



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) 
CỤC SỞ HỮU TRÍ TUỆ
1-0043603
(51)^{2020.01} **G10L 21/0308; G10L 25/60; G10L** (13) **B**
25/30

(21) 1-2021-04286 (22) 20/12/2019
(86) PCT/EP2019/086565 20/12/2019 (87) WO2020/127900 25/06/2020
(30) 18215707.3 21/12/2018 EP
(45) 25/02/2025 443 (43) 25/11/2021 404A
(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) UHLE, Christian (DE); TORCOLI, Matteo (IT); DISCH, Sascha (DE); PAULUS,
Jouni (FI); HERRE, Jürgen (DE); HELLMUTH, Oliver (DE); FUCHS, Harald (DE).
(74) CÔNG TY LUẬT TRÁCH NHIỆM HỮU HẠN AMBYS HÀ NỘI (AMBYS
HANOI)

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP TẠO RA TÍN HIỆU ÂM THANH ĐƯỢC PHÂN
TÁCH TỪ TÍN HIỆU ĐẦU VÀO ÂM THANH

(21) 1-2021-04286

(57) Sáng chế đề xuất thiết bị và phương pháp tạo ra tín hiệu âm thanh được phân tách từ tín hiệu đầu vào âm thanh. Thiết bị tạo ra tín hiệu âm thanh được phân tách từ tín hiệu âm thanh đầu vào âm thanh được đề xuất. Tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư. Phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích. Thiết bị bao gồm bộ phân tách nguồn (110), môđun xác định (120) và bộ xử lý tín hiệu (130). Bộ phân tách nguồn (110) được tạo cấu hình để xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích. Môđun xác định (120) được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được. Bộ xử lý tín hiệu (130) được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư.

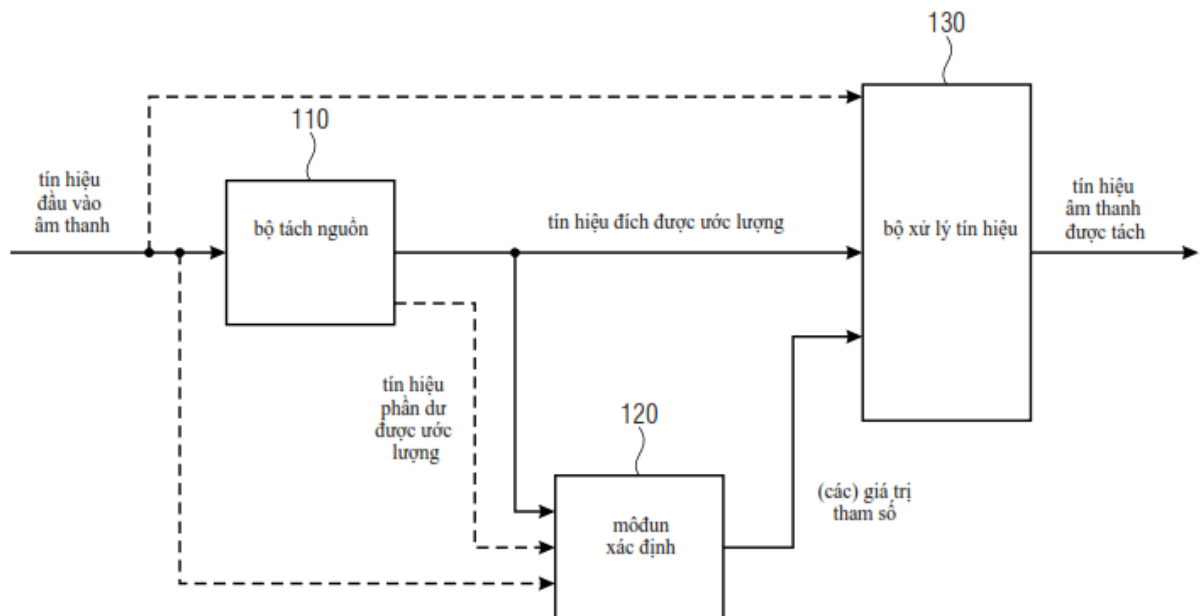


Fig. 1a

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến sự phân tách nguồn của các tín hiệu âm thanh, cụ thể, sự điều khiển thích ứng tín hiệu âm thanh của chất lượng âm của các tín hiệu đầu ra được phân tách, và cụ thể, đến thiết bị và phân tách phân tách nguồn sử dụng sự ước lượng và điều khiển chất lượng âm.

Tình trạng kỹ thuật của sáng chế

Trong sự phân tách nguồn, chất lượng của các tín hiệu đầu ra bị suy biến và sự suy biến này tăng một cách đơn điệu với sự suy giảm các tín hiệu nhiễu.

Sự phân tách nguồn các tín hiệu âm thanh đã được thực hiện trước đây.

Sự phân tách nguồn các tín hiệu âm thanh nhằm mục đích thu được tín hiệu đích $s(n)$ đã cho tín hiệu trộn $x(n)$,

$$x(n) = s(n) + b(n) \quad (1)$$

trong đó $b(n)$ bao gồm tất cả các tín hiệu nhiễu và sau đây được đề cập đến như "tín hiệu nhiễu". Đầu ra của sự phân tách $h(\cdot)$ là sự ước lượng tín hiệu đích $\hat{s}_n$,

$$\hat{s}(n) = h(x(n)) \quad (2)$$

và cách khác có khả năng sự ước lượng của tín hiệu nhiễu $\hat{b}_n$,

$$\hat{s}(n), \hat{b}(n) = h(x(n)) \quad (3)$$

Sự xử lý như vậy thường đưa các thành phần lạ trong tín hiệu đầu ra mà làm hỏng chất lượng âm. Sự suy biến chất lượng âm này tăng một cách đơn điệu với lượng phân tách, sự suy giảm các tín hiệu nhiễu. Nhiều ứng dụng không yêu cầu tổng sự phân tách mà sự nâng cao riêng phần, các âm nhiễu bị suy giảm nhưng vẫn có trong tín hiệu đầu ra.

Điều này có lợi ích bổ sung rằng chất lượng âm cao hơn trong các tín hiệu được tách, vì ít thành phần lạ được đưa vào và sự rõ ràng các tín hiệu nhiều chắn một phần các thành phần lạ được nhận biết.

Sự chặn một phần tín hiệu âm thanh có nghĩa là độ to của nó (ví dụ, mật độ được nhận biết) được giảm một phần. Hơn nữa, nó còn được mong muốn và yêu cầu rằng, hơn là đạt được sự suy giảm lớn, chất lượng âm của đầu ra không rơi xuống dưới mức chất lượng âm được xác định.

Ví dụ cho ứng dụng này là sự nâng cao hội thoại. Các tín hiệu âm thanh trong TV và phát rộng vô tuyến và âm điện ảnh thường là sự trộn các tín hiệu tiếng nói và tín hiệu nền, ví dụ, các âm môi trường và âm nhạc. Khi các tín hiệu này được trộn sao cho mức của tiếng nói quá thấp so với mức của nền, người nghe có thể gặp khó khăn để hiểu những gì đã nói, hoặc sự hiểu yêu cầu nỗ lực nghe cao và điều này dẫn đến sự mệt mỏi của người nghe. Các phương pháp tự động giảm mức nền có thể được áp dụng cho các ngữ cảnh này, nhưng kết quả sẽ là chất lượng âm cao.

Các phương pháp khác nhau để tách nguồn có trong tình trạng kỹ thuật. Sự phân tách tín hiệu đích từ sự trộn các tín hiệu đã được thảo luận trong tình trạng kỹ thuật. Các phương pháp này có thể được phân loại thành hai phương pháp. Loại thứ nhất của các phương pháp dựa trên các giả định được tạo công thức về mô hình tín hiệu và/hoặc mô hình trộn. Mô hình tín hiệu mô tả đặc điểm của các tín hiệu đầu vào, ở đây là $s(n)$ và $b(n)$. Mô hình trộn mô tả các đặc điểm làm thế nào các tín hiệu đầu vào được tổ hợp để tạo ra tín hiệu trộn $x(n)$, ở đây bằng phép cộng.

Dựa trên các giả định này, phương pháp được thiết kế theo cách phân tích hoặc khám phá. Ví dụ, phương pháp phân tích thành phần độc lập có thể được suy ra bằng cách giả định rằng sự trộn bao gồm hai tín hiệu nguồn mà độc lập theo thống kê, sự trộn đã được ghi bởi hai micrô, và hỗn hợp đã được suy ra bằng cách cộng cả hai tín hiệu (tạo ra sự trộn tức thời). Quy trình nghịch đảo của sự trộn sau đó được suy ra theo toán học như sự nghịch đảo của ma trận trộn và các phần tử của ma trận không trộn này được tính toán theo phương pháp cụ thể. Các phương pháp được suy ra theo cách phân tích nhất được suy ra bằng cách tạo công thức vấn đề

phân tách như sự tối ưu số học của tiêu chí, ví dụ, sai số được bình phương trung bình giữa đích đúng và đích được ước lượng.

Loại thứ hai là hướng dữ liệu. Ở đây, sự biểu diễn của các tín hiệu đích được ước lượng, hoặc tập hợp các tham số để truy hồi các tín hiệu đích từ hỗn hợp đầu vào được ước lượng. Sự ước lượng dựa trên mô hình đã được đào tạo trên tập hợp dữ liệu đào tạo, ở đây gọi là “hướng dữ liệu”. Sự ước lượng được suy ra bằng cách tối ưu hóa tiêu chí, ví dụ, bằng cách tối thiểu hóa sai số bình phương trung bình giữa đích đúng và đích được ước lượng, đưa ra dữ liệu đào tạo. Ví dụ về loại này là mạng nơron nhân tạo (Artificial Neural Networks - ANN) mà đã được đào tạo để xuất ra sự ước lượng tín hiệu tiếng nói đưa ra hỗn hợp tín hiệu tiếng nói và tín hiệu nhiễu. Trong khi đào tạo, các tham số có thể điều chỉnh của mạng nơron nhân tạo được xác định sao cho tiêu chí hiệu suất được tính toán cho tập hợp dữ liệu đào tạo được tối ưu hóa - trên trung bình so với tập hợp dữ liệu đầy đủ.

Về sự phân tách nguồn, giải pháp mà tối ưu về sai số bình phương trung bình hoặc tối ưu đối với bất kỳ tiêu chí số học khác không cần thiết là giải pháp với chất lượng âm cao nhất mà được thực hiện bởi người nghe.

Vấn đề thứ hai bắt nguồn từ thực tế rằng sự phân tách nguồn luôn dẫn đến hai hiệu quả, thứ nhất là sự suy giảm mong muốn về các âm nhiễu và thứ hai là sự suy giảm không mong muốn về chất lượng âm. Cả hai ảnh hưởng là tương quan, ví dụ, sự tăng ảnh hưởng mong muốn dẫn đến sự tăng ảnh hưởng không mong muốn. Mục đích cuối cùng là điều khiển sự cân bằng giữa cả hai.

Chất lượng âm có thể được ước lượng, ví dụ, được lượng tử hóa bằng sự kiểm tra nghe hoặc bằng các mô hình tính toán của chất lượng âm. Chất lượng âm có nhiều khía cạnh, sau đây đề cập đến như các thành phần chất lượng âm (Sound Quality Component - SQC).

Ví dụ, chất lượng âm được xác định bởi mật độ các thành phần lạ được nhận biết (đây là các thành phần tín hiệu mà đã được đưa ra bởi sự xử lý tín hiệu, ví dụ, sự phân tách nguồn, và giảm chất lượng âm).

Hoặc, ví dụ, chất lượng âm được xác định bởi mật độ được nhận biết các tín hiệu nhiễu, hoặc, ví dụ, bởi độ rõ tiếng nói (khi tín hiệu đích là tiếng nói), hoặc, ví dụ, bởi chất lượng âm toàn phần.

Các mô hình tính toán khác nhau của chất lượng âm đang có mà tính toán (ước lượng) các thành phần chất lượng âm q_m , $1 \leq m \leq M$, trong đó M biểu thị các thành phần chất lượng âm.

Các phương pháp này thường ước lượng thành phần chất lượng âm đưa ra tín hiệu đích và sự ước lượng cho tín hiệu đích,

$$q_m = f(s(n), \hat{s}(n)) \quad (4)$$

hoặc cũng đưa ra tín hiệu nhiễu,

$$q_m = f(s(n), b(n), \hat{s}(n)) \quad (5)$$

Trong các ứng dụng thực tế, các tín hiệu đích $s(n)$ (và các tín hiệu nhiễu $b(n)$) không có sẵn, nếu không sự phân tách có sẽ không được yêu cầu. Khi chỉ tín hiệu đầu vào $x(n)$ và ước lượng tín hiệu đích $\hat{s}(n)$ có sẵn, các thành phần chất lượng âm không thể được tính toán với các phương pháp này.

Trong lĩnh vực kỹ thuật, các mô hình tính toán khác nhau để ước lượng các khía cạnh chất lượng âm, bao gồm độ rõ, đã được mô tả.

Sự đánh giá phân tách nguồn mù (Blind Source Separation Evaluation - BSSEval) (xem tài liệu [1]) là hộp công cụ đánh giá hiệu quả đa tiêu chí. Tín hiệu được ước lượng được khai triển bởi phần nhô thẳng góc vào thành phần tín hiệu đích, giao với các nguồn khác, và các thành phần lạ. Các metric được tính toán như các tỉ lệ năng lượng của các thành phần này và được biểu diễn bằng dB. Chúng là: Tỉ lệ nguồn trên độ méo (Source to Distortion Ratio - SDR), tỉ lệ nguồn trên nhiễu (Source to Interference Ratio - SIR), và tỉ lệ nguồn trên thành phần lạ (Source to Artifact Ratio - SAR).

Các phương pháp đánh giá cảm quan cho sự phân tách nguồn âm (Perceptual Evaluation methods for Audio Source Separation - PEASS) (xem tài liệu [2]) được

thiết kế như sự kế thừa kích thích cảm quan của BSSEval. Sự nhô ra tín hiệu được thực hiện trên các phân đoạn thời gian và với giàn lọc gammatone.

PEMO-Q (xem tài liệu [3]) được sử dụng để cung cấp nhiều dấu hiệu. Bốn điểm cảm quan thu được từ các dấu hiệu này sử dụng mạng nơron được đào tạo với tỉ lệ chủ quan. Điểm là: Điểm tổng cảm quan (Overall Perceptual Score - OPS), điểm cảm quan về nhiễu (Interference-related Perceptual Score - IPS), điểm cảm quan về thành phần lạ (Artifact-related Perceptual Score - APS), và điểm cảm quan về đích (Target-related Perceptual Score - TPS).

Sự đánh giá cảm quan về chất lượng âm thanh (Perceptual Evaluation of Audio Quality - PEAQ) (xem tài liệu [4]) là metric được thiết kế để mã hóa âm thanh. Nó sử dụng mô hình tai ngoại vi để tính toán sự biểu diễn màng nền của tài liệu tham khảo và tín hiệu thử nghiệm. Các khía cạnh của sự chênh lệch giữa các sự biểu diễn được lượng tử hóa bởi một số biến số đầu ra. Bằng mạng nơron được đào tạo với dữ liệu chủ quan, các biến số này được tổ hợp để cho đầu ra chính, ví dụ, cấp chênh lệch tổng (Overall Difference Grade - ODG).

Sự đánh giá cảm quan về chất lượng tiếng nói (Perceptual Evaluation of Speech Quality - PESQ) (xem tài liệu [5]) là metric được thiết kế để tiếng nói được truyền dẫn qua mạng viễn thông. Do đó, phương pháp bao gồm sự xử lý trước mà mô phỏng ống nghe điện thoại. Các ước số cho sự nhiễu loạn có thể nghe thấy được tính toán từ độ to cụ thể của các tín hiệu và được tổ hợp trong các điểm PESQ. Từ chúng, điểm MOS được dự báo bằng hàm ánh xạ đa thức (xem tài liệu [6])

ViSQOLAudio (xem tài liệu [7]) là metric được thiết kế để nhạc được mã hóa tại tốc độ bit thấp được phát triển từ người nghe khách quan chất lượng tiếng nói ảo (Virtual Speech Quality Objective Listener - ViSQOL). Cả hai metric dựa trên mô hình của chế độ hệ thống nghe ngoại biên để tạo ra sự biểu diễn bên trong của các tín hiệu được gọi là các dấu vết thần kinh. Những điều này được so sánh qua sự thích ứng của chỉ số tương tự kết cấu, được phát triển ban đầu để đánh giá chất lượng ảnh được nén.

Chỉ số chất lượng âm thanh hỗ trợ nghe (Hearing-Aid Audio Quality Index - HAAQI) (xem [8]) là chỉ số được thiết kế để dự báo chất lượng âm nhạc cho các cá nhân nghe qua vật hỗ trợ nghe. Chỉ số dựa trên mô hình của ngoại biên thính học, được mở rộng để bao gồm các hiệu quả tổn hao nghe. Điều này khớp với cơ sở dữ liệu gồm các đánh giá chất lượng được thực hiện bởi người nghe có khả năng nghe bình thường hoặc khiếm khuyết. Sự mô phỏng tổn hao nghe có thể được bỏ qua và chỉ số có hiệu lực cũng cho người nghe bình thường. Dựa trên mô hình thính giác giống nhau, các tác giả của HAAQI cũng được đề xuất chỉ số cho chất lượng tiếng nói, chỉ số chất lượng tiếng nói hỗ trợ nghe (Hearing-Aid Speech Quality Index - HASQI) (xem [9]) và độ rõ cảm quan tiếng nói hỗ trợ nghe (Hearing-Aid Speech Perception Index - HASPI) (xem [10]).

Độ rõ khách quan thời gian ngắn (Short-Time Objective Intelligibility - STOI) (xem [11]) là ước số mà được mong đợi có mối quan hệ đơn điệu với độ rõ tiếng nói trung bình. Nó chỉ rõ cụ thể tiếng nói được xử lý bởi một số loại gán trọng số thời gian - tần số.

Trong tài liệu [12] mạng nơron nhân tạo được đào tạo để ước lượng tỉ lệ nguồn trên độ méo đã cho chỉ tín hiệu đầu vào và tín hiệu đích được ước lượng đầu ra, trong đó sự tính toán tỉ lệ nguồn trên độ méo sẽ thường được coi như các đầu vào cũng là đích đúng và tín hiệu nhiễu. Vùng gồm các thuật toán phân tách được vận hành song song trên tín hiệu đầu vào giống nhau. Các ước lượng tỉ lệ nguồn trên độ méo được sử dụng để lựa chọn cho khung thời gian đầu ra từ thuật toán với tỉ lệ nguồn trên độ méo tốt nhất. Do đó, không có sự điều khiển độ cân bằng giữa chất lượng âm và sự phân tách được tạo công thức, và không có sự điều khiển các tham số của thuật toán phân tách được đề xuất. Hơn nữa, tỉ lệ nguồn trên độ méo được sử dụng, mà không được kích thích về cảm quan và đã được thể hiện tương quan kém với chất lượng được nhận biết, ví dụ, trong tài liệu [13].

Hơn nữa, các nghiên cứu gần đây về sự nâng cao tiếng nói bởi sự học hỏi được giám sát liệu các ước lượng thành phần chất lượng âm được tích hợp trong các hàm giá trị, trong khi, về truyền thống, các mô hình nâng cao tiếng nói được tối ưu dựa trên độ sai bình phương trung bình (mean square error - MSE) giữa tiếng nói

được ước lượng và sạch. Ví dụ, trong tài liệu [14], [15], [16] các hàm giá trị dựa trên STOI thay vì MSE được sử dụng. Trong tài liệu [17] sự học hỏi được củng cố dựa trên PESQ hoặc PEASS được sử dụng. Tuy nhiên, không có sự điều khiển về sự cân bằng giữa chất lượng âm và sự phân tách có sẵn.

Trong tài liệu [18] thiết bị xử lý âm thanh được đề xuất trong đó ước số độ nghe rõ sử dụng cùng với ước số nhận biết thành phần lạ để điều khiển các độ khuếch đại thời gian - tần số được áp dụng bởi sự xử lý. Điều này cung cấp, ví dụ, lượng giảm nhiễu âm tại mức tối đa phải chịu giới hạn mà không có thành phần lạ được đưa vào, sự cân bằng giữa chất lượng âm và sự phân tách được cố định. Hơn nữa, hệ thống không bao gồm sự học hỏi được giám sát. Để nhận biết thành phần lạ, tỉ lệ Kurtosis được sử dụng, ước số mà trực tiếp so sánh các tín hiệu đầu ra và đầu vào (có khả năng trong các phân đoạn mà tiếng nói không có), không cần đích đúng và tín hiệu nhiễu. Ước số đơn giản này được làm giàu bởi ước số độ nghe rõ.

Tài liệu tham khảo

[1] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[2] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, “Subjective and objective quality assessment of audio source separation,” *IEEE Trans. Audio, Speech and Language Process.*, vol. 19, no. 7, 2011.

[3] R. Huber and B. Kollmeier, “PEMO-Q - a new method for objective audio quality assessment using a model of auditory perception,” *IEEE Trans. Audio, Speech and Language Process.*, vol. 14, 2006.

[4] ITU-R Rec. BS.1387-1, “Method for objective measurements of perceived audio quality,” 2001.

[5] ITU-T Rec. P.862, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” 2001.

[6] ITU-T Rec. P.862.1, “Mapping function for transforming P.862 raw results scores to MOS-LQO,” 2003.

[7] A. Hines, E. Gillen et al., “ViSQOLAudio: An Objective Audio Quality Metric for Low Bitrate Codecs,” *J. Acoust. Soc. Am.*, vol. 137, no. 6, 2015.

[8] J. M. Kates and K. H. Arehart, “The Hearing-Aid Audio Quality Index (HAAQI),” *IEEE Trans. Audio, Speech and Language Process.*, vol. 24, no. 2, 2016, evaluation code kindly provided by Prof. J.M. Kates.

[9] J. M. Kates and K. H. Arehart, “The Hearing-Aid Speech Quality Index (HASQI) version 2,” *Journal of the Audio Engineering Society*, vol. 62, no. 3, pp. 99–117, 2014.

[10] J. M. Kates and K. H. Arehart, “The Hearing-Aid Speech Perception Index (HASPI),” *Speech Communication*, vol. 65, pp. 75–93, 2014.

[11] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time-frequency weighted noisy speech,” *IEEE Trans. Audio, Speech and Language Process.*, vol. 19, no. 7, 2011.

[12] E. Manilow, P. Seetharaman, F. Pishdadian, and B. Pardo, “Predicting algorithm efficacy for adaptive multi-cue source separation,” in *Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017 IEEE Workshop on, 2017, pp. 274–278.

[13] M. Cartwright, B. Pardo, G. J. Mysore, and M. Hoffman, “Fast and easy crowdsourced perceptual audio evaluation,” in *Acoustics, Speech and Signal Processing (ICASSP)*, 2016 IEEE International Conference on, 2016.

[14] S.-W. Fu, T.-W. Wang, Y. Tsao, X. Lu, and H. Kawai, “End-to-end waveform utterance enhancement for direct evaluation metrics optimization by fully convolutional neural networks,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 9, 2018.

[15] Y. Koizumi, K. Niwa, Y. Hioka, K. Kobayashi, and Y. Haneda, “Dnn-based source enhancement to increase objective sound quality assessment score,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.

[16] Y. Zhao, B. Xu, R. Giri, and T. Zhang, “Perceptually guided speech enhancement using deep neural networks,” in *Acoustics, Speech and Signal Processing (ICASSP)*, 2018 IEEE International Conference on, 2018.

[17] Y. Koizumi, K. Niwa, Y. Hioka, K. Kobayashi, and Y. Haneda, “Dnn-based source enhancement self-optimized by reinforcement learning using sound quality measurements,” in *Acoustics, Speech and Signal Processing (ICASSP)*, 2017 IEEE International Conference on, 2017.

[18] J. Jensen and M. S. Pedersen, “Audio processing device comprising artifact reduction,” *US Patent US 9,432,766 B2*, Aug. 30, 2016.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất các khái niệm được cải thiện cho sự phân tách nguồn. Mục đích của sáng chế được giải quyết bởi thiết bị theo điểm 1, bởi phương pháp mã hóa theo điểm 16 và bởi chương trình máy tính theo điểm 17.

Thiết bị tạo ra tín hiệu âm thanh được phân tách từ tín hiệu âm thanh đầu vào âm thanh được đề xuất. Tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư. Phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích. Thiết bị bao gồm bộ phân tách nguồn, môđun xác định và bộ xử lý tín hiệu. Bộ phân tách nguồn được tạo cấu hình để xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích. Môđun xác định được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được. Bộ xử lý tín hiệu được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và

phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư.

Tuy nhiên, phương pháp tạo ra tín hiệu âm thanh được phân tách từ tín hiệu âm thanh đầu vào âm thanh được đề xuất. Tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư. Phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích. Phương pháp bao gồm:

- Xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích.

- Xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được. Và:

- Tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư.

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

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 tạo ra tín hiệu âm thanh được phân tách từ tín hiệu đầu vào âm thanh theo phương án,

Fig.1b minh họa thiết bị tạo ra tín hiệu âm thanh được phân tách theo phương án khác, còn bao gồm mạng nơron nhân tạo,

Fig.2 minh họa thiết bị theo phương án mà được tạo cấu hình để sử dụng sự ước lượng chất lượng âm và được tạo cấu hình để thực hiện sự xử lý sau,

Fig.3 minh họa thiết bị theo phương án khác, trong đó sự ước lượng trực tiếp của các tham số xử lý sau được thực hiện,

Fig.4 minh họa thiết bị theo phương án khác, trong đó sự ước lượng chất lượng âm và sự phân tách thứ cấp được thực hiện, và

Fig.5 minh họa thiết bị theo phương án khác, trong đó sự ước lượng trực tiếp của các tham số phân tách được thực hiện.

Mô tả chi tiết sáng chế

Fig.1a minh họa thiết bị tạo ra tín hiệu âm thanh được phân tách từ tín hiệu đầu vào âm thanh theo phương án. Tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư. Phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích.

Thiết bị bao gồm bộ phân tách nguồn 110, môđun xác định 120 và bộ xử lý tín hiệu 130.

Bộ phân tách nguồn 110 được tạo cấu hình để xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích.

Môđun xác định 120 được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được.

Bộ xử lý tín hiệu 130 được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng. Tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư.

Một cách tùy chọn, trong phương án, môđun xác định 120 có thể, ví dụ, được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín hiệu đích được ước lượng và phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

Các phương án cung cấp sự điều khiển kích thích cảm quan và thích ứng tín hiệu trên sự cân bằng giữa chất lượng âm và sự phân tách sử dụng sự học tập được giám sát. Điều này có thể đạt được theo hai cách. Phương pháp thứ nhất ước lượng chất lượng âm của tín hiệu đầu ra và sử dụng sự ước lượng để thích ứng các tham số của sự phân tách hoặc sự xử lý sau của các tín hiệu được phân tách. Trong phương án thứ hai, phương pháp hồi quy xuất ra trực tiếp các tham số điều khiển sao cho chất lượng âm của tín hiệu đầu ra đáp ứng các yêu cầu được định nghĩa.

Theo các phương án, sự phân tích tín hiệu đầu vào và tín hiệu đầu ra của sự phân tách được thực hiện để tạo ra ước lượng của chất lượng âm q_m , và xác định các tham số xử lý dựa trên q_m sao cho chất lượng âm của đầu ra (khi sử dụng các tham số xử lý được xác định) không thấp hơn giá trị chất lượng được định nghĩa.

Trong một số phương án, sự phân tích xuất ra ước số chất lượng q_m trong công thức (9). Từ ước số chất lượng, tham số điều khiển $p1$ trong công thức (13) sau đây được tính toán (ví dụ (thừa số định tỉ lệ), và đầu ra cuối cùng thu được bằng cách trộn đầu ra ban đầu và đầu ra như trong công thức (13) sau đây. Sự tính toán $p1 = f(q_m)$ có thể được thực hiện lặp đi lặp lại hoặc bằng cách hồi quy, các tham số hồi quy được học hỏi từ tập hợp các tín hiệu đào tạo, xem Fig 2. Trong các phương án, thay vì thừa số định tỉ lệ, tham số điều khiển có thể, ví dụ, là tham số làm nhấn hoặc tương tự.

Trong một số phương án, sự phân tích tạo ra tham số điều khiển $p1$ trong công thức (13) một cách trực tiếp, xem Fig 3.

Fig.4 và Fig.5 định nghĩa các phương án khác.

Một số phương án đạt được sự điều khiển chất lượng âm trong bước xử lý sau, như được mô tả dưới đây.

Tập con của các phương án được mô tả có thể được áp dụng độc lập với phương pháp phân tách. Một số phương án được mô tả ở đây điều khiển tham số của quy trình phân tách.

Sự phân tách nguồn sử dụng các quy trình gán trọng số phổ tạo tín hiệu trong miền thời gian - tần số hoặc miền phổ ngắn hạn. Tín hiệu đầu vào $x(n)$ được biến đổi bằng phép biến đổi Fourier ngắn hạn (short-time Fourier transform - STFT) hoặc được xử lý bằng giàn lọc, tạo ra các hệ số STFT giá trị phức hoặc các tín hiệu băng phụ $X(m, k)$, trong đó m biểu thị chỉ số khung thời gian, k biểu thị chỉ số ngắn tần số hoặc chỉ số băng phụ. Các hệ số STFT giá trị phức hoặc các tín hiệu băng phụ của tín hiệu mong muốn là $S(m, k)$, và tín hiệu nhiễu là $B(m, k)$.

Các tín hiệu đầu ra được phân tách được tính toán bằng cách gán trọng số phổ như

$$\hat{S}(m, k) = G(m, k)X(m, k) \quad (6)$$

trong đó các trọng số phổ $G(m, k)$ được nhân theo từng phần tử với tín hiệu đầu vào. Mục đích là để giảm các phần tử trong $X(m, k)$ trong đó nguồn nhiễu $B(m, k)$ là lớn. Để đạt được điều này, các trọng số phổ có thể được tính toán dựa trên ước lượng của đích $\hat{S}(m, k)$ hoặc ước lượng của nguồn nhiễu $\hat{B}(m, k)$ hoặc ước lượng của tỉ lệ tín hiệu trên nguồn nhiễu, ví dụ,

$$G(m, k) = \frac{(|\hat{X}(m, k)|^a - |\hat{B}(m, k)|^a)^c}{|X(m, k)|^a} \quad (7)$$

hoặc

$$G(m, k) = \frac{|\hat{S}(m, k)|^c}{|X(m, k)|} \quad (8)$$

trong đó a và c là các tham số điều khiển sự phân tách. Ví dụ, tăng c có thể dẫn đến sự suy giảm lớn hơn của nguồn nhiễu nhưng cũng dẫn đến sự suy giảm chất

lượng âm. Các trọng số phổ có thể còn được biến đổi, ví dụ, bằng ngưỡng sao cho G lớn hơn ngưỡng. Các độ khuếch đại được biến đổi G_m được tính toán như

$$G_m(m, k) = \begin{cases} G(m, k) & \text{nếu } G(m, k) > v \\ v & \text{otherwise} \end{cases}$$

Tăng ngưỡng v giảm sự suy giảm nguồn nhiễu và giảm sự suy giảm có tiềm năng của chất lượng âm.

Sự ước lượng số lượng được yêu cầu (đích $\hat{S}(m, k)$ hoặc nguồn nhiễu $\hat{B}(m, k)$ hoặc tỉ lệ tín hiệu trên nguồn nhiễu) là lỗi của các phương pháp này đã được phát triển trong quá khứ. Chúng đều theo sau một trong số hai phương pháp được mô tả trên đây.

Tín hiệu đầu ra $\hat{s}(n)$ sau đó được tính toán sử dụng sự xử lý nghịch đảo của STFT hoặc giàn lọc.

Sau đây, sự phân tách nguồn sử dụng ước lượng của tín hiệu đích theo các phương án được mô tả.

Sự biểu diễn của tín hiệu đích có thể cũng được ước lượng trực tiếp từ tín hiệu đầu vào, ví dụ, bằng mạng nơron nhân tạo. Các phương pháp khác nhau đã được đề xuất trong đó mạng nơron nhân tạo đã được đào tạo để ước lượng tín hiệu thời gian đích, hoặc các hệ số STFT, hoặc các cường độ của các hệ số STFT.

Về chất lượng âm, thành phần chất lượng âm (Sound Quality Component - SQC) thu được bằng cách áp dụng mô hình học hỏi được giám sát $g(\cdot)$ để ước lượng các đầu ra của mô hình tính toán,

$$\hat{q}_m = g(x(n), \hat{s}(n)) \quad (9)$$

Phương pháp học hỏi được giám sát $g(\cdot)$ được nhận biết bởi:

1. Tạo cấu hình mô hình học hỏi được giám sát $g(\cdot)$ với các tham số có thể đào tạo, N_i biến số đầu vào và N_o biến số đầu ra,
2. Tạo ra tập hợp dữ liệu với các tín hiệu ví dụ cho đích $s(n)$ và hỗn hợp $x(n)$,

3. Tính toán các ước lượng cho các tín hiệu đích bằng sự phân tách nguồn, $\hat{s}_n = h(x(n))$,

4. Tính toán các thành phần chất lượng âm q_m từ các tín hiệu thu được bằng các mô hình tính toán chất lượng âm theo (9) hoặc (10),

5. Đào tạo mô hình học hồi được giám sát $g(\cdot)$ sao cho mô hình xuất ra các ước lượng $\hat{q}_m$ đã cho các tín hiệu ví dụ tương ứng cho đích được ước lượng $\hat{s}(n)$ (đầu ra của sự phân tách nguồn) và hỗn hợp $x(n)$. Cách khác, đào tạo mô hình học hồi được giám sát $g(\cdot)$ sao cho mô hình xuất ra các ước lượng $\hat{q}_m$ đã cho $\hat{s}(n)$ và $\hat{b}(n)$ (nếu $x(n) = \hat{s}(n) + \hat{b}(n)$).

6. Theo sáng chế, mô hình được đào tạo được cấp đích được ước lượng $\hat{s}(n)$ (đầu ra của sự phân tách nguồn) thu được từ hỗn hợp $x(n)$ sử dụng phương pháp phân tách nguồn cùng với hỗn hợp $x(n)$.

Sự ứng dụng các phương pháp học hồi được giám sát để quản lý chất lượng của tín hiệu đầu ra được phân tách được cung cấp.

Sau đây, sự ước lượng chất lượng âm sử dụng sự học hồi được giám sát theo các phương án được mô tả.

Fig.1b minh họa phương án, trong đó mô đun xác định 120 bao gồm mạng nơron nhân tạo 125. Mạng nơron nhân tạo 125 có thể, ví dụ, được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín hiệu đích được ước lượng. Mạng nơron nhân tạo 125 có thể, ví dụ, được tạo cấu hình để nhận nhiều giá trị đầu vào, mỗi giá trị trong số nhiều giá trị đầu vào phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh. Mạng nơron nhân tạo 125 có thể, ví dụ, được tạo cấu hình để xác định một hoặc nhiều giá trị thu như một hoặc nhiều giá trị đầu ra của mạng nơron nhân tạo 125.

Một cách tùy chọn, trong phương án, mạng nơron nhân tạo 125 có thể, ví dụ, được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín

hiệu đích được ước lượng và ít nhất một tín hiệu trong số tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

Trong phương án, mỗi giá trị trong số nhiều giá trị đầu vào có thể, ví dụ, phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh. Một hoặc nhiều giá trị thu được biểu thị chất lượng âm được ước lượng của tín hiệu đích được ước lượng.

Theo phương án, mỗi giá trị trong số nhiều giá trị đầu vào có thể, ví dụ, phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh. Một hoặc nhiều giá trị thu được có thể, ví dụ, là một hoặc nhiều giá trị tham số.

Trong phương án, mạng nơron nhân tạo 125 có thể, ví dụ, được tạo cấu hình để được đào tạo bằng cách nhận nhiều tập hợp đào tạo, trong đó mỗi tập hợp trong số nhiều tập hợp đào tạo bao gồm nhiều giá trị đào tạo đầu vào của mạng nơron nhân tạo 125 và một hoặc nhiều giá trị đào tạo đầu ra của mạng nơron nhân tạo 125, trong đó mỗi giá trị trong số nhiều giá trị đào tạo đầu ra có thể, ví dụ, phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đích đào tạo và tín hiệu phần dư đào tạo và tín hiệu đầu vào đào tạo, trong đó mỗi giá trị trong số một hoặc nhiều giá trị đào tạo đầu ra có thể, ví dụ, phụ thuộc vào sự ước lượng của chất lượng âm của tín hiệu đích đào tạo.

Trong các phương án, ước lượng của thành phần chất lượng âm thu được bằng sự học hỏi được giám sát sử dụng mô hình học hỏi được giám sát (supervised learning model - SLM), ví dụ, mạng nơron nhân tạo (Artificial Neural Network - ANN) 125. Mạng nơron nhân tạo 125 có thể, ví dụ, là mạng nơron nhân tạo được nối toàn bộ 125 mà bao gồm lớp đầu vào với A đơn vị, ít nhất một lớp ẩn với ít nhất hai đơn vị mỗi lớp, và lớp đầu ra với một hoặc nhiều đơn vị.

Mô hình học hỏi được giám sát có thể được thực hiện như mô hình hồi quy hoặc mô hình phân loại. Mô hình hồi quy ước lượng một giá trị đích tại đầu ra của một đơn vị trong lớp đầu ra. Cách khác, vấn đề hồi quy có thể được tạo công thức

như vấn đề phân loại bằng cách lượng tử hóa giá trị đầu ra thành ít nhất 3 bước và sử dụng lớp đầu ra với C đơn vị trong đó C bằng số lượng các bước lượng tử hóa.

Đối với bước lượng tử hóa, một đơn vị đầu ra được sử dụng.

Mô hình học hỏi được giám sát đầu tiên được đào tạo với tập hợp dữ liệu mà chứa nhiều ví dụ về tín hiệu hỗn hợp x , đích được ước lượng $\hat{s}$, và thành phần chất lượng âm q_m , trong đó thành phần chất lượng âm đã được tính toán từ đích được ước lượng $\hat{s}$, và đích đúng s , ví dụ. Một mục của tập hợp dữ liệu được biểu thị bởi $\{x_i, \hat{s}_i, q_i\}$. Đầu ra của mô hình học hỏi được giám sát ở đây được biểu thị bởi q_i .

Số lượng các đơn vị trong lớp đầu vào A tương ứng với số lượng các giá trị đầu vào. Các đầu vào cho các mô hình được tính toán từ tín hiệu đầu vào. Mỗi tín hiệu có thể được xử lý tùy chọn bằng giàn lọc phép biến đổi thời gian - tần số, ví dụ, phép biến đổi Fourier ngắn hạn (a short-term Fourier transform - STFT). Ví dụ, đầu vào có thể được dựng bằng cách nối chuỗi các hệ số STFT được tính toán từ D khung liên tiếp từ x_i và $\hat{s}_i$, trong đó $D = 3$ hoặc $D = 7$. Với B là tổng số lượng các hệ số phổ mỗi khung, tổng số lượng hệ số đầu vào là $2 \cdot B \cdot D$.

Mỗi đơn vị của mạng nơron nhân tạo 125 tính toán đầu ra của mình như sự tổ hợp tuyến tính của các giá trị đầu vào mà khi đó được xử lý với hàm nén phi tuyến tính.

$$u = h\left(\sum_i^K w_i v_i + o_i\right) \quad (10)$$

trong đó u biểu thị đầu ra của nơron đơn, v_i biểu thị K giá trị đầu vào, w_i biểu thị K gán trọng số cho sự tổ hợp tuyến tính và o_i biểu thị K thời hạn lệch bổ sung. Đối với các đơn vị trong lớp ảnh thứ nhất, số lượng các giá trị đầu vào K bằng số lượng các hệ số đầu vào D . Tất cả w_i và o_i là các tham số của mạng nơron nhân tạo 125 mà được xác định trong quy trình đào tạo.

Các đơn vị của một lớp được nối với các đơn vị của lớp sau đây, các đầu ra của các đơn vị của lớp trước là các đầu vào cho các đơn vị của lớp tiếp theo.

Sự đào tạo được thực hiện bằng cách giảm thiểu hóa sai số dự báo sử dụng phương pháp tối ưu hóa bằng số, ví dụ, phương pháp giảm độ chênh lệch. Sai số dự báo cho mục đơn là hàm của hiệu số $e_i = f(q_i - \hat{q}_i)$. Sai số dự báo trên tập hợp dữ liệu đầy đủ hoặc tập con của tập hợp dữ liệu mà được sử dụng tiêu chí tối ưu hóa là cho, ví dụ, sai số được bình phương trung bình (mean squared error - MSE) hoặc sai số tuyệt đối trung bình (mean absolute error - MAE), trong đó N biểu thị số lượng các mục trong tập hợp dữ liệu.

$$MSE = \frac{1}{N} \sum_i^N e_i^2 \quad (11)$$

$$MAE = \frac{1}{N} \sum_i^N |e_i| \quad (12)$$

Các metric sai số khác là khả thi cho mục đích đào tạo nếu chúng là hàm đơn điệu của e_i và có thể phân biệt. Tương tự, các kết cấu và phần tử để tạo kết cấu mạng nơron nhân tạo tồn tại, ví dụ, các lớp mạng nơron nhân chập hoặc các lớp mạng nơron hồi quy.

Tất cả những điều chung đó là chúng thực thi sự ánh xạ từ đầu vào đa kích thước thành đầu ra một kích thước hoặc đa kích thước trong đó hàm ánh xạ được điều khiển bởi tập hợp các tham số (w_i và o_i) mà được xác định trong quy trình đào tạo bằng cách tối ưu hóa tiêu chí về hướng.

Sau khi đào tạo, mô hình học hỏi được giám sát có thể được sử dụng cho sự ước lượng của chất lượng âm của đích được ước lượng chưa biết $\hat{s}$ đã cho hỗn hợp mà không cần cho đích đúng s .

Với các mô hình tính toán chất lượng âm, các mô hình tính toán khác nhau để ước lượng các khía cạnh của chất lượng âm (bao gồm độ rõ) đã được sử dụng thành công trong các thí nghiệm theo các phương án, như các mô hình tính toán được mô tả trong [1] – [11], cụ thể, sự đánh giá phân tách nguồn mù (Blind Source Separation Evaluation - BSSEval) (xem [1]), các phương pháp đánh giá cảm qua

cho sự phân tách nguồn âm thanh (Perceptual Evaluation methods for Audio Source Separation - PEASS) (xem [2]), PEMO-Q (xem [3]), sự đánh giá cảm quan về chất lượng âm thanh (Perceptual Evaluation of Audio Quality - PEAQ) (xem [4]), sự đánh giá cảm quan về chất lượng tiếng nói (Perceptual Evaluation of Speech Quality - PESQ) (xem [5] và [6]), ViSQOLAudio (xem [7]), chỉ số chất lượng âm thanh hỗ trợ nghe (Hearing-Aid Audio Quality Index - HAAQI) (xem [8]), chỉ số chất lượng tiếng nói hỗ trợ nghe (Hearing-Aid Speech Quality Index - HASQI) (xem [9]), chỉ số nhận biết tiếng nói hỗ trợ nghe (Hearing-Aid Speech Perception Index - HASPI) (xem [10]), và độ rõ khách quan ngắn hạn (Short-Time Objective Intelligibility - STOI) (xem [11]).

Do đó, theo phương án, sự ước lượng chất lượng âm của tín hiệu đích đào tạo có thể, ví dụ, phụ thuộc vào một hoặc nhiều mô hình tính toán chất lượng âm.

Ví dụ, trong phương án, sự ước lượng chất lượng âm của tín hiệu đích đào tạo có thể, ví dụ, phụ thuộc vào một hoặc nhiều mô hình tính toán chất lượng âm sau đây:

Sự đánh giá phân tách nguồn mù,

Các phương pháp đánh giá cảm quan cho sự phân tách nguồn âm thanh,

Sự đánh giá cảm quan chất lượng âm thanh,

Sự đánh giá cảm quan chất lượng tiếng nói,

Âm thanh người nghe khách quan chất lượng tiếng nói ảo,

Chỉ số chất lượng âm thanh hỗ trợ nghe,

Chỉ số chất lượng tiếng nói hỗ trợ nghe,

Chỉ số cảm quan tiếng nói hỗ trợ nghe, và

Độ rõ khách quan thời gian ngắn.

Các mô hình tính toán khác của chất lượng âm có thể, ví dụ, cũng được sử dụng trong các phương án khác.

Sau đây, sự điều khiển chất lượng âm được mô tả.

Sự điều khiển chất lượng âm có thể được thực thi bằng cách ước lượng thành phần chất lượng âm và tính toán các tham số xử lý dựa trên ước lượng thành phần chất lượng âm, hoặc bằng cách ước lượng trực tiếp các tham số xử lý tối ưu sao cho thành phần chất lượng âm đáp ứng giá trị đích q_0 (hoặc không nằm dưới đích đó).

Sự ước lượng thành phần chất lượng âm đã được mô tả trên đây. Theo cách tương tự, các tham số xử lý tối ưu có thể được ước lượng bằng cách đào tạo phương pháp hồi quy với các giá trị mong muốn cho các tham số xử lý tối ưu. Các tham số xử lý tối ưu được tính toán như được mô tả dưới đây. Sự xử lý được đề cập đến như môđun ước lượng tham số (Parameter Estimation Module - PEM) sau đây.

Giá trị đích cho chất lượng âm q_0 sẽ xác định sự cân bằng giữa sự phân tách và chất lượng âm. Tham số này có thể được điều khiển bởi người dùng, hoặc phụ thuộc cụ thể vào ngữ cảnh tái tạo âm. Sự tái tạo âm tại nhà trong môi trường yên lặng trên trang thiết bị chất lượng cao có thể thu lợi từ chất lượng âm cao hơn và sự phân tách thấp hơn. Sự tái tạo âm trong phương tiện trong môi trường ồn ào qua loa phát thanh được dựng trong điện thoại thông minh có thể có lợi từ chất lượng âm thấp hơn nhưng sự phân tách cao hơn và độ rõ tiếng nói.

Tương tự, các số lượng được ước lượng (hoặc thành phần chất lượng âm hoặc các tham số xử lý) có thể còn được áp dụng để điều khiển sự xử lý sau hoặc để điều khiển sự phân tách thứ cấp.

Tiếp theo, bốn khái niệm khác nhau có thể được sử dụng cho sự thực thi phương pháp được đề xuất. Các khái niệm này được minh họa trên Fig.2, Fig.3, Fig.4 và Fig.5 và được mô tả sau đây.

Fig.2 minh họa thiết bị theo phương án mà được tạo cấu hình để sử dụng sự ước lượng chất lượng âm và được tạo cấu hình để thực hiện sự xử lý sau.

Theo phương án, môđun xác định 120 có thể, ví dụ, được tạo cấu hình để xác định, phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, giá trị chất lượng âm như một hoặc nhiều giá trị thu được, trong đó giá trị chất lượng âm biểu thị chất lượng âm được ước lượng của tín hiệu đích được ước lượng. Môđun xác định 120 có thể,

ví dụ, được tạo cấu hình để xác định một hoặc nhiều giá trị tham số phụ thuộc vào giá trị chất lượng âm.

Do đó, theo phương án, môđun xác định 120 có thể, ví dụ, được tạo cấu hình để xác định, phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng, tham số điều khiển là một hoặc nhiều giá trị tham số. Bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để xác định tín hiệu âm thanh được phân tách phụ thuộc vào tham số điều khiển và phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

Các phương án cụ thể, được mô tả sau đây:

Trong bước thứ nhất, sự phân tách được áp dụng. Tín hiệu được phân tách và tín hiệu chưa được xử lý là các đầu vào cho môđun ước lượng chất lượng (Quality Estimation Module (QEM)). QEM tính toán ước lượng cho các thành phần chất lượng âm, $\hat{q}(n)$.

Các thành phần chất lượng âm được ước lượng $\hat{q}(n)$ được sử dụng để tính toán tập hợp các tham số $\hat{p}(n)$ để điều khiển sự xử lý sau.

Các biến số $q(n)$, $\hat{q}(n)$, $p(n)$, và $\hat{p}(n)$ có thể biến đổi thời gian, nhưng sự phụ thuộc thời gian được bỏ qua trong phần mô tả sau đây để chú giải rõ ràng.

Sự xử lý sau như vậy, ví dụ, cộng bản sao được định tỉ lệ hoặc được lọc của tín hiệu đầu vào cho bản sao được định tỉ lệ hoặc được lọc của tín hiệu đầu ra, và nhờ đó giảm sự suy giảm của tín hiệu nhiễu (ví dụ, ảnh hưởng của sự phân tách).

$$y(n) = p_1 \hat{s}(n) + (1 - p_1)x(n) \quad (13)$$

trong đó tham số p_1 điều khiển lượng phân tách.

Trong các phương án khác, công thức:

$$y(n) = p_1 \hat{s}(n) + (1 - p_1) \hat{b}(n)$$

có thể, ví dụ, được sử dụng, trong đó $\hat{b}$ là tín hiệu phần dư được ước lượng.

Giảm sự phân tách dẫn đến

1) lượng thành phần lạ giảm và

2) sự rõ ràng các âm nhiều tăng mà chặn các thành phần lạ phân tách.

Do đó, trong phương án, bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để xác định tín hiệu âm thanh được phân tách, phụ thuộc vào công thức (13), trong đó y là tín hiệu âm thanh được phân tách, trong đó $\hat{s}$ là tín hiệu đích được ước lượng, trong đó x là tín hiệu đầu vào âm thanh, trong đó p_1 là tham số điều khiển, và trong đó n là chỉ số.

Tham số được tính toán đã cho ước lượng của chất lượng âm $\hat{q}$ và ước số chất lượng đích q_0 ,

$$\hat{p} = f(\hat{q}, q_0) \quad (14)$$

Hàm f này có thể, ví dụ, tìm kiếm bao quát lặp đi lặp lại, như được minh họa bởi mã giả sau đây.

Thuật toán 1 Sự tìm kiếm tăng cường lặp đi lặp lại của $\hat{p}$

```

1:  $k \leftarrow 0$ 
2:  $p_k \leftarrow p_{init}$ 
3:  $\hat{q} \leftarrow QEM(x(n), \hat{s}(n))$ 
4:  $\nabla Q \leftarrow q_0 - \hat{q}$ 
5: while  $|\nabla Q| > Q_{thr}$  AND  $k < k_{max}$  do
6:    $k \leftarrow k + 1$ 
7:    $p_k \leftarrow p_{k-1} - \gamma_k \nabla Q$ 
8:    $p_k \leftarrow \min(1, \max(0, p_k))$ 
9:    $y(n) \leftarrow p_k \hat{s}(n) + (1 - p_k)x(n)$ 
10:   $\hat{q} \leftarrow QEM(x(n), y(n))$ 
11:   $\nabla Q \leftarrow q_0 - \hat{q}$ 
12: end while
13:  $\hat{p} \leftarrow p_k$ 

```

Cách khác, mối quan hệ $\hat{p} = f(\hat{q})$ có thể được tính toán bởi

1. Tính toán $\hat{q}_k$ cho tập hợp các giá trị p_k , $k = 1 \dots K$
2. Tính toán các giá trị còn lại của $\hat{q}$ bằng phép nội suy và phép ngoại suy.

Ví dụ, khi tham số xử lý p điều khiển sự xử lý sau như trong phương trình (13), $\hat{q}$ được tính toán số cố định của các giá trị của p_1 , ví dụ, tương ứng với 18, 12, và 6 dB của độ khuếch đại tương ứng là $\hat{s}$.

Do đó, sự ánh xạ $\hat{p} = f(\hat{q})$ được làm gần đúng và $\hat{p} = f(q_0)$ có thể được lựa chọn.

Tóm tắt lại, trong phương án, bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách bằng cách xác định phiên bản thứ nhất của tín hiệu âm thanh được phân tách và bằng cách biến đổi tín hiệu âm thanh được phân tách một hoặc nhiều lần để thu được một hoặc nhiều phiên bản trung gian của tín hiệu âm thanh được phân tách. Môđun xác định 120 có thể, ví dụ, được tạo cấu hình để biến đổi giá trị chất lượng âm phụ thuộc vào một giá trị trong số một hoặc nhiều giá trị trung gian của tín hiệu âm thanh được phân tách. Bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để dừng biến đổi tín hiệu âm thanh được phân tách, nếu giá trị chất lượng âm lớn hơn hoặc bằng giá trị chất lượng được định nghĩa.

Fig.3 minh họa thiết bị theo phương án khác, trong đó sự ước lượng trực tiếp của các tham số xử lý sau được thực hiện.

Đầu tiên, sự phân tách được áp dụng. Các tín hiệu được phân tách được nhập vào môđun ước lượng tham số (Parameter Estimation Module - PEM). Các tham số được ước lượng được áp dụng để điều khiển sự xử lý sau. PEM đã được đào tạo để ước lượng một cách trực tiếp $\hat{s}(n)$ và tín hiệu đầu vào $x(n)$. Điều này có nghĩa là thuật toán trong phương trình 14 được chuyển đến pha đào tạo và phương pháp hồi quy được đào tạo để ước lượng $\hat{p}$ thay vì $\hat{q}$. Do đó, hàm sau đây được học hỏi.

$$\hat{p} = z(x(n), \hat{s}(n)) \quad (15)$$

Rõ ràng rằng quy trình có lợi ích yêu cầu ít sự tính toán ngược lại với quy trình được mô tả trên đây. Điều này sẽ xảy ra tại giá trị có ít độ đàn hồi, vì mô hình được đào tạo cho thiết lập cố định là q_0 . Tuy nhiên, một số mô hình có thể được đào

tạo trên các giá trị khác nhau là q_0 . Sau đây, độ linh hoạt cuối cùng theo sự lựa chọn của q_0 có thể được tiếp tục sử dụng.

Theo phương án, bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự xử lý sau của tín hiệu đích được ước lượng.

Fig.4 minh họa thiết bị theo phương án khác, trong đó sự ước lượng chất lượng âm và sự phân tách thứ cấp được thực hiện.

Đầu tiên, sự phân tách được áp dụng. Các tín hiệu được phân tách được nhập vào QEM. Các thành phần chất lượng âm được ước lượng được sử dụng để tính toán tập hợp các tham số để điều khiển sự phân tách thứ cấp. Sự nhập vào sự phân tách thứ cấp $z(n)$ là tín hiệu đầu vào $x(n)$ hoặc sự xuất ra của sự phân tách thứ nhất $\hat{s}(n)$, sự tổ hợp tuyến tính của cả $z(n) = a x(n) + b \hat{s}(n)$ trong đó a và b là các tham số gán trọng số hoặc kết quả trung gian từ sự phân tách thứ nhất.

Do đó, trong phương án như vậy, bộ xử lý 130 có thể, ví dụ, được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh; hoặc bộ xử lý tín hiệu 130 có thể, ví dụ, được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng.

Các tham số phù hợp để điều khiển sự phân tách thứ cấp là, ví dụ, các tham số biến đổi các trọng số phổ.

Fig.5 minh họa thiết bị theo phương án khác, trong đó sự ước lượng trực tiếp của các tham số phân tách được thực hiện.

Đầu tiên, sự phân tách được áp dụng. Các tín hiệu được phân tách được nhập vào PEM. Các tham số được ước lượng điều khiển sự phân tách thứ cấp.

Sự nhập vào sự phân tách thứ cấp $z(n)$ là tín hiệu đầu vào $x(n)$ hoặc sự xuất ra của sự phân tách thứ nhất $\hat{s}(n)$, sự tổ hợp tuyến tính của cả $z(n) = a x(n) + b \hat{s}(n)$

trong đó a và b là các tham số gán trọng số hoặc trung gian thu được từ sự phân tách thứ nhất.

Ví dụ, các tham số sau được điều khiển: a , và c từ các phương trình (5), (6) và v như được mô tả trên đây.

Về sự xử lý lặp đi lặp lại theo các phương án, Fig.4 và Fig.5 biểu thị sự xử lý lặp đi lặp lại với một sự lặp đi lặp lại. Nhìn chung, điều này có thể được lặp lại nhiều lần, và được thực thi trong vòng lặp.

Sự xử lý lặp đi lặp lại (không có sự ước lượng chất lượng giữa chúng) là rất giống với các phương pháp ưu tiên khác mà nối chuỗi nhiều sự phân tách.

Phương pháp như vậy có thể, ví dụ, phù hợp để tổ hợp nhiều phương pháp khác nhau (mà tốt hơn việc lặp lại một phương pháp).

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 phạm vi 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. Trong một số phương án, một hoặc nhiều trong số các bước phương pháp quan trọng nhất có thể được thực hiện bằng các thiết bị như vậy.

Phụ thuộc vào các yêu cầu thực thi nhất định, các phương án của sáng chế có thể được thực thi trong phần cứng hoặc trong phần mềm hoặc ít nhất trong phần cứng hoặc ít nhất 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 bằng điện tử được lưu trữ trên đó, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính khả trình sao cho phương pháp tương ứng được thực hiện. Do đó, vật ghi lưu trữ số có thể 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 khả trình, 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 thi như sản phẩm chương trình máy tính với mã chương trình, mã chương trình hoạt động để thực hiện một trong 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 ở đây, được lưu trữ trên vật mang có thể đọc được bằng máy.

Nói cách khác, do đó, phương án của phương pháp theo sáng chế là, chương trình máy tính có mã chương trình để thực hiện một trong số các phương pháp được mô tả ở đây, 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, vật ghi đọc được bằng máy tính) gồm có, đã được ghi lại trên đó, chương trình máy tính để thực hiện một trong các phương pháp được mô tả ở đây. Vật mang dữ liệu, vật ghi lưu trữ số hoặc vật ghi được ghi lại dạng hữu hình điện hình và/hoặc không tạm thời.

Do đó, phương án khác của phương pháp theo sáng chế là dòng dữ liệu hoặc chuỗi các tín hiệu biểu diễn chương trình máy tính để thực hiện một trong số các phương pháp được mô tả ở đây. Ví dụ, có thể tạo cấu hình dòng dữ liệu hoặc chuỗi tín hiệu để được truyền dẫn 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 tạo 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 bao gồm máy tính có chương trình máy tính được cài đặt trên máy tính để thực hiện một trong số các phương pháp được mô tả ở đây.

Phương án khác theo sáng chế gồm có thiết bị hoặc hệ thống được cấu hình để truyền tải (ví dụ, bằng điện tử hoặc quang học) chương trình máy tính để thực hiện một trong số 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, bao gồm máy chủ tệp tin để truyền tải chương trình máy tính đến bộ nhận.

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

Thiết bị được mô tả ở đây có thể được thực hiện sử dụng thiết bị phần cứng, hoặc sử dụng máy tính, hoặc sử dụng tổ hợp của thiết bị phần cứng và máy tính.

Các phương pháp được mô tả ở đây có thể được thực hiện sử dụng thiết bị phần cứng, hoặc sử dụng máy tính, hoặc sử dụng tổ hợp của thiết bị phần cứng và máy tính.

Các phương án được mô tả bên trên chỉ mang tính minh họa cho các nguyên lý của sáng chế. Cần hiểu rằng các cải biên và biến đổi của các phương án và các chi tiết được mô tả ở đây sẽ là rõ ràng đối với người có hiểu biết 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. Thiết bị tạo ra tín hiệu âm thanh được phân tách từ tín hiệu đầu vào âm thanh, trong đó tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư, trong đó phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích, trong đó thiết bị bao gồm:

bộ phân tách nguồn (110) được tạo cấu hình để xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích,

môđun xác định (120), trong đó môđun xác định (120) được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được, và

bộ xử lý tín hiệu (130) để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư,

trong đó bộ xử lý (130) được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh; hoặc trong đó bộ xử lý tín hiệu (130) được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng.

2. Thiết bị theo điểm 1,

trong đó môđun xác định (120) được tạo cấu hình để xác định, phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng, tham số điều khiển là một hoặc nhiều giá trị tham số, và

trong đó bộ xử lý tín hiệu được tạo cấu hình để xác định tín hiệu âm thanh được phân tách phụ thuộc vào tham số điều khiển và phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

3. Thiết bị theo điểm 2,

trong đó bộ xử lý tín hiệu (130) được tạo cấu hình để xác định tín hiệu âm thanh được phân tách phụ thuộc vào:

$$y(n) = p_1 \hat{s}(n) + (1 - p_1) x(n),$$

hoặc phụ thuộc vào:

$$y(n) = p_1 \hat{s}(n) + (1 - p_1) \hat{b}(n),$$

trong đó y là tín hiệu âm thanh được phân tách,

trong đó $\hat{s}$ là tín hiệu đích được ước lượng,

trong đó x là tín hiệu đầu vào âm thanh,

trong đó $\hat{b}$ là tín hiệu phần dư được ước lượng,

trong đó p_1 là tham số điều khiển, và

trong đó n là chỉ số.

4. Thiết bị theo điểm 2 hoặc 3,

trong đó môđun xác định (120) được tạo cấu hình để xác định, phụ thuộc vào ít nhất một trong số tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, giá trị chất lượng âm như một hoặc nhiều giá trị thu được, trong đó giá trị chất lượng âm biểu thị chất lượng âm được ước lượng của tín hiệu đích được ước lượng, và

trong đó môđun xác định (120) được tạo cấu hình để xác định một hoặc nhiều giá trị tham số phụ thuộc vào giá trị chất lượng âm.

5. Thiết bị theo điểm 4,

trong đó bộ xử lý tín hiệu (130) được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách bằng cách xác định phiên bản thứ nhất của tín hiệu âm thanh được phân tách và bằng cách biến đổi tín hiệu âm thanh được phân tách một hoặc nhiều lần để thu được một hoặc nhiều phiên bản trung gian của tín hiệu âm thanh được phân tách,

trong đó môđun xác định (120) được tạo cấu hình để biến đổi giá trị chất lượng âm phụ thuộc vào một giá trị trong số một hoặc nhiều giá trị trung gian của tín hiệu âm thanh được phân tách, và

trong đó bộ xử lý tín hiệu (130) được tạo cấu hình để dùng biến đổi tín hiệu âm thanh được phân tách, nếu giá trị chất lượng âm lớn hơn hoặc bằng giá trị chất lượng được định nghĩa.

6. Thiết bị theo một trong số các điểm đã nêu,

trong đó môđun xác định (120) được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín hiệu đích được ước lượng và phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

7. Thiết bị theo một trong số các điểm đã nêu,

trong đó môđun xác định (120) bao gồm mạng nơon nhân tạo (125) để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín hiệu đích được ước lượng, trong đó mạng nơon nhân tạo (125) được tạo cấu hình để nhận nhiều giá trị đầu vào, mỗi giá trị trong số nhiều giá trị đầu vào phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh, và trong đó mạng nơon nhân tạo (125) được tạo cấu hình để xác định một hoặc nhiều giá trị thu được như một hoặc nhiều giá trị đầu ra của mạng nơon nhân tạo (125).

8. Thiết bị theo điểm 7,

trong đó mỗi giá trị trong số nhiều giá trị đầu vào phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh, và

trong đó một hoặc nhiều giá trị thu được biểu thị chất lượng âm được ước lượng của tín hiệu đích được ước lượng.

9. Thiết bị theo điểm 7,

trong đó mỗi giá trị trong số nhiều giá trị đầu vào phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng và tín hiệu đầu vào âm thanh, và

trong đó một hoặc nhiều giá trị thu được là một hoặc nhiều giá trị tham số.

10. Thiết bị theo một trong số các điểm từ 7 đến 9,

trong đó mạng nơron nhân tạo (125) được tạo cấu hình để được đào tạo bằng cách nhận nhiều tập hợp đào tạo, trong đó mỗi tập hợp trong số nhiều tập hợp đào tạo bao gồm nhiều giá trị đào tạo đầu vào của mạng nơron nhân tạo (125) và một hoặc nhiều giá trị đào tạo đầu ra của mạng nơron nhân tạo (125), trong đó mỗi giá trị trong số nhiều giá trị đào tạo đầu ra phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đích đào tạo và tín hiệu phần dư đào tạo và tín hiệu đầu vào đào tạo, trong đó mỗi giá trị trong số một hoặc nhiều giá trị đào tạo đầu ra phụ thuộc vào sự ước lượng của chất lượng âm của tín hiệu đích đào tạo.

11. Thiết bị theo điểm 10,

trong đó sự ước lượng chất lượng âm của tín hiệu đích đào tạo phụ thuộc vào một hoặc nhiều mô hình tính toán chất lượng âm.

12. Thiết bị theo điểm 11,

trong đó một hoặc nhiều mô hình tính toán chất lượng âm là ít nhất một trong số:

sự đánh giá phân tách nguồn mù,

các phương pháp đánh giá cảm quan cho sự phân tách nguồn âm thanh,

sự đánh giá cảm quan chất lượng âm thanh,
 sự đánh giá cảm quan chất lượng tiếng nói,
 âm thanh người nghe khách quan chất lượng tiếng nói ảo,
 chỉ số chất lượng âm thanh hỗ trợ nghe,
 chỉ số chất lượng tiếng nói hỗ trợ nghe,
 chỉ số cảm quan tiếng nói hỗ trợ nghe, và
 độ rõ khách quan thời gian ngắn.

13. Thiết bị theo một trong số các điểm từ 7 đến 12,

trong đó mạng neuron nhân tạo (125) được tạo cấu hình để xác định một hoặc nhiều giá trị thu được phụ thuộc vào tín hiệu đích được ước lượng và phụ thuộc vào ít nhất một tín hiệu trong số tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng.

14. Thiết bị theo một trong số các điểm đã nêu,

trong đó bộ xử lý tín hiệu (130) được tạo cấu hình để tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự xử lý sau của tín hiệu đích được ước lượng.

15. Phương pháp tạo ra tín hiệu âm thanh được phân tách từ tín hiệu đầu vào âm thanh, trong đó tín hiệu đầu vào âm thanh bao gồm phần tín hiệu âm thanh đích và phần tín hiệu âm thanh phần dư, trong đó phần tín hiệu âm thanh phần dư biểu thị phần dư giữa tín hiệu đầu vào âm thanh và phần tín hiệu âm thanh đích, trong đó phương pháp bao gồm:

xác định tín hiệu đích được ước lượng mà phụ thuộc vào tín hiệu đầu vào âm thanh, tín hiệu đích được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh đích,

xác định một hoặc nhiều giá trị thu được phụ thuộc vào chất lượng âm được ước lượng của tín hiệu đích được ước lượng để thu được một hoặc nhiều giá trị

tham số, trong đó một hoặc nhiều giá trị tham số là một hoặc nhiều giá trị thu được hoặc phụ thuộc vào một hoặc nhiều giá trị thu được, và

tạo ra tín hiệu âm thanh được phân tách phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào ít nhất một tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh và tín hiệu phần dư được ước lượng, tín hiệu phần dư được ước lượng là ước lượng của tín hiệu mà chỉ bao gồm phần tín hiệu âm thanh phần dư,

trong đó sự tạo ra tín hiệu âm thanh được phân tách được thực hiện phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu đầu vào âm thanh; hoặc trong đó sự tạo ra tín hiệu âm thanh được phân tách được thực hiện phụ thuộc vào một hoặc nhiều giá trị tham số và phụ thuộc vào sự tổ hợp tuyến tính của tín hiệu đích được ước lượng và tín hiệu phần dư được ước lượng.

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

1/6

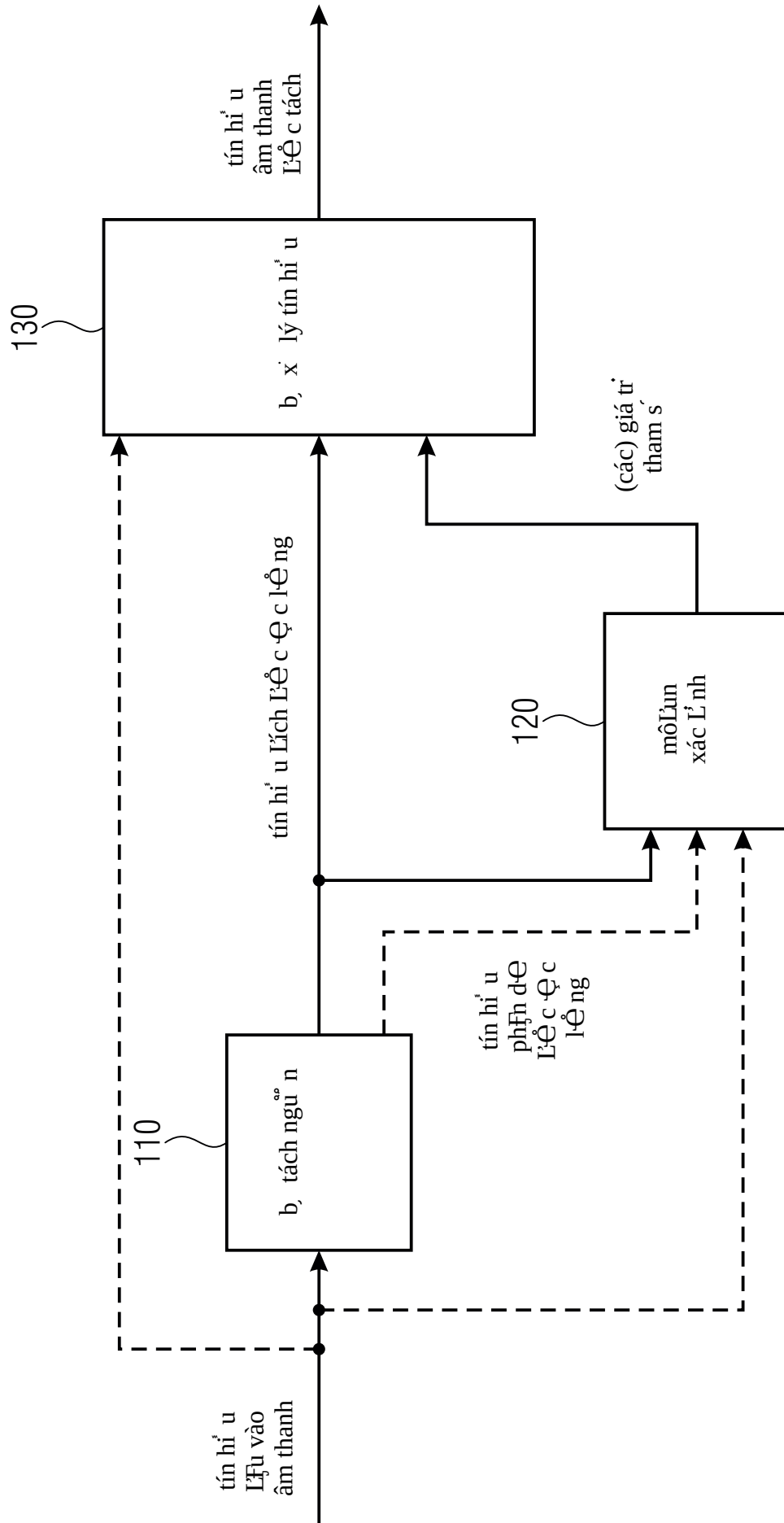


Fig. 1a

2/6

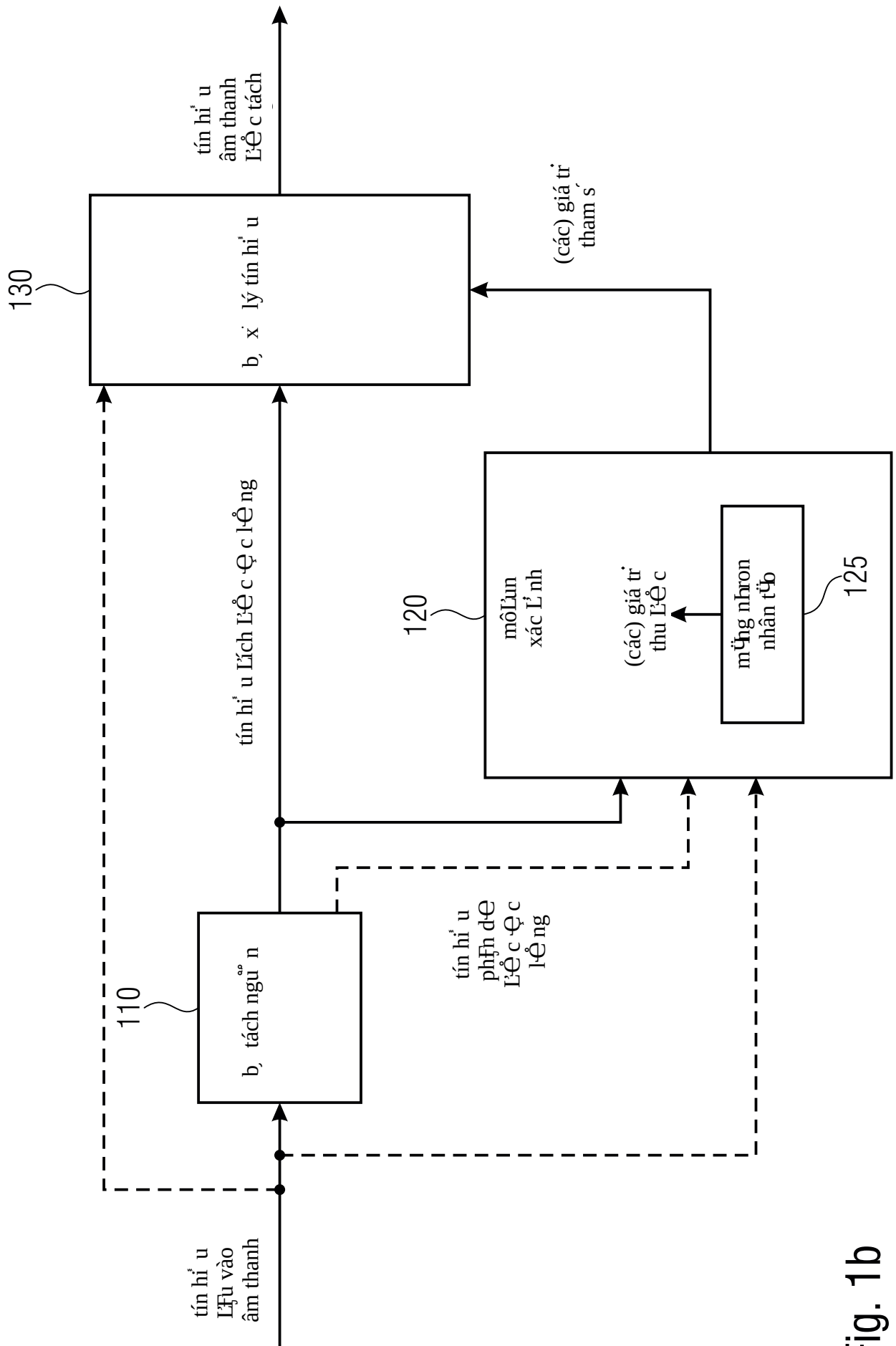


Fig. 1b

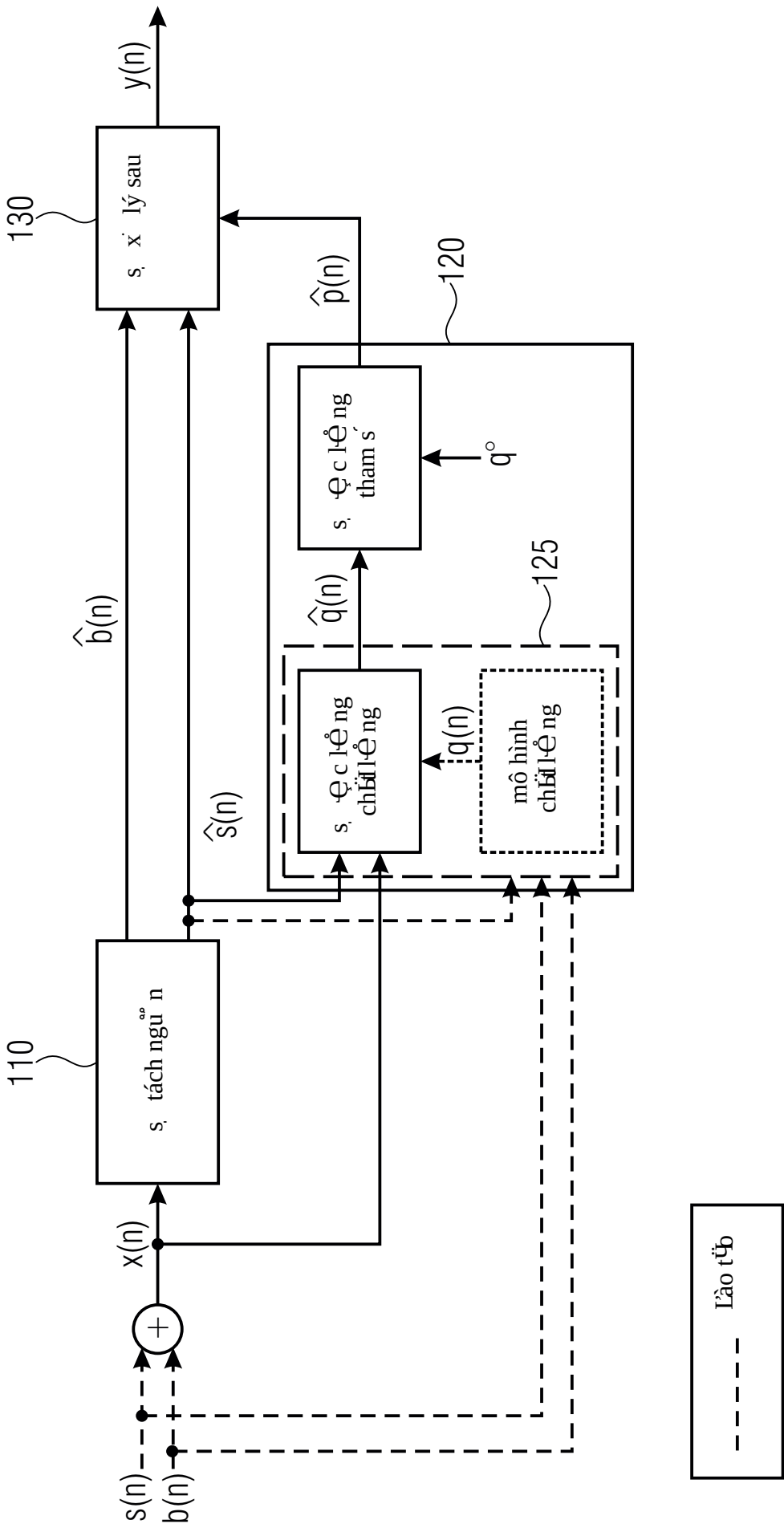


Fig. 2

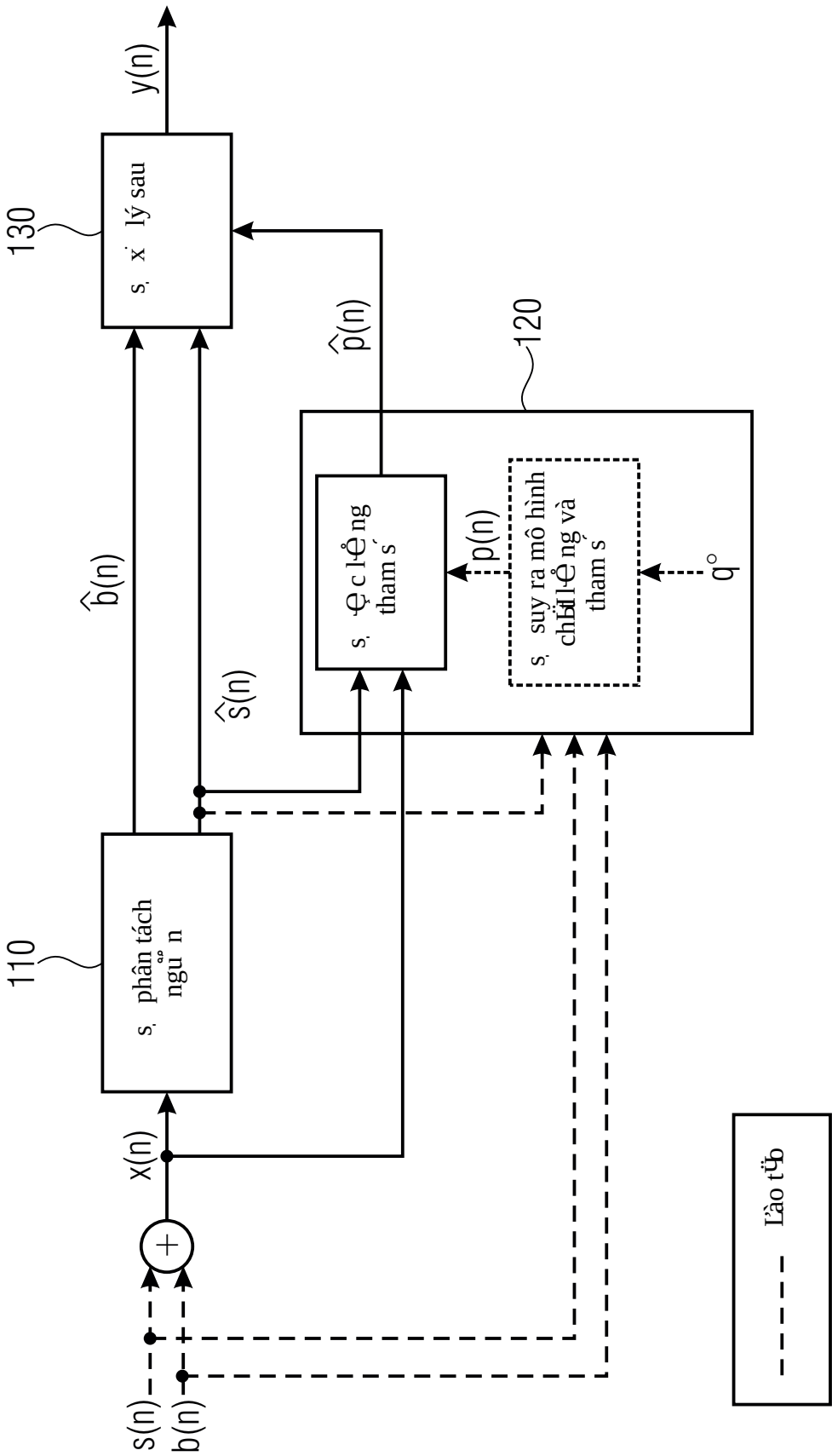


Fig. 3

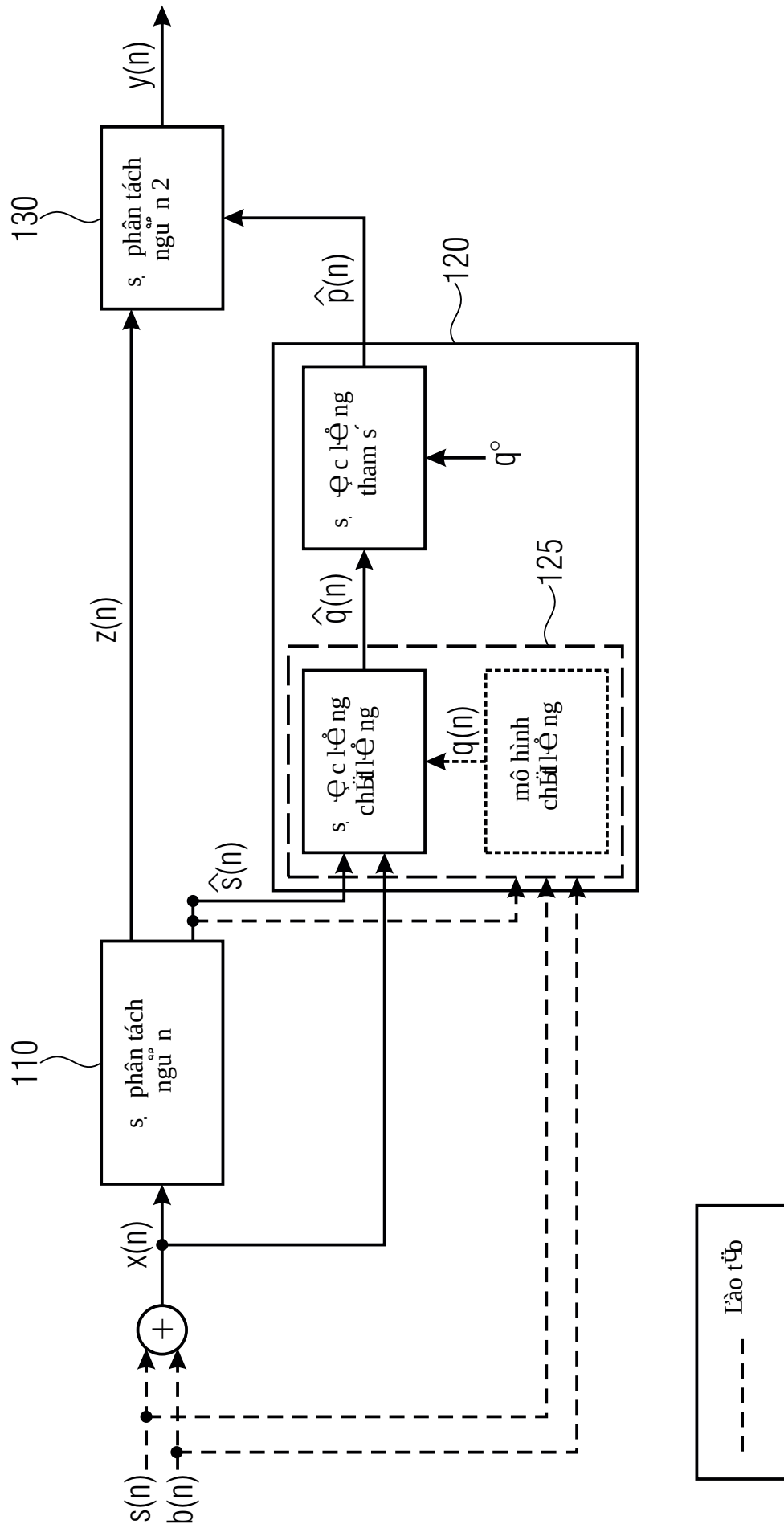


Fig. 4

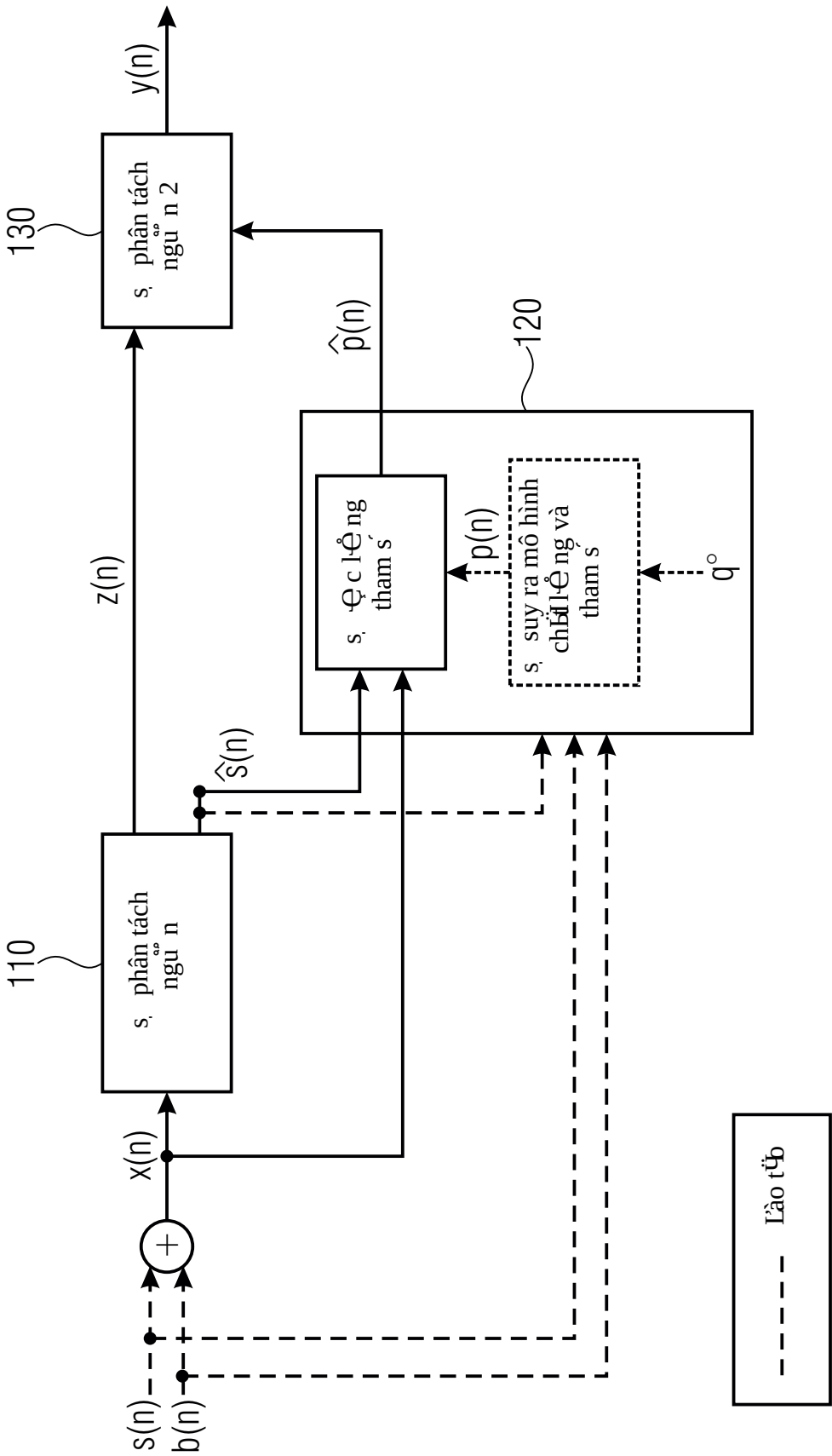


Fig. 5