồm các bộ lọc miền thời gian được điều khiển bởi tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC để thu được tín hiệu dư từ tín hiệu âm thanh;

bộ giải mã được tạo cấu hình để giải mã tín hiệu âm thanh được mã hóa,

trong đó mỗi bộ mã hóa và bộ giải mã được thực hiện bằng thiết bị phần cứng hoặc sử dụng máy tính hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

12. Phương pháp mã hóa tín hiệu âm thanh với tạp âm nền được giảm sử dụng mã hóa dự báo tuyến tính, phương pháp bao gồm:

ước lượng phép biểu diễn tạp âm nền của tín hiệu âm thanh;

tạo phép biểu diễn tín hiệu âm thanh được giảm tạp âm nền bằng việc trừ phép biểu diễn của tạp âm nền được ước lượng của tín hiệu âm thanh từ phép biểu diễn của tín hiệu âm thanh;

đưa phép biểu diễn của tín hiệu âm thanh trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (LPC) và đưa phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ hai gồm các hệ số của bộ lọc dự báo tuyến tính (LPC); và

điều khiển tầng nối tiếp gồm các bộ lọc miền thời gian bằng tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC để thu tín hiệu dư từ tín hiệu âm thanh.

13. Vật ghi lưu trữ kỹ thuật số không chuyên tiếp chứa chương trình máy tính được

lưu trữ trên đó để thực hiện phương pháp mã hóa tín hiệu âm thanh với tập âm nền được giảm sử dụng mã hóa dự báo tuyến tính, phương pháp này bao gồm:

ước lượng phép biểu diễn tập âm nền của tín hiệu âm thanh;

tạo phép biểu diễn tín hiệu âm thanh được giảm tạp âm nền bằng việc trừ phép biểu diễn của tập âm nền được ước lượng của tín hiệu âm thanh từ phép biểu diễn của tín hiệu âm thanh;

đưa phép biểu diễn của tín hiệu âm thanh trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ nhất gồm các hệ số của bộ lọc dự báo tuyến tính (LPC) và đưa phép biểu diễn của tín hiệu âm thanh được giảm tạp âm nền trải qua phép phân tích dự báo tuyến tính để thu được tập hợp thứ hai của các hệ số của bộ lọc dự báo tuyến tính (LPC); và

điều khiển tầng nối tiếp gồm các bộ lọc miền thời gian bằng tập hợp thứ nhất được thu gồm các hệ số LPC và tập hợp thứ hai được thu gồm các hệ số LPC để thu tín hiệu dư từ tín hiệu âm thanh,

trong đó chương trình máy tính đã nêu được chạy bởi máy tính.

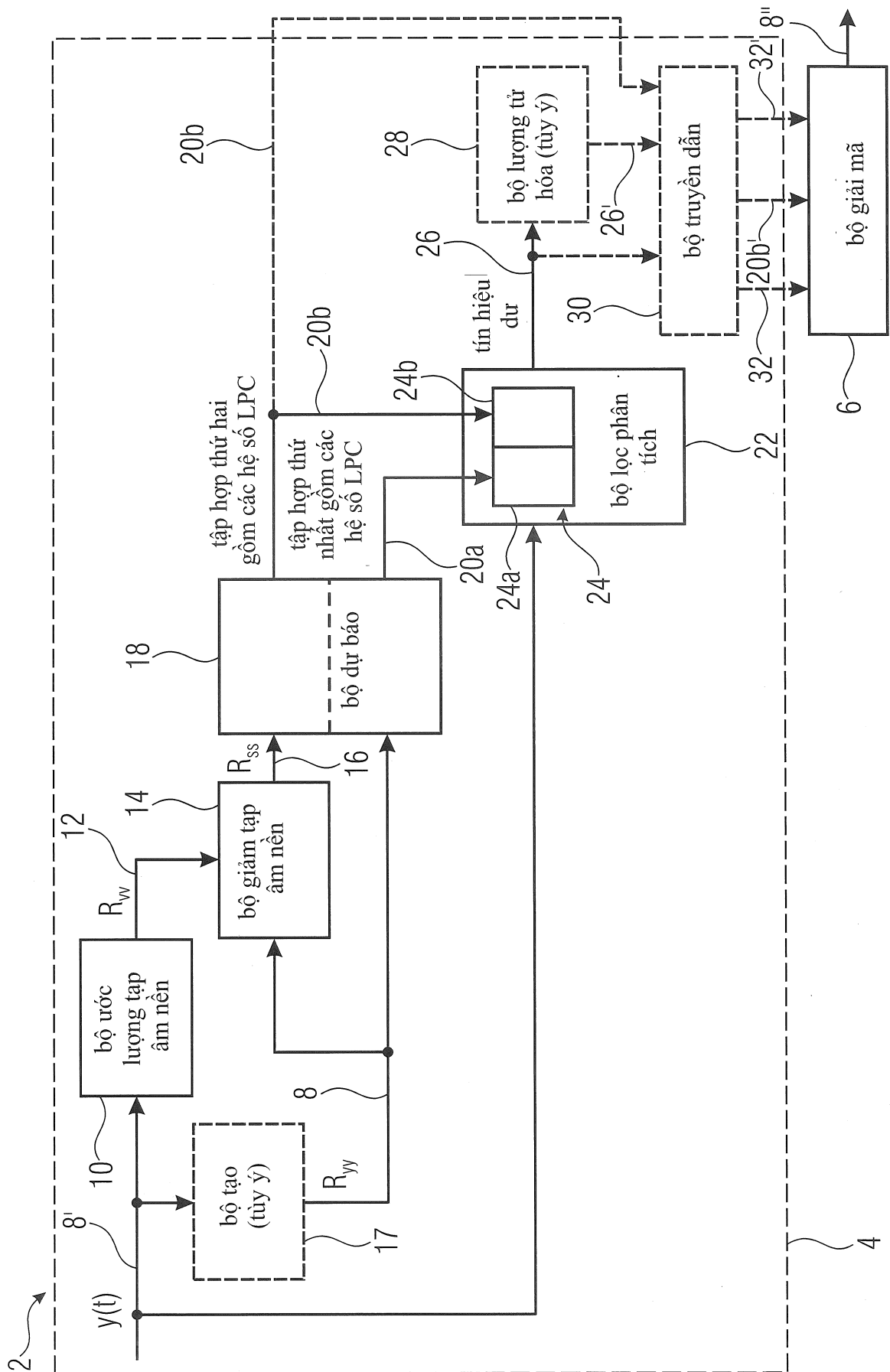
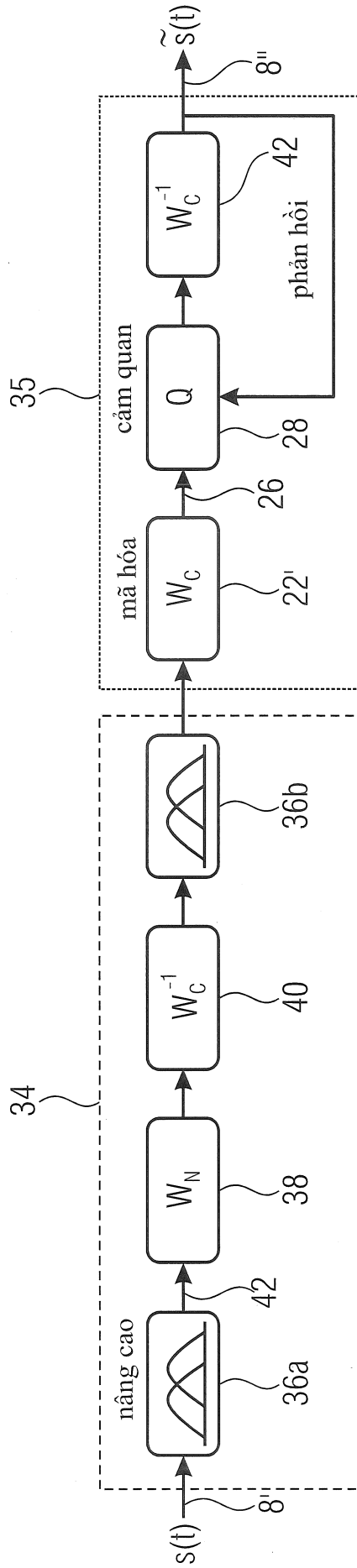
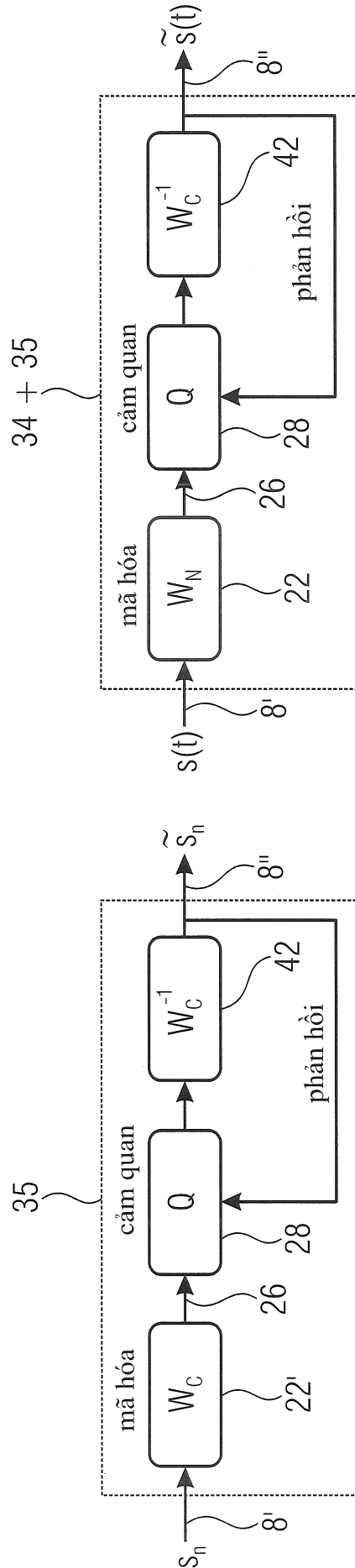


Fig. 1



nâng cao và mã hóa tầng nối tiếp

Fig. 2a

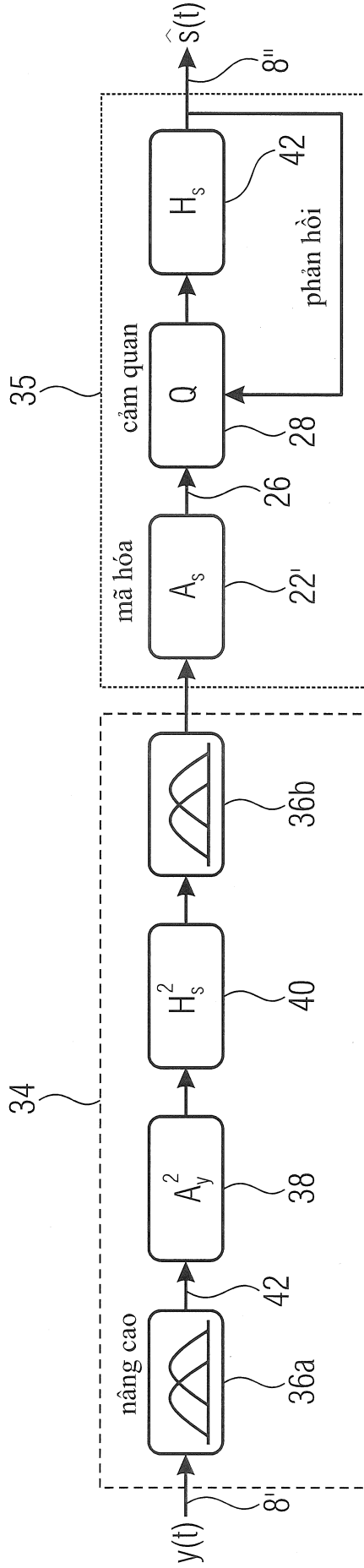


mã hóa tiếng nói CELP

Fig. 2b

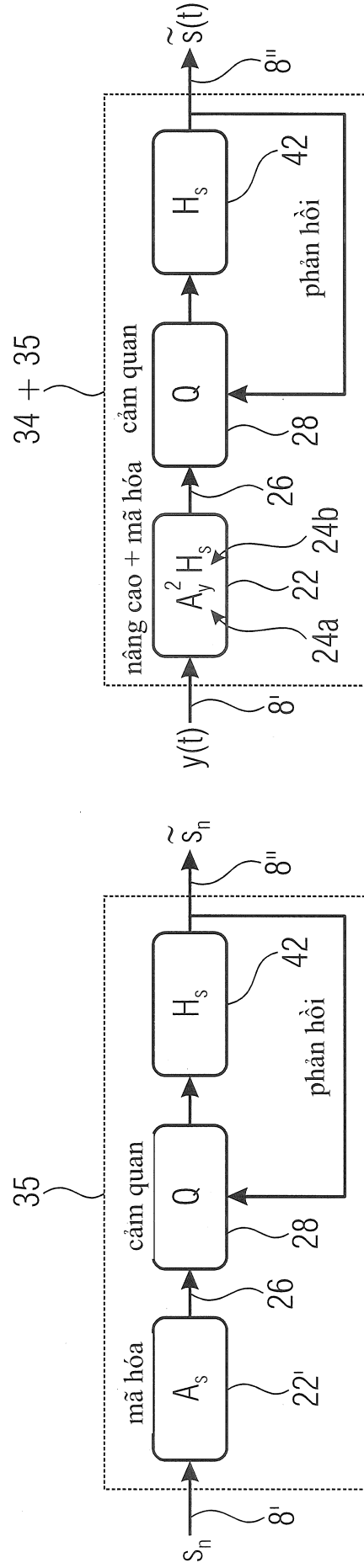
nâng cao và mã hóa liên kết

Fig. 2c



nâng cao và mã hóa tăng nối tiếp

Fig. 3a



mã hóa tiếng nói CELP

Fig. 3b

nâng cao và mã hóa liên kết

Fig. 3c

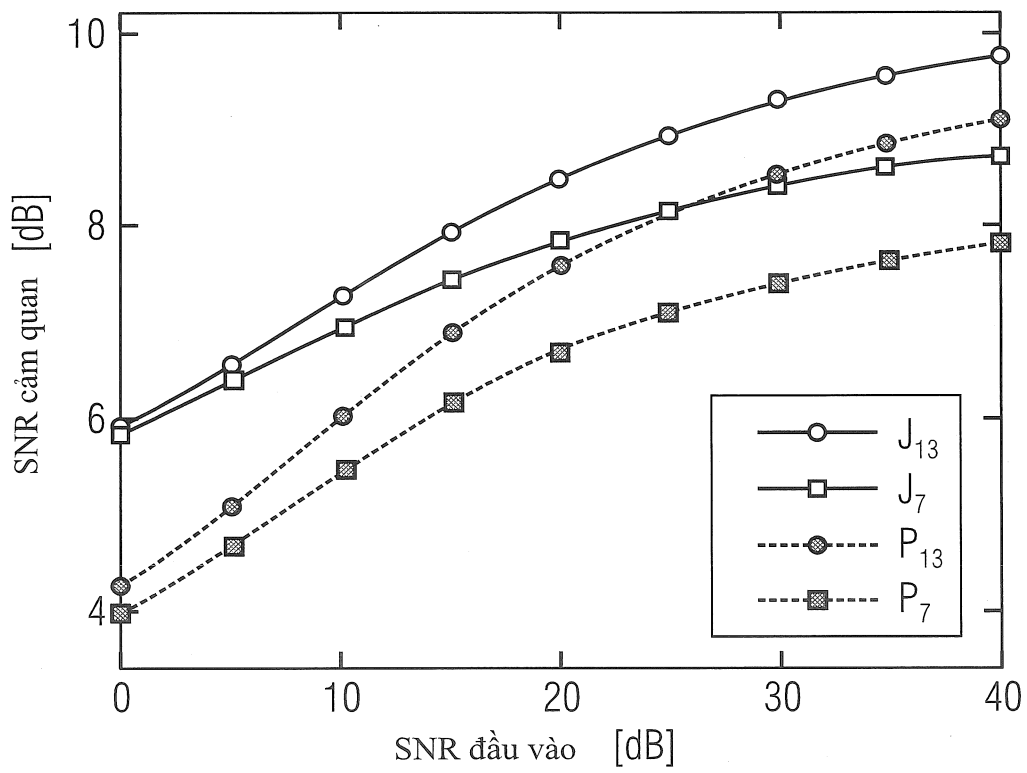


Fig. 4

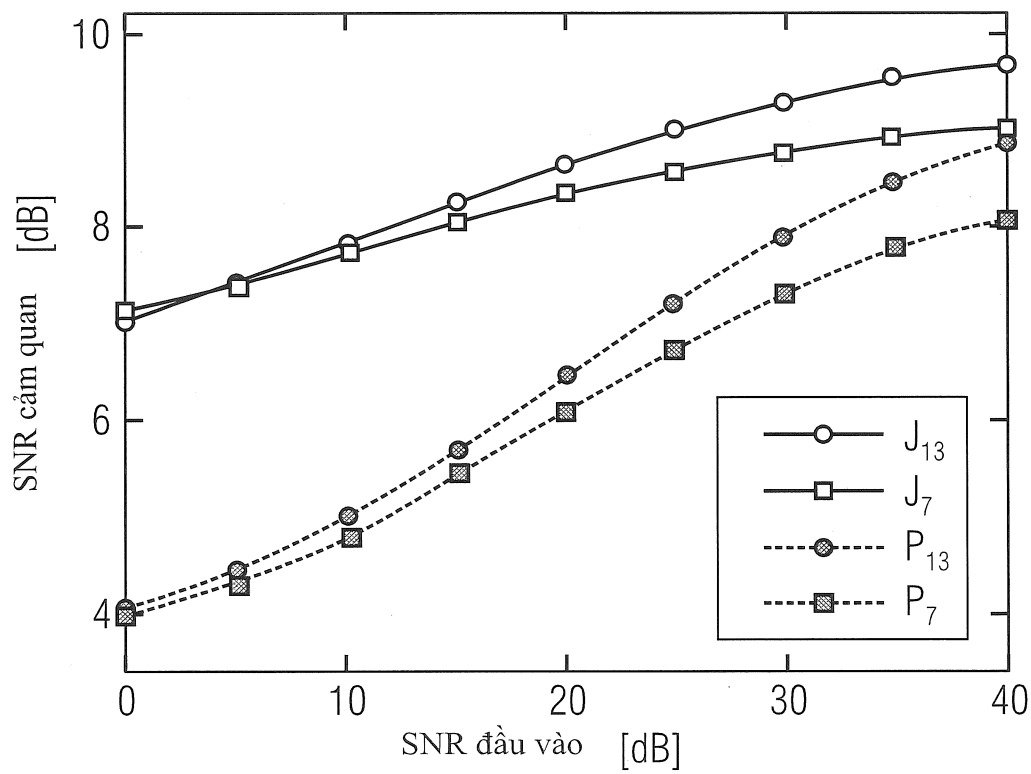


Fig. 5

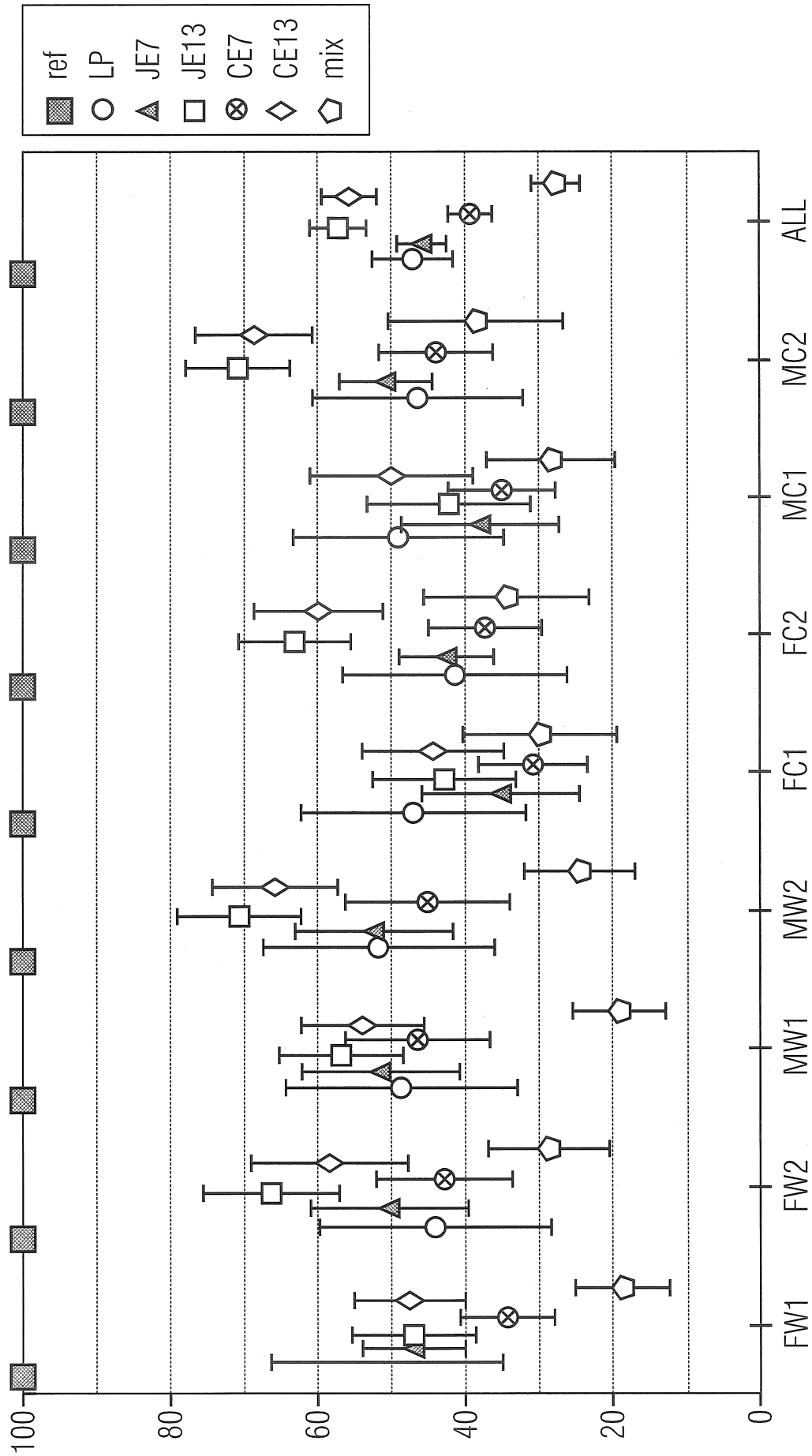
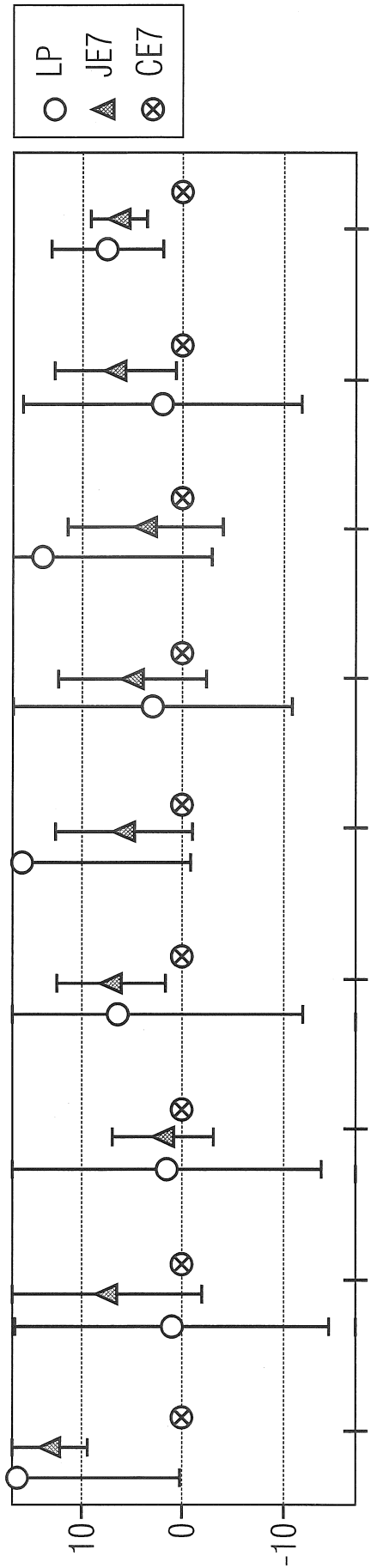
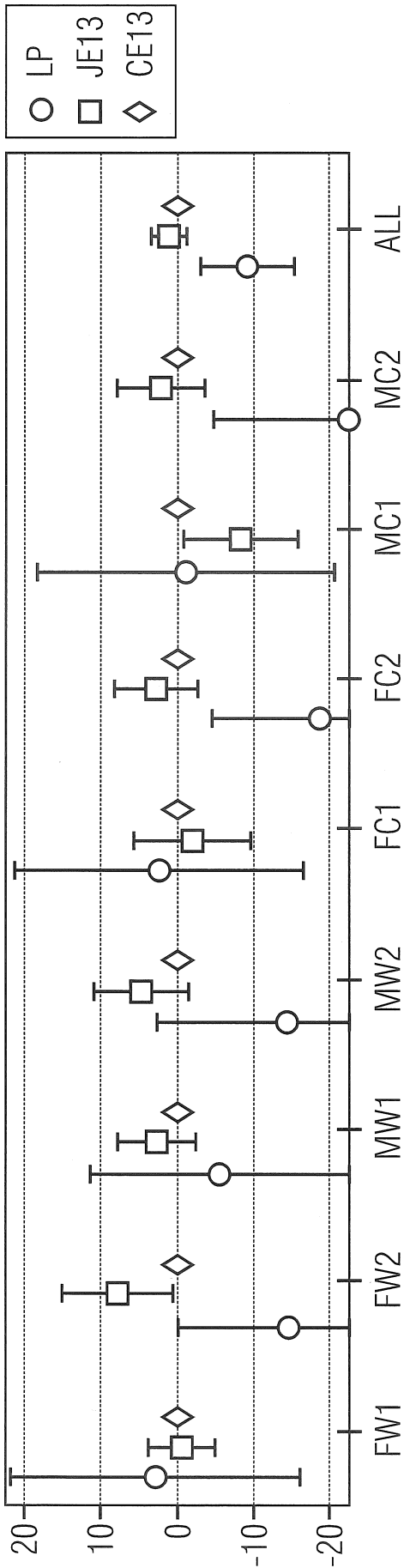


Fig. 6



Kết quả đối với 7,2kHz

Fig. 7a



Kết quả đối với 13,2kHz

Fig. 7b

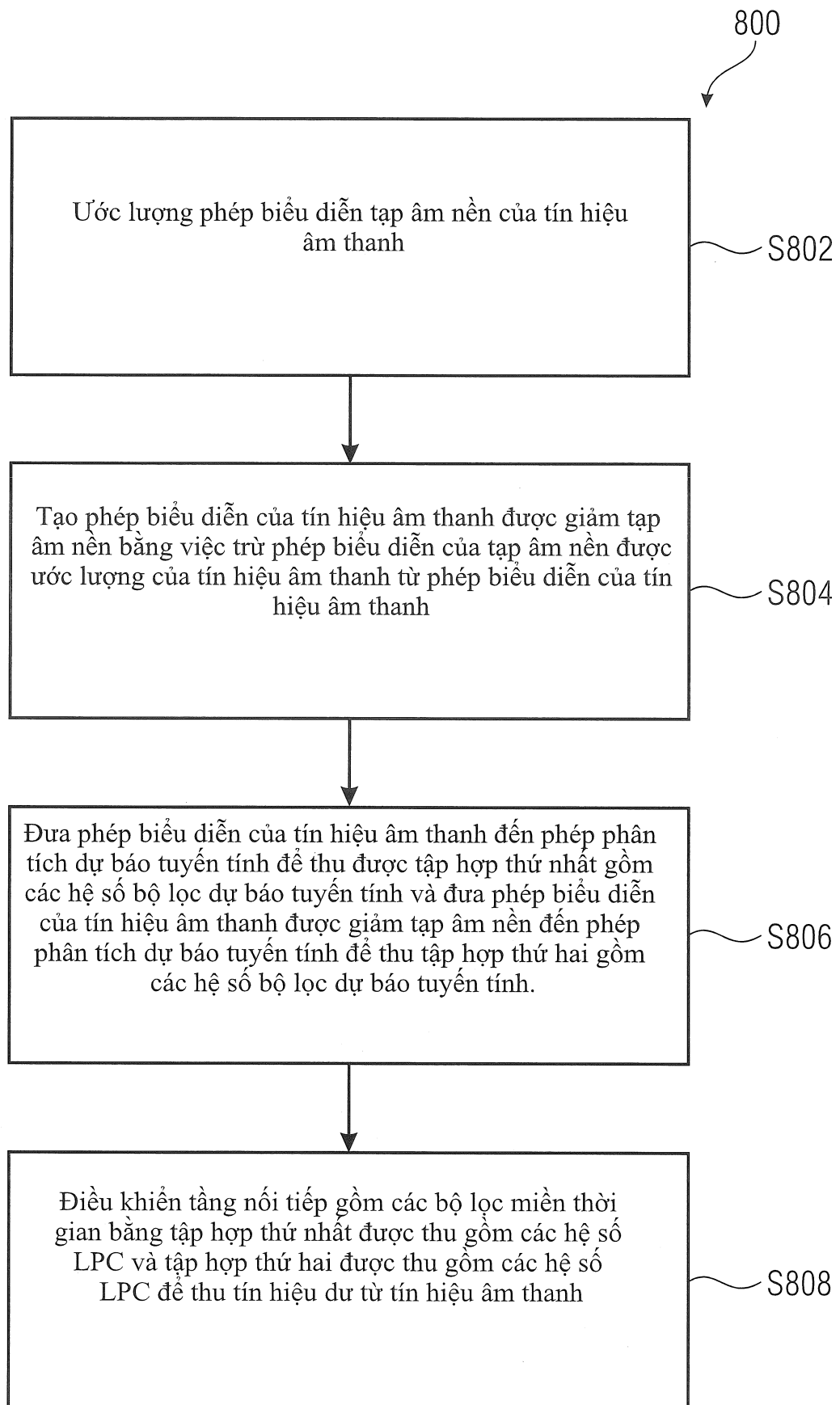


Fig. 8