



(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)



1-0042101

(51)⁷ H04R 3/02; H04M 9/08

(13) B

(21) 1-2019-06106

(22) 19/04/2018

(86) PCT/EP2018/060006 19/04/2018

(87) WO 2018/193028 25/10/2018

(30) 17167304.9 20/04/2017 EP; 17196416.6 13/10/2017 EP

(45) 25/12/2024 441

(43) 30/01/2020 382A

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

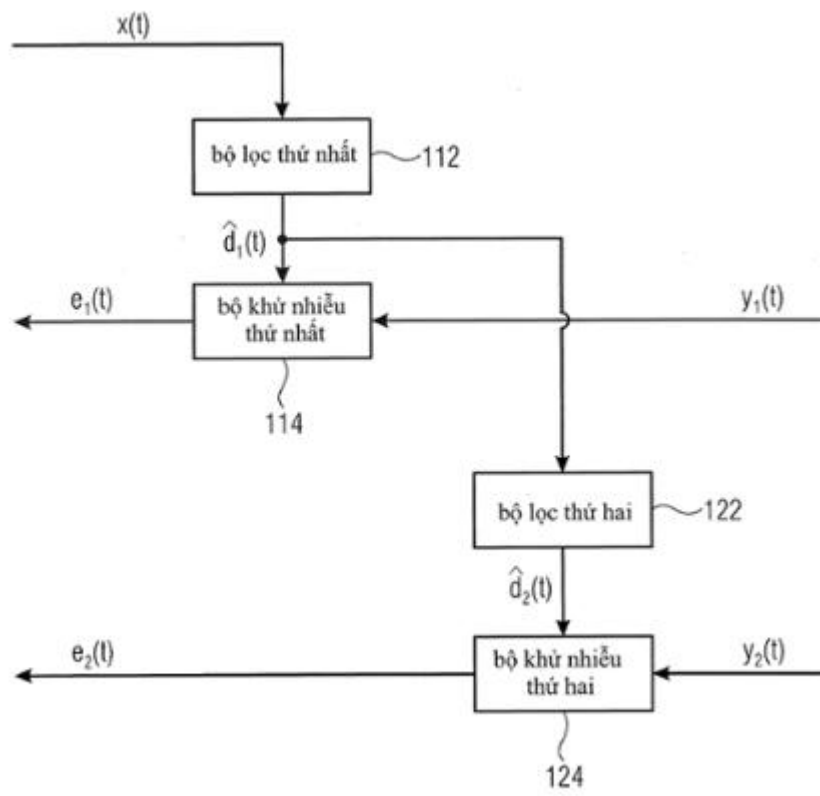
Hansastrasse 27c, 80686 Muenchen, Germany

(72) LUIS VALERO, Maria (ES); HABETS, Emanuel (NL); ANNIBALE, Paolo (IT);
LOMBARD, Anthony (FR); WILD, Moritz (DE); RUTHA, Marcel (DE).

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

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP KHỬ NHIỀU ĐA KÊNH

(57) Sáng chế đề cập đến thiết bị và phương pháp khử nhiễu đa kênh. Thiết bị khử nhiễu đa kênh trong tín hiệu âm thanh nhận được bao gồm hai hoặc nhiều hơn hai kênh âm thanh nhận được để thu được tín hiệu âm thanh được biến đổi bao gồm hai hoặc nhiều hơn hai kênh âm thanh được biến đổi. Thiết bị bao gồm bộ lọc thứ nhất (112) được tạo cấu hình để tạo ước lượng thứ nhất của tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu. Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ nhất (114) được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất. Hơn nữa, thiết bị bao gồm một bộ lọc thứ hai (122) được tạo cấu hình để tạo ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất. Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ hai (124) được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai của tín hiệu nhiễu thứ hai.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc xử lý tín hiệu âm thanh, và, cụ thể là đề cập đến thiết bị và phương pháp làm giảm độ phức tạp của việc khử nhiễu đa kênh và để khử nhiễu đa kênh có độ phức tạp thấp.

Tình trạng kỹ thuật của sáng chế

Các thiết bị truyền thông rảnh tay hiện đại sử dụng nhiều tín hiệu micrô, ví dụ, để tăng cường giọng nói, suy luận hình dạng phòng hoặc nhận dạng giọng nói tự động. Các thiết bị này cùng loại với các trợ lý kích hoạt bằng giọng nói, các thiết bị nhà thông minh và các loa thông minh, cho đến các điện thoại thông minh, các máy tính bảng hoặc các máy tính cá nhân. Nhiều thiết bị thông minh, chẳng hạn như các trợ lý kích hoạt bằng giọng nói, các điện thoại thông minh, các máy tính bảng hoặc các máy tính cá nhân, được trang bị các loa. Xem xét các thiết bị như vậy, ví dụ, thiết bị mà cũng tích hợp ít nhất một loa, bộ khử nhiễu âm thanh được bố trí cho mỗi đầu ra của micrô để giảm sự ghép điện thanh.

Khử tiếng vang âm thanh (Acoustic echo cancellation - AEC) (ví dụ, xem tài liệu [1]) là kỹ thuật được sử dụng rộng rãi nhất để làm giảm sự ghép điện thanh giữa (các) loa và (các) micrô trong các thiết lập truyền thông rảnh tay. Với thiết lập như vậy, các micrô thu được, ngoài tiếng nói gần cuối mong muốn, còn có các tiếng vang âm thanh và tiếng ồn nền. AEC sử dụng các kỹ thuật lọc thích ứng (ví dụ, xem tài liệu [2]) để ước lượng các đáp ứng xung âm (acoustic impulse response - AIR) giữa (các) loa và (các) micrô. Sau đó, các ước lượng tiếng vang âm thanh được tính toán bằng cách lọc tín hiệu loa có sẵn với các AIR ước lượng. Cuối cùng, các tiếng vang âm thanh được ước lượng được loại trừ khỏi các tín hiệu micrô, dẫn đến việc khử các tiếng vang âm thanh.

Trong trường hợp cụ thể của việc khử tiếng vang âm thanh (AEC), sự ghép điện thanh gây ra bởi tín hiệu loa ở xa được tái tạo bởi loa. Tuy nhiên, trong các thiết bị truyền thông rảnh tay đã nói ở trên, sự ghép điện thanh cũng có thể gây ra bởi phản hồi, âm nhạc hoặc trợ lý giọng nói của chính thiết bị. Giải pháp đơn giản nhất để giảm

sự ghép điện thanh giữa loa và micrô là đặt bộ khử nhiễu âm thanh ở đầu ra của mỗi micrô (ví dụ, xem tài liệu [3]).

Các hàm chuyển tương đối mô hình mối quan hệ giữa các AIR miền tần số, thường được ký hiệu là các hàm chuyển âm (acoustic transfer function - ATF). Các RTF (RTF - relative transfer function có nghĩa là hàm chuyển tương đối) thường được sử dụng trong bối cảnh tăng cường giọng nói đa micrô (ví dụ, xem các tài liệu [5], [8], [12]). Xem xét thêm các ứng dụng liên quan, việc ước lượng các hàm chuyển tương đối tiếng vang dư được sử dụng trong các tài liệu [13], [14] để ước lượng mật độ phổ công suất của tiếng vang dư, ví dụ, các thành phần tiếng vang còn lại sau khi khử, của kênh chính. Để tăng cường quá trình ước lượng, tín hiệu micrô thứ hai được sử dụng. Phương pháp được đề xuất trong các tài liệu [13], [14] ước lượng mối quan hệ giữa tín hiệu chính sau khi khử và tín hiệu micrô thứ cấp, cung cấp mối quan hệ giữa lỗi trong việc ước lượng AIR sơ cấp và AIR thứ cấp. Cuối cùng, hàm chuyển tương đối tiếng vang dư được sử dụng để tính toán mật độ phổ công suất của tiếng vang âm thanh dư sơ cấp.

Xem xét ứng dụng cụ thể của việc xử lý mảng micrô, một vài phương pháp đã được trình bày nhằm mục đích giảm độ phức tạp của toàn bộ thuật toán tăng cường giọng nói, ví dụ, lọc không gian kết hợp với AEC. Ví dụ, việc sử dụng một AEC duy nhất được đặt ở đầu ra của bộ lọc theo không gian được nghiên cứu đầu tiên trong các tài liệu [3], [15]. Một số phương pháp thay thế nhằm tích hợp khử tiếng vang âm thanh và xử lý mảng micrô đã được đề xuất trong các tài liệu [8], [16], [18].

Do độ phức tạp của bộ khử nhiễu âm thanh đa micrô tỷ lệ thuận với số lượng micrô, đối với nhiều thiết bị hiện đại, không thể đạt được mức tăng độ phức tạp như vậy.

Do đó sẽ được đánh giá cao hơn nếu có các giải pháp có độ phức tạp thấp để khử nhiễu đa kênh.

Tài liệu tham khảo

[1] E. Hänsler and G. Schmidt, "Acoustic Echo and Noise Control: A practical Approach" New Jersey, USA: Wiley, 2004.

- [2] S. Haykin, "Adaptive Filter Theory", 4th ed. New Jersey, USA: Prentice-Hall, 2001.
- [3] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays", in Proc. IEEE ICASSP, Munich, Germany, Apr. 1997, pp. 219-222.
- [4] O. Shalvi and E. Weinstein, "System identification using nonstationary signals", IEEE Trans. Signal Process., vol. 44, no. 8, pp. 2055-2063, 1996.
- [5] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans. Signal Process., vol. 49, no. 8, pp. 1614-1626, Aug. 2001.
- [6] I. Cohen, "Relative transfer function identification using speech signals," IEEE Trans. Speech Audio Process., vol. 12, no. 5, pp. 451-459, Sep. 2004.
- [7] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 4, pp. 546-555, May 2009.
- [8] G. Reuven, S. Gannot, and I. Cohen, "Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller," Speech Communication, vol. 49, no. 7-8, pp. 623-635, Aug. 2007.
- [9] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 7, pp. 1420-1434, Sep. 2009.
- [10] T. Dvorkind and S. Gannot, "Speaker localization in a reverberant environment," in Proc. the 22nd convention of Electrical and Electronics Engineers in Israel (IEEEI), Tel-Aviv, Israel, Dec. 2002, pp. 7-7.
- [11] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," Signal Processing, vol. 85, no. 1, pp. 177-204, Jan. 2005.

- [12] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of the direct-path relative transfer function for supervised sound-source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 4, no. 11, pp. 2171 - 2186, Nov. 2016.
- [13] C. Yemdji, M. Mossi Idrissa, N. Evans, C. Beaugeant, and P. Vary, "Dual channel echo postfiltering for hands-free mobile terminals," in *Proc. IWAENC*, Aachen, Germany, Sep. 2012, pp. 1-4.
- [14] C. Yemdji, L. Lepauloux, N. Evans, and C. Beaugeant, "Method for processing an audio signal and audio receiving circuit," U.S. Patent 2014/0 334 620, 2014.
- [15] W. Kellermann, "Joint design of acoustic echo cancellation and adaptive beamforming for microphone arrays," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, London, UK, 1997, pp. 81-84.
- [16] W. Herbordt and W. Kellermann, "GSAEC - acoustic echo cancellation embedded into the generalized sidelobe canceller," in *Proc. European Signal Processing Conf. (EUSIPCO)*, vol. 3, Tampere, Finland, Sep. 2000, pp. 1843-1846.
- [17] W. Herbordt, W. Kellermann, and S. Nakamura, "Joint optimization of LCMV beamforming and acoustic echo cancellation," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 2003-2006.
- [18] K.-D. Kammeyer, M. Kallinger, and A. Mertins, "New aspects of combining echo cancellers with beamformers," in *Proc. IEEE ICASSP*, vol. 3, Philadelphia, USA, Mar. 2005, pp. 137-140.
- [19] Y. Avargel and I. Cohen, "Adaptive system identification in the short-time fourier transform domain using cross-multiplicative transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 6, no. 1, pp. 162 - 173, Jan. 2008.
- [20] "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305-1319, May 2007.

- [21] "On multiplicative transfer function approximation in the short-time fourier transform domain," IEEE Signal Process. Lett., vol. 14, no. 5, pp. 337 - 340, May 2007.
- [22] I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator," IEEE Signal Process. Lett., vol. 11, no. 9, pp. 725-728, Sep. 2004.
- [23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am., vol. 65, no. 4, pp. 943-950, Apr. 1979.
- [24] P. C. W. Sommen, "Partitioned frequency-domain adaptive filters," in Proc. Asilomar Conf. on Signals, Systems and Computers, 1989, pp. 677-681.
- [25] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," IEEE Signal Process. Mag., vol. 9, no. 1, pp. 14-37, Jan. 1992.
- [26] S. Haykin, "Adaptive Filter Theory", 4th ed. Prentice-Hall, 2002.
- [27] M. Dentino, J. McCool, and B. Widrow, "Adaptive filtering in the frequency domain," Proc. of the IEEE, vol. 66, no. 12, pp. 1658 - 1659, Dec. 1978.
- [28] G. A. Clark, S. R. Parker, and S. K. Mitra, "A unified approach to time- and frequency-domain realization of FIR adaptive digital filters", IEEE Trans. Acoust., Speech, Signal Process., vol. 31, no. 5, pp. 1073 - 1083, Oct. 1983.
- [29] A. Oppenheim and R. W. Schaffer, "Digital Signal Processing", 2nd ed. Prentice-Hall Inc., Englewood Cliff, NJ, 1993.
- [30] R. M. M. Derkx, G. P. M. Engelmeers, and P. C. W. Sommen, "New constraining method for partitioned block frequency-domain adaptive filters", IEEE Trans. Signal Process., vol. 50, no. 3, pp. 2177-2186, 2002.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất các giải pháp có độ phức tạp thấp để khử nhiễu đa kênh. Mục đích của sáng chế được giải quyết bởi các đối tượng của các điểm yêu cầu bảo hộ độc lập.

Thiết bị khử nhiễu đa kênh trong tín hiệu âm thanh nhận được bao gồm hai hoặc nhiều hơn hai kênh âm thanh nhận được để thu được tín hiệu âm thanh được biến đổi

bao gồm hai hoặc nhiều hơn hai kênh âm thanh được biến đổi theo phương án được đề xuất.

Thiết bị bao gồm bộ lọc thứ nhất được tạo cấu hình để tạo ước lượng thứ nhất về tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu.

Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ nhất được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất.

Hơn nữa, thiết bị bao gồm bộ lọc thứ hai được tạo cấu hình để tạo ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất.

Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ hai được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai của tín hiệu nhiễu thứ hai.

Các phương án đề xuất các giải pháp, ví dụ, thiết bị và phương pháp, để khử nhiễu đa kênh sử dụng các hàm truyền tương đối.

Ví dụ, đối với AEC, các giải pháp theo các phương án sử dụng ước lượng của tín hiệu tiếng vang âm thanh sơ cấp để tính toán các ước lượng của các tín hiệu tiếng vang âm thanh còn lại, hay thứ cấp. Để làm được điều này, mối quan hệ giữa các đáp ứng xung âm sơ cấp (AIR), ví dụ, các AIR giữa loa và các micrô sơ cấp, và các AIR thứ cấp, ví dụ, các AIR giữa loa và các micrô sơ cấp, được xác định. Sau đó, các tín hiệu tiếng vang âm thanh thứ cấp được tính toán bằng cách lọc tín hiệu tiếng vang âm thanh sơ cấp với mối quan hệ ước lượng giữa các AIR. Cuối cùng, sự khử được áp dụng cho từng và mọi tín hiệu micrô. Nếu khoảng cách giữa các micrô nhỏ, các mối quan hệ này có thể được mô hình hóa bằng các bộ lọc tương đối ngắn. Do đó, độ phức tạp tính toán có thể được giảm.

Ngoài ra, phương pháp khử nhiễu đa kênh trong tín hiệu âm thanh nhận được bao gồm hai hoặc nhiều hơn hai kênh âm thanh nhận được để thu được tín hiệu âm

thanh được biến đổi bao gồm hai hoặc nhiều hơn hai kênh âm thanh được biến đổi theo phương án được đề xuất.

Phương pháp bao gồm các bước:

- Tạo ước lượng thứ nhất của tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu.

- Tạo kênh âm thanh được biến đổi thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất.

- Tạo ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất. Và:

- Tạo kênh âm thanh được biến đổi thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai của tín hiệu nhiễu thứ hai.

Hơn nữa, chương trình máy tính được đề xuất, trong đó chương trình máy tính được tạo cấu hình để thực hiện phương pháp được mô tả ở trên khi được chạy trên máy tính hoặc bộ xử lý tín hiệu.

Mô tả vắn tắt các hình vẽ

Sau đây, các phương án của sáng chế được mô tả chi tiết hơn với sự tham chiếu đến các hình vẽ, trong đó:

Fig.1a minh họa thiết bị khử nhiễu đa kênh theo phương án,

Fig.1b minh họa thiết bị khử nhiễu đa kênh theo một phương án khác,

Fig.1c minh họa thiết bị khử nhiễu đa kênh theo phương án khác nữa,

Fig.2 minh họa AEC đa micro,

Fig.3 minh họa AEC đa micro theo phương án.

Fig.4 minh họa AEC đa micrô trong miền biến đổi Fourier thời gian ngắn (short-time Fourier transform - STFT),

Fig.5 minh họa AEC đa micrô trong miền STFT theo phương án,

Fig.6 mô tả các kết quả tương ứng với các mô phỏng với các AIR bị cắt ngắn,

Fig.7 mô tả sự so sánh giữa hàm chuyển tiếng vang âm thanh (acoustic echo transfer function - AETF) và AEC dựa trên hàm chuyển tiếng vang tương đối (relative echo transfer function - RETF) với $T_{60} = 0.15s$ và $L = 256$ tiếng gõ nhẹ, và

Fig.8 minh họa sự so sánh giữa AETF và AEC trên cơ sở RETF với $T_{60} = 0.35s$ và $L = 1024$ tiếng gõ nhẹ.

Mô tả chi tiết sáng chế

Fig.1a minh họa thiết bị khử nhiễu đa kênh theo phương án.

Thiết bị bao gồm bộ lọc thứ nhất được tạo cấu hình để tạo ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu $x(t)$.

Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ nhất 114 được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ nhất $e_1(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất $y_1(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Hơn nữa, thiết bị bao gồm một bộ lọc thứ hai 122 được tạo cấu hình để tạo ước lượng thứ hai $\hat{d}_2(t)$ của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Ngoài ra, thiết bị bao gồm bộ khử nhiễu thứ hai 124 được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ hai $e_2(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai $y_2(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai $\hat{d}_2(t)$ của tín hiệu nhiễu thứ hai.

Các phương án dựa trên việc phát hiện rằng ước lượng thứ nhất của tín hiệu nhiều thứ nhất có thể được sử dụng để tạo ra ước lượng thứ hai của tín hiệu nhiều thứ hai. Việc sử dụng lại ước lượng thứ nhất của tín hiệu nhiều thứ nhất để xác định ước lượng thứ hai của tín hiệu nhiều thứ hai làm giảm độ phức tạp tính toán so với các giải pháp tạo ra ước lượng thứ hai của tín hiệu nhiều thứ hai bằng cách sử dụng tín hiệu tham chiếu thay vì sử dụng ước lượng thứ nhất của tín hiệu nhiều thứ nhất.

Một số phương án liên quan đến sự khử tiếng vang âm thanh (AEC).

Trong phương án, ước lượng thứ nhất của tín hiệu nhiều thứ nhất có thể, ví dụ, là ước lượng thứ nhất của tín hiệu tiếng vang âm thanh thứ nhất, ước lượng thứ hai của tín hiệu nhiều thứ hai là ước lượng thứ hai của tín hiệu tiếng vang âm thanh thứ hai.

Bộ khử nhiễu thứ nhất 114 có thể, ví dụ, được tạo cấu hình để thực hiện sự khử tiếng vang âm thanh trên kênh âm thanh nhận được thứ nhất (ví dụ, bằng cách trừ ước lượng thứ nhất của tín hiệu tiếng vang âm thanh thứ nhất từ kênh âm thanh nhận được thứ nhất) để thu được kênh âm thanh được biến đổi thứ nhất.

Bộ khử nhiễu thứ hai 124 có thể, ví dụ, được tạo cấu hình để thực hiện khử tiếng vang âm thanh trên kênh âm thanh nhận được thứ hai (ví dụ, bằng cách trừ ước lượng thứ hai của tín hiệu tiếng vang âm thanh thứ hai từ kênh âm thanh nhận được thứ hai) để thu được kênh âm thanh được biến đổi thứ hai.

Fig.1b minh họa thiết bị khử nhiễu đa kênh theo một phương án khác.

So với thiết bị theo Fig.1a, thiết bị theo Fig.1b còn bao gồm bộ lọc thứ ba 132 và bộ khử nhiễu thứ ba 134.

Trong phương án theo Fig.1b, tín hiệu âm thanh nhận được bao gồm ba hoặc nhiều hơn ba kênh âm thanh nhận được, và tín hiệu âm thanh được biến đổi bao gồm ba hoặc nhiều hơn ba kênh âm thanh được biến đổi.

Bộ lọc thứ ba 132 được tạo cấu hình để tạo ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiều thứ ba tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiều thứ nhất.

Bộ khử nhiễu thứ ba 134 được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ ba $e_3(t)$ trong số ba hoặc nhiều hơn ba kênh âm thanh được biến đổi từ kênh âm

thanh nhận được thứ ba $y_3(t)$ trong số ba hoặc nhiều hơn ba kênh âm thanh nhận được tùy thuộc vào ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba.

Fig.1c minh họa thiết bị khử nhiễu đa kênh theo phương án khác nữa.

So với thiết bị theo Fig.1a, thiết bị theo Fig.1c còn bao gồm bộ lọc thứ ba 132 và bộ khử nhiễu thứ ba 134.

Trong phương án theo Fig.1c, tín hiệu âm thanh nhận được bao gồm ba hoặc nhiều hơn ba kênh âm thanh nhận được, và tín hiệu âm thanh được biến đổi bao gồm ba hoặc nhiều hơn ba kênh âm thanh được biến đổi.

Bộ lọc thứ ba 132 được tạo cấu hình để tạo ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba tùy thuộc vào ước lượng thứ hai $\hat{d}_2(t)$ của tín hiệu nhiễu thứ hai. Do đó, phương án theo Fig.1c khác với phương án theo Fig.1b ở chỗ tạo ra ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba được thực hiện tùy thuộc vào ước lượng thứ hai $\hat{d}_2(t)$ của tín hiệu nhiễu thứ hai thay vì phụ thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Bộ khử nhiễu thứ ba 134 được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ ba $e_3(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ ba $y_3(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba.

Trong các phương án khác (thực hiện đường nét đứt tùy chọn 199 trên Fig.1c), bộ lọc thứ ba 132 được tạo cấu hình để tạo ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba tùy thuộc vào ước lượng thứ hai $\hat{d}_2(t)$ của tín hiệu nhiễu thứ hai và tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Fig.2 minh họa AEC đa micrô theo kỹ thuật trước đây. Theo cách tiếp cận của kỹ thuật trước đây, bộ lọc thứ nhất 282 được sử dụng để tạo ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất từ tín hiệu tham chiếu $x(t)$.

Bộ khử nhiễu thứ nhất 284 sau đó tạo ra kênh âm thanh được biến đổi thứ nhất $e_1(t)$ từ kênh âm thanh nhận được thứ nhất $y_1(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Theo cách tiếp cận của kỹ thuật trước đây theo Fig.2, bộ lọc thứ hai 292 tạo ra ước lượng thứ hai $\hat{d}_N(t)$ của tín hiệu nhiễu thứ hai từ tín hiệu tham chiếu $x(t)$ cũng được sử dụng bởi bộ lọc thứ nhất 282.

Bộ khử nhiễu thứ hai 294 sau đó tạo ra kênh âm thanh được biến đổi thứ hai $e_N(t)$ từ kênh âm thanh nhận được thứ hai $y_N(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai $\hat{d}_N(t)$ của tín hiệu nhiễu thứ hai.

Một số phương án làm giảm độ phức tạp của sự khử tiếng vang âm thanh (AEC) đa micrô được mô tả trên Fig.2, bằng cách sử dụng cách tiếp cận dựa trên hàm chuyển tương đối (RTF), như được mô tả trên Fig.3. Các hàm chuyển tương đối được mô tả trong các tài liệu [4], [7].

Fig.3 minh họa sự khử tiếng vang âm thanh (AEC) đa micrô theo các phương án. Trên Fig.3, bộ lọc thứ nhất 312 được sử dụng để tạo ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất từ tín hiệu tham chiếu $x(t)$.

Bộ khử nhiễu thứ nhất 314 sau đó tạo ra kênh âm thanh được biến đổi thứ nhất $e_1(t)$ từ kênh âm thanh nhận được thứ nhất $y_1(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất.

Thiết bị theo Fig.3 hiện tại khác với Fig.2 ở chỗ, bộ lọc thứ hai 322 tạo ra ước lượng thứ hai $\hat{d}_N(t)$ của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất $\hat{d}_1(t)$ của tín hiệu nhiễu thứ nhất mà đã được tạo bởi bộ lọc thứ nhất 312.

Bộ khử nhiễu thứ hai 324 sau đó tạo ra kênh âm thanh được biến đổi thứ hai $e_N(t)$ từ kênh âm thanh nhận được thứ hai $y_N(t)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai $\hat{d}_N(t)$ của tín hiệu nhiễu thứ hai.

Một số phương án làm giảm độ phức tạp của sự khử tiếng vang âm thanh (AEC) đa micrô được mô tả trên Fig.2, bằng cách sử dụng cách tiếp cận dựa trên hàm chuyển tương đối (RTF), như được mô tả trên Fig.3. Các hàm chuyển tương đối được mô tả trong các tài liệu [4], [7].

Các phương án sử dụng ước lượng của tín hiệu nhiễu sơ cấp, để tính toán các ước lượng của các tín hiệu nhiễu còn lại, hay thứ cấp. Để ước lượng tín hiệu nhiễu sơ cấp, bộ lọc sơ cấp, đặc trưng cho mối quan hệ giữa tín hiệu tham chiếu và tín hiệu nhận được sơ cấp, được xác định. Ước lượng của tín hiệu nhiễu sơ cấp sau đó thu được bằng cách lọc tín hiệu tham chiếu với ước lượng của bộ lọc sơ cấp. Sau đó, các bộ lọc thứ cấp, ví dụ, các bộ lọc đặc trưng cho mối quan hệ giữa tín hiệu nhiễu sơ cấp đã ước lượng và các tín hiệu nhận được thứ cấp, được xác định. Tiếp theo, các ước lượng của các tín hiệu nhiễu thứ cấp được tính toán bằng cách lọc ước lượng của tín hiệu nhiễu sơ cấp với các bộ lọc thứ cấp đã ước lượng. Cuối cùng, sự khử được áp dụng để làm giảm sự ghép điện thanh. Nếu khoảng cách giữa các micrô nhỏ, các bộ lọc thứ cấp ngắn hơn các bộ lọc sơ cấp (ví dụ, xem các tài liệu [10], [19]), dẫn đến giảm độ phức tạp tính toán.

Một số phương án được sử dụng để khử tiếng vang âm thanh. Để đạt được mục đích này, Fig.3 mô tả viễn cảnh truyền thông rảnh tay với một loa (một bộ phát) và N micrô (bộ thu). Trong trường hợp cụ thể này, tín hiệu tham chiếu là tín hiệu loa $x(t)$, tín hiệu micrô sơ cấp là $y_1(t)$, không mất tính tổng quát, và t biểu thị hệ số thời gian riêng rẽ. Hơn nữa, ước lượng của bộ lọc sơ cấp được ký hiệu là $\hat{h}_1(t)$, là ước lượng của tín hiệu (nhiều) tiếng vang âm thanh sơ cấp $\hat{d}_1(t)$, và tín hiệu sau khi khử $e_1(t) = y_1(t) - \hat{d}_1(t)$. Như có thể quan sát thấy, tín hiệu tiếng vang âm thanh thứ cấp $\hat{d}_N(t)$ được tính toán bằng cách lọc ước lượng của tín hiệu tiếng vang âm thanh sơ cấp $\hat{d}_1(t)$ với ước lượng của bộ lọc thứ cấp $\hat{a}_N(t)$. Cần lưu ý rằng độ trễ của $D \geq 0$ mẫu được đưa vào tín hiệu micrô thứ cấp. Điều này được thực hiện để đảm bảo rằng D hệ số phi nhân quả của các bộ lọc thứ cấp được ước lượng. Trong trường hợp micro cần được đồng bộ hóa, tín hiệu sơ cấp sau khi khử cũng cần phải được làm trễ bởi D mẫu. Ngược lại, các sơ đồ khử nhiễu cổ điển (như được mô tả trên Fig.2), tính toán các ước

lượng của N tín hiệu nhận được bằng cách lọc tín hiệu $x(t)$ tham chiếu với N bộ lọc sơ cấp đã ước lượng.

Sau đây, cách tiếp cận từng bước theo một số phương án được đề xuất:

1.) Tín hiệu nhiễu sơ cấp được ước lượng bằng cách sử dụng tín hiệu tham chiếu. Trong ứng dụng cụ thể của sự khử tiếng vang âm thanh, trước là tín hiệu tiếng vang âm thanh, và sau là tín hiệu loa. Để làm được như vậy:

1.1.) bộ lọc sơ cấp, đặc trưng cho mối quan hệ giữa tín hiệu tham chiếu và tín hiệu nhận sơ cấp, đây là một trong hai

(a) một tín hiệu nhận duy nhất,

(b) sự kết hợp tuyến tính của các tín hiệu nhận,

được xác định bằng cách sử dụng, ví dụ, các kỹ thuật lọc thích ứng.

1.2.) tín hiệu tham chiếu được lọc với ước lượng của bộ lọc sơ cấp để tính toán ước lượng của tín hiệu nhiễu sơ cấp.

1.3.) khử nhiễu được áp dụng bằng cách trừ đi ước lượng của tín hiệu nhiễu sơ cấp từ tín hiệu nhận được sơ cấp, đây là một trong hai

(a) một tín hiệu nhận duy nhất.

(b) sự kết hợp tuyến tính của các tín hiệu nhận.

2.) Tín hiệu nhiễu thứ cấp được ước lượng dựa trên ước lượng của tín hiệu nhiễu sơ cấp. Để làm được như vậy:

2.1.) bộ lọc thứ cấp, đặc trưng cho mối quan hệ giữa ước lượng của tín hiệu nhiễu sơ cấp và tín hiệu nhận được thứ cấp, được xác định bởi, ví dụ:

i.) tối ưu hóa hàm giá trị hoặc chỉ tiêu lỗi (ví dụ, lỗi bình phương trung bình, lỗi bình phương nhỏ nhất (có trọng số), v.v.)

ii.) kỹ thuật lọc thích ứng trong miền thời gian, tần số hoặc băng con.

sử dụng tín hiệu nhận thứ cấp hoặc tín hiệu thứ cấp sau khi khử, và ước lượng của tín hiệu nhiễu sơ cấp. (Bộ lọc thứ cấp có thể, ví dụ, được coi là cấu hình bộ lọc.)

2.2.) ước lượng của tín hiệu nhiều sơ cấp được lọc với ước lượng của bộ lọc thứ cấp, để tính toán ước lượng của tín hiệu nhiều thứ cấp.

2.3.) khử nhiễu được áp dụng bằng cách trừ đi ước lượng của tín hiệu nhiều thứ cấp từ tín hiệu nhận thứ cấp.

3.) Lặp lại bước 2 cho mỗi tín hiệu nhiều thứ cấp.

4.) Lặp lại các bước 1, 2, 3 cho mỗi tín hiệu nhiễu.

5.) Trong đó bộ phát là loa và bộ nhận là micro.

6.) Trong đó ước lượng của tín hiệu nhiều thứ cấp có thể được sử dụng làm ước lượng của tín hiệu nhiễu sơ cấp dẫn đến cấu hình xếp tầng.

7.) Trong trường hợp có nhiều hơn hai bộ nhận, các tập hợp con của các bộ nhận có thể được xác định, mỗi tập hợp con trong số các tập hợp con có một bộ nhận sơ cấp.

Các phương án tiếp theo chỉ áp dụng một số bước ở trên và/hoặc áp dụng các bước theo một thứ tự khác.

Sau đây, các phương án sử dụng các bộ lọc thích ứng miền STFT được mô tả (STFT có nghĩa là biến đổi Fourier thời gian ngắn).

Cho thiết lập truyền thông rảnh tay với một loa và N micro, tín hiệu micro thứ n có thể được thể hiện trong miền STFT là

$$Y_n(\ell, k) = D_n(\ell, k) + R_n(\ell, k), \quad n \in \{1, \dots, N\}, \quad (1)$$

trong đó ℓ và k tương ứng là các hệ số khung thời gian và tần số. Ngoài ra, $R_n(\ell, k)$ là tín hiệu gần cuối, bao gồm tiếng nói gần cuối và tiếng ồn nền, và $D_n(\ell, k)$ là tiếng vang âm thanh thứ n . $D_n(\ell, k)$ là kết quả của tín hiệu loa $X(\ell, k)$ được truyền qua phòng, và được thu bởi micro thứ n . Công thức chính xác của nó trong miền STFT (xem ví dụ tài liệu [20]) là

$$D_n(\ell, k) = \sum_{b=-\infty}^{\infty} h_n^H(b, k) x(\ell - b), \quad (2)$$

trong đó $x(\ell) = [X(\ell, 0), \dots, X(\ell, K-1)]^T$, các chỉ số trên $\cdot^T$ và $\cdot^H$ tương ứng biểu thị chuyển vế và chuyển vế liên hợp và K là chiều dài biến đổi. Hơn nữa, phân vùng thứ b của AETF thứ n là $h_n(b, k) = [H_n(b, k, 0), \dots, H_n(b, k, K-1)]^T$ là vector chứa tất cả các sự phụ thuộc tần số $H_n(b, k, k')$, với $k' \in \{0, \dots, K-1\}$ (AETF có nghĩa là hàm truyền tiếng vang âm thanh).

Cần lưu ý rằng các AETF trong miền STFT mà được phân tích xuyên suốt trong tài liệu [20], là phi nhân quả. Ngoài ra, số lượng các phân vùng hay các khung đầu vào cần được ước lượng các hệ số AIR L là $B = \lceil (L + K - 1) / R \rceil + \lceil K / R \rceil - 1$, trong đó R biểu thị việc chuyển khung giữa các khung đầu vào tiếp theo. Do tính phi nhân quả của các AETF, các khung nhìn về phía trước $B_{nc} = \lceil K / R \rceil - 1$ của $X(\ell, k)$ là cần thiết để tính toán các tín hiệu tiếng vang.

Giả sử rằng độ chọn lọc tần số của các cửa sổ phân tích và tổng hợp STFT là đủ để có thể bỏ qua các sự phụ thuộc tần số. Ngoài ra, giải thích một cách ngắn gọn, theo các phương án, các tác giả sáng chế cho rằng độ trễ của các khung B_{nc} được đưa vào đường dẫn lặp lại như được mô tả trên Fig.4. Trong thực tế, đường dẫn bất mục tiêu thường bị trễ thay vào đó, xem ví dụ các tài liệu [7], [20].

Các tín hiệu trên Fig.4 là các tín hiệu trong miền biến đổi. Cụ thể, các tín hiệu trên Fig.4 là các tín hiệu trong miền biến đổi Fourier thời gian ngắn (miền STFT).

Trên Fig.4, bộ lọc thứ nhất 482 được sử dụng để tạo ước lượng thứ nhất $\hat{D}_1(\ell, k)$ của tín hiệu nhiễu thứ nhất từ tín hiệu tham chiếu $X(\ell, k)$.

Bộ khử nhiễu thứ nhất 484 sau đó tạo ra kênh âm thanh được biến đổi thứ nhất $E_1(\ell, k)$ từ kênh âm thanh nhận được thứ nhất $Y_1(\ell, k)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất $\hat{D}_1(\ell, k)$ của tín hiệu nhiễu thứ nhất.

Theo cách tiếp cận theo Fig.4, bộ lọc thứ hai 492 tạo ra ước lượng thứ hai $\hat{D}_N(\ell, k)$ của tín hiệu nhiễu thứ hai từ tín hiệu tham chiếu $X(\ell, k)$ cũng được sử dụng bởi bộ lọc thứ nhất 482.

Bộ khử nhiễu thứ hai 494 sau đó tạo ra kênh âm thanh được biến đổi thứ hai $E_N(\ell, k)$ từ kênh âm thanh nhận được thứ hai $Y_N(\ell, k)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai $\hat{D}_N(\ell, k)$ của tín hiệu nhiễu thứ hai.

Fig.4 minh họa AEC đa micro trong miền STFT. Trong thực tế, đường dẫn bất mục tiêu thường bị trễ thay vào đó, xem ví dụ các tài liệu [7], [20]. Bây giờ, bằng cách sử dụng phép tính gần đúng hàm chuyển tích hợp (convolutive transfer function - CTF) (xem ví dụ tài liệu [7]), có thể viết

$$D_n(\ell, k) \approx \sum_{b=-B_{nc}}^{B-B_{nc}-1} H_n^*(b, k) X(\ell - B_{nc} - b, k), \quad (3)$$

trong đó $*$ biểu thị sự liên hợp phức tạp và, để ngắn gọn, $H_n(b, k) = H_n(b, k, k)$. Các thuật toán thích ứng trong AEC được điều khiển bởi tín hiệu lỗi sau khi khử, ví dụ:

$$E_n(\ell, k) = Y_n(\ell, k) - \hat{D}_n(\ell, k) = Y_n(\ell, k) - \hat{\mathbf{h}}_n^H(\ell, k) \mathbf{x}(\ell, k), \quad (4)$$

trong đó $\hat{\mathbf{h}}_n$ được sử dụng để biểu thị các ước lượng, $\hat{\mathbf{h}}_n(\ell, k) = [\hat{H}_n(\ell, -B_{nc}, k), \dots, \hat{H}_n(\ell, B - B_{nc} - 1, k)]^T$ và $\mathbf{x}(\ell, k) = [X(\ell, k), \dots, X(\ell - B + 1, k)]^T$. Chỉ số trên ^H chỉ ra Hermit. Hầu hết các bộ lọc thích ứng được sử dụng trong AEC thuộc loại gradient descent (xem ví dụ tài liệu [2]), do đó, phương trình điều chỉnh chung được cho bởi

$$\hat{\mathbf{h}}_n(\ell + 1, k) = \hat{\mathbf{h}}_n(\ell, k) + \mathbf{M}_n(\ell, k) \mathbf{x}(\ell, k) E_n^*(\ell, k), \quad (5)$$

trong đó $\mathbf{M}_n(\ell, k)$ là ma trận bước kích thước của bộ lọc thích ứng, công thức của nó phụ thuộc vào thuật toán thích ứng cụ thể được sử dụng.

Sau đây, việc sử dụng các hàm chuyển tiếng vang tương đối theo các phương án được mô tả.

Do các hạn chế về độ phức tạp tính toán, việc triển khai AEC đa micrô như được mô tả trên Fig.4 không phải lúc nào cũng khả thi.

Theo các phương án, các tác giả sáng chế đề xuất giảm độ phức tạp bằng cách sử dụng cách tiếp cận trên cơ sở RETF, như được mô tả trên Fig.5 (RETF có nghĩa là hàm chuyển tiếng vang tương đối). Fig.5 minh họa AEC đa micrô trong miền STFT theo phương án.

Cũng như vậy, các tín hiệu trên Fig.5 là các tín hiệu trong miền biến đổi. Cụ thể, các tín hiệu trên Fig.5 là các tín hiệu trong miền biến đổi Fourier thời gian ngắn (miền STFT).

Trên Fig.5, bộ lọc thứ nhất 512 được sử dụng để tạo ước lượng thứ nhất $\hat{D}_1(\ell, k)$ của tín hiệu nhiễu thứ nhất từ tín hiệu tham chiếu $X(\ell, k)$.

Bộ khử nhiễu thứ nhất 514 sau đó tạo ra kênh âm thanh được biến đổi thứ nhất $E_1(\ell, k)$ từ kênh âm thanh nhận được thứ nhất $Y_1(\ell, k)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ nhất $\hat{D}_1(\ell, k)$ của tín hiệu nhiễu thứ nhất.

Thiết bị theo Fig.5 hiện tại khác với Fig.4 ở chỗ, bộ lọc thứ hai 522 tạo ra ước lượng thứ hai $\hat{D}_N(\ell, k)$ của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất $\hat{D}_1(\ell, k)$ của tín hiệu nhiễu thứ nhất mà đã được tạo bởi bộ lọc thứ nhất 512.

Bộ khử nhiễu thứ hai 524 sau đó tạo ra kênh âm thanh được biến đổi thứ hai $E_N(\ell, k)$ từ kênh âm thanh nhận được thứ hai $Y_N(\ell, k)$ trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai $\hat{D}_N(\ell, k)$ của tín hiệu nhiễu thứ hai.

Trong các phương án, bộ lọc thứ hai 122 có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào kênh âm thanh nhận được thứ hai, và bộ lọc thứ hai 122 có thể được tạo cấu hình để xác định ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào cấu hình bộ lọc.

Ví dụ, bộ lọc thứ hai 122 được tạo cấu hình để xác định cấu hình bộ lọc bằng cách giảm thiểu hàm giá trị hoặc bằng cách giảm thiểu chỉ tiêu lỗi, ví dụ, bằng cách giảm thiểu lỗi bình phương trung bình.

Trong phần sau đây, các cấu hình bộ lọc như vậy sẽ được xác định, ví dụ, có thể là $A_n(p, k)$ và/hoặc $a_n(k)$ và/hoặc $\hat{a}_n^{opt}(\ell, k)$. Việc đặt vấn đề được suy ra giả định rằng các bộ lọc là bất biến theo thời gian, trong khi các ước lượng là những bộ lọc thay đổi theo thời gian.

Ví dụ cụ thể cho phương án như vậy được đưa ra sau đây.

Không mất tính tổng quát, tín hiệu tiếng vang sơ cấp được ký hiệu là $D_1(\ell, k)$ - được xác định như trong công thức (3). Theo các giả định được thực hiện trước đó về các sự phụ thuộc tần số, có thể viết,

$$D_n(\ell, k) = \sum_{p=-\infty}^{\infty} A_n^*(p, k) D_1(\ell - p, k), \quad n \in \{2, \dots, N\} \quad (6)$$

trong đó $A_n(p, k)$ là phân vùng thứ p của hàm chuyển tiếng vang tương đối thứ n (RETF). Với điều kiện là khoảng cách giữa micrô sơ cấp và micrô thứ cấp tương đối nhỏ, có thể giả định rằng các phân vùng phi nhân quả của $A_n(p, k) \forall n$ là không đáng kể. Điều đáng nói là một vài hệ số miền thời gian phi nhân quả vẫn được mô hình hóa bởi $A_n(0, k)$. Theo giả định này, không cần nhìn về phía trước, và do đó, không phát sinh thêm sự trễ nào.

Cuối cùng, sử dụng phép tính gần đúng CTF dẫn đến

$$D_n(\ell, k) \approx \sum_{p=0}^{P-1} A_n^*(p, k) D_1(\ell - p, k), \quad n \in \{2, \dots, N\} \quad (7)$$

trong đó P là số phân vùng RETF.

Vì $D_1(\ell, k)$ không thể quan sát được, theo các phương án, các tác giả sáng chế đề xuất thay thế $D_1(\ell, k)$ bởi $\hat{D}_1(\ell, k)$ trong công thức (7) mà có thể nhận được bằng

cách sử dụng AEC trong phần tình trạng kỹ thuật. Để ước tính $A_n(p, k) \forall p$, theo các phương án, tín hiệu lỗi được giảm thiểu:

$$E_n(\ell, k) = Y_n(\ell, k) - \hat{\mathbf{a}}_n^H(\ell, k) \hat{\mathbf{d}}_1(\ell, k), \quad (8)$$

trong đó $\hat{\mathbf{a}}_n(\ell, k) = [\hat{A}_n(\ell, 0, k), \dots, \hat{A}_n(\ell, P-1, k)]^T$ là vector chùm thứ n của các phân vùng RETF và $\hat{\mathbf{d}}_1(\ell, k) = [\hat{D}_1(\ell, k), \dots, \hat{D}_1(\ell - P + 1, k)]^T$. Bộ lọc tối ưu theo phương sai số toàn phương trung bình mà nhận được bằng cách giảm thiểu hàm giá trị bậc hai $J(\ell, k) = E \|E_n(\ell, k)\|^2$ bằng với

$$\hat{\mathbf{a}}_n^{\text{opt}}(\ell, k) = \mathbf{\Psi}_1(\ell, k)^{-1} \psi_{1n}(\ell, k), \quad (9)$$

trong đó $\psi_1(\ell, k)$ là ma trận hiệp phương sai của $\hat{\mathbf{d}}_1(\ell, k)$, và $\psi_{1n}(\ell, k)$ là vector tương quan chéo giữa $\hat{\mathbf{d}}_1(\ell, k)$ và $Y_n(\ell, k)$, ví dụ,

$$\begin{aligned} \mathbf{\Psi}_1(\ell, k) &= E \left\{ \hat{\mathbf{d}}_1(\ell, k) \hat{\mathbf{d}}_1^H(\ell, k) \right\}, \\ \psi_{1n}(\ell, k) &= E \left\{ \hat{\mathbf{d}}_1(\ell, k) Y_n^*(\ell, k) \right\}, \end{aligned}$$

trong đó $E \{\cdot\}$ biểu thị kỳ vọng toán. Cần lưu ý rằng $\psi_{1n}(\ell, k) = E \{ \hat{\mathbf{d}}_1(\ell, k) D_n^*(\ell, k) \}$, theo giả định rằng $E \{ \hat{\mathbf{d}}_1(\ell, k) R_n^*(\ell, k) \} = \mathbf{0}$. Có nghĩa là $\hat{\mathbf{a}}_n^{\text{opt}}(\ell, k)$ mô hình hóa mối quan hệ giữa AETF sơ cấp được ước lượng và AETF thứ cấp được ước lượng thứ n . Chẳng hạn, xem xét trường hợp thông thường $B = P = 1$, với $B_{nc} = 0$, ví dụ, phép tính gần đúng hàm chuyển nhân bội (xem ví dụ tài liệu [21]), với điều kiện RETF được ước lượng thứ n bằng

$$\hat{A}_n^{\text{opt}}(\ell, 0, k) = \hat{H}_1(\ell, 0, k)^{-1} H_n(\ell, 0, k) \quad (10)$$

tức là, một khi bộ khử tiếng vang âm thanh sơ cấp đã hội tụ, bằng với $A_n(\ell, 0, k)$ như được xác định trong công thức (7).

So với vấn đề ước lượng các RTF từ các sự quan sát nhiễu (xem ví dụ các tài liệu [4], [7], [22]) trong công thức của chúng tôi, không có sai lệch bổ sung nào do các thành phần nhiễu có tương quan giữa các kênh.

Hơn nữa, như tín hiệu loa đã biết, việc thực hiện các bộ cảm biến hoạt động giọng nói (voice activity detector - VAD) để kiểm soát quá trình ước lượng được đơn giản hóa rất nhiều. Ngược lại, bộ cảm biến trò chuyện đôi là cần thiết do thực tế là trong thực tế $\mathbf{d}_1^*(\ell, k)R_n(\ell, k)$ được tính gần đúng là $\frac{1}{T} \sum_{t=0}^{T-1} \hat{\mathbf{d}}_1^*(\ell - t, k)R_n(\ell - t, k)$, và do đó, giả định trước đó về mối quan hệ thống kê giữa $\hat{\mathbf{d}}_1(\ell, k)$ và $R_n(\ell, k)$ không phải lúc nào cũng đúng.

Sau đây, các phương án sử dụng ước lượng RETF thích ứng được đề xuất.

Trong các phương án như vậy, bộ lọc thứ hai 522 theo Fig.5 có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai bằng cách sử dụng ma trận bước kích thước. Ví dụ, bộ lọc thứ hai 522 theo Fig.5, có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc tùy thuộc vào cấu hình bộ lọc cho hệ số thời gian thứ nhất trước hệ số thời gian thứ hai, tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất, và tùy thuộc vào mẫu kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất.

Trong các phương án cụ thể, bộ lọc thứ hai 522 có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai theo

$$\hat{\mathbf{a}}_n(\ell + 1, k) = \hat{\mathbf{a}}_n(\ell, k) + \mathbf{C}_n(\ell, k)\hat{\mathbf{d}}_1(\ell, k)E_n^*(\ell, k)$$

trong đó $\ell + 1$ biểu thị hệ số thời gian thứ hai, và trong đó ℓ biểu thị hệ số thời gian thứ nhất, và trong đó k biểu thị hệ số tần số, trong đó $\hat{\mathbf{a}}_n(\ell + 1, k)$ là cấu hình bộ lọc cho hệ số thời gian thứ hai, và trong đó $\hat{\mathbf{a}}_n(\ell, k)$ là cấu hình bộ lọc cho hệ số thời gian thứ nhất, trong đó $\hat{\mathbf{d}}_1(\ell, k)$ là ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất, trong đó $E_n^*(\ell, k)$ là kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất, và trong đó $\mathbf{C}_n(\ell, k)$ là ma trận bước kích thước (ví dụ, nghịch đảo của ma trận hiệp phương sai của $\hat{\mathbf{d}}_1(\ell, k)$).

Mô tả chi tiết hơn, các bộ lọc thích ứng có thể được sử dụng để theo dõi các RETF biến thiên theo thời gian chậm. Do thực tế $\hat{\mathbf{d}}_1(\ell, k)$ là ước lượng của tín hiệu tiếng vang được thu bởi micro sơ cấp, không thể coi đó là không tương quan theo thời gian. Chính xác hơn, các ngoại chéo của $\Psi_1(\ell, k)$ là không đáng kể nếu các cửa sổ STFT ngắn, hoặc nếu sự chồng lấp giữa chúng lớn. Cân nhắc điều này, phương pháp của Newton (xem ví dụ tài liệu [2]),

$$\hat{\mathbf{a}}_n(\ell + 1, k) = \hat{\mathbf{a}}_n(\ell, k) + \eta \Psi_1(\ell, k)^{-1} \hat{\mathbf{d}}_1(\ell, k) E_n^*(\ell, k), \quad (11)$$

đảm bảo sự hội tụ nhanh và ổn định đối với bộ lọc tối ưu. Trong công thức (11), η là bước kích thước cố định được sử dụng để kiểm soát quá trình thích ứng. Trong thực tế, ma trận hiệp phương sai $\Psi_1(\ell, k)$ được tính gần đúng bằng cách tính trung bình theo thời gian, ví dụ, bằng cách sử dụng bộ lọc đệ quy bậc nhất:

$$\tilde{\Psi}_1(\ell, k) = \beta \tilde{\Psi}_1(\ell - 1, k) + (1 - \beta) \hat{\mathbf{d}}_1(\ell, k) \hat{\mathbf{d}}_1^H(\ell, k),$$

trong đó các trung bình thời gian được biểu thị bởi $\sim$ và β là hệ số quên.

Sau đây, sự đánh giá việc thực hiện được mô tả.

Để đánh giá phương pháp đề xuất, ba bộ thí nghiệm đã được tiến hành, trong đó thiết lập mô phỏng được thiết kế như sau. Các tín hiệu tiếng vang được tạo ra bằng cách tạo tín hiệu giọng nói sạch với các AIR mô phỏng. Tín hiệu giọng nói sạch được tạo bằng cách sử dụng phương pháp hình ảnh (xem ví dụ tài liệu [23]) cho căn phòng có các kích thước $3 \times 4 \times 2.5m^3$, và thời gian vang $T_{60} = 0.15$ và $0.35s$. Độ dài của các AIR mô phỏng là $L = 4096$ tiếng gõ nhẹ, tại tần số lấy mẫu $F_s = 16kHz$. Các AIR được tạo ra để thiết lập với hai micro và một loa. Thiết lập đường cơ sở đã sử dụng khoảng cách giữa loa và micro sơ cấp là $l_1 = 10cm$ và giữa các micro là $\Delta = 1.5cm$. Khoảng cách giữa loa và micro thứ cấp là $l_2 = l_1 + \Delta cm$. Tác động của các thông số này đến việc thực hiện cũng được phân tích. Để đạt được mục đích này, $\Delta = 3cm$ và $l_1 = 20cm$ cũng được đánh giá.

Các tín hiệu được chuyển đổi sang miền STFT bằng cách sử dụng các cửa sổ phân tích và tổng hợp Hamming có độ dài $K = 512$ với 75% chồng lấp, do đó $R = 128$

mẫu. Thuật toán thích ứng được sử dụng để ước lượng cả các AETF (5) và các RETF (11) là phương pháp của Newton. Do đó, ma trận bước kích thước trong công thức (5) $\mathbf{M}_n(\ell, k) = \mu \Psi_x(\ell, k)^{-1}$. Vì thực tế cho rằng tín hiệu loa không bị nhiễu theo thời gian, ma trận hiệp phương sai của nó đã được đơn giản hóa bằng cách:

$$\tilde{\Psi}_x(\ell, k) = \beta \tilde{\Psi}_x(\ell - 1, k) + (1 - \beta) \mathbf{I} \odot \mathbf{x}(\ell, k) \mathbf{x}^H(\ell, k),$$

trong đó $\odot$ biểu thị phép nhân theo từng phần tử, và $\mathbf{I}$ là ma trận đồng nhất $B \times B$. Cần lưu ý rằng mặc dù thực hiện việc đơn giản hóa này, các hệ số chuẩn hóa vẫn phụ thuộc vào phân vùng.

Các hệ số bước kích thước là $\mu = 0.5/B$ và $\eta = 0.225/P$, và hệ số quên là $\beta = 0.9$. Ngoài ra, các bộ lọc thích ứng và các ma trận hiệp phương sai không được điều chỉnh trong khi tạm dừng giọng nói, và sự chính quy hóa được sử dụng để đảm bảo tính không đơn lẻ của các ma trận hiệp phương sai. Cuối cùng, nhiễu Gaussian trắng đã được thêm vào các tín hiệu micro để mô phỏng tỷ lệ nhiễu âm vang phân đoạn (segmental echoto-noise ratio - SegENR) cố định. Để làm cho sự khác biệt trong việc thực hiện đáng chú ý, SegENR 60 dexiben (decibel - dB) đã được sử dụng. Ba bộ thí nghiệm đã được tiến hành:

1. Các AIR được tạo để mô phỏng $T_{60} = 0.15s$ đã bị cắt ngắn tới độ dài 256 tiếng gõ nhẹ, và được sử dụng để tạo các tín hiệu tiếng vang. Độ dài của AIR sơ cấp được ước lượng là $L = 256$.

2. Môi trường được mô phỏng với $T_{60} = 0.15s$, là độ dài của AIR sơ cấp được ước lượng $L = 256$ tiếng gõ nhẹ.

3. Môi trường được mô phỏng với $T_{60} = 0.35s$, là độ dài của AIR sơ cấp được ước lượng $L = 1024$ tiếng gõ nhẹ.

Cần nhắc lại rằng số lượng phân vùng AETF cần thiết để ước lượng hoàn toàn các hệ số L AIR là $B = B_{nc} + \left\lceil \frac{L + K - 1}{R} \right\rceil$, do đó, ít nhất các hệ số bộ lọc tiếp theo K cũng được ước lượng một phần.

Trong tất cả các mô phỏng, B phân vùng của AETF sơ cấp được ước lượng, trong khi các AETF và RETF thứ cấp được ước lượng bằng cách sử dụng số lượng phân vùng khác nhau tương ứng $B_{nc} < B' \leq B$ và P . Các tín hiệu tiếng vang thứ cấp sau đó thu được bằng cách kết hợp trong miền STFT các AETF thứ cấp với tín hiệu loa, và các RETF với tín hiệu tiếng vang sơ cấp được ước lượng. Tăng cường giảm dội lại tiếng vang (echo return loss enhancement - ERLE) đã được sử dụng để đo mức giảm tiếng vang trong kênh thứ cấp, với

$$\text{ERLE}(\ell) = 10 \log_{10} \frac{\|\mathbf{d}_2(\ell)\|_2^2}{\|\mathbf{d}_2(\ell) - \hat{\mathbf{d}}_2(\ell)\|_2^2} \quad (12)$$

trong đó $\|\cdot\|_2$ là tiêu chuẩn l_2 , và $\mathbf{d}_2(\ell) = [d_2(\ell R + 1), \dots, d_2(\ell R + K)]$ là khung thứ ℓ của tiếng vang âm thanh thứ cấp trong miền thời gian. Kết quả của những mô phỏng này được mô tả trên các hình vẽ từ Fig. 5 đến Fig.7, trong đó các phép đo ERLE được tính trung bình trên 60 khung để rõ ràng. Trong đó, AEC trên cơ sở RETF đã đề xuất được so sánh với AEC trong phần tình trạng kỹ thuật bằng cách sử dụng B và $B' = B_{nc} + P$ phân vùng để ước lượng AETF. Điều kiện phía sau được đưa vào để thể hiện sự so sánh với AEC trên cơ sở AETF sử dụng ít phân vùng CTF nhân quả hơn, điều này cũng sẽ làm giảm độ phức tạp tính toán tổng thể.

Fig.6 mô tả các kết quả tương ứng với các mô phỏng với các AIR được cắt ngắn, Cụ thể, Fig.6 mô tả sự so sánh giữa AETF và AEC trên cơ sở RETF với các AIR được cắt ngắn và $L = 256$ tiếng gõ nhẹ. Việc giảm tiếng vang thu được với $P = 1$ và 2, các hình phụ bên trái và bên phải, được thể hiện cho tất cả các điều kiện thử nghiệm. Có thể thấy rằng với $P = 1$, cách tiếp cận dựa trên RETF hội tụ đến giá trị ERLE cao hơn so với cách tiếp cận trên cơ sở AETF với B' phân vùng, ví dụ, chỉ với P phân vùng nhân quả.

Hơn nữa, hiệu suất chỉ kém hơn so với cách tiếp cận trên cơ sở AETF với B phân vùng. Đối với $P = 2$, tất cả các điều kiện thử nghiệm được thực hiện tương tự.

Sự so sánh hiệu suất với $T_{60} = 0.15s$ được thể hiện trên Fig.7. Cụ thể, Fig.7 mô tả sự so sánh giữa AETF và AEC trên cơ sở RETF với $T_{60} = 0.15s$ và $L = 256$ tiếng gõ

nhẹ. Các kết quả được mô tả trên các hình phụ phía trên bên trái và phía trên bên phải tương ứng với $P = 1$ và 2 cho thiết lập đường cơ sở. Có thể thấy rằng với $P = 1$, cách tiếp cận trên cơ sở RETF vượt trội hơn so với cách tiếp cận trên cơ sở AETF với cùng số lượng phân vùng nhân quả. Với $P = 2$, hiệu suất của cách tiếp cận trên cơ sở AETF được nâng cao rõ rệt, và hiệu quả thu được bằng cách sử dụng cách tiếp cận trên cơ sở RETF bị giảm đi.

Tuy nhiên, cách tiếp cận trên cơ sở RETF vẫn hoạt động tốt hơn, và gần như là cách tiếp cận trên cơ sở AETF với $B = 9$ phân vùng. Ở phía dưới, sự so sánh cho các thiết lập mô phỏng khác nhau được cung cấp với $P = 1$. Ở bên trái, các kết quả với các khoảng cách giữa các micrô khác nhau được thể hiện. Trong khi, ở bên phải, các khoảng cách khác nhau giữa loa và micrô sơ cấp được đánh giá. Có thể thấy rằng, đối với tất cả các điều kiện được thử nghiệm, việc mở rộng bất kỳ tham số nào trong các tham số này sẽ tác động tiêu cực đến hiệu suất của bộ khử. Cần lưu ý rằng việc tăng khoảng cách giữa các micrô có tác động cao hơn đến cách tiếp cận được đề xuất, và nói chung, l_1 có tác động cao hơn đến hiệu suất của bộ khử. Tuy nhiên, đối với các tham số được sử dụng trong các mô phỏng này, cách tiếp cận đề xuất có khả năng vượt trội hơn AEC trên cơ sở AETF với số lượng phân vùng nhân quả bằng nhau.

Cuối cùng, các kết quả được thể hiện trên Fig.8 tương ứng với thiết lập mô phỏng với $T_{60} = 0.35s$. Cụ thể, Fig.8 minh họa sự so sánh giữa AETF và AEC trên cơ sở RETF với $T_{60} = 0.35s$ và $L = 1024$ tiếng gõ nhẹ. Các kết quả thu được với $P = 1$ và 4 phân vùng được mô tả trên các hình phụ bên trái và bên phải.

Có thể thấy rằng phương pháp được đề xuất vượt trội hơn, trong cả hai trường hợp thử nghiệm, cách tiếp cận trên cơ sở AETF với cùng số phân vùng nhân quả. Hơn nữa, với $P = 4$, phương pháp này chỉ hoạt động kém hơn so với AEC trên cơ sở AETF với $B = 15$.

Tóm lại, có thể thấy cách tiếp cận được đề xuất có khả năng vượt trội hơn AEC trên cơ sở AET trong phần tình trạng kỹ thuật với số lượng phân vùng nhân quả bằng nhau. Hơn nữa, các tác giả sáng chế đã chứng minh rằng bằng cách sử dụng AEC trên cơ sở RETF, số lượng phân vùng ước lượng có thể được giảm xuống, với giá trị giảm hiệu suất vừa phải.

Sau đây, việc sử dụng các bộ lọc thích ứng miền tần số theo các phương án được mô tả.

Cụ thể, sự mô tả được đề xuất bằng cách sử dụng các bộ lọc thích ứng miền tần số khối được phân vùng (partitioned-block frequency-domain adaptive filter - PB-FDAF) (xem ví dụ tài liệu [24]). Cụ thể, việc triển khai hiệu quả các bộ lọc thích ứng miền tần số (frequency-domain adaptive filter - FDAF) (xem ví dụ các tài liệu [24], [26]), mà là bản sao tần số của các bộ lọc thích ứng miền thời gian khối (xem ví dụ các tài liệu [27], [28]), rất khác so với bộ lọc STFT. Để biết thêm thông tin về điều này (xem ví dụ tài liệu [20]) và các tài liệu tham khảo trong đó.

Theo một số phương án, hai hoặc nhiều hơn hai kênh âm thanh nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi có thể, ví dụ, là các kênh của miền tần số khối được phân vùng, trong đó hai hoặc nhiều hơn hai kênh âm thanh nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi bao gồm nhiều phân vùng. Tín hiệu tham chiếu và các tín hiệu nhiễu thứ nhất và thứ hai có thể, ví dụ, là các tín hiệu của miền tần số khối được phân vùng, trong đó mỗi tín hiệu trong số tín hiệu tham chiếu và các tín hiệu nhiễu thứ nhất và thứ hai bao gồm nhiều phân vùng.

Trong một số phương án, bộ lọc thứ hai 122; 322; 522 có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào kênh âm thanh nhận được thứ hai. Hơn nữa, bộ lọc thứ hai 122; 322; 522 có thể, ví dụ, được tạo cấu hình để xác định ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào cấu hình bộ lọc. Hơn nữa, bộ lọc thứ hai 122; 322; 522 có thể, ví dụ, được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai tùy thuộc vào cấu hình bộ lọc cho hệ số thời gian thứ nhất trước hệ số thời gian thứ hai, tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất, và tùy thuộc vào mẫu kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất.

Sau đó, sự mô tả các phương án sử dụng các PB-FDAF được phác thảo bằng cách sử dụng kỹ thuật lưu chồng lấp (xem ví dụ các tài liệu [25], [29]). Công thức miền tần số khối được phân vùng của tín hiệu micrô là

$$\mathbf{y}_n(\ell) = \mathbf{d}_n(\ell) + \mathbf{r}_n(\ell), \quad n \in \{1, \dots, N\}, \quad (13)$$

trong đó tín hiệu tiếng vang trong miền tần số thu được sau khi tuyến tính hóa kết quả của tích chập tròn có độ dài K :

$$\mathbf{d}_n(\ell) = \mathbf{F} \text{diag}\{\underbrace{[0, \dots, 0]}_Q, \underbrace{[1, \dots, 1]}_V\} \mathbf{F}^{-1} \sum_{b=0}^{B-1} \mathbf{X}(\ell, b) \mathbf{h}_n(b). \quad (14)$$

trong đó $\mathbf{F}$ là ma trận biến đổi Fourier rời rạc (discrete Fourier transform - DFT) có kích thước $K \times K$, và sự biểu diễn miền tần số của phân vùng AIR thứ b được cho bởi:

$$\mathbf{h}_n(b) = \mathbf{F}[h_n(bQ), \dots, h_n((b+1)Q-1), \mathbf{0}_{1 \times V}]^T,$$

trong đó Q là chiều dài phân vùng, và V là chiều dài của phần đệm không. Ngoài ra, tín hiệu loa đầu vào được xây dựng dưới dạng ma trận đường chéo $K \times K$ có dạng (xem ví dụ tài liệu [25]),

$$\mathbf{X}(\ell, b) = \text{diag}\{\mathbf{F}[x(\ell R - bQ - K + 1), \dots, x(\ell R)]^T\}.$$

Cần phải đề cập rằng tổng số thành phần tuyến tính tạo ra từ tích chập tròn trong công thức (14) là $K - Q + 1$, tuy nhiên, để đơn giản hóa các đạo hàm tiếp theo, theo các phương án, $V = K - Q$ thành phần tuyến tính được chọn trong công thức (14). Bây giờ, có thể suy ra rằng V là độ dài khung tín hiệu đầu ra, và $Q = K - V$ là độ dài của sai số bao quanh, sao cho công thức miền tần số chung của các tín hiệu đầu ra $a \in \{y, d, r, e\}$; ví dụ, bằng:

$$\alpha_n(\ell) = \mathbf{F}[\underbrace{[0, \dots, 0]}_Q, \alpha_n(\ell R - V + 1), \dots, \alpha_n(\ell R)]^T,$$

trong đó các mẫu tín hiệu miền thời gian được biểu thị bởi $a_n(t)$, với t biểu thị hệ số thời gian rời rạc. Để giải thích ngắn gọn, theo các phương án, nó được định nghĩa:

$$\underline{\mathbf{X}}(\ell) = [\mathbf{X}(\ell, 0), \dots, \mathbf{X}(\ell, B-1)],$$

$$\underline{\mathbf{h}}_n = [\mathbf{h}_n^T(0), \dots, \mathbf{h}_n^T(B-1)]^T.$$

mà tương ứng là ma trận chòm của các ma trận đầu vào miền tần số và là vector chòm của các phân vùng AIR miền tần số. Sau đây, có thể xây dựng chính xác công thức (14) dưới dạng $\mathbf{d}_n(\ell) = \mathbf{G}\underline{\mathbf{X}}(\ell)\underline{\mathbf{h}}_n$, trong đó $\mathbf{G}$ là ma trận ràng buộc tích chập tròn trong miền tần số.

Việc áp dụng công thức này tương đương với việc áp dụng DFT nghịch đảo, loại bỏ các thành phần vòng tròn trong miền thời gian bằng cách nhân kết quả của phép tích chập tròn với cửa sổ ràng buộc tích chập tròn $\mathbf{g} = [\mathbf{0}_{1 \times Q}, \mathbf{1}_{1 \times V}]$, và chuyển kết quả của phép tuyến tính hóa trở lại miền tần số. Điều quan trọng là phải nhấn mạnh thực tế rằng công thức trong miền tần số là nhân quả, vì không cần phải tính toán trước để ước lượng các AETF. Trong miền tần số, tín hiệu lỗi sau khi khử là

$$\mathbf{e}_n(\ell) = \mathbf{y}_n(\ell) - \hat{\mathbf{d}}_n(\ell), \quad n \in \{1, \dots, N\}, \quad (15)$$

và phương trình điều chỉnh PB-FDAF chung là

$$\hat{\underline{\mathbf{h}}}_n(\ell+1) = \hat{\underline{\mathbf{h}}}_n(\ell) + \underline{\mathbf{M}}_n \underline{\bar{\mathbf{G}}} \underline{\mathbf{X}}^H(\ell) \mathbf{e}_n(\ell), \quad (16)$$

trong đó

$$\underline{\bar{\mathbf{G}}} = \begin{bmatrix} \bar{\mathbf{G}} & \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} \\ \mathbf{0}_{K \times K} & \bar{\mathbf{G}} & \dots & \mathbf{0}_{K \times K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & \bar{\mathbf{G}} \end{bmatrix}, \text{ and}$$

$$\bar{\mathbf{G}} = \mathbf{F} \text{diag}\{[\mathbf{1}_{1 \times Q}, \mathbf{0}_{1 \times V}]\} \mathbf{F}^{-1} = \mathbf{F} \text{diag}\{\bar{\mathbf{g}}\} \mathbf{F}^{-1}$$

biểu thị ma trận ràng buộc tương quan tròn, trong đó $\bar{\mathbf{g}}$ là cửa sổ ràng buộc tương quan tròn miền thời gian, và toán tử $\text{diag}\{\mathbf{v}\}$ tạo ra ma trận đường chéo với các phần tử $\mathbf{v}$ trên đường chéo chính của nó.

Theo cách tương tự, đối với công thức sử dụng các RETF, người ta có thể xác định các tín hiệu tiếng vang thứ cấp như sau

$$\mathbf{d}_n(\ell) = \mathbf{G} \sum_{p=-P_{nc}}^{P-P_{nc}-1} \mathbf{D}_1(\ell, p) \mathbf{a}_n(p) = \mathbf{G} \underline{\mathbf{D}}_1(\ell) \underline{\mathbf{a}}_n, \quad n \in \{2, \dots, N\}, \quad (17)$$

trong đó không mất tính tổng quát $\mathbf{D}_1(\ell, p) = \text{diag}\{\mathbf{F}[d_1(\ell R - pQ - K + 1), \dots, d_1(\ell R)]^T\}$ là tín hiệu, sơ cấp, hoặc tham chiếu, tiếng vang và $\underline{\mathbf{D}}_1(\ell)$ và $\underline{\mathbf{a}}_n$ được định nghĩa tương tự $\underline{\mathbf{X}}(\ell)$ và $\underline{\mathbf{h}}_n$.

Cần nhớ rằng, trái ngược với công thức miền STFT, các AETF và RETF miền tần số là nhân quả, ví dụ, $\mathbf{h}_n(0)$ và $\mathbf{a}_n(0)$ không mô hình hóa bất kỳ hệ số phi nhân quả nào. Tuy nhiên, tùy thuộc vào vị trí tương đối của micrô sơ cấp so với các micrô thứ cấp, các RETF có thể là nhân quả hoặc phi nhân quả. Do đó, cần phải nhìn trước các phân vùng P_{nc} của tín hiệu tiếng vang sơ cấp để tính đến tính phi nhân quả có thể có của các RETF miền tần số $\mathbf{a}_n(p)$.

Trong thực tế, điều này có thể được khắc phục bằng cách làm trễ các tín hiệu micrô thứ cấp, như được mô tả trên Fig.3 trong miền thời gian hoặc tần số. Để đồng bộ hóa, tín hiệu lỗi sơ cấp sau khi khử cũng phải được trì hoãn. Để giải thích ngắn gọn, theo các phương án, bây giờ giả định rằng $P_{nc} = 0$.

Như trong công thức (8), theo các phương án, $\underline{\mathbf{D}}_1(\ell)$ được tính gần đúng bằng $\hat{\underline{\mathbf{D}}}_1(\ell)$ để tính toán các ước lượng của các tín hiệu tiếng vang thứ cấp:

$$\hat{\mathbf{d}}_n(\ell) = \mathbf{G} \hat{\underline{\mathbf{D}}}_1(\ell) \hat{\underline{\mathbf{a}}}_n(\ell), \quad n \in \{2, \dots, N\}. \quad (18)$$

Tín hiệu lỗi sau khi khử sau đó bằng

$$\mathbf{e}_n(\ell) = \mathbf{y}_n(\ell) - \mathbf{G} \hat{\underline{\mathbf{D}}}_1(\ell) \hat{\underline{\mathbf{a}}}_n(\ell), \quad n \in \{2, \dots, N\}, \quad (19)$$

và giảm thiểu hàm giá trị $\mathbf{E} \{ \mathbf{e}_n^H(\ell) \mathbf{e}_n(\ell) \}$ dẫn đến biểu thức sau cho các RETF tối ưu trong miền tần số:

$$\hat{\mathbf{a}}_n^{\text{opt}}(\ell) = \mathbf{E} \left\{ \hat{\mathbf{D}}_1^H(\ell) \mathbf{G} \hat{\mathbf{D}}_1(\ell) \right\}^{-1} \mathbf{E} \left\{ \hat{\mathbf{D}}_1^H(\ell) \mathbf{y}_n(\ell) \right\} = \underline{\Psi}_1(\ell)^{-1} \underline{\psi}_{1n}(\ell, k). \quad (20)$$

Do đó, phương pháp của Newton có dạng sau

$$\hat{\mathbf{a}}_n(\ell + 1) = \hat{\mathbf{a}}_n(\ell) + \tilde{\Psi}_1(\ell)^{-1} \bar{\mathbf{G}} \hat{\mathbf{D}}_1^H(\ell) \mathbf{e}_n(\ell), \quad (21)$$

nếu một công thức tạo ra thuật toán thích ứng trong miền tần số khối được phân vùng, với

$$\tilde{\Psi}_1(\ell + 1) = \beta \tilde{\Psi}_1(\ell) + (1 - \beta) \bar{\mathbf{G}} \hat{\mathbf{D}}_1^H(\ell) \mathbf{G} \hat{\mathbf{D}}_1(\ell). \quad (22)$$

Trong phương án tổng quát hơn, bộ lọc thứ hai 122; 322; 522 được tạo cấu hình để xác định cấu hình bộ lọc trong miền tần số khối được phân vùng theo

$$\hat{\mathbf{a}}_n(\ell + 1) = \hat{\mathbf{a}}_n(\ell) + \mathbf{C}_n \bar{\mathbf{G}} \hat{\mathbf{D}}_1^H(\ell) \mathbf{e}_n(\ell)$$

trong đó $\ell + 1$ biểu thị hệ số thời gian thứ hai, và trong đó ℓ biểu thị hệ số thời gian thứ nhất, và trong đó k biểu thị hệ số tần số, trong đó $\hat{\mathbf{a}}_n(\ell + 1)$ là cấu hình bộ lọc cho hệ số thời gian thứ hai, và trong đó $\hat{\mathbf{a}}_n(\ell)$ là cấu hình bộ lọc cho hệ số thời gian thứ nhất, trong đó $\hat{\mathbf{D}}_1(\ell)$ là ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất, trong đó $\mathbf{C}_n$ là ma trận bước kích thước, trong đó $\mathbf{e}_n(\ell)$ là kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất, và trong đó $\bar{\mathbf{G}}$ là ma trận ràng buộc tích chập tròn.

Sau đây, các khía cạnh thực hiện và đồng bộ hóa của các phương án được xem xét.

Cụ thể, sự mô tả chi tiết về các việc thực hiện phi nhân quả ($P_{nc} > 0$) theo các phương án được cung cấp.

Do tính phi nhân quả của các bộ lọc RETF, nên cần phải làm trễ các tín hiệu micro thứ cấp, như được mô tả trên Fig.3, để đảm bảo rằng các hệ số phi nhân quả

cũng được mô hình hóa bởi các RETF ước lượng trong (PB) miền tần số. Có hai chiến lược để làm như vậy:

- Đệm các tín hiệu đầu vào cho các micrô thứ cấp trên cơ sở mẫu, ví dụ, trong miền thời gian. Điều này cho phép người dùng giữ độ trễ thấp nhất có thể. Tuy nhiên, để đồng bộ hóa, tín hiệu sơ cấp sau khi khử phải được làm trễ tương ứng, và điều này có nghĩa là phải chuyển $\mathbf{e}_1(\ell)$ trở lại miền thời gian.
- Đệm các tín hiệu đầu vào cho các micrô thứ cấp trong miền tần số. Theo đó, những tín hiệu này phải được làm trễ trên cơ sở khung, dẫn đến độ trễ cao hơn độ trễ được đưa vào trong miền thời gian. Ưu điểm của lựa chọn này nằm ở chỗ không cần thiết phải chuyển tín hiệu sơ cấp sau khi khử sang miền thời gian. Do đó, bộ khử nhiễu đa kênh có thể được thao tác trực tiếp trong miền tần số đến bộ xử lý sau.

Sau đây, hai cách thực hiện có thể có được mô tả chi tiết.

Trước hết, các phương án với tín hiệu micrô thứ cấp được làm trễ được xem xét.

Từ công thức (17), rõ ràng là độ trễ của P_{nc} phân vùng được thêm vào tất cả các tín hiệu micrô thứ cấp cho phép ước lượng tính phi nhân quả tiềm tàng của các RETF $\mathbf{a}_n(p)$. Việc thực hiện tương ứng tương tự như việc thực hiện được mô tả trên Fig.3 với D là bội số nguyên của kích thước phân vùng Q . Theo cách này, P_{nc} phân vùng đầu tiên của bộ lọc thích ứng được sử dụng để mô hình hóa $Q \cdot P_{nc}$ hệ số RETF phi nhân quả. Với cách tiếp cận đơn giản này, cần ít nhất 2 phân vùng để ước lượng các hệ số RETF nhân quả cũng như phi nhân quả, trong trường hợp đơn giản này, phân vùng bộ lọc thứ nhất mô hình hóa các hệ số phi nhân quả của $\mathbf{a}_n(-1)$.

Bây giờ, các phương án với ràng buộc gradien đối xứng được xem xét.

Sự cải tiến của phương pháp được mô tả ở trên xem xét việc biến đổi ràng buộc gradien $\overline{\mathbf{G}}$ để giữ lại tối đa $\frac{Q}{2}$ hệ số nhân quả cũng như $\frac{Q}{2}$ hệ số phi nhân quả của sự tương quan tròn miền thời gian. Để đạt được mục đích này, theo các phương án, ràng buộc $\overline{\mathbf{G}}$ từ công thức (16) được tính gần đúng theo cách này

$$\bar{\mathbf{G}}^{\text{sym}} = \mathbf{F} \text{diag}\{[1_{1 \times \frac{Q}{2}}, 0_{1 \times V}, 1_{1 \times \frac{Q}{2}}]\} \mathbf{F}^{-1} = \mathbf{F} \text{diag}\{\bar{\mathbf{g}}^{\text{sym}}\} \mathbf{F}^{-1}. \quad (23)$$

Để đảm bảo đầu ra không có bí danh sau khi lọc, ràng buộc tích chập trong công thức (14) cũng phải được biến đổi cho phù hợp:

$$\mathbf{G}^{\text{sym}} = \mathbf{F} \text{diag}\{\underbrace{[0, \dots, 0]_{\frac{Q}{2}}}, \underbrace{[1, \dots, 1]_V}, \underbrace{[0, \dots, 0]_{\frac{Q}{2}}}\} \mathbf{F}^{-1} = \mathbf{F} \text{diag}\{\mathbf{g}^{\text{sym}}\} \mathbf{F}^{-1}. \quad (24)$$

Lưu ý rằng các ràng buộc trên loại bỏ $\frac{Q}{2}$ mẫu trong quá khứ cũng như $\frac{Q}{2}$ mẫu đầu ra gần đây nhất của tích chập tròn để cung cấp đầu ra tích chập tuyến tính, điều này gây ra độ trễ của $\frac{Q}{2}$ mẫu trong các ước lượng của các tín hiệu tiếng vang thứ cấp.

Các ràng buộc đối xứng này không là gì ngoài các ràng buộc miền thời gian ban đầu $\mathbf{g}$ và $\bar{\mathbf{g}}$, được dịch chuyển theo vòng tròn bởi $\frac{Q}{2}$ mẫu. Do đó, các phép biểu diễn miền tần số tương ứng sau đó tương ứng là $\bar{\mathbf{G}}^{\text{sym}} = \mathbf{J}\bar{\mathbf{G}}$ và $\mathbf{G}^{\text{sym}} = \mathbf{J}\mathbf{G}$, trong đó ma trận không đổi

$$\mathbf{J} = \text{diag}\{[e^{j \cdot \frac{2\pi}{K} \frac{Q}{2} 0}, e^{j \cdot \frac{2\pi}{K} \frac{Q}{2} 1}, \dots, e^{j \cdot \frac{2\pi}{K} \frac{Q}{2} (K-1)}]\}, \quad (25)$$

là miền tần số tương đương với sự dịch chuyển tròn. Để triển khai thực tế, ma trận trên không được quan tâm vì các ràng buộc thường được áp dụng trong miền thời gian.

Tuy nhiên, ma trận tương tự $\mathbf{J}$ có thể được xác định để điều khiển các tín hiệu trong miền tần số, trước và sau các ràng buộc thông thường, có cùng lựa chọn các hệ số tuyến tính được cung cấp bởi các công thức (23) và (24). Để điều chỉnh trọng lượng mong muốn, sử dụng phương pháp của Newton, có thể được lấy là

$$\hat{\underline{\mathbf{a}}}_n^{\text{sym}}(\ell + 1) = \hat{\underline{\mathbf{a}}}_n^{\text{sym}}(\ell) + \tilde{\underline{\Psi}}_1(\ell)^{-1} \underline{\mathbf{J}}^H \left(\bar{\underline{\mathbf{G}}} \underline{\mathbf{J}} (\underline{\mathbf{D}}_1^H(\ell) \mathbf{e}_n(\ell)) \right). \quad (26)$$

Bằng cách sử dụng công thức trên theo các phương án, tính linh hoạt có được khi lựa chọn các hệ số tuyến tính được xác định theo định nghĩa của $\mathbf{J}$. Trong thực tế, $\mathbf{J}$ có thể được điều chỉnh cho các trường hợp rất cụ thể, ví dụ, nó có thể thực hiện sự

dịch chuyển ngắn hơn $\frac{Q}{2}$ làm giảm số lượng các hệ số phi nhân quả và do đó hệ thống bị trễ.

Bây giờ, chi tiết về các cách tiếp cận triển khai sử dụng PB-FDAF được cung cấp.

Việc lựa chọn sử dụng cách triển khai nào phụ thuộc vào ngữ cảnh ứng dụng. Rõ ràng là không có giả định về các vị trí micro nguồn tương đối, việc đưa ra độ trễ nhất định là cần thiết để đạt được đầu ra bộ lọc chất lượng cao. Bảng sau đây tóm tắt các phương pháp thực hiện được trình bày.

Cách tiếp cận	Độ trễ bổ sung	Số lượng phân vùng tối thiểu	Các lưu ý khi triển khai
Các tín hiệu micro thứ hai được làm trễ	$\geq Q$	2	PB-FDAF ban đầu
Ràng buộc của bộ lọc đối xứng	$\leq \frac{Q}{2}$	1	độ trễ được giảm

Sau đây, phép phân tích độ phức tạp được mô tả cho trường hợp cụ thể trong đó có 1 kênh sơ cấp và $N-1$ kênh thứ cấp.

Trước hết, miền thời gian được xem xét.

Sau đó, phép phân tích độ phức tạp mẫu được cung cấp liên quan tới các phép cộng và các phép nhân. Để làm được điều này, việc biểu thị độ dài của bộ lọc sơ cấp đã ước lượng là L và độ dài của $N-1$ bộ lọc thứ cấp đã ước lượng là P , và giả sử rằng các bộ lọc sơ cấp và thứ cấp được ước lượng bằng các kỹ thuật lọc thích ứng. Độ phức tạp trên mỗi mẫu tín hiệu đầu vào của bộ lọc thích ứng trong miền thời gian là

$$\mathcal{O}(AF) \approx 2M + M\mathcal{O}(\text{Update}, M),$$

trong đó $M \in \{L, N\}$, và độ phức tạp của phương trình điều chỉnh $\mathcal{O}(\text{Update}, M)$ phụ thuộc vào thuật toán thích ứng được sử dụng và, trong nhiều trường hợp, cũng phụ thuộc vào độ dài bộ lọc. Do đó, nếu N bộ lọc thích ứng được sử dụng song song (một bộ lọc cho mỗi micro), thì độ phức tạp thuật toán của AEC đa micro là $N\mathcal{O}(AF)$.

Phương pháp đề xuất có khả năng giảm độ phức tạp thuật toán bằng cách giảm độ dài của các bộ lọc thích ứng. Việc giảm độ phức tạp thuật toán sau đó được cho theo tỷ lệ

$$\frac{\mathcal{O}(\text{Proposed})}{N\mathcal{O}(\text{AF})} = \frac{[2L + L\mathcal{O}(\text{Update}, L)] + (N-1)[2P + P\mathcal{O}(\text{Update}, P)]}{N[2L + L\mathcal{O}(\text{Update}, L)]} \quad (27)$$

Nói chung, nếu sử dụng cùng một thuật toán thích ứng để ước lượng cả bộ lọc sơ cấp và các bộ lọc thứ cấp, thì tỷ lệ này được cho bởi

$$\frac{\mathcal{O}(\text{Proposed})}{N\mathcal{O}(\text{AF})} \approx \frac{1}{N} + \frac{N-1}{N} \frac{P}{L} \left[\frac{2 + \mathcal{O}(\text{Update}, P)}{2 + \mathcal{O}(\text{Update}, L)} \right] \quad (28)$$

Ví dụ đơn giản nhất sẽ là: nếu thuật toán bình phương trung bình tối thiểu (least-mean squares - LMS) (xem ví dụ tài liệu [2]) được sử dụng cho các bộ khử tiếng vang sơ cấp và thứ cấp, $\mathcal{O}(\text{Update})=1$ độc lập với độ dài của bộ lọc, và tỷ lệ được cho bởi

$$\frac{\mathcal{O}(\text{Proposed})}{N\mathcal{O}(\text{AF})} \approx \frac{1}{N} + \frac{N-1}{N} \frac{P}{L} \quad (29)$$

Nếu các bộ lọc thích ứng khác nhau được sử dụng để ước lượng các bộ lọc sơ cấp và thứ cấp thì độ phức tạp tính toán của các thuật toán riêng lẻ phải được xem xét cẩn thận.

Bây giờ, miền STFT được xem xét.

Sau đây, độ phức tạp về các phép cộng và các phép nhân được phân tích. Để đạt được mục đích này, trước hết xem xét độ phức tạp trên mỗi phân vùng của bộ lọc thích ứng trong miền STFT, đó là

$$\mathcal{O}(\text{AF}) = \frac{3}{Q} \mathcal{O}(\text{FFT}) + \underbrace{\mathcal{O}(\text{CplxMult}) + 2K}_{\text{filtering \& cancellation}} + \mathcal{O}(\text{Update}),$$

trong đó $Q \in \{P, B\}$, $\mathcal{O}(\text{FFT}) \approx 2K \log_2(K) - 4K$ là độ phức tạp của biến đổi Fourier nhanh (fast Fourier transform - FFT), $\mathcal{O}(\text{CplxMult}) = 6K$ là độ phức tạp của phép nhân phức của độ dài K (xem ví dụ tài liệu [30]) và độ phức tạp của phương trình điều chỉnh $\mathcal{O}(\text{Update})$ phụ thuộc vào thuật toán thích ứng được sử dụng. Do đó,

nếu N bộ lọc thích ứng được sử dụng song song (một bộ lọc cho mỗi micrô), thì độ phức tạp thuật toán của AEC đa micrô trên mỗi phân vùng là $NO(AF)$.

Phương pháp đã đề xuất có khả năng làm giảm độ phức tạp thuật toán nếu $P < B$. Việc giảm độ phức tạp thuật toán sau đó được cho theo tỷ lệ

$$\frac{\mathcal{O}(\text{Proposed})}{NB\mathcal{O}(AF)} = \frac{B\mathcal{O}(AF) + (N-1)P\mathcal{O}(AF)}{NB\mathcal{O}(AF)}.$$

Do đó, nếu sử dụng cùng một bộ lọc thích ứng cho các bộ khử tiếng vang sơ cấp và thứ cấp, thì tỷ lệ này được cho bởi

$$\frac{\mathcal{O}(\text{Proposed})}{NB\mathcal{O}(AF)} \approx \frac{1}{N} + \frac{N-1}{N} \frac{P}{B}.$$

Nếu các bộ lọc thích ứng khác nhau được sử dụng cho việc ước lượng AETF và RETF, thì độ phức tạp tính toán của các thuật toán riêng lẻ phải được xem xét cẩn thận.

Các ứng dụng cụ thể của các phương án có thể, ví dụ, triển khai giải pháp có độ phức tạp thấp cho MC-AEC cho các ứng dụng sau:

- Điện thoại thông minh, máy tính bảng và máy tính cá nhân.
- Trợ lý kích hoạt bằng giọng nói, loa thông minh và thiết bị nhà thông minh.
- Tivi thông minh.

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 xử lý của phương pháp cũng thể hiện mô tả của khối hoặc mục tương ứng hoặc đặc điểm của thiết bị tương ứng. Một số hoặc tất cả các bước phương pháp có thể được thực hiện (hoặc sử dụng) bởi thiết bị phần cứng, như ví dụ, bộ vi xử lý, máy tính có thể lập trình hoặc mạch điện tử. Trong một số phương án, một hoặc nhiều bước phương pháp trong số các bước phương pháp quan trọng nhất có thể được thực hiện bởi các thiết bị như vậy.

Tùy thuộc vào các yêu cầu thực hiện nhất định, các phương án của sáng chế có thể được thực hiện trong phần cứng hoặc phần mềm hoặc ít nhất là một phần trong phần cứng hoặc ít nhất là một phần trong phần mềm. Phương án có thể được thực hiện bằng cách 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 trong tài liệu này 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 các phương pháp được mô tả trong tài liệu này, được lưu trữ trên vật mang có thể đọc được bằng máy.

Nói cách khác, một phương án của phương pháp theo sáng chế là, chương trình máy tính có mã chương trình để thực hiện một trong các phương pháp được mô tả trong tài liệu này, khi chương trình máy tính chạy trên máy tính.

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) bao gồm chương trình máy tính được ghi lại trên đó để thực hiện một trong các phương pháp được mô tả trong tài liệu này. Vật mang dữ liệu, vật ghi lưu trữ số hoặc vật ghi được ghi lại có dạng hữu hình điển hình và/hoặc không chuyển tiếp.

Do đó, phương án khác theo phương pháp của sáng chế là dòng dữ liệu hoặc chuỗi các tín hiệu biểu diễn chương trình máy tính để thực hiện một trong các phương

pháp được mô tả trong tài liệu này. Dòng dữ liệu hoặc chuỗi tín hiệu có thể, ví dụ, được tạo cấu hình để được truyền thông qua kết nối truyền thông dữ liệu, ví dụ, thông qua Liên mạng.

Phương án khác bao gồm các phương tiện xử lý, ví dụ, máy tính, hoặc thiết bị logic có thể lập trình, được tạo cấu hình hoặc được làm thích ứng với một trong các phương pháp được mô tả trong tài liệu này.

Phương pháp nữa bao gồm máy tính có chương trình máy tính được cài đặt trên đó để thực hiện một trong các phương pháp được mô tả trong tài liệu này.

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

Theo một số phương án, thiết bị logic lập trình được (ví dụ, mảng cổng có thể 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 trong tài liệu này. Theo một số phương án, mảng cổng có thể lập trình được dạng trường có thể liên kết với bộ vi xử lý để thực hiện một trong các phương pháp được mô tả trong tài liệu này. Thông thường, các phương pháp tốt hơn là được thực hiện bởi thiết bị phần cứng bất kỳ.

Thiết bị được mô tả trong tài liệu này có thể được thực hiện bằng cách sử dụng thiết bị phần cứng, hoặc sử dụng máy tính, hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

Các phương pháp được mô tả trong tài liệu này có thể được thực hiện bằng cách sử dụng thiết bị phần cứng, hoặc sử dụng máy tính, hoặc sử dụng kết hợp thiết bị phần cứng và máy tính.

Các phương án nêu trên chỉ đơn thuần minh họa các nguyên lý của sáng chế. Sẽ được hiểu rằng các biến thể và các biến đổi về các cách bố trí và các chi tiết được mô tả trong tài liệu này sẽ là rõ ràng đối với người có trình độ trung bình trong lĩnh vực kỹ

thuật tương ứng. Do đó, mục đích của sáng chế chỉ bị giới hạn bởi phạm vi của các yêu cầu bảo hộ của sáng chế và không bởi phần mô tả chi tiết cụ thể được thể hiện theo cách mô tả và giải thích của các phương án trong tài liệu này.

YÊU CẦU BẢO HỘ

1. Thiết bị khử nhiễu đa kênh trong tín hiệu âm thanh nhận được bao gồm hai hoặc nhiều hơn hai kênh âm thanh nhận được để thu được tín hiệu âm thanh được biến đổi bao gồm hai hoặc nhiều hơn hai kênh âm thanh được biến đổi, trong đó thiết bị bao gồm:

bộ lọc thứ nhất (112; 312; 512) được tạo cấu hình để tạo ước lượng thứ nhất của tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu,

bộ khử nhiễu thứ nhất (114; 314; 514) được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được sử dụng ước lượng thứ nhất của tín hiệu nhiễu thứ nhất,

bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để tạo ước lượng thứ hai của tín hiệu nhiễu thứ hai sử dụng ước lượng thứ nhất của tín hiệu nhiễu thứ nhất, và

bộ khử nhiễu thứ hai (124; 324; 524) được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai của tín hiệu nhiễu thứ hai.

2. Thiết bị theo điểm 1,

trong đó ước lượng thứ nhất của tín hiệu nhiễu thứ nhất là ước lượng thứ nhất của tín hiệu tiếng vang âm thanh thứ nhất,

trong đó ước lượng thứ hai của tín hiệu nhiễu thứ hai là ước lượng thứ hai của tín hiệu tiếng vang âm thanh thứ hai,

trong đó bộ khử nhiễu thứ nhất (114; 314; 514) được tạo cấu hình để thực hiện việc khử tiếng vang âm thanh trên kênh âm thanh nhận được thứ nhất để thu được kênh âm thanh được biến đổi thứ nhất, và

trong đó bộ khử nhiễu thứ hai (124; 324; 524) được tạo cấu hình để thực hiện việc khử tiếng vang âm thanh trên kênh âm thanh nhận được thứ hai để thu được kênh âm thanh được biến đổi thứ hai.

3. Thiết bị theo điểm 1 hoặc điểm 2, trong đó hai hoặc nhiều hơn hai kênh âm thanh nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi là các kênh của miền biến đổi, và trong đó tín hiệu tham chiếu và các tín hiệu nhiễu thứ nhất và thứ hai là các tín hiệu của miền biến đổi.

4. Thiết bị theo một điểm trong số các điểm trên, trong đó hai hoặc nhiều hơn hai kênh âm thanh nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi là các kênh của miền biến đổi Fourier thời gian ngắn, và trong đó tín hiệu tham chiếu và các tín hiệu nhiễu thứ nhất và thứ hai là các tín hiệu của miền biến đổi Fourier thời gian ngắn.

5. Thiết bị theo một điểm trong số các điểm trên,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào kênh âm thanh nhận được thứ hai, và

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định ước lượng thứ hai của tín hiệu nhiễu thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và tùy thuộc vào cấu hình bộ lọc.

6. Thiết bị theo điểm 5,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc bằng cách giảm thiểu hàm giá trị hoặc bằng cách giảm thiểu chỉ tiêu lỗi.

7. Thiết bị theo điểm 5 hoặc điểm 6,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc $\hat{\mathbf{a}}_n^{opt}(\ell, k)$ theo

$$\hat{\mathbf{a}}_n^{opt}(\ell, k) = \mathbf{\Psi}_1(\ell, k)^{-1} \psi_{1n}(\ell, k),$$

trong đó $\mathbf{\Psi}_1(\ell, k)$ là ma trận hiệp phương sai của $\hat{\mathbf{d}}_1(\ell, k)$,

trong đó $\psi_{1n}(\ell, k)$ là vectơ tương quan chéo giữa $\hat{\mathbf{d}}_1(\ell, k)$ và $Y_n(\ell, k)$,

trong đó $\hat{\mathbf{d}}_1(\ell, k)$ biểu thị ước lượng thứ nhất của tín hiệu nhiễu thứ nhất,

trong đó $Y_n(\ell, k)$ biểu thị kênh âm thanh nhận được thứ hai,

trong đó ℓ biểu thị hệ số thời gian, và trong đó k biểu thị hệ số tần số.

8. Thiết bị theo một trong số các điểm từ 1 đến 4,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai tùy thuộc vào cấu hình bộ lọc cho hệ số thời gian thứ nhất trước hệ số thời gian thứ hai theo thời gian, tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất, và tùy thuộc vào mẫu kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất.

9. Thiết bị theo điểm 8,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai theo

$$\hat{\mathbf{a}}_n(\ell + 1, k) = \hat{\mathbf{a}}_n(\ell, k) + \mathbf{C}_n(\ell, k) \hat{\mathbf{d}}_1(\ell, k) E_n^*(\ell, k)$$

trong đó $\ell + 1$ biểu thị hệ số thời gian thứ hai, và trong đó ℓ biểu thị hệ số thời gian thứ nhất, và trong đó k biểu thị hệ số tần số,

trong đó $\hat{\mathbf{a}}_n(\ell + 1, k)$ là cấu hình bộ lọc cho hệ số thời gian thứ hai, và trong đó $\hat{\mathbf{a}}_n(\ell, k)$ là cấu hình bộ lọc cho hệ số thời gian thứ nhất,

trong đó $\hat{\mathbf{d}}_1(\ell, k)$ là ước lượng thứ nhất của tín hiệu nhiễu thứ nhất cho hệ số thời gian thứ nhất,

trong đó $E_n^*(\ell, k)$ là kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất,

trong đó $\mathbf{C}_n(\ell, k)$ là ma trận bước kích thước.

10. Thiết bị theo một trong số các điểm từ 1 đến 3,

trong đó hai hoặc nhiều hơn hai kênh âm thanh nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi là các kênh của miền tần số khối được phân vùng, trong đó mỗi kênh âm thanh trong số hai hoặc nhiều hơn hai kênh âm thanh

nhận được và hai hoặc nhiều hơn hai kênh âm thanh được biến đổi bao gồm nhiều phân vùng, và

trong đó tín hiệu tham chiếu và các tín hiệu nhiều thứ nhất và thứ hai là các tín hiệu của miền tần số khối được phân vùng, trong đó mỗi tín hiệu trong số tín hiệu tham chiếu và các tín hiệu nhiều thứ nhất và thứ hai bao gồm nhiều phân vùng.

11. Thiết bị theo điểm 10,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiều thứ nhất và tùy thuộc vào kênh âm thanh nhận được thứ hai,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định ước lượng thứ hai của tín hiệu nhiều thứ hai tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiều thứ nhất và tùy thuộc vào cấu hình bộ lọc,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc cho hệ số thời gian thứ hai tùy thuộc vào cấu hình bộ lọc cho hệ số thời gian thứ nhất trước hệ số thời gian thứ hai theo thời gian, tùy thuộc vào ước lượng thứ nhất của tín hiệu nhiều thứ nhất cho hệ số thời gian thứ nhất, và tùy thuộc vào mẫu kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất.

12. Thiết bị theo điểm 11,

trong đó bộ lọc thứ hai (122; 322; 522) được tạo cấu hình để xác định cấu hình bộ lọc trong miền tần số khối được phân vùng theo

$$\hat{\mathbf{a}}_n(\ell + 1) = \hat{\mathbf{a}}_n(\ell) + \mathbf{C}_n \bar{\mathbf{G}} \hat{\mathbf{D}}_1^H(\ell) \mathbf{e}_n(\ell)$$

trong đó $\ell + 1$ biểu thị hệ số thời gian thứ hai, và trong đó ℓ biểu thị hệ số thời gian thứ nhất, và trong đó k biểu thị hệ số tần số,

trong đó $\hat{\mathbf{a}}_n(\ell + 1)$ là cấu hình bộ lọc cho hệ số thời gian thứ hai, và trong đó $\hat{\mathbf{a}}_n(\ell)$ là cấu hình bộ lọc cho hệ số thời gian thứ nhất,

trong đó $\hat{\mathbf{D}}_1(\ell)$ là ước lượng thứ nhất của tín hiệu nhiều thứ nhất cho hệ số thời gian thứ nhất,

trong đó C_n là ma trận bước kích thước,

trong đó $e_n(\ell)$ là kênh âm thanh được biến đổi thứ hai cho hệ số thời gian thứ nhất, và

trong đó $\overline{G}$ là ma trận ràng buộc tích chập tròn.

13. Thiết bị theo một điểm trong số các điểm trên,

trong đó tín hiệu âm thanh nhận được bao gồm ba hoặc nhiều hơn ba kênh âm thanh nhận được, và trong đó tín hiệu âm thanh được biến đổi bao gồm ba hoặc nhiều hơn ba kênh âm thanh được biến đổi,

trong đó thiết bị còn bao gồm bộ lọc thứ ba (132) và bộ khử nhiễu thứ ba (134),

trong đó bộ lọc thứ ba (132) được tạo cấu hình để tạo ước lượng thứ ba của tín hiệu nhiễu thứ ba tùy thuộc vào ít nhất một trong các ước lượng thứ nhất của tín hiệu nhiễu thứ nhất và ước lượng thứ hai của tín hiệu nhiễu thứ hai,

trong đó bộ khử nhiễu thứ ba (134) được tạo cấu hình để tạo kênh âm thanh được biến đổi thứ ba $e_3(t)$ trong số ba hoặc nhiều hơn ba kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ ba $y_3(t)$ trong số ba hoặc nhiều hơn ba kênh âm thanh nhận được tùy thuộc vào ước lượng thứ ba $\hat{d}_3(t)$ của tín hiệu nhiễu thứ ba.

14. Phương pháp khử nhiễu đa kênh trong tín hiệu âm thanh nhận được bao gồm hai hoặc nhiều hơn hai kênh âm thanh nhận được để thu được tín hiệu âm thanh được biến đổi bao gồm hai hoặc nhiều hơn hai kênh âm thanh được biến đổi, trong đó phương pháp bao gồm các bước:

tạo ước lượng thứ nhất của tín hiệu nhiễu thứ nhất tùy thuộc vào tín hiệu tham chiếu,

tạo kênh âm thanh được biến đổi thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ nhất trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được sử dụng ước lượng thứ nhất của tín hiệu nhiễu thứ nhất,

tạo ước lượng thứ hai của tín hiệu nhiễu thứ hai sử dụng ước lượng thứ nhất của tín hiệu nhiễu thứ nhất, và

tạo kênh âm thanh được biến đổi thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh được biến đổi từ kênh âm thanh nhận được thứ hai trong số hai hoặc nhiều hơn hai kênh âm thanh nhận được tùy thuộc vào ước lượng thứ hai của tín hiệu nhiễu thứ hai.

15. Vật ghi có thể đọc được bằng máy tính chứa chương trình máy tính để thực hiện phương pháp theo điểm 14 khi được chạy trên máy tính hoặc bộ xử lý tín hiệu.

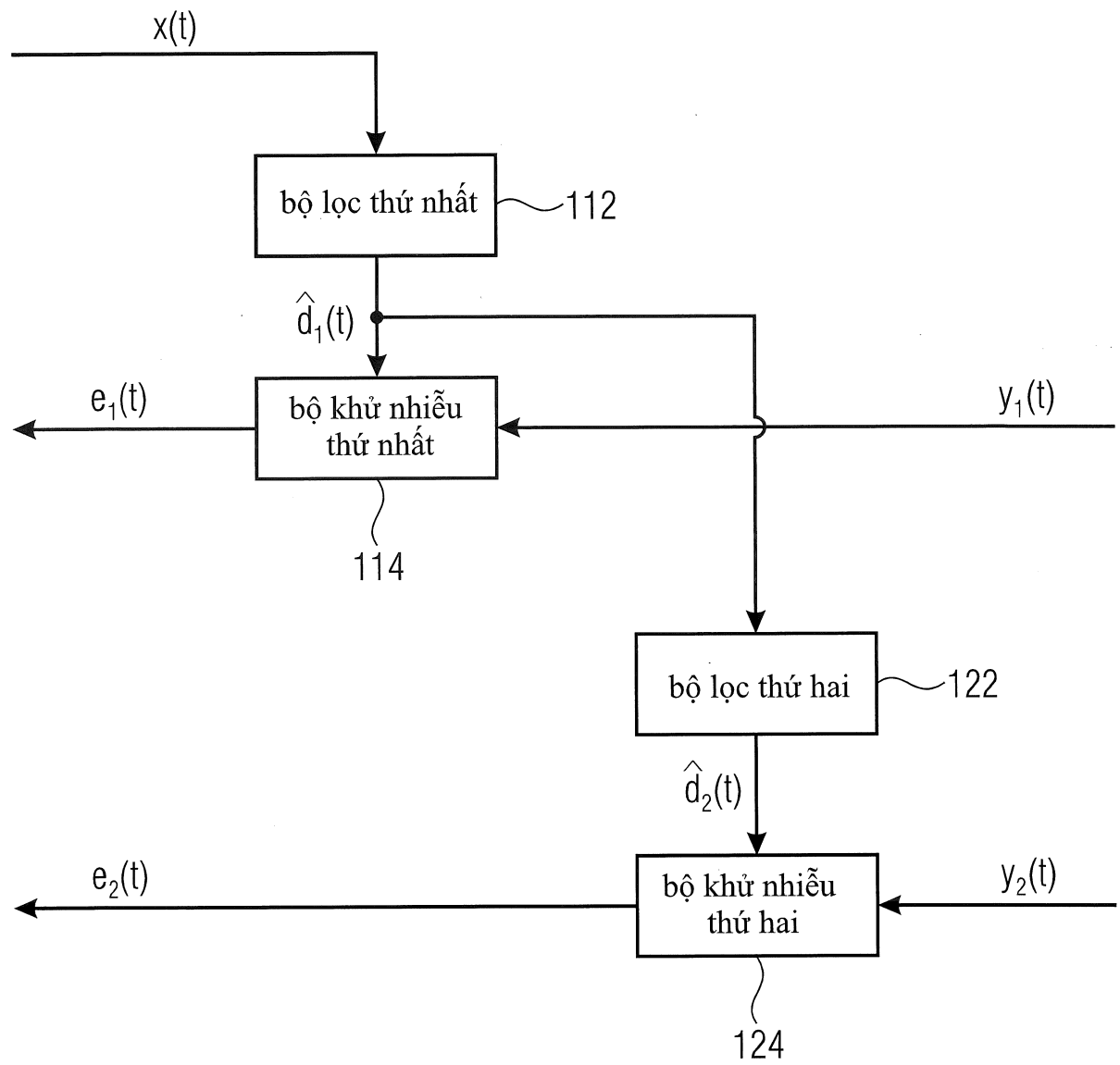


Fig. 1a

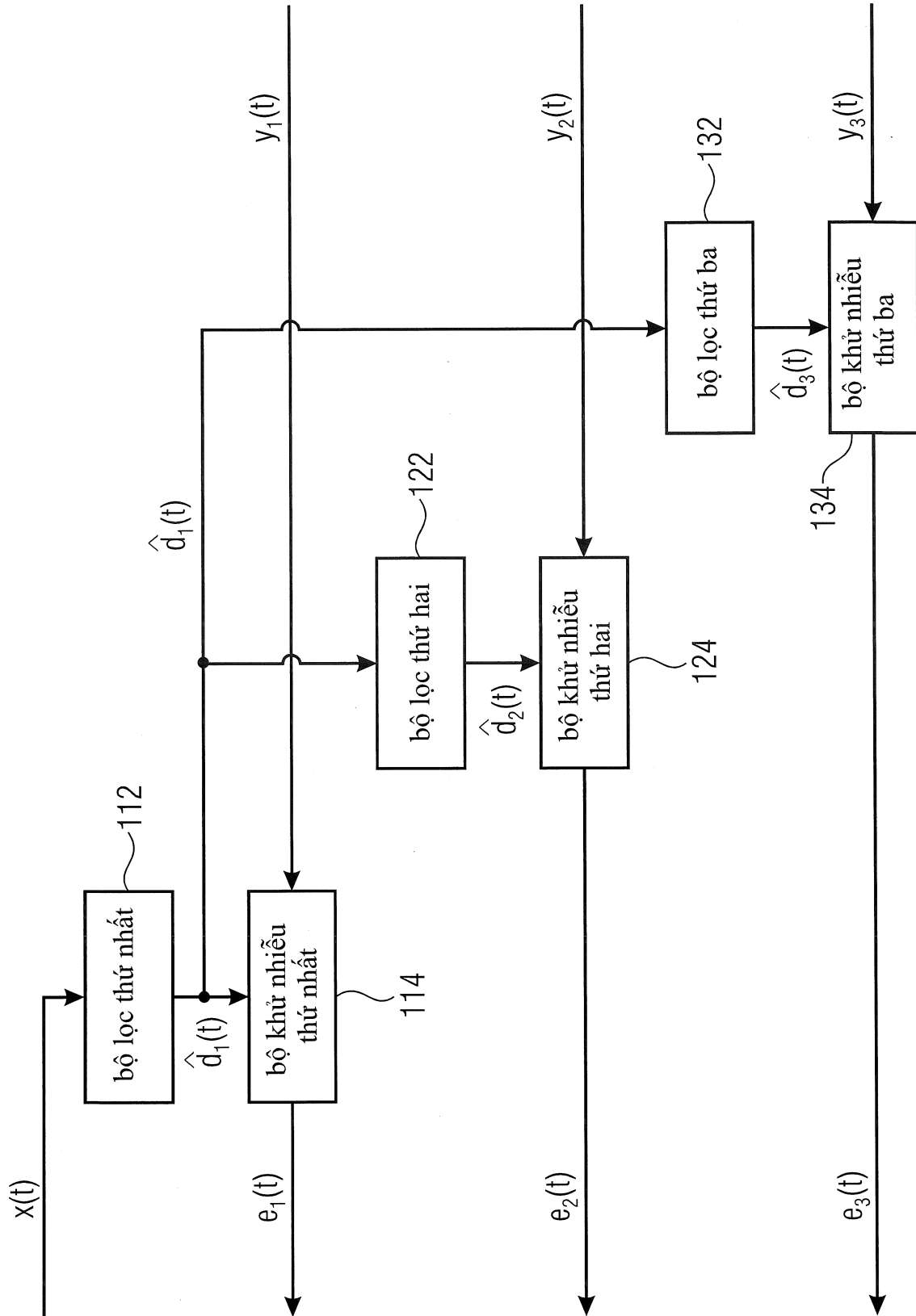


Fig. 1b

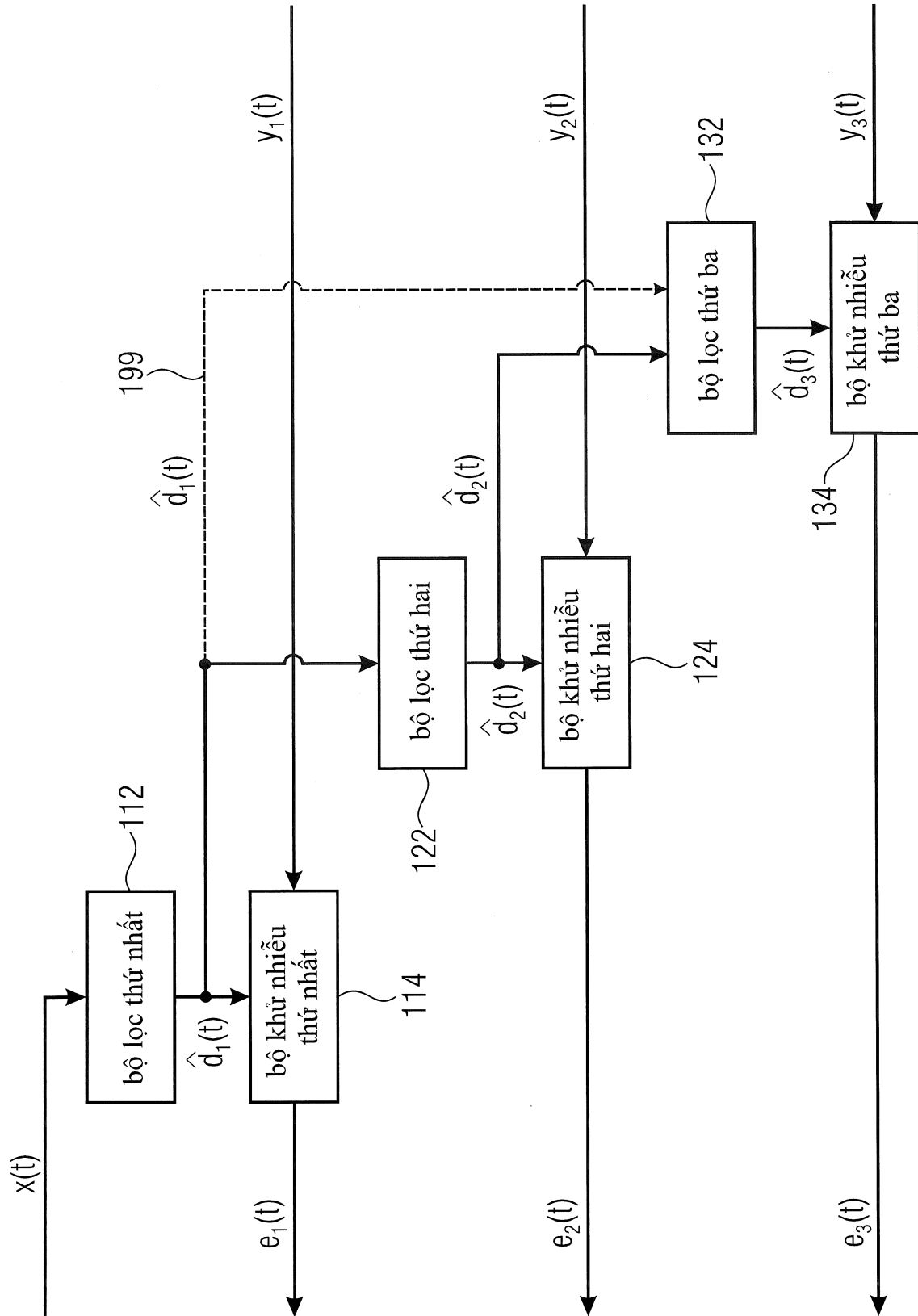


Fig. 1c

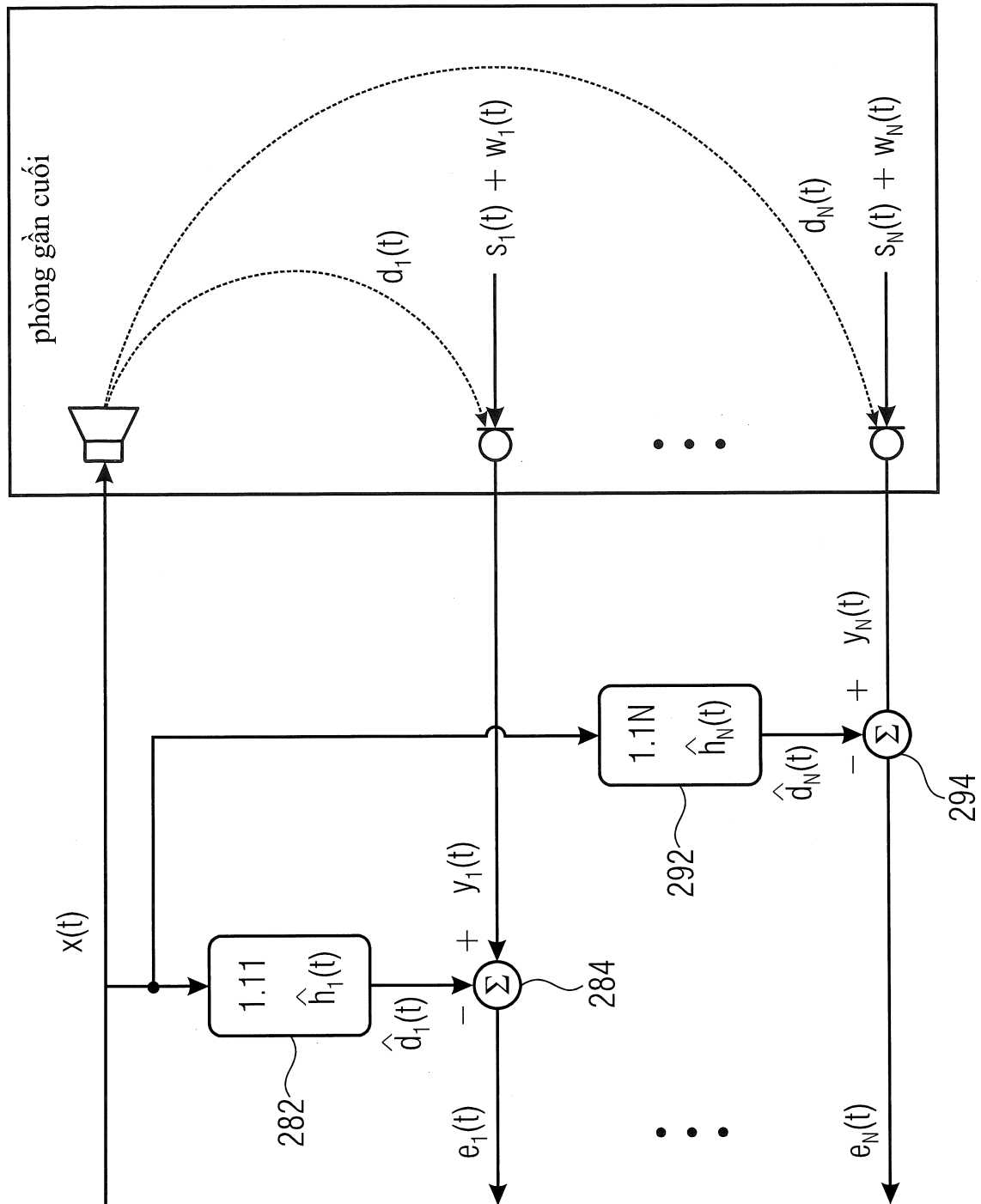


Fig. 2

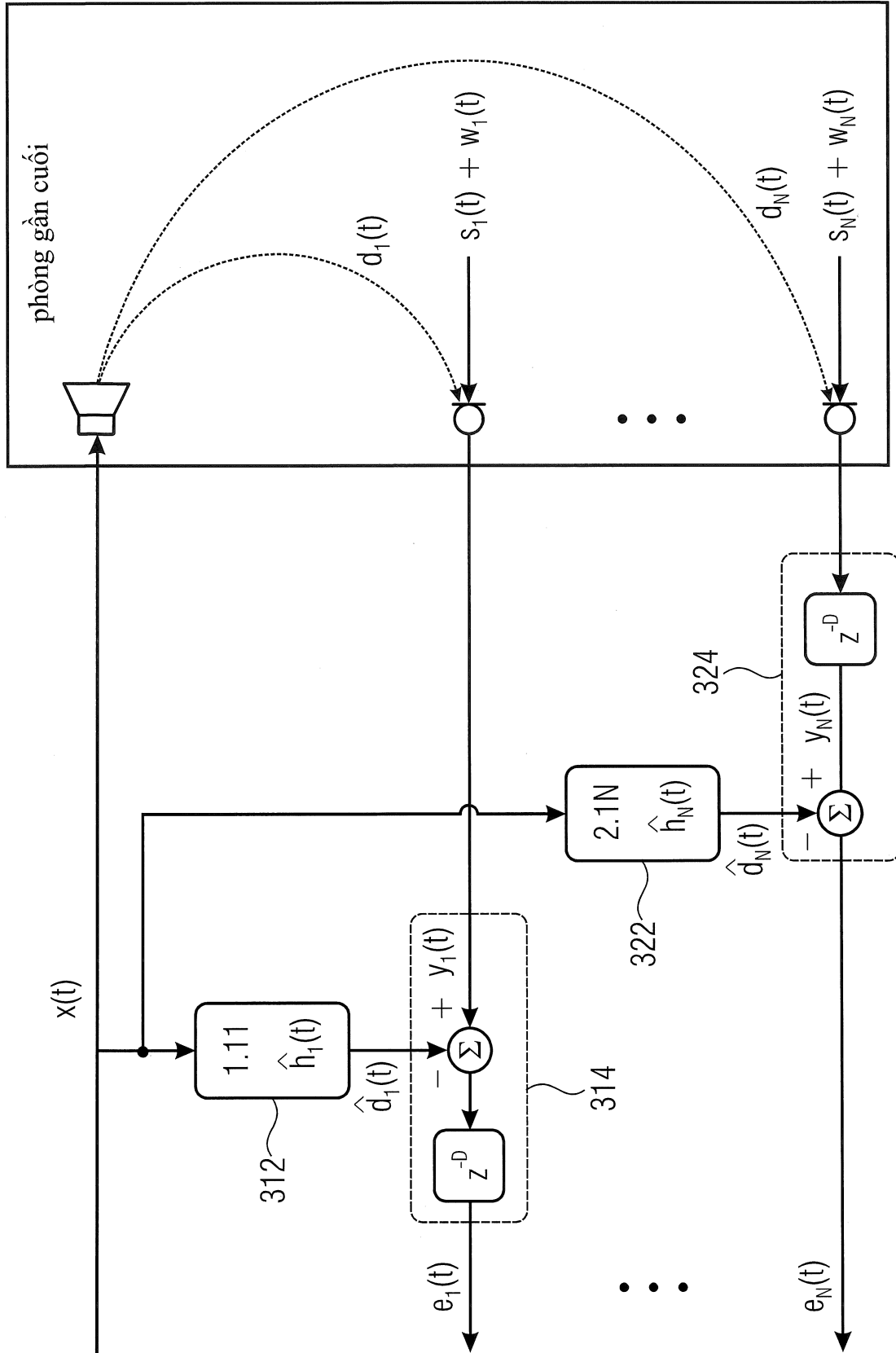


Fig. 3

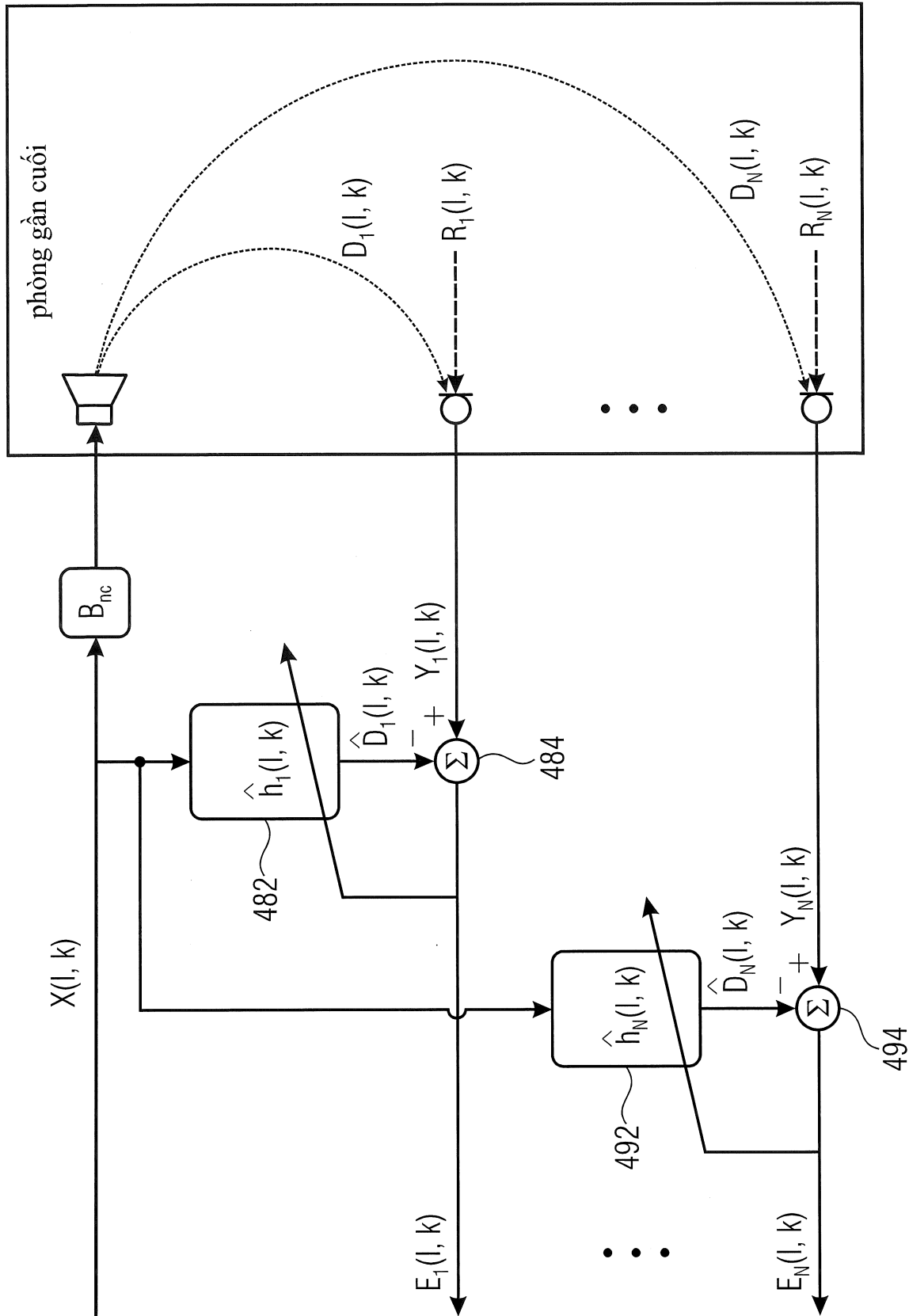


Fig. 4

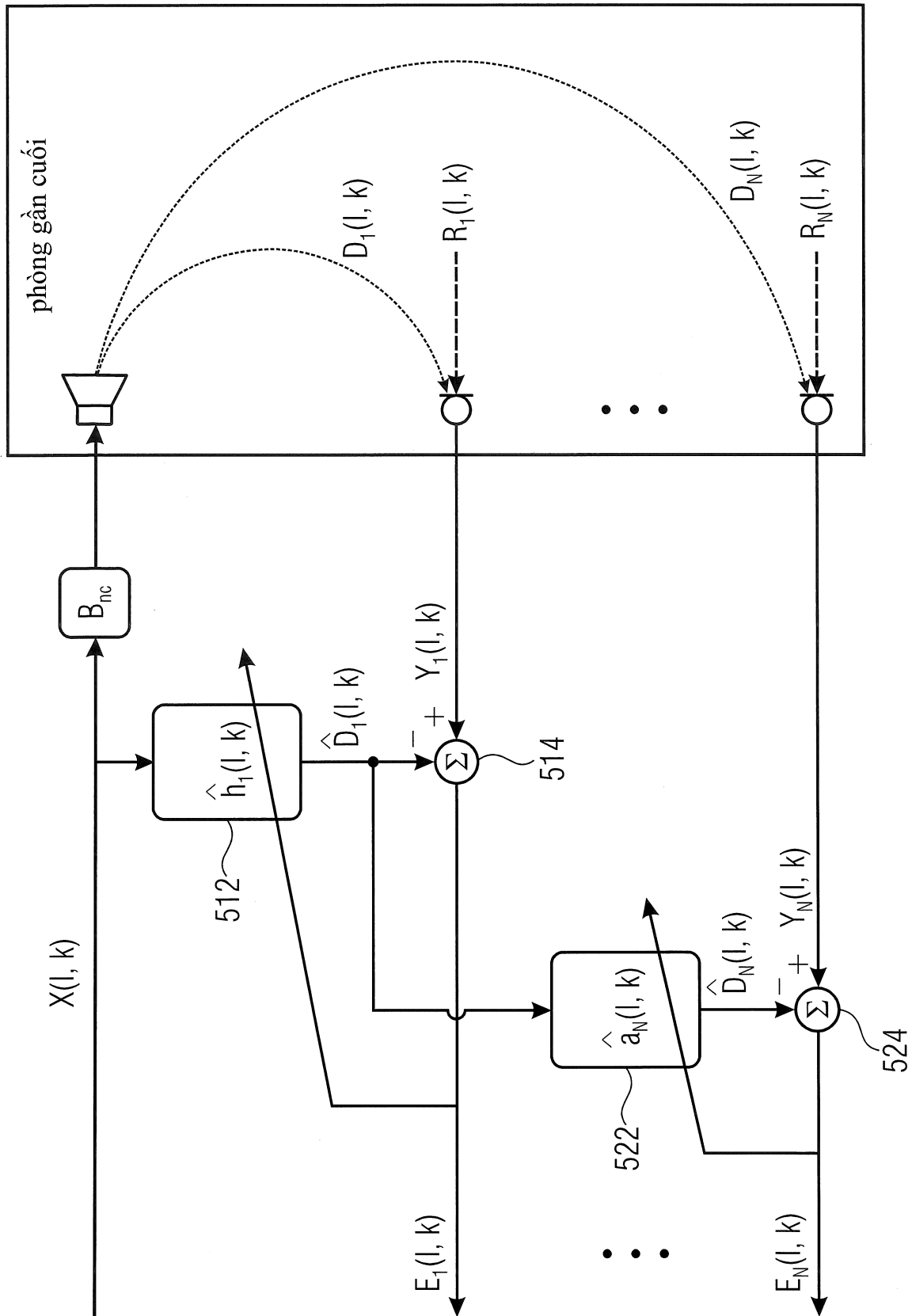


Fig. 5

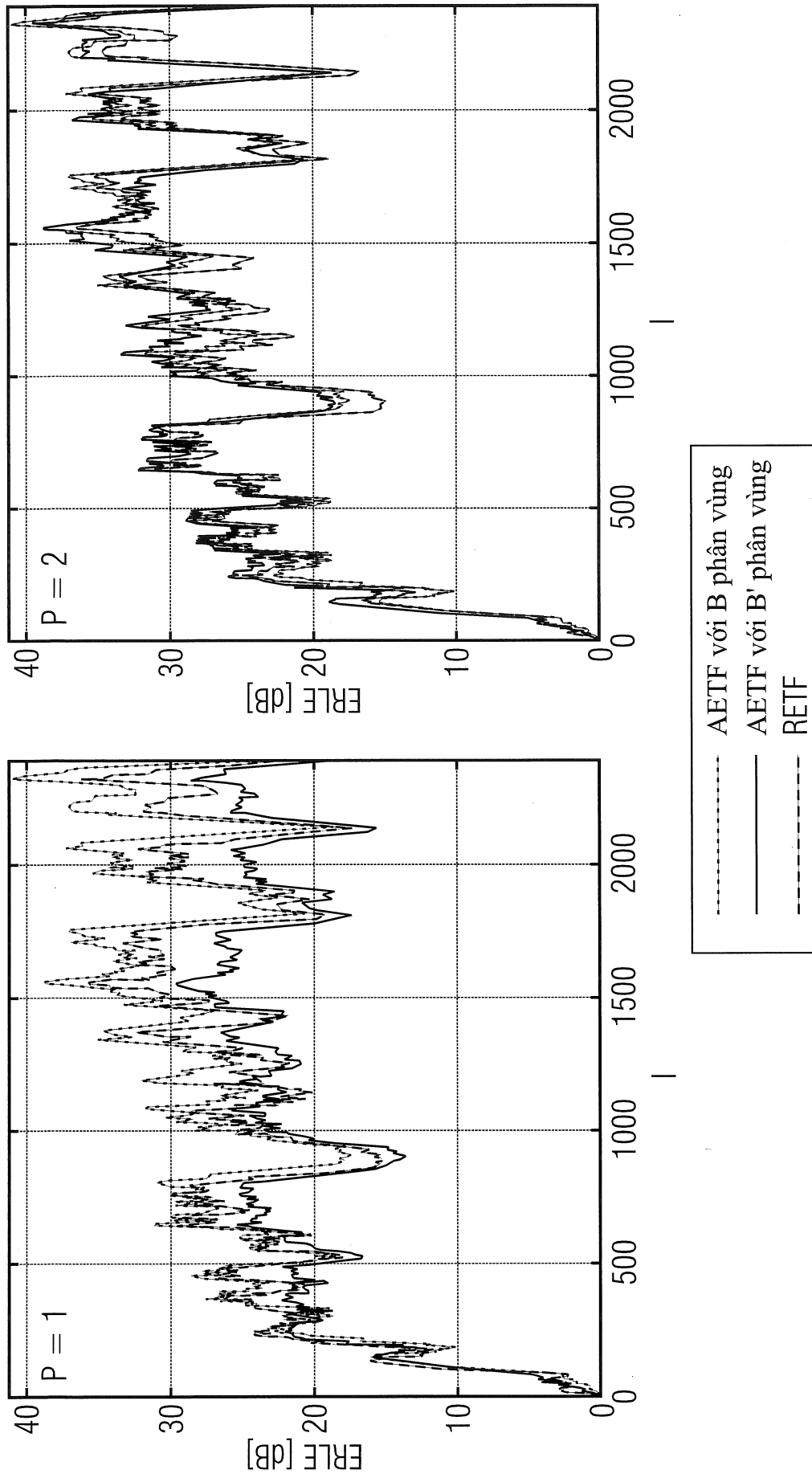


Fig. 6

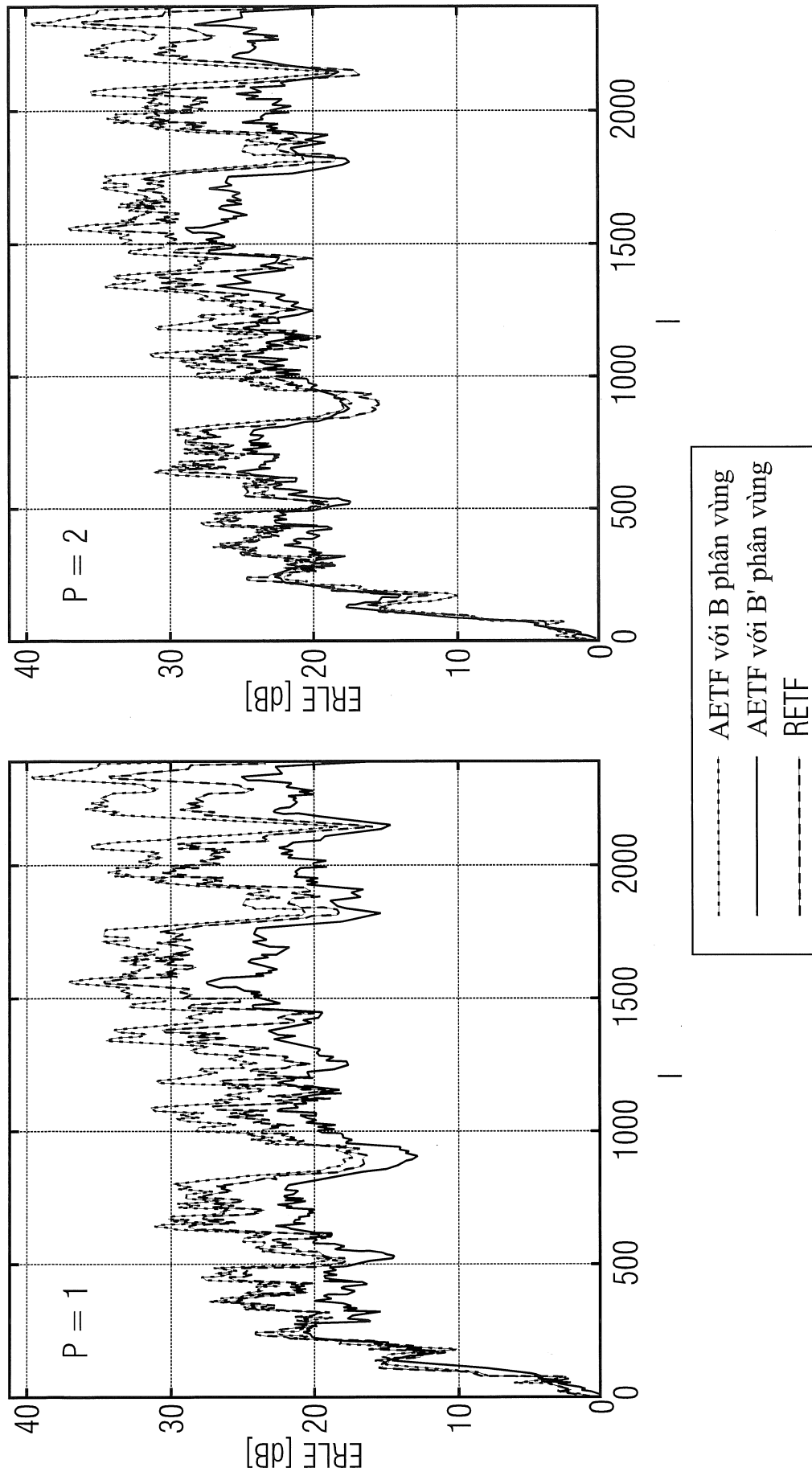


Fig. 7
(Phần 1)

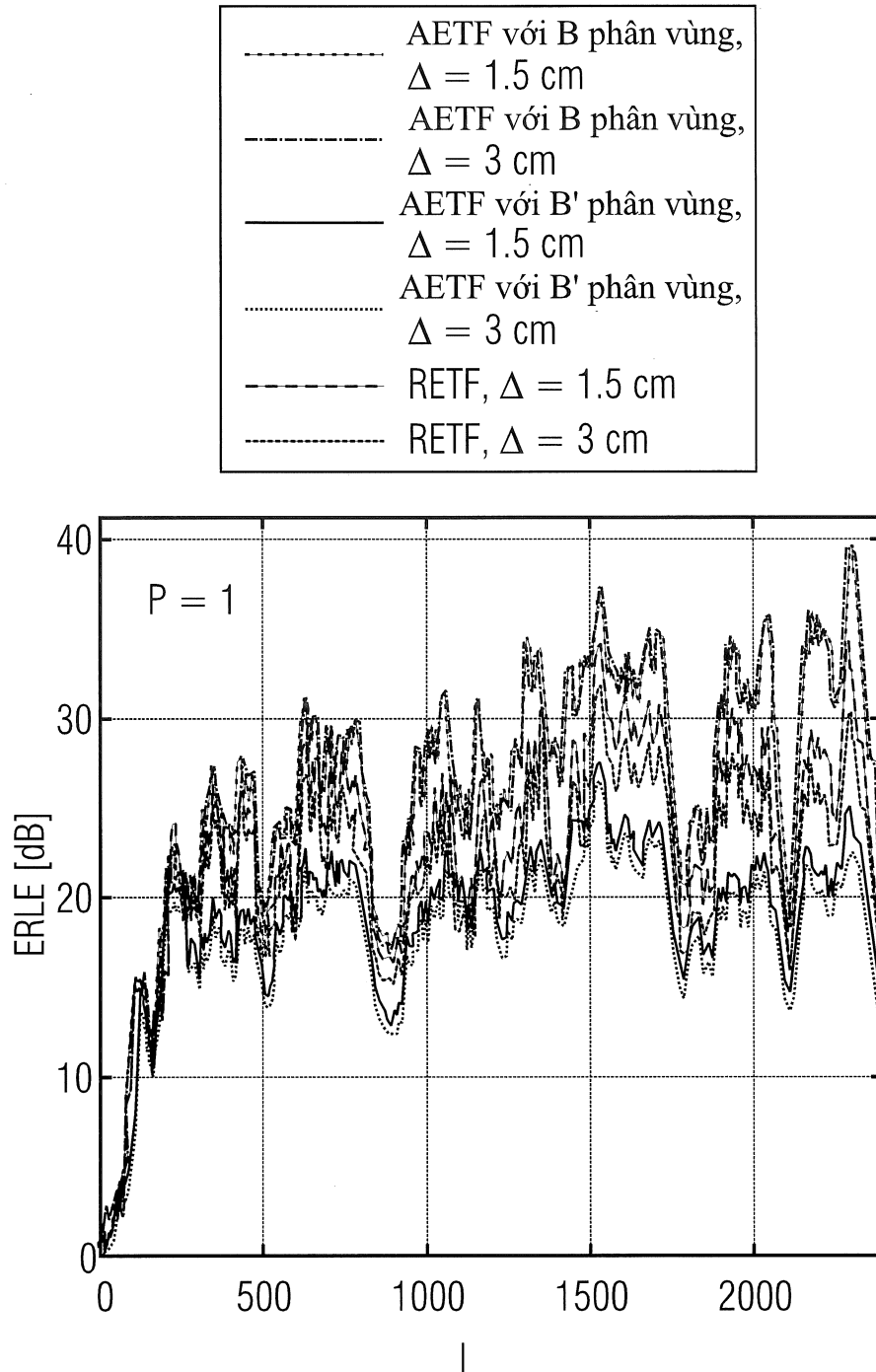


Fig. 7
(Phần 2)

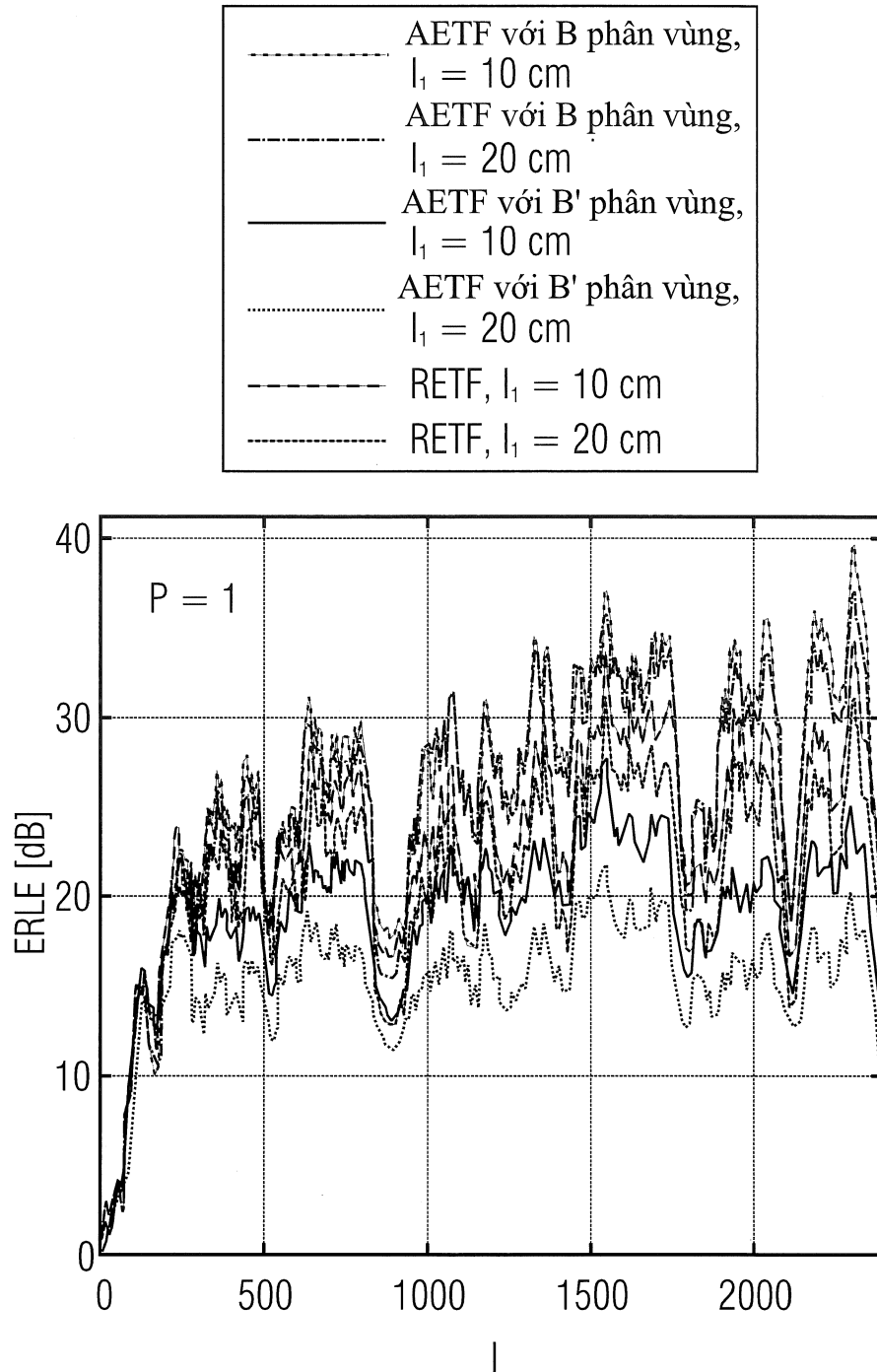


Fig. 7
(Phần 3)

12/12

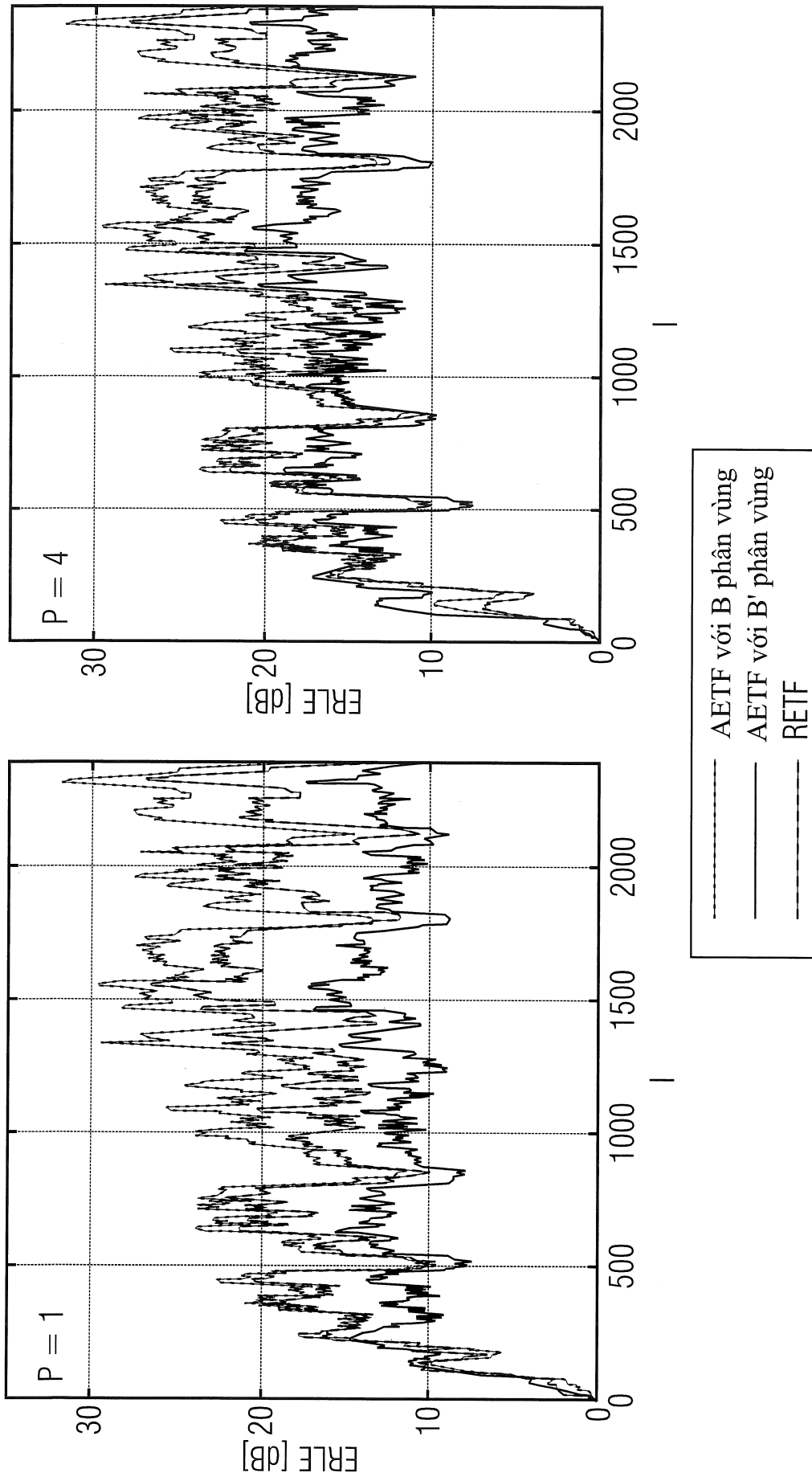


Fig. 8