(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) **1-0035918**

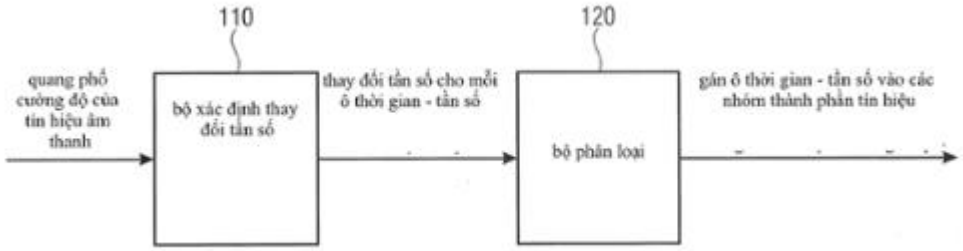
(51)[7] **G10H 1/40**; G10H 1/00; G06F 17/10; G06F 17/30

(13) **B**

(21) 1-2018-04596
(22) 16/03/2017
(86) PCT/EP2017/056257 16/03/2017
(87) WO/2017/158102 21/09/2017
(30) 16161251.0 18/03/2016 EP
(45) 26/06/2023 423
(43) 25/12/2018 369A
(73) Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e. V. (DE)
Hansastrasse 27c, 80686 Muenchen, Germany
(72) NIEDERMEIER, Andreas (DE); FUEG, Richard (DE); DISCH, Sascha (DE); MUELLER, Meinard (DE); DRIEDGER, Jonathan (DE).
(74) Công ty Luật TNHH Ambys Hà Nội (AMBYS HANOI)

(54) THIẾT BỊ VÀ PHƯƠNG PHÁP PHÂN TÍCH ẢNH PHỔ CƯỜNG ĐỘ CỦA TÍN HIỆU ÂM THANH

(57) Sáng chế đề cập đến thiết bị và phương pháp phân tích ảnh phổ cường độ của tín hiệu âm thanh. Thiết bị bao gồm bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định sự thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh. Hơn nữa, thiết bị bao gồm bộ phân loại (120) được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số và nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số đã nêu.



**Lĩnh vực kỹ thuật được đề cập**

Sáng chế đề cập đến xử lý tín hiệu âm thanh và cụ thể hơn là thiết bị và phương pháp phân tách âm thanh sóng hài - bộ gõ - phần dư sử dụng tenxơ cấu trúc trên các ảnh phổ.

**Tình trạng kỹ thuật của sáng chế**

Việc có thể phân tách âm thanh thành sóng hài của nó và thành phần bộ gõ là bước xử lý trước hiệu quả cho nhiều ứng dụng.

Trong khi "phân tách sóng hài - bộ gõ (- phần dư)" là thuật ngữ thông thường thì thuật ngữ này thường bị nhầm lẫn là cấu trúc sóng hài với các đường hình sin có tần số là bội số nguyên của tần số cơ bản. Mặc dù thuật ngữ chính xác cần là "Phân tách âm - bộ gõ (dư)", thuật ngữ "sóng hài" thay "âm" được sử dụng sau đây để dễ hiểu hơn.

Sử dụng thành phần bộ gõ được phân tách trong việc ghi âm nhạc, ví dụ, có thể mang đến sự cải thiện về chất lượng đối với việc theo dõi nhịp (xem [1]) phân tích và phiên âm nhịp điệu của các nhạc cụ nhịp điệu. Thành phần sóng hài được phân tách thích hợp cho việc phiên âm các nhạc cụ có cao độ và phát hiện hợp âm (xem [3]). Hơn nữa, phân tách sóng hài - bộ gõ có thể được sử dụng cho các mục đích phối lại như thay đổi tỉ số mức giữa cả hai thành phần tín hiệu (xem [4]) mà mang đến cảm quan âm thanh tổng thể "mượt hơn" và "mạnh mẽ hơn".

Một số phương pháp phân tách âm thanh sóng hài - bộ gõ dựa trên giả thuyết rằng các âm sóng hài có cấu cấu trúc theo chiều ngang trong ảnh phổ cường độ của tín hiệu đầu vào (theo hướng thời gian), trong khi các âm bộ gõ xuất hiện theo cấu trúc theo chiều dọc (theo hướng tần số). Ono và cộng sự đã trình bày phương pháp mà đầu tiên tạo ra các ảnh phổ được nâng cao theo sóng hài/bộ gõ bằng việc truyền theo hướng thời gian/tần số (xem [5]). Bằng việc so sánh các phép biểu diễn được nâng cao sau đó, quyết định liệu âm thanh là sóng hài hoặc âm bộ gõ có thể được suy ra.

Phương pháp tương tự được công bố bởi Fitzgerald, trong đó các ảnh phổ được

nâng cao được tính toán bằng việc sử dụng bộ lọc trung bình theo các hướng vuông góc thay vì truyền (xem [6]) mà mang đến kết quả tương tự trong khi giảm độ phức tạp tính toán.

Lấy cảm hứng bởi mô hình tín hiệu hình sin + chuyển tiếp + nhiễu âm (S+T+N) (xem [7], [8], [9]), kết cấu nhằm mô tả các thành phần tín hiệu tương ứng bằng tập hợp nhỏ gồm các tham số. Phương pháp Fitzgerald sau đó được mở rộng để phân tách sóng hài - bộ gõ - phần dư (harmonic - percussive - residual - HPR) trong [10]. Do các tín hiệu âm thanh thường bao gồm các âm thanh mà không phải sóng hài hoặc bộ gõ, quy trình này thu các âm thanh này trong thành phần dư thứ ba. Trong khi một vài tín hiệu dư này rõ ràng có cấu trúc đẳng hướng, không phải cấu trúc theo chiều ngang hoặc theo chiều dọc (ví dụ, nhiễu âm), tồn tại các âm thanh mà không có cấu trúc theo chiều ngang rõ ràng nhưng tuy nhiên vẫn mang thông tin âm và có thể được coi như là phần sóng hài của âm thanh. Ví dụ là các âm được điều biến tần số như các âm có thể xảy ra trong khi ghi âm tiếng đàn viôlông hoặc tiếng hát, trong đó chúng được coi là có "âm rung". Do chiến lược nhận diện cấu trúc theo chiều ngang hoặc theo chiều dọc, các phương pháp đã nêu thường không thể thu được các âm thanh này trong thành phần sóng hài của chúng.

Quy trình phân tách sóng hài - bộ gõ dựa trên việc tìm thừa số ma trận không âm mà có khả năng thu âm sóng hài với cấu trúc quang phổ không theo chiều ngang trong thành phần sóng hài được đề xuất trong [1]. Tuy nhiên, nó không gồm có thành phần dư thứ ba.

Tóm tắt các phương pháp ở trên, các phương pháp gần đây dựa trên sự quan sát rằng phép biểu diễn quang phổ, các âm sóng hài dẫn đến cấu trúc theo chiều ngang và các âm bộ gõ dẫn đến cấu trúc theo chiều dọc. Ngoài ra, các phương pháp kết hợp các cấu trúc mà không theo chiều ngang hoặc chiều dọc (tức là các âm thanh không phải sóng hài, không phải bộ gõ) với phần dư. Tuy nhiên, giả thiết này không áp dụng cho các tín hiệu như các âm được điều biến tần số mà thể hiện cấu trúc quang phổ dao động, trong khi vẫn mang thông tin âm.

Tenxơ cấu trúc, công cụ được sử dụng trong xử lý hình ảnh (xem [12], [13]) được áp dụng vào các hình ảnh định tỉ lệ xám cho các đường biên cho việc phát hiện

đường biên và góc (xem [14]) hoặc để ước lượng phương hướng của đối tượng. Tenxơ cấu trúc đã được sử dụng trong việc xử lý trước và chiết xuất đặc tính trong xử lý âm thanh (xem [15], [16]).

*Tài liệu tham khảo:*

[1] Aggelos Gkiokas, Vassilios Katsouros, George Carayannis, and Themos Stafylakis, "*Music tempo estimation and beat tracking by applying source separation and metrical relations*", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2012, pp. 421–424.

[2] Bernhard Lehner, Gerhard Widmer, and Reinhard Sonnleitner, "*On the reduction of false positives in singing voice detection*", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 7480–7484.

[3] Yushi Ueda, Yuuki Uchiyama, Takuya Nishimoto, Nobutaka Ono, and Shigeki Sagayama, "*HMM - based approach for automatic chord detection using refined acoustic features*", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, Texas, USA, 2010, pp. 5518–5521.

[4] Nobutaka Ono, Kenichi Miyamoto, Hirokazu Kameoka, and Shigeki Sagayama, "*A real - time equalizer of harmonic and percussive components in music signals*", in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Philadelphia, Pennsylvania, USA, 2008, pp. 139–144.

[5] Nobutaka Ono, Kenichi Miyamoto, Jonathan LeRoux, Hirokazu Kameoka, and Shigeki Sagayama, "*Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram*", in European Signal Processing Conference, Lausanne, Switzerland, 2008, pp. 240–244.

[6] Derry Fitzgerald, "*Harmonic/percussive separation using median filtering*", in Proceedings of the International Conference on Digital Audio Effects (DAFX), Graz, Austria, 2010, pp. 246–253.

[7] Scott N. Levine and Julius O. Smith III, "*A sines+transients+noise audio representation for data compression and time/pitch scale modications*", in Proceedings

of the AES Convention, 1998.

[8] Tony S. Verma and Teresa H.Y. Meng, "*An analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio*", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Seattle, Washington, USA, May 1998, pp. 3573–3576.

[9] Laurent Daudet, "*Sparse and structured decompositions of signals with the molecular matching pursuit*", IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 5, pp. 1808–1816, September 2006.

[10] Jonathan Driedger, Meinard Müller, and Sascha Disch, "*Extending harmonic - percussive separation of audio signals*", in Proceedings of the International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014, pp. 611–616.

[11] Jeongsoo Park and Kyogu Lee, "*Harmonic - percussive source separation using harmonicity and sparsity constraints*", in Proceedings of the International Conference on Music Information Retrieval (ISMIR), Málaga, Spain, 2015, pp. 148–154.

[12] Josef Bigun and Gösta H. Granlund, "*Optimal orientation detection of linear symmetry*", in Proceedings of the IEEE First International Conference on Computer Vision, London, UK, 1987, pp. 433–438.

[13] Hans Knutsson, "R*epresenting local structure using tensors*", in 6th Scandinavian Conference on Image Analysis, Oulu, Finland, 1989, pp. 244–251.

[14] Chris Harris and Mike Stephens, "*A combined corner and edge detector*", in Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 1988, pp. 147–151.

[15] Rolf Bardeli, "*Similarity search in animal sound databases*", IEEE Transactions on Multimedia, vol. 11, no. 1, pp. 68–76, January 2009.

[16] Matthias Zeppelzauer, Angela S. Stöger, and Christian Breiteneder, "*Acoustic detection of elephant presence in noisy environments*", in Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data, Barcelona, Spain, 2013, pp4. 3–8.

[17] Hanno Scharr, "*Optimale Operatoren in der digitalen Bildverarbeitung*", Dissertation, IWR, Fakultät für Physik und Astronomie, Universität Heidelberg, Heidelberg, Germany, 2000.

[18] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, "*Performance measurement in blind audio source separation*", IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1462–1469, 2006.

**Mục đích của sáng chế**

Mục tiêu của sáng chế là đề xuất khái niệm cải thiện cho xử lý tín hiệu âm thanh.

**Bản chất kỹ thuật của sáng chế**

Mục tiêu này của sáng chế được giải quyết bởi thiết bị theo điểm 1, phương pháp theo điểm 16, và chương trình máy tính theo điểm 17.

Thiết bị phân tích ảnh phổ cường độ của tín hiệu âm thanh được cung cấp. Thiết bị bao gồm bộ xác định thay đổi tần số được tạo cấu hình để xác định sự thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh. Hơn nữa, thiết bị bao gồm bộ phân loại được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số đã nêu.

Hơn nữa, phương pháp phân tích ảnh phổ cường độ của tín hiệu âm thanh được cung cấp. Phương pháp bao gồm:

- Xác định sự thay đổi tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh. Và:

- Gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số

đã nêu.

Hơn nữa, chương trình máy tính được cung cấp, 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 cụ thể với sự viện dẫn đến các hình vẽ, trong đó:

Fig.1 minh họa thiết bị phân tích ảnh phổ cường độ của tín hiệu âm thanh theo sáng chế,

Fig.2 minh họa ảnh phổ của hỗn hợp gồm tiếng hát, đàn catanet và tiếng vỗ tay với vùng được phóng đại theo phương án, trong đó định hướng của các mũi tên biểu thị hướng và trong đó độ dài của các mũi tên biểu thị ước số không đẳng hướng,

Fig.3 minh họa khoảng giá trị định hướng/không đẳng hướng được tính toán bằng việc sử dụng tenxơ cấu trúc theo phương án,

Fig.4 minh họa sự so sánh giữa phương pháp HPR - M và HPR - ST đối với đoạn trích của tín hiệu đầu vào tổng hợp,

Fig.5 minh họa thiết bị theo phương án của sáng chế, trong đó thiết bị bao gồm bộ tạo tín hiệu, và

Fig.6 minh họa thiết bị theo phương án của sáng chế, trong đó thiết bị bao gồm một hoặc nhiều ống phóng thanh để ghi âm tín hiệu âm thanh.

**Mô tả chi tiết sáng chế**

Fig.1 minh họa thiết bị phân tích ảnh phổ cường độ của tín hiệu âm thanh theo các phương án.

Thiết bị bao gồm bộ xác định thay đổi tần số 110. Bộ xác định thay đổi tần số 110 được tạo cấu hình để xác định sự thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh.

Hơn nữa, thiết bị bao gồm bộ phân loại 120. Bộ phân loại 120 được tạo cấu

hình để gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số đã nêu.

Theo một phương án, bộ xác định thay đổi tần số 110 có thể, ví dụ, được tạo cấu hình để xác định sự thay đổi tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào góc $\alpha(b, k)$ cho ô thời gian - tần số đã nêu. Góc $\alpha(b, k)$ cho ô thời gian - tần số đã nêu phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh.

Trong một phương án, bộ xác định thay đổi tần số 100 có thể, ví dụ được tạo cấu hình để xác định sự thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số còn phụ thuộc vào tần số lấy mẫu $f_s$ của tín hiệu âm thanh và phụ thuộc vào chiều dài $N$ của cửa sổ phân tích và phụ thuộc vào kích thước bước nhảy $H$ của cửa sổ phân tích.

Theo một phương án, bộ xác định thay đổi 110 được tạo cấu hình để xác định thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào công thức

$$R(b,k) = \frac{{f_s}^2}{HN} \cdot \tan(\alpha(b,k))$$

$(b, k)$ biểu thị ô thời gian - tần số trong số nhiều ô thời gian - tần số, trong đó $R(b, k)$ biểu thị sự thay đổi của tần số cho ô thời gian - tần số đã nêu $(b, k)$, trong đó $b$ biểu thị thời gian, trong đó $k$ biểu thị tần số, trong đó $f_s$ biểu thị tần số lấy mẫu tủa tín hiệu âm thanh, trong đó $N$ chiều dài của cửa sổ phân tích, trong đó $H$ biểu thị kích thước bước của cửa sổ phân tích, và trong đó $\alpha(b, k)$ biểu thị góc cho ô thời gian - tần số đã nêu $(b, k)$, trong đó góc $\alpha(b, k)$ phụ thuộc vào ảnh phổ cường độ.

Trong một phương án, bộ xác định thay đổi tần số 110 có thể, ví dụ, được tạo cấu hình để xác định đạo hàm riêng $\mathbf{S}_b$ của ảnh phổ cường độ $\mathbf{S}$ của tín hiệu âm thanh đối với chỉ số thời gian. Trong phương án này, bộ xác định thay đổi tần số 110, ví dụ, có thể được tạo cấu hình để xác định đạo hàm riêng $\mathbf{S}_k$ của ảnh phổ cường độ $\mathbf{S}$ của tín hiệu âm thanh đối với chỉ số thời gian.

Hơn nữa, trong một phương án, bộ xác định thay đổi tần số 110 được tạo cấu hình để xác định tenxơ cấu trúc $T(b, k)$ cho mỗi ô thời gian - tần số $(b, k)$ trong số nhiều ô thời gian - tần số phụ thuộc vào đạo hàm riêng $S_b$ của ảnh phổ cường độ S của tín hiệu âm thanh đối với chỉ số thời gian và phụ thuộc vào đạo hàm riêng $S_k$ của ảnh phổ cường độ S của tín hiệu âm thanh đối với chỉ số thời gian.

Hơn nữa, trong phương một phương án, bộ xác định thay đổi tần số 110, ví dụ, có thể được tạo cấu hình để xác định góc $\alpha(b, k)$ cho mỗi ô thời gian - tần số $(b, k)$ của nhiều ô thời gian - tần số phụ thuộc vào tenxơ cấu trúc $T(b, k)$ cho ô thời gian - tần số $(b, k)$.

Theo một phương án, bộ xác định thay đổi tần số 110 có thể, ví dụ, được tạo cấu hình để xác định góc $\alpha(b, k)$ cho mỗi ô thời gian - tần số $(b, k)$ trong số nhiều ô thời gian - tần số bằng việc xác định hai thành phần của $v_1(b, k)$ và $v_2(b, k)$ véctơ riêng $v(b, k)$ của tenxơ cấu trúc ($T(b, k)$) của ô thời gian - tần số đã nêu $(b, k)$, và bằng việc xác định góc ($\alpha(b, k)$) cho ô thời gian - tần số đã nêu ($(b, k)$) theo

$$\alpha(b,k) = \operatorname{atan}\left(\frac{v_2(b,k)}{v_1(b,k)}\right) \in [-\pi/2; \pi/2]$$

.

$\alpha(b, k)$ biểu thị góc cho ô thời gian - tần số đã nêu ($(b, k)$), trong đó $b$ biểu thị thời gian, trong đó $k$ biểu thị tần số, và trong đó atan() biểu thị hàm tan nghịch đảo.

Trong một phương án, bộ phân loại 120, ví dụ, có thể được tạo cấu hình để xác định ước số không đẳng hướng cho mỗi ô thời gian - tần số $(b, k)$ của nhiều ô thời gian - tần số phụ thuộc vào ít nhất một trong các công thức:

$$\left(\frac{\mu(b,k)-\lambda(b,k)}{\mu(b,k)+\lambda(b,k)}\right)^2$$

và

$$\mu(b,k) + \lambda(b,k) \geq e,$$

$\mu(b, k)$ là giá trị riêng thứ nhất $\lambda(b, k)$ là giá trị riêng thứ hai của tenxơ cấu

trúc ($T(b,\ k)$) của ô thời gian - tần số đã nêu $(b,\ k)$, và $e \in \mathbb{R}^{>0}$.

Trong phương án này, bộ phân loại 120 ví dụ, có thể được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu còn phụ thuộc vào sự thay đổi của ước số không đẳng hướng.

Theo một phương án, bộ phân loại 120, ví dụ, có thể được tạo cấu hình để xác định ước số không đẳng hướng cho ô thời gian - tần số đã nêu $(b,\ k)$ phụ thuộc vào công thức:

$$C(b,k) = \begin{cases} \left(\frac{\mu(b,k)-\lambda(b,k)}{\mu(b,k)+\lambda(b,k)}\right)^2 & , \mu(b,k)+\lambda(b,k) \geq e \\ 0 & , \text{ else} \end{cases}$$

$C(b,\ k)$ là ước số không đẳng hướng phụ thuộc vào ô thời gian - tần số đã nêu $(b,\ k)$, và trong đó bộ phân loại 120 được tạo cấu hình để gán ô thời gian - tần số đã nêu $(b,\ k)$ vào nhóm thành phần dư trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu, nếu ước số không đẳng hướng $C(b,\ k)$ là nhỏ hơn giá trị ngưỡng thứ nhất $c$, hoặc trong đó bộ phân loại 120 được tạo cấu hình để gán ô thời gian - tần số đã nêu $(b,\ k)$ vào nhóm thành phần dư trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu, nếu ước số không đẳng hướng $C(b,\ k)$ nhỏ hơn hoặc bằng với giá trị ngưỡng thứ nhất $c$, trong đó $c \in \mathbb{R}^{>0}$.

Trong một phương án, bộ phân loại 120, ví dụ, có thể được tạo cấu hình để gán mỗi ô thời gian - tần số I trong số nhiều ô thời gian tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi $R(b,\ k)$ của tần số được xác định cho ô thời gian - tần số đã nêu $(b,\ k)$, sao cho bộ phân loại 120 gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu sóng hài trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào liệu giá trị tuyệt đối $|R(b,\ k)|$ của sự thay đổi $R(b,\ k)$ của tần số được xác định cho ô thời gian - tần số $(b,\ k)$ có nhỏ hơn giá trị ngưỡng thứ hai $r_h$ hay không, hoặc phụ thuộc vào liệu giá trị tuyệt đối $|R(b,\ k)|$ của sự thay đổi $R(b,\ k)$ của

tần số được xác định cho ô thời gian - tần số đã nêu $(b,\ k)$ có nhỏ hơn hoặc bằng với giá trị ngưỡng thứ hai $r_h$ hay không, trong đó $r_h \in \mathbb{R}^{>0}$.

Theo một phương án, bộ phân loại 120, ví dụ, có thể được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi $R(b,\ k)$ của tần số được xác định cho ô thời gian - tần số đã nêu $(b,\ k)$, sao cho bộ phân loại 120 gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu bộ gõ trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào liệu giá trị tuyệt đối $|R(b,\ k)|$ của sự thay đổi $R(b,\ k)$ của tần số được xác định cho ô thời gian - tần số $(b,\ k)$ lớn hơn giá trị ngưỡng thứ ba $r_p$ hay không, hoặc phụ thuộc vào liệu giá trị tuyệt đối $|R(b,\ k)|$ của sự thay đổi $(R(b,\ k))$ của tần số được xác định cho ô thời gian - tần số đã nêu $(b,\ k)$ là lớn hơn hoặc bằng giá trị ngưỡng thứ ba $r_p$ hay không, trong đó $r_p \in \mathbb{R}^{>0}$.

Sau đây, mô tả chi tiết của các phương án được cung cấp.

Các phương án đề xuất các khái niệm được nâng cao cho việc phân tách âm thanh sóng hài - bộ gõ - phần dư (Harmonic - percussive - residual - HPR) dựa trên cấu trúc tenxơ. Một số phương án bao gồm các âm thanh được điều biến tần số mà giữ thông tin âm trong thành phần sóng hài bằng việc sử dụng thông tin về định hướng của cấu trúc quang phổ được cung cấp bởi cấu trúc tenxơ.

Các phương án dựa trên việc tìm ra rằng việc phân loại chặt chẽ theo chiều ngang và dọc là không phù hợp cho các tín hiệu này và có thể dẫn đến sự rò rỉ thông tin âm vào thành phần dư. Các phương án liên quan đến phương án mới mà thay vì sử dụng cấu trúc tenxơ, công cụ toán học, để tính toán góc định hướng nổi bật trong ảnh phổ cường độ. Các phương án sử dụng thông tin định hướng để phân biệt giữa các thành phần sóng hài, âm bộ gõ và tín hiệu dư, thậm chí trong trường hợp các tín hiệu điều biến tần số. Cuối cùng, hiệu quả của khái niệm của các phương án được xác minh bằng các phép đo đánh giá chủ quan cũng như các ví dụ âm thanh.

Hơn nữa, một số phương án dựa trên việc tìm ra rằng tenxơ cấu trúc có thể được coi là hộp đen, trong đó đầu vào là hình ảnh tỉ lệ xám và các đầu ra là các góc n

cho mỗi pixel tương ứng với hướng của thay đổi nhỏ nhất và phép đo chắc chắn hoặc đẳng hướng đối với hướng này cho mỗi pixel. Ngoài ra tenxơ cấu trúc cung cấp khả năng có thể được làm nhẵn, mà giảm ảnh hưởng của nhiễu âm đối với độ mạnh được cải thiện. Hơn nữa, phép đo chắc chắn có thể được sử dụng để xác định chất lượng của góc được ước lượng. Giá trị thấp của phép đo chắc chắn biểu thị rằng lớp pixel trong vùng độ sáng không đổi mà không có bất kỳ hướng nào rõ ràng.

Thay đổi tần số cục bộ, ví dụ, có thể được chiết xuất từ các góc được thu bằng tenxơ cấu trúc. Tạo thành các góc này, có thể được xác định, liệu ô thời gian - tần số trong ảnh phổ thuộc về thành phần sóng hài (= thay đổi tần số cục bộ thập) hay thành phần bộ gõ (= thay đổi tần số cục bộ cao).

Các phương án được cải thiện cho việc tách và phân loại sóng hài - bộ gõ - phần dư được cung cấp.

Việc phân tách âm thanh sóng hài - bộ gõ - phần dư là công cụ xử lý trước hữu ích cho các ứng dụng ví dụ như phiên âm các nhạc cụ hoặc chiết xuất nhịp điệu. Thay vì chỉ tìm kiếm các cấu trúc theo chiều ngang về theo chiều dọc một cách chặt chẽ, một số phương án xác định góc định hướng chiếm ưu thế cũng như không đẳng hướng cục bộ trong ảnh phổ bằng việc sử dụng tenxơ cấu trúc đã biết từ việc xử lý hình ảnh.

Trong các phương án, thông tin được cung cấp về định hướng của cấu trúc quang phổ sau đó có thể được sử dụng để phân biệt các thành phần tín hiệu sóng hài, bộ gõ và phần dư bằng việc thiết lập các ngưỡng phù hợp, xem Fig.2.

Fig.2 minh họa ảnh phổ của hỗn hợp gồm tiếng hát, tiếng đàn catanhet và tiếng vỗ tay với vùng được phóng to ngoài ra thể hiện hướng ( hướng các mũi tên) và phép đo không đẳng hướng (độ dài của mũi tên) được thu bởi tenxơ cấu trúc. Màu của các mũi tên biểu thị liệu các ô thời gian - tần số tương ứng có được gán cho thành phần sóng hài (các vùng 210), cho thành phần bộ gõ (các vùng 230) hoặc cho thành phần dư (các vùng 220) dựa trên thông tin định hướng và không đẳng hướng.

Tất cả các ô đều không có tốc độ thay đổi tần số cục bộ cao hoặc thấp hoặc ước số chắc chắn mà biểu thị vùng không đổi được gán thuộc về thành phần dư. Ví dụ về việc phân tách ảnh phổ có thể được thấy trên Fig.2. Các phương án thực hiện tốt cho

các tín hiệu âm thanh chưa các âm thanh được điều biến tần số hơn so với các phương pháp tương tự thực hiện trên ảnh phổ cường độ.

Đầu tiên, khái niệm tenxơ cấu trúc được mô tả và khái niệm chung này được mở rộng để có thể ứng dụng trong ngữ cảnh xử lý âm thanh.

Sau đó, các ma trận và các véctơ được viết bằng các chữ cái in đậm để thuận tiện cho việc ký pháp. Hơn nữa, bộ vận hành (.) được sử dụng để chỉ ra thành phần cụ thể. Trong trường hợp này, ma trận hoặc véctơ được viết như các chữ cái không được in đậm để thể hiện việc sử dụng vô hướng của các véctơ.

Đầu tiên, việc tính toán ảnh phổ theo các phương án được mô tả. Tín hiệu âm thanh, ví dụ, có thể là tín hiệu âm thanh đầu vào (rời rạc).

Tenxơ cấu trúc thể được áp dụng cho phép biểu diễn ảnh phổ của tín hiệu âm thanh của tín hiệu âm thanh đầu vào rời rạc $\boldsymbol{x} \in \mathbb{R}^M$với tần số lấy mẫu của $f_s$. Đối với phân tích quang phổ của $\boldsymbol{x}$, biến đổi Fourier thời gian ngắn (short - time Fourier - transform - STFT)

$$X(b,k) := \sum_{n=0}^{N-1} w(n)x(n+Hb)\exp\left(-i2\pi nk/N\right) \tag{1}$$

được sử dụng, trong đó $X(b,k) \in \mathbb{C}$, $b$ biểu thị chỉ số khung, $k$ là chỉ số tần số và $w \in \mathbb{R}^N$ là hàm cửa sổ của độ dài $N$ (Nói cách khác: $N$ là độ dài của cửa sổ phân tích). $H \in \mathbb{N}$, $H \leq N$ thể hiện kích thước bước phân tích của cửa sổ. Cần lưu ý rằng do quang phổ STFT có đối xứng nhất định quanh điểm Nyquist tại $\frac{N}{2}$, quy trình, ví dụ, có thể bị giới hạn tới $0 \leq k \leq \frac{N}{2}$, do đối xứng này có thể được khôi phục trong STFT nghịch đảo.

Bằng việc sử dụng công thức trên (1), ảnh phổ có thể được thu được. Ảnh phổ bao gồm nhiều quang phổ, trong đó nhiều quang phổ nối tiếp nhau theo thời gian. Quang phổ thứ hai trong số nhiều quang phổ nối tiếp quang phổ thứ nhất theo thời gian, nếu ít nhất một số mẫu miền thời gian thứ hai tồn tại được sử dụng để tạo ra quang phổ thứ hai và không được sử dụng để tạo ra quang phổ thứ nhất, và là các mẫu

miền thời gian mà liên quan đến thời điểm sau các mẫu miền thời gian thứ nhất mà được sử dụng để tạo ra quang phổ thứ nhất. Các cửa sổ của các mẫu miền thời gian được sử dụng để tạo quang phổ liền kề theo thời gian, ví dụ, có thể chồng lấp.

Trong các phương án, độ dài cửa sổ phân tích $N$, ví dụ, có thể được xác định bằng:

$$256 \text{ mẫu} \leq N \leq 2048 \text{ mẫu}.$$

Trong một số phương án, độ dài cửa sổ phân tích, ví dụ, có thể bằng 2048. Trong một số phương án khác, độ dài cửa sổ phân tích ví dụ, có thể bằng 1024 mẫu. Trong các phương án khác, độ dài cửa sổ phân tích, ví dụ, có thể bằng 768 mẫu. Trong các phương án khác, độ dài cửa sổ phân tích, ví dụ, có thể bằng 256 mẫu.

Trong các phương án, kích thước bước phân tích H, ví dụ, có thể nằm trong khoảng từ 35% đến 75% của cửa sổ phân tích. Trong các phương án này:

$$0{,}25\ N \leq H \leq 0{,}75\ N.$$

Do đó, trong các phương án này, nếu cửa sổ phân tích, ví dụ, có 2048 mẫu (N=2048), kích thước cửa sổ phân tích, ví dụ, có thể nằm trong khoảng:

$$512 \text{ mẫu} \leq H \leq 1536 \text{ mẫu}.$$

Nếu kích thước cửa sổ phân tích, ví dụ, bằng 256 mẫu ($N$ = 256), kích thước bước phân tích, ví dụ, có thể nằm trong khoảng:

$$64 \text{ mẫu} \leq H \leq 192 \text{ mẫu}.$$

Trong các phương án được ưu tiên, kích thước bước phân tích, ví dụ, có thể bằng 50% cửa sổ phân tích. Việc này tương ứng với sự chồng lấp cửa sổ của hai cửa sổ phân tích liền kề bằng 50%.

Trong một số phương án, kích thước bước phân tích, ví dụ, có thể bằng 45% của cửa sổ phân tích. Việc này tương ứng với sự chồng lấp cửa sổ của hai cửa sổ phân tích liền kề bằng 75%.

Trong các phương án khác, kích thước bước phân tích, ví dụ, có thể bằng 75% cửa sổ phân tích. Việc này tương ứng với sự chồng lấp cửa sổ của hai cửa sổ phân tích liền kề bằng 25 %.

Cần lưu ý rằng các khái niệm của sáng chế có thể được áp dụng đối với bất kỳ loại miền thời gian nào cho phép biến đổi miền quang phổ, ví dụ, cho MDCT (biến đổi cosin rời rạc cải biên - Modified Discrete Cosine Transform), MDST (Biến đổi Sin rời rạc cải biên - Modified Discrete Sine Transform), DSTFT (Biến đổi Fourier thời gian ngán rời rạc - Discrete Short - Time Fourier Transform), v.v.

Ảnh phổ logarit giá trị thực, ví dụ, có thể được tính toán như:

$$S(b,k) = 20\log_{10}|X(b,k)| \tag{2}$$

Ảnh phổ cường độ của tín hiệu âm thanh có thể được để cập đến như là **S** và giá trị ảnh phổ đối với ô thời gian - tần số $(b,\ k)$ có thể được đề cập đến như là $S(b,\ k)$.

Sau đây, việc tính toán tenxơ cấu trúc theo các phương án được mô tả.

Đối với việc tính toán tenxơ cấu trúc, đạo hàm riêng của S là cần thiết. Đạo hàm riêng liên quan đến chỉ số thời gian *b* được đưa ra bởi

$$\mathbf{S}_b = \mathbf{S} * \mathbf{d} \tag{3}$$

trong khi đạo hàm riêng liên quan đến chỉ số tần số *k* được xác định là

$$\mathbf{S}_k = \mathbf{S} * \mathbf{d}^T \tag{4}$$

trong đó **d** là toán chênh lệch rời rạc (ví dụ, đối với toán tử chênh lệch trung tâm chọn $\mathbf{d} = [-1,\ 0,\ 1]\ /\ 2$ ) và $*$ biểu thị phép nhân chập hai chiều.

Hơn nữa, nó còn được xác định bởi:

$$\mathbf{T}_{11} = (\mathbf{S}_b \odot \mathbf{S}_b) * \mathbf{G} \tag{5}$$

$$\mathbf{T}_{21} = \mathbf{T}_{12} = (\mathbf{S}_k \odot \mathbf{S}_b) * \mathbf{G} \tag{6}$$

$$\mathbf{T}_{22} = (\mathbf{S}_k \odot \mathbf{S}_k) * \mathbf{G} \tag{7}$$

trong đó $\odot$ là phép nhân ma trận theo điểm, mà cũng được biết đến như là tích Hadamard và **G** là bộ lọc làm nhẵn Gaussian 2-D có độ lệch tiêu chuẩn $\sigma_b$ theo hương chỉ số thời gian và $\sigma_k$ theo hướng chỉ số tần số. Tenxơ cấu trúc $T(b,\ k)$ mà sau đó

được đưa ra bởi ma trận nửa xác định đối xứng và dương $2 \times 2$

$$\mathbf{T}(b,k) = \begin{bmatrix} T_{11}(b,k) & T_{12}(b,k) \\ T_{21}(b,k) & T_{22}(b,k) \end{bmatrix} \quad (8)$$

Tenxơ cấu trúc chứa thông tin về hướng chủ yếu của ảnh phổ tại $(b,\ k)$. Cần lưu ý rằng trong trường hợp đặc biệt trong đó **G** là vô hướng, $\boldsymbol{T}(b,\ k)$ không chứa nhiều thông tin hơn gradient tại điểm này của ảnh phổ. Tuy nhiên, trái ngược với gradient, tenxơ cấu trúc có thể được làm nhẵn bởi **G** mà không có hiệu ứng hủy bỏ mà làm chúng mạnh hơn trong việc tương phản với nhiễu âm.

Cần lưu ý rằng, tenxơ cấu trúc $\boldsymbol{T}(b,\ k)$ được xác định cho mỗi ô thời gian - tần số $(b,\ k)$ trong số nhiều ô thời gian - tần số. Khi nhiều ô thời gian tần số được xem xét, ví dụ, các ô thời gian tần số $(0,\ 0);\ (0,\ 1);\ (0,\ 2);\ \ldots (1,\ 0);\ (1,\ 1);\ (1,\ 2);\ \ldots$ thì tồn tại nhiều tenxơ cấu trúc $\boldsymbol{T}(0,\ 0);\ \boldsymbol{T}(0,\ 1);\ \boldsymbol{T}(0,\ 2);\ \ldots \boldsymbol{T}(1,\ 0);\ \boldsymbol{T}(1,\ 1);\ \boldsymbol{T}(1,\ 2);\ \ldots$ Ví dụ, cho mỗi ô thời gian - tần số $(b,\ k)$ trong số nhiều ô thời gian - tần số, một tenxơ cấu trúc $\boldsymbol{T}(b,\ k)$ được xác định.

Sau đây, việc tính toán các góc và ước số không đẳng hướng theo các phương án được mô tả.

Thông tin về hướng cho mỗi ô trong ảnh phổ được thu được bằng việc tính toán các giá trị riêng $\lambda(b,\ k)$, $\mu(b,\ k)$ với $\lambda(b,\ k) \leq \mu(b,\ k)$ và các véctơ riêng tương ứng $\boldsymbol{v}(b,\ k) = [v_1(b,\ k), v_2(b,\ k)]^T$ và $\boldsymbol{w}(b,\ k) = [w_1(b,\ k), w_2(b,\ k)]^T$ của tenxơ cấu trúc $\boldsymbol{T}(b,\ k)$. Lưu ý rằng $\boldsymbol{v}(b,\ k)$, véctơ riêng tương ứng với giá trị riêng nhỏ hơn $\lambda(b,\ k)$, là chỉ theo hướng thay đổi thấp nhất của ảnh phổ tại chỉ số $(b,\ k)$, trong khi $\boldsymbol{w}(b,\ k)$ là chỉ theo hướng cao nhất. Do đó, góc của hướng tại ô nhất định có thể được thu được bởi

$$\alpha(b,k) = \operatorname{atan}\left(\frac{v_2(b,k)}{v_1(b,k)}\right) \in [-\pi/2; \pi/2] \quad (9)$$

$v_1(b,\ k)$ và $v_2(b,\ k)$ là các thành phần của vectơ riêng $\boldsymbol{v}(b,\ k)$.

atan() biểu thị hàm tan nghịch đạo.

Ngoài ra, ước số không đẳng hướng

$$C(b,k) = \begin{cases} \left(\frac{\mu(b,k)-\lambda(b,k)}{\mu(b,k)+\lambda(b,k)}\right)^2 & , \mu(b,k)+\lambda(b,k) \geq e \\ 0 & , \text{ else} \end{cases} \quad (10)$$

với $e \in \mathbb{R}^{>0}$ có thể được xác định cho mỗi ô. Cần lưu ý rằng $C(b,\ k) \in [0;\ 1]$. Các giá trị $C(b,\ k)$ gần bằng 1 biểu thị tính không đẳng hướngg cao của ảnh phổ tại chỉ số $(b,\ k)$, trong khi các vùng liền kề không đổi dẫn đến các giá trị gần bằng 0. Ngưỡng $e$, mà xác định giới hạn mà trên đó cái được coi là không đẳng hướng, có thể được lựa chọn để tăng thêm độ mạnh tương phản với nhiễu âm.

Ý nghĩa vật lý của góc $\alpha(b,\ k)$ có thể được hiểu bởi việc xem xét tín hiệu liên tục với sự thay đổi của tần số tức thời $\Delta f$ trong xuất khoảng thời gian $\Delta t$. Do đó tốc độ thay đổi tần số tức thời $R$ được chỉ ra bởi

$$R = \frac{\Delta f}{\Delta t} \quad (11)$$

Ví dụ, theo các phương án, các góc (được biểu thị bằng hướng của các mũi tên trên Fig.2) được thu bởi tenxơ cấu trúc, có thể, ví dụ được chuyển đổi thành tốc độ thay đổi tần số cục bộ

$$R = \frac{\Delta f}{\Delta t}[Hz/s] \quad (11a)$$

đối với mỗi ô thời gian - tần số của ảnh phổ.

Việc thay đổi tần số của ô tần số đối với mỗi ô thay đổi thời gian - tần số, ví dụ, có thể được đề cập đến như là tần số thay đổi tần số nhất thời.

Xem xét tốc độ lấy mẫu chiều dài và kích thước bước của phân tích STFT, mối liên hệ giữa các góc trong ảnh phổ và tốc độ thay đổi tần số tức thời $R(b,\ k)$ cho mỗi ô có thể được suy ra từ

$$R(b,k) = \frac{{f_s}^2}{HN} \cdot \tan(\alpha(b,k)) \qquad (12)$$

Hơn nữa, các độ lệch tiêu chuẩn G trong miền rời rạc $\sigma_b$ và $\sigma_k$ có thể được chuyển đổi thành các tham số vật lý liên tiếp $\sigma_t$ và $\sigma_f$ bởi

$$\sigma_t = \frac{H}{f_s}\sigma_b, \qquad \sigma_f = \frac{f_s}{N}\sigma_k \qquad (13)$$

Sau đây, việc phân tách sóng hài - bộ gõ - phần dư sử dụng tenxơ cấu trúc được mô tả.

Thông tin được thu được qua tenxơ cấu trúc có thể được áp dụng cho vấn đề phân tách HPR này, ví dụ, để phân loại mỗi ô của phân loại HPR, ví dụ, để phân loại mỗi ô trong số nhiều ảnh phổ như là một phần của sóng hài, của âm bộ gõ và thành phần dư đầu vào.

Các phương án dựa trên việc tìm ra rằng các ô được gán vào các thành phần sóng hài mà thuộc về cấu trúc theo chiều ngang hơn, trong khi các ô thuộc về cấu trúc theo chiều dọc hơn cần được gán cho thành phần bộ gõ. Hơn nữa, các ô không thuộc về bất kỳ cấu trúc được định hướng nào cần được gán cho thành phần dư.

Theo các phương án, ô $(b,\ k)$ ví dụ, có thể được gán cho thành phần sóng hài, nếu nó thỏa mãn điều kiện đầu tiên trong hai điều kiện sau đây.

Theo các phương án được ưu tiên, ô $(b,\ k)$, ví dụ, có thể được gán cho thành phần sóng hài, nếu nó thỏa mãn cả hai điều kiện sau đây:

Điều kiện thứ nhất, ví dụ, có thể là giá trị tuyệt đối của góc $\alpha(b,\ k)$ nhỏ hơn (hoặc bằng) ngưỡng $\alpha_h$. Ví dụ, ngưỡng $\alpha_h$ có thể nằm trong khoảng $\alpha_h \in [0;\ \pi/2]$. Điều này có nghĩa là ô cần là một phần của một số cấu trúc quang phổ mà không có độ nghiêng lơn hơn hoặc nhỏ hơn $\alpha_h$. Điều này có nghĩa là các âm thanh được điều biến tần số có thể được coi là một phần của thành phần sóng hài phụ thuộc vào tham số $\alpha_h$.

Điều kiện thứ hai, ví dụ, có thể là ước số không đẳng hướng $C(b,\ k)$

chứng minh rằng ô $(b,\ k)$ là một phần của một số cấu trúc không đẳng hướng, được định hướng, và do đó vượt quá ngưỡng thứ hai $c$. Cần lưu ý rằng đối với ngưỡng đã cho $(b,\ k)$, góc $\alpha(b,\ k)$ và ước số không đẳng hướng $C(b,\ k)$ cùng nhau xác định điểm trong t$\mathbb{R}^2$ được đưa ra trong hệ tọa độ cực.

Tương tự, trong các phương án, ngưỡng góc khác $\alpha_p$ được gán để xác định khi ô cần được gán cho thành phần bộ gõ (các vùng 330 với các được dọc trên Fig.3).

Do đó, theo các phương án, ô $(b,\ k)$ ví dụ, có thể được gán cho thành phần bộ gõ, nếu nó thỏa mãn điều kiện thứ nhất trong số hai điều kiện sau đây.

Theo các phương án được ưu tiên, ô $(b,\ k)$ ví dụ, có thể được gán cho thành phần bộ gõ, nếu nó thỏa mãn cả hai điều kiện sau đây:

Điều kiện thứ nhất có thể là, ví dụ, giá trị tuyệt đối của góc $\alpha(b,\ k)$ lớn hơn (hoặc bằng) với ngưỡng $\alpha_p$. Ví dụ, ngưỡng $\alpha_p$ có thể nằm trong khoảng $\alpha_p \in [0;\ \pi/2]$. Điều này có nghĩa là ô cần là một phần của một số cấu trúc quang phổ mà không có độ nghiêng lơn hơn hoặc nhỏ hơn $\alpha_p$. Điều này có nghĩa là các âm thanh được điều biến tần số có thể được coi là một phần của thành phần sóng hài phụ thuộc vào tham số $\alpha_p$.

Điều kiện thứ hai, ví dụ, có thể là ước số không đẳng hướng $C(b,\ k)$ chứng minh rằng ô $(b,\ k)$ là một phần của một số cấu trúc không đẳng hướng, được định hướng, và do đó vượt quá ngưỡng thứ hai $c$. Cần lưu ý rằng đối với ngưỡng đã cho $(b,\ k)$, góc $\alpha(b,\ k)$ và ước số không đẳng hướng $C(b,\ k)$ cùng nhau xác định điểm trong t$\mathbb{R}^2$ được đưa ra trong hệ tọa độ cực.

Cuối cùng, trong các phương án, tất cả các ô mà không được gán vào sóng hài hoặc thành phần bộ gõ, ví dụ, có thể được gán vào thành phần dư.

Quy trình gán được mô tả ở trên có thể được biểu thị bằng việc xác định mặt nạ cho thành phần sóng hài $\mathbf{M}_h$, mặt nạ cho thành phần bộ gõ $\mathbf{M}_p$ và mặt nạ cho thành phần dư $\mathbf{M}_r$.

Cần lưu ý rằng, thay vì sử dụng ngưỡng $\alpha_h$ và ngưỡng $\alpha_p$ các ngưỡng trong các

phương án, ví dụ, có thể được xác định trên tốc độ thay đổi tần số tuyệt đối cực đại $r_h$, $r_p \in \mathbb{R}^{>0}$ với $r_p \geq r_h$ để đưa diễn giải vật lý tốt hơn cho các tham số. Mặt nạ sau đó được đưa ra bởi:

$$M_h(b,k) = \begin{cases} 1 & , |R(b,k)| \leq r_h \ \wedge \ C(b,k) > c \\ 0 & , \text{else} \end{cases} \tag{14}$$

$$M_p(b,k) = \begin{cases} 1 & , |R(b,k)| > r_p \ \wedge \ C(b,k) > c \\ 0 & , \text{else} \end{cases} \tag{15}$$

$$M_r(b,k) = 1 - M_h(b,k) - M_p(b,k) \tag{16}$$

Cuối cùng, STFT của thành phần sóng hài $\mathbf{X}_h$, thành phần bộ gõ $\mathbf{X}_p$ và thành phần dư $\mathbf{X}_r$ được thu bởi

$$\mathbf{X}_h = \mathbf{M}_h \odot \mathbf{X} \tag{17}$$

$$\mathbf{X}_p = \mathbf{M}_p \odot \mathbf{X} \tag{18}$$

$$\mathbf{X}_r = \mathbf{M}_r \odot \mathbf{X} \tag{19}$$

Các tín hiệu thời gian tương ứng sau đó có thể được tính toán qua STFT nghịch đảo.

Fig.3 minh họa khoảng giá trị định hướng/ không đẳng hướng được tính toán bởi tenxơ cấu trúc.

Cụ thể, Fig.3 mô tả tập con gồm tất cả các điểm mà dẫn đến việc gán vào thành phần sóng hài. Cụ thể, các giá trị trong các vùng 310 với các đường sóng dẫn đến việc gán vào thành phần sóng hài.

Các giá trị trong vùng 330 với các đường dọc dẫn đến việc gán vào thành phần bộ gõ.

Các giá trị trong vùng 330 mà được tạo chấm dẫn đến việc gán vào thành phần phần dư.

Ngưỡng $\alpha_h$ xác định đường trên Fig.3, và ngưỡng $\alpha_p$ xác định đường trên Fig.3.

Fig.5 minh họa thiết bị theo sáng chế trong đó thiết bị bao gồm bộ tạo tín hiệu 130 được tạo cấu hình để tạo ra tín hiệu đầu ra âm thanh phụ thuộc vào việc gán nhiều ô thời gian - tần số vào hai hoặc nhiều nhóm ô thời gian hoặc tín hiệu.

Ví dụ, bộ tạo tín hiệu có thể lọc các thành phần khác nhau của tín hiệu âm thanh bằng việc áp dụng các thừa số gán trọng số khác nhau trên các giá trị cường độ của các ô thời gian - tần số của các nhóm thành phần tín hiệu khác nhau. Ví dụ, thành phần tín hiệu sóng hài có thể có thừa số gán trọng số thứ nhất $w_h$, nhóm thành phần tín hiệu bộ gõ có thể có thừa số gán trọng số thứ hai $w_p$, và nhóm thành phần tín hiệu dư có thể có thừa số gán trọng số thứ nhất $w_r$, và giá trị cường độ của mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số, ví dụ, có thể được gán trọng số với thừa số gán trọng số của nhóm thành phần tín hiệu, ô thời gian - tần số được gán trọng số vào.

Ví dụ, để nhấn mạnh các thành phần tín hiệu sóng hài, trong phương án, trong đó, các thừa số gán trọng số được nhân, ví dụ, với các giá trị cường độ tuyến tính.

$$w_h = 1,3,\ w_p = 0,7,\ \text{và}\ w_r = 0,2$$

Ví dụ, để nhấn mạnh các thành phần tín hiệu sóng hài, trong phương án, trong đó các thừa số gán trọng số, ví dụ, được thêm vào các giá trị cường độ logarit.

$$w_h = +0,26,\ w_p = -0,35,\ \text{và}\ w_r = -1,61$$

Ví dụ để nhấn mạnh các thành phần tín hiệu bộ gõ, trong phương án, trong đó các thừa số gán trọng số, ví dụ, được nhân với các giá trị cường độ tuyến tính.

$$w_h = 0,7,\ w_p = 1,3,\ \text{và}\ w_r = 0,2$$

Ví dụ, để nhấn mạnh các thành phần tín hiệu bộ gõ, trong phương án, trong đó các thừa số gán trọng số được thêm vào các giá trị cường độ logarit

$$w_h = -0,35,\ w_p = +0,26,\ \text{và}\ w_r = -1,61$$

Do đó, bộ tạo tín hiệu 130 được tạo cấu hình để áp dụng thừa số gán trọng số trên giá trị cường độ của ô thời gian - tần số trong số nhiều ô thời gian - tần số để thu tín hiệu đầu ra âm thanh, trong đó thừa số gán trọng số mà được áp dụng cho ô thời gian - tần số đã nêu phụ thuộc vào nhóm thành phần tín hiệu mà ô thời gian - tần số đã nêu được gán vào.

Theo phương án cụ thể của Fig.5, bộ tạo tín hiệu 120, ví dụ, có thể là bộ trộn tăng được tạo cấu hình để trộn tăng tín hiệu âm thanh để thu được tín hiệu đầu ra âm thanh bao gồm hai hoặc nhiều hơn hai kênh đầu ra âm thanh. Bộ trộn tăng, ví dụ, có thể được tạo cấu hình để tạo hai hoặc nhiều hơn hai kênh đầu ra âm thanh phụ thuộc vào việc gán nhiều ô thời gian - tần số vào hai hoặc nhiều hơn hai nhóm thành phần tín hiệu.

Ví dụ, hai hoặc nhiều hơn hai kênh đầu ra âm thanh có thể được tạo ra từ bộ lọc tín hiệu âm thanh, các thành phần khác nhau của tín hiệu âm thanh bằng việc áp dụng các thừa số gán trọng số khác nhau trên các giá trị cường độ của các ô thời gian - tần số của các nhóm thành phần tín hiệu khác nhau như được mô tả ở trên.

Tuy nhiên, để tạo các kênh âm thanh khác nhau, các trọng số khác nhau cho các nhóm thành phần tín hiệu có thể được sử dụng mà, ví dụ, có thể là cụ thể cho mỗi kênh trong số các kênh đầu ra âm thanh khác nhau.

Ví dụ, đối với kênh đầu ra âm thanh thứ nhất, các trọng số sẽ được thêm vào các giá trị cường độ logarit, ví dụ, có thể là

$$w_{1h} = +0{,}26,\ w_{1p} = -0{,}35, \text{ và } w_{1r} = -1{,}61.$$

Và đối với kênh đầu ra âm thanh thứ hai, các trọng số sẽ được thêm vào các giá trị cường độ logarit, ví dụ, có thể là

$$w_{2h} = +0{,}35,\ w_{2p} = -0{,}26, \text{ và } w_{2r} = -1{,}61.$$

Ví dụ, khi trộn tăng tín hiệu âm thanh để thu năm kênh đầu ra âm thanh bên trai phía trước, trung tâm, bên phải, xung quanh bên trái và xung quanh bên phải:

Thừa số gán trọng số $w_{1h}$ có thể lớn hơn để tạo ra các kênh đầu ra âm thanh bên trái, trung tâm và bên phải so với thừa số gán trọng số sóng hài $w_{2h}$ để tạo ra các kênh đầu ra âm thanh xung quanh bên trái và xung quanh bên phải.

- Thừa số gán trọng số bộ gõ $w_{1p}$ có thể nhỏ hơn để tạo ra các kênh đầu ra âm thanh bên trái, trung tâm và bên phải so với thừa số gán trọng số $w_{2p}$ để tạo ra các kênh đầu ra âm thanh xung quanh bên trái và xung quanh bên phải.

Các thừa số gán trọng số riêng biệt có thể được sử dụng cho mỗi kênh đầu ra âm thanh sẽ được tạo ra.

Fig.6 minh hoạ thiết bị theo phương án, trong đó thiết bị bao gồm một hoặc nhiều ống phóng thanh 171, 172 để ghi tín hiệu âm thanh.

Trên Fig.5, ống phóng thanh thứ nhất 171 ghi kênh âm thanh thứ nhất của tín hiệu âm thanh. Ống phóng thanh thứ hai tùy chọn 172 ghi kênh âm thanh thứ hai tùy chọn của tín hiệu âm thanh.

Hơn nữa, thiết bị theo Fig.6 còn bao gồm bộ tạo ảnh phổ cường độ 180 để tạo ảnh phổ cường độ của tín hiệu âm thanh từ tín hiệu âm thanh mà bao gồm kênh âm thanh thứ nhất và bao gồm một cách tùy chọn kênh âm thanh thứ hai tùy chọn. Việc tạo ra ảnh phổ cường độ từ tín hiệu âm thanh là khái niệm phổ biến đối với người có hiểu biết trong lĩnh vực kỹ thuật.

Tiếp theo, việc đánh giá các phương án được xem xét.

Để thể hiện hiệu quả của các phương ánh trong việc thu các âm được điều biến tần số trong thành phần sóng hài, phương pháp HPR dựa trên tenxơ cấu trúc (HPR - ST) theo các phương án được so sánh với phương pháp không lặp lại dựa trên việc lọc trung bình được biểu diễn trong [10] (HPR - M). Ngoài ra, các số đo cũng được tính toán cho các kết quả phân tách với mặt nạ nhị phân lý lưởng (ideal binary masks - IBM) mà đóng vai trò như là tham chiếu cho chất lượng phân tách có thể đạt được cực đại.

Xem xét các tham số hệ thống thử nghiệm, đối với cả HPR - ST cũng như HPR - M, các tham số STFT được chọn là $f_s$ = 22050Hz, $N$ = 1024 và $H$ = 256, sử dụng 9 cửa sổ cho **w**. Các tham số phân tách cho HPR - M được lựa chọn như trong các thí nghiệm được thực hiện trong [10]. Theo các phương án, tenxơ cấu trúc được tính toán sử dụng toán tử chênh lệch, ví dụ, toán tử Scharr [17] như là toán tử chênh lệch rời rạc **d**. Việc làm nhẵn được thực hiện sử dụng bộ lọc Gauss đẳng hướng $9 \times 9$ với các

độ lệch tiêu chuẩn $\sigma_b = \sigma_k = 1{,}4$ mà dẫn đến $\sigma_t \approx 16ms$ và $\sigma_f \approx 30Hz$. Cuối cùng, các ngưỡng cho việc phân tách được thiết lập bằng $e = 20, c = 0{,}2$ và $r_h = r_p = 10000Hz/s$.

Cần lưu ý rằng bằng việc lựa chọn $r_h$ và $r_p$ theo các phương án, thậm chí mỗi cấu trúc rất dốc trong ảnh phổ được gán cho thành phần sóng hài. Các phương án sử dụng việc quan sát về các âm rung của thế giới thực tại như là ví dụ thể hiện trên Fig.2. Ở đây, có thể thấy rằng tại một số thời khác, âm rung trong giọng hát có tốc độ thay đổi tần số tức thời rất cao. Hơn nữa, cần lưu ý rằng bằng việc lựa chọn $r_h = r_p$, việc gán ô trong ảnh phổ vào thành phần dư chỉ phụ thuộc vào ước số không đẳng hướng.

Hiệu quả của HPR - ST theo các phương án được đánh giá bằng việc so sánh nó với bộ lọc trung bình theo tình trạng kỹ thuật dựa trên phương pháp HPR - M được biểu diễn trong [10] bằng cả hai ước số đánh giá khách quan cũng như các mẫu âm thanh.

Để so sánh các trạng thái của HPR - ST theo các phương án và HPR - M theo tình trạng kỹ thuật khi được áp dụng vào các tín hiệu chứa các âm thanh được điều biến tần số để thu được các kết quả khách quan, hai mục thử nghiệm được tạo ra.

Mục thử nghiệm 1 gồm có sự chồng chấp của các âm thanh tổng hợp hoàn toàn. Các nguồn sóng hài được lựa chọn để có âm rung với tần số cơ bản bằng 1000Hz, tần số rung bằng 3Hz, phạm vi bằng 50Hz và 4 họa âm cao. Đối nguồn bộ gõ, một số xung được sử dụng, trong khi nhiễu âm trắng biểu diễn cho nguồn không phải sóng hài không phải nguồn dư bộ gõ.

Mục thực nghiêm 2 được tạo ra bằng việc chồng chất các tín hiệu thế giới thực của giọng hát với âm rung (sóng hài), catanhet (bộ gõ) và tiếng vỗ tay) không phải sóng hài hoặc bộ gõ).

Diễn giải việc phân tách HPR của các mục này như là vấn đề phân tách nguồn, các số đo đánh giá phân tách nguồn tiêu chuẩn được tính toán (tỉ số nguồn trên biến dạng SDR, tỉ số nguồn trên nhiễu SIR, và tỉ số nguồn trên thành phần lạ SAR, được giới thiệu trong [18]) cho các kết quả phân tách của cả hai quy trình. Các kết quả được thể hiện trong bảng 1.

Bảng 1 mô tả các ước số đánh giá khác quang, trong đó tất cả các giá trị được đưa ra theo đơn vị dB:

| | | SDR | | | SIR | | | SAR | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | IBM | HPR-M | HPR-ST | IBM | HPR-M | HPR-ST | IBM | HPR-M | HPR-ST |
| Mục 1 | Âm rung | 29,43 | 11,51 | 21,25 | 34,26 | 27,94 | 30,01 | 31,16 | 11,61 | 21,88 |
| | Các xung | 8,56 | -10,33 | -1,47 | 20,31 | -7,96 | 12,03 | 8,90 | 2,02 | -1,00 |
| | Nhiễu âm | 8,49 | -13,53 | 2,58 | 24,70 | -11,99 | 14,12 | 8,61 | 3,97 | 3,06 |
| Mục 2 | Tiếng hát | 14,82 | 6,48 | 9,18 | 22,75 | 20,83 | 15,61 | 15,60 | 6,68 | 10,42 |
| | Catanet | 8,48 | 3,79 | 2,37 | 21,59 | 16,09 | 17,96 | 8,73 | 4,16 | 2,56 |
| | Tiếng vỗ tay | 7,39 | -2,03 | -0,37 | 20,31 | 1,11 | 6,34 | 7,66 | 3,33 | 1,58 |

(Bảng 1)

Đối với mục 1 HPR - ST tạo ra ra SDR bằng 21,25dB cho âm rung, và do đó gần với kết quả phân tách tối ưu của IBM (29,43dB) hơn là kết quả phân tách của HPR - M (11,51dB). Điều này biểu thị rằng HPR - ST cải thiện việc thu âm thanh được điều biến tần số này trong thành phần sóng hài trong việc so sánh với HPRM. Việc này được thể hiện trên Fig.4.

Fig.4 minh họa sự so sánh giữa HPR - M và phương pháp HPR - ST đối với đoạn trích của tín hiệu đầu vào tổng hợp (mục 1). Để khả năng hiển thị được tăng cường, các ảnh phổ được tính toán với các tham số STFT khác nhau so với được sử dụng cho các thuật toán phân tách.

Fig.4(a) minh họa tần số của tín hiệu đầu vào theo thời gian. Trên Fig.4, các ảnh phổ của các thành phần sóng hài và tổng của thành phần bộ gõ và dư được tính toán cho cả hai quy trình được vẽ sơ đồ. Có thể thấy răng đối với HPR - M, các độ nghiêng dốc của âm rung bị rò rỉ vào thành phần dư (Fig.4(b) và (c)), trong khi HPR - ST (Fig.4(d) và (e)) tạo ra sự phân tách tốt. Điều này cũng giải thích các giá trị SIR rất

thấp cho thành phần dư so sánh với HPR - ST( ( - 11,99dB so với 14,12dB).

Cần lưu ý rằng giá trị SIR cao của HPR - M đối với thành phần sóng hài chỉ phản xạ rằng ít âm nhiễu từ các thành phần khác, mà không phải các âm thanh của âm rung mà được thu như một khối. Nói chung, hầu hết các quan sát đối với mục 1 là ít rõ rằng hơn, nhưng cũng hợp lệ đối với hỗn hợp các âm thanh trong thế giới thực trong mục 2. Đối với mục này, giá trị SIR của HPR - M đối với giọng hát thậm chí vượt quá giá trị SỈ của HPR - ST (20,83dB vs. 15,61dB). Một lần nữa, giá trị SIR thấp đối với tiếng vỗ tay chứng minh rằng các phần của âm rung trong tiếng hát bị rò rỉ vào các thành phần dư đối với HPR - M (1,1dB) trong khi thành phần dư HPR - ST chứa ít âm nhiễu hơn (6,34dB). Việc này biểu thị rằng các phương án có thể thu các cấu trúc được điều biến tần số của tiếng hát tốt hơn HPR - M.

Tóm tắt các kết quả, đối với các tín hiệu mà chứa các âm được điều biến tần số, khái niệm HPR - ST theo các phương án cung cấp các kết quả phân tách tốt hơn so với HPR - M.

Một số phương án sử dụng tenxơ cấu trúc để phát hiện tiếng hát. (Việc phát hiện giọng hát theo kỹ thuật đã biết được thể hiện trên [2]).

Mặc dù một vài khía cạnh đã được mô tả trong ngữ cảnh của thiết bị, rõ ràng rằng các khía cạnh này cũng biểu diễn sự mô tả của phương pháp tương ứng, mà khối hoặc thiết bị tương ứng với bước phương pháp hoặc dấu hiệu của bước phương pháp. Tương tự, các khía cạnh được mô tả trong ngữ cảnh của bước phương pháp cũng biểu diễn sự mô tả của khối hoặc mục hoặc dấu hiệu tương ứng của thiết bị tương ứng. Một số hoặc tất cả các bước phương pháp có thể được thực hiện bởi (hoặc sử dụng) thiết bị phần cứng, như ví dụ, bộ vi xử lý, máy tính có thể lập trình hoặc mạch điện tử. Trong một số phương án, một số 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 và các yêu cầu thực thi nhất định, các phương án của sáng chế có thể được thực hiện trong phần cứng hoặc trong phần mềm hoặc ít nhất một phần trong phần cứng hoặc một phần trong phần mềm. Phương án có thể được thực hiện sử dụng vật ghi lưu trữ số, ví dụ, đĩa mềm, DVD, Blu - Ray, CD, ROM, PROM, EPROM, EEPROM hoặc bộ nhớ FLASH, có các tín hiệu điều khiển có thể đọc được bằng điện

tử được lưu trữ trên đó, mà kết hợp (hoặc có thể kết hợp) với hệ thống máy tính có thể lập trình sao cho phương pháp tương ứng được thực hiện. Do đó, vật ghi lưu trữ số có thể có khả năng đọc được bằng máy tính.

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

Nói chung, các phương án của sáng chế có thể được thực hiện như sản phẩm chương trình máy tính với mã chương trình, mã chương trình có tác dụng để thực hiện một trong số các phương pháp khi sản phẩm chương trình máy tính chạy trên máy tính. Mã chương trình có thể ví dụ được lưu trữ trên sóng mang máy đọc được.

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

Do đó, nói cách khác, 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.

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

Do đó, phương án 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 cấu hình dòng dữ liệu hoặc chuỗi các tín hiệu để được truyền thông qua sự kết nối truyền thông dữ liệu, ví dụ thông qua Liên mạng.

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

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

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

Theo một số phương án, thiết bị lôgic 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 hoạ cho các nguyên lý của sáng chế. Cần hiểu rằng các sự biến thể và biến đổi của các phương án và các chi tiết được mô tả ở đây sẽ là rõ ràng đối với người có trình độ trong lĩnh vực kỹ thuật tương ứng. Do đó, mục đích của sáng chế là chỉ 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 cách mô tả và giải thích của các phương án ở đây.

## YÊU CẦU BẢO HỘ

1. Thiết bị phân tích ảnh phổ cường độ của tín hiệu âm thanh, thiết bị bao gồm:

bộ xác định sự thay đổi tần số (110) được tạo cấu hình để xác định sự thay đổi của tần số đối với mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh, và

bộ phân loại (120) được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số đã nêu.

2. Thiết bị theo điểm 1,

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định sự thay đổi tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào góc ($\alpha(b,\ k)$) cho ô thời gian - tần số đã nêu, trong đó góc ($\alpha(b,\ k)$) cho ô thời gian - tần số đã nêu phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh.

3. Thiết bị theo điểm 2,

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định sự thay đổi của tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số còn phụ thuộc vào tần số lấy mẫu ($f_s$) của tín hiệu âm thanh và phụ thuộc và độ dài ($N$) của cửa sổ phân tích và phụ thuộc vào kích thước bước ($H$) của cửa sổ phân tích.

4. Thiết bị theo điểm 3,

trong đó bộ xác định thay đổi tần số của thiết bị (110) được tạo cấu hình để xác định sự thay đổi tần số của mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số phụ thuộc vào công thức

$$R(b,k) = \frac{{f_s}^2}{HN} \cdot \tan(\alpha(b,k))$$

trong đó $(b,\ k)$ biểu thị ô thời gian - tần số trong số nhiều ô thời gian - tần số,

trong đó $R(b, k)$ biểu thị sự thay đổi của tần số đối với ô thời gian - tần số đã nêu $(b, k)$,

trong đó $b$ biểu thị thời gian,

trong đó $k$ biểu thị tần số,

trong đó $f_s$ biểu thị tần số lấy mẫu của tín hiệu âm thanh,

trong đó $N$ biểu thị độ dài của cửa sổ phân tích,

trong đó $H$ biểu thị kích thước bước của cửa sổ phân tích, và

trong đó $\alpha(b, k)$ biểu thị góc cho ô thời gian - tần số $(b, k)$, trong đó góc $\alpha(b, k)$ phụ thuộc vào ảnh phổ cường độ.

5. Thiết bị theo một điểm trong số các điểm từ điểm 2 đến 4,

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định đạo hàm riêng ($\mathbf{S}_b$) của ảnh phổ cường độ (**S**) của tín hiệu âm thanh đối với chỉ số thời gian,

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định đạo hàm riêng ($\mathbf{S}_k$) của ảnh phổ cường độ (**S**) của tín hiệu âm thanh đối với chỉ số thời gian, và

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định tenxơ cấu trúc ($\boldsymbol{T}(b, k)$) cho mỗi ô thời gian - tần số ($(b, k)$) trong số nhiều ô thời gian - tần số phụ thuộc vào đạo hàm riêng ($\mathbf{S}_b$) của ảnh phổ cường độ (**S**) của tín hiệu âm thanh đối với chỉ số thời gian và phụ thuộc vào đạo hàm riêng ($\mathbf{S}_k$) của ảnh phổ cường độ (**S**) của tín hiệu âm thanh đối với chỉ số tần số, và

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định góc ($\alpha(b, k)$) cho mỗi ô thời gian - tần số ($(b, k)$) trong số nhiều ô thời gian - tần số phụ thuộc vào tenxơ cấu trúc ($\boldsymbol{T}(b, k)$) đối với ô thời gian - tần số đã nêu ($(b, k)$).

6. Thiết bị theo điểm 5,

trong đó bộ xác định thay đổi tần số (110) được tạo cấu hình để xác định góc ($\alpha(b, k)$) cho mỗi ô thời gian - tần số ($(b, k)$) trong số nhiều ô thời gian - tần số bằng

việc xác định hai thành phần của $v_1(b,\ k)$ và $v_2(b,\ k)$ véctơ riêng $v(b,\ k)$ của tenxơ cấu trúc ($T(b,\ k)$) của ô thời gian - tần số đã nêu ($(b,\ k)$), và bằng việc xác định góc ($\alpha(b,\ k)$) cho ô thời gian - tần số đã nêu ($(b,\ k)$) theo công thức

$$\alpha(b,k) = \operatorname{atan}\left(\frac{v_2(b,k)}{v_1(b,k)}\right) \in [-\pi/2;\pi/2]$$

trong đó $\alpha(b,\ k)$ biểu thị góc cho ô thời gian - tần số đã nêu ($(b,\ k)$),

trong đó $b$ biểu thị thờigian

trong đó $k$ biểu thị tần số, và

trong đó atan() biểu thị hàm tiếp tuyến nghịch đảo.

7. Thiết bị theo điểm 5 hoặc điểm 6,

trong đó bộ phân loại (120) được tạo cấu hình để xác định ước số không đẳng hướng cho mỗi ô thời gian - tần số $(b,\ k)$ trong số nhiều ô thời gian - tần số phụ thuộc vào ít nhất một công thức:

$$\left(\frac{\mu(b,k)-\lambda(b,k)}{\mu(b,k)+\lambda(b,k)}\right)^2$$

và

$$\mu(b,k)+\lambda(b,k) \geq e,$$

trong đó $\mu(b,\ k)$ là véctơ riêng $\lambda(b,\ k)$ là véctơ thứ hai của tenxơ cấu trúc ($T(b,\ k)$) của ô thời gian - tần số đã nêu $(b,\ k)$, và $e \in \mathbb{R}^{>0}$,

trong đó bộ phân loại (120) được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều nhóm thành phần tín hiệu còn phụ thuộc vào sự thay đổi của ước số không đẳng hướng.

8. Thiết bị theo điểm 7,

trong đó bộ phân loại (120) được tạo cấu hình để xác định ước số không đẳng

hướng cho ô thời gian – tần số đã nêu $(b,\ k)$ phụ thuộc vào công thức:

$$C(b,k) = \begin{cases} \left(\frac{\mu(b,k)-\lambda(b,k)}{\mu(b,k)+\lambda(b,k)}\right)^2 & , \mu(b,k)+\lambda(b,k) \geq e \\ 0 & , \text{ else} \end{cases}$$

trong đó $C(b,\ k)$ là ước số không đẳng hướng phụ thuộc vào ô thời gian - tần số đã nêu $(b,\ k)$, và

trong đó bộ xác định (120) được tạo cấu hình để gán ô thời gian - tần số $(b,\ k)$ đã nêu vào nhóm thành phần dư trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu, nếu ước số không đẳng hướng $C(b,\ k)$ nhỏ hơn giá trị ngưỡng thứ nhất $c$, hoặc trong đó bộ phân loại (120) được tạo cấu hình để gán ô thời gian - tần số đã nêu $(b,\ k)$ vào nhóm thành phần dư trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu, nếu ước số không đẳng hướng $C(b,\ k)$ nhỏ hơn hoặc bằng với giá trị ngưỡng thứ nhất $c$,

trong đó $c \in \mathbb{R}^{>0}$.

9. Thiết bị theo một trong số các phương án đã nêu, trong đó bộ phân loại (120) được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi $(R(b,\ k))$ của tần số được xác định cho ô thời gian - tần số đã nêu $((b,\ k))$, sao cho bộ phân loại (120) gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu sóng hài trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào liệu giá trị tuyệt đối $(|R(b,\ k)|)$ của sự thay đổi $(R(b,\ k))$ tần số được xác định cho ô thời gian - tần số đã nêu $((b,\ k))$ có nhỏ hơn giá trị ngưỡng thứ hai $r_h$ hay không, hoặc phụ thuộc vào liệu giá trị tuyệt đối $(|R(b,\ k)|)$ của sự thay đổi $(R(b,\ k))$ của tần số được xác định cho ô thời gian - tần số đã nêu $((b,\ k))$ có nhỏ hơn hoặc bằng giá trị ngưỡng thứ hai $r_h$ hay không,

trong đó $r_h \in \mathbb{R}^{>0}$.

10. Thiết bị theo một điểm trong số các điểm từ 1 đến 8, trong đó bộ phân loại

(120) được tạo cấu hình để gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$), để bộ phân loại (120) gán ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào liệu giá trị tuyệt đối ($|R(b,\ k)|$) của sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$) có lớn hơn giá trị ngưỡng thứ ba $r_p$ hay không, hoặc phụ thuộc vào liệu giá trị tuyệt đối ($|R(b,\ k)|$) của sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$) có lớn hoặc hoặc bằng giá trị ngưỡng thứ ba $r_p$ hay không,

trong đó $r_p \in \mathbb{R}^{>0}$.

11. Thiết bị theo điểm 9, trong đó bộ phân loại (120) được tạo cấu hình để gán mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$), để bộ phân loại (120) gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu bộ gõ trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào liệu giá trị tuyệt đối ($|R(b,\ k)|$) của sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$) có lớn hơn giá trị ngưỡng thứ ba $r_p$ hay không, hoặc phụ thuộc liệu giá trị tuyệt đối ($|R(b,\ k)|$) của sự thay đổi ($R(b,\ k)$) của tần số được xác định cho ô thời gian - tần số đã nêu ($(b,\ k)$) có lớn hơn hoặc bằng giá trị ngưỡng thứ ba $r_p$ hay không,

trong đó $r_p \in \mathbb{R}^{>0}$.

12. Thiết bị theo một trong số các điểm đã nêu, trong đó thiết bị bao gồm bộ tạo tín hiệu (130) được tạo cấu hình để tạo tín hiệu đầu ra âm thanh phụ thuộc vào việc gán nhiều ô thời gian - tần số vào hai hoặc nhiều hơn hai nhóm thành phần tín hiệu.

13. Thiết bị theo điểm 12, trong đó bộ tạo tín hiệu (130) được tạo cấu hình để áp dụng thừa số gán trọng số trên giá trị cường độ ($w_h$, $w_p$, $w_r$) của mỗi ô thời gian -

quang phổ trong số nhiều ô thời gian - quang phổ để thu tín hiệu đầu ra âm thanh, trong đó thừa số gán trọng số ($w_h$, $w_p$, $w_r$) mà được áp dụng trên ô thời gian - tần số đã nêu phụ thuộc vào nhóm thành phần tín hiệu mà ô thời gian - tần số được gán vào.

14. Thiết bị theo điểm 12 hoặc điểm 13,

trong đó bộ xử lý tín hiệu (130) là bộ trộn tăng được tạo cấu hình để trộn tăng tín hiệu âm thanh để thu được tín hiệu đầu ra âm thanh bao gồm hai hoặc nhiều hơn hai kênh đầu ra âm thanh,

trong đó bộ trộn tăng được tạo cấu hình để tạo ra hai hoặc nhiều hơn hai kênh đầu ra âm thanh phụ thuộc vào việc gán nhiều ô thời gian - tần số vào hai hoặc nhiều hơn hai nhóm thành phần tín hiệu.

15. Thiết bị theo một điểm bất kì trong số các điểm nêu trên,

trong đó thiết bị bao gồm một hoặc nhiều ống phóng thanh (171, 172) để ghi âm tín hiệu âm thanh, và

trong đó thiết bị còn bao gồm bộ tạo ảnh phổ cường độ (180) để tạo ảnh phổ cường độ của tín hiệu âm thanh từ tín hiệu âm thanh.

16. Phương pháp phân tích ảnh phổ cường độ của tín hiệu âm thanh, phương pháp bao gồm:

xác định sự thay đổi tần số cho mỗi ô thời gian - tần số trong số nhiều ô thời gian - tần số của ảnh phổ cường độ của tín hiệu âm thanh phụ thuộc vào ảnh phổ cường độ của tín hiệu âm thanh, và

gán ô thời gian - tần số trong số nhiều ô thời gian - tần số vào nhóm thành phần tín hiệu trong số hai hoặc nhiều hơn hai nhóm thành phần tín hiệu phụ thuộc vào sự thay đổi của tần số được xác định cho ô thời gian - tần số đã nêu.

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

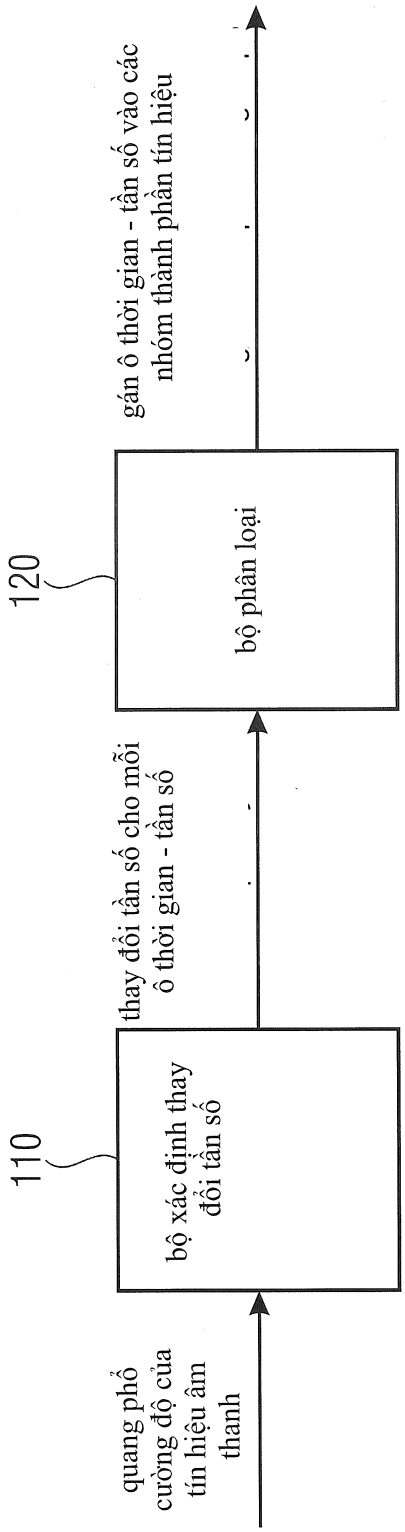

quang phổ cường độ của tín hiệu âm thanh
110
bộ xác định thay đổi tần số
thay đổi tần số cho mỗi ô thời gian - tần số
120
bộ phân loại
gán ô thời gian - tần số vào các nhóm thành phần tín hiệu

Fig. 1

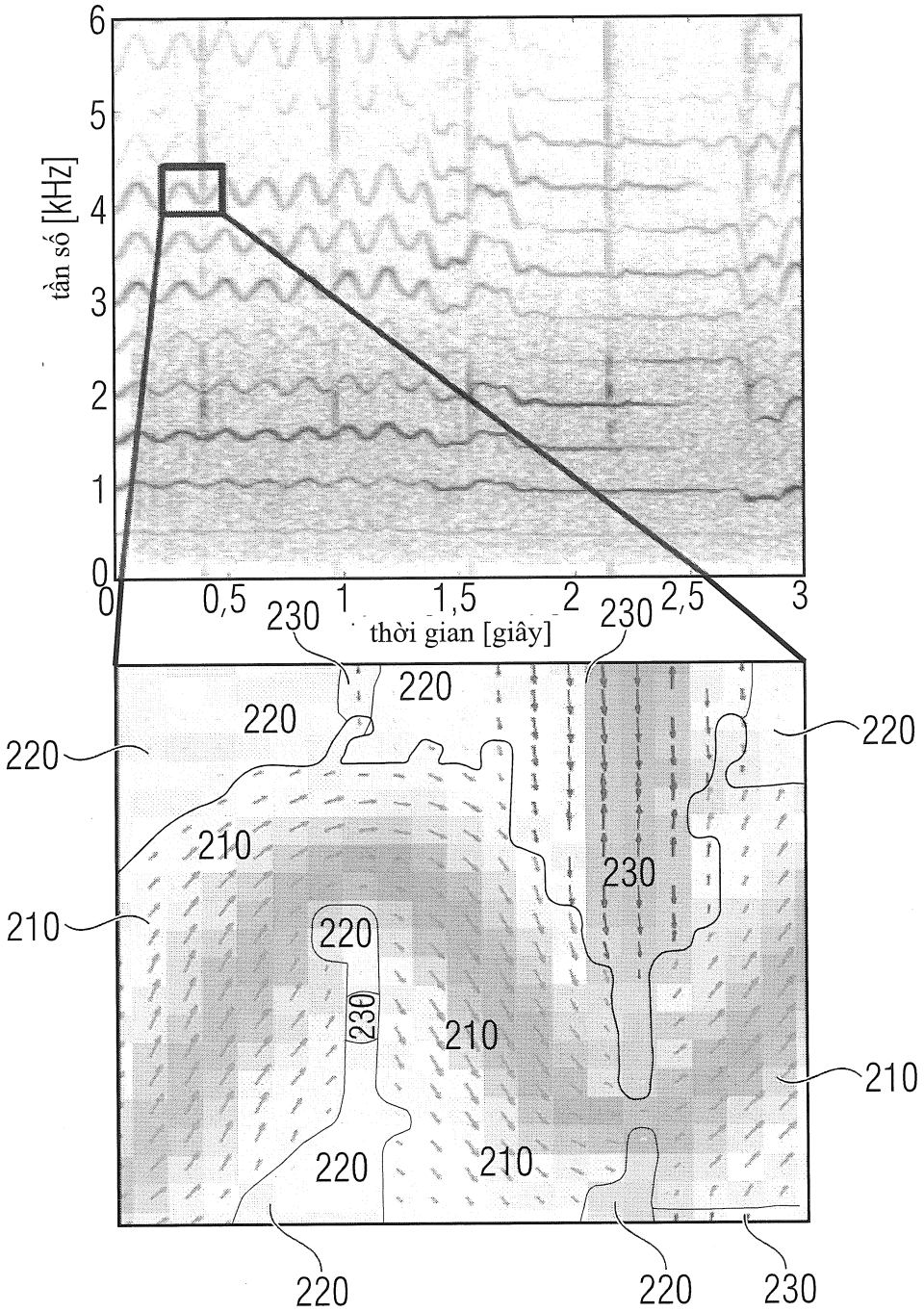

tần số [kHz]
6
5
4
3
2
1
0
0
0,5
1
1,5
2
2,5
3
thời gian [giây]
230
230
220
220
220
220
210
210
230
220
230
210
210
220
210
220
220
230

Fig. 2

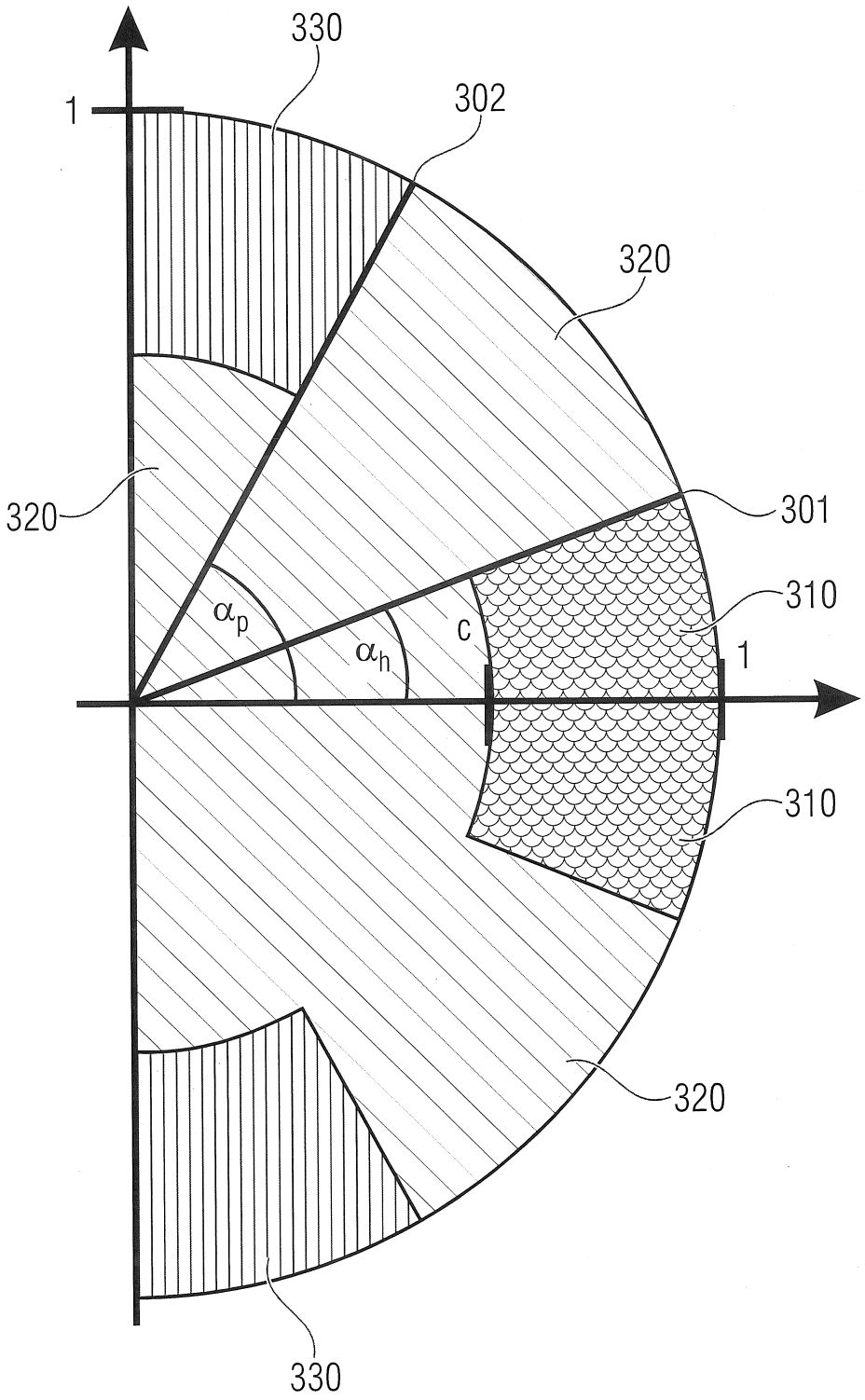



Fig. 3

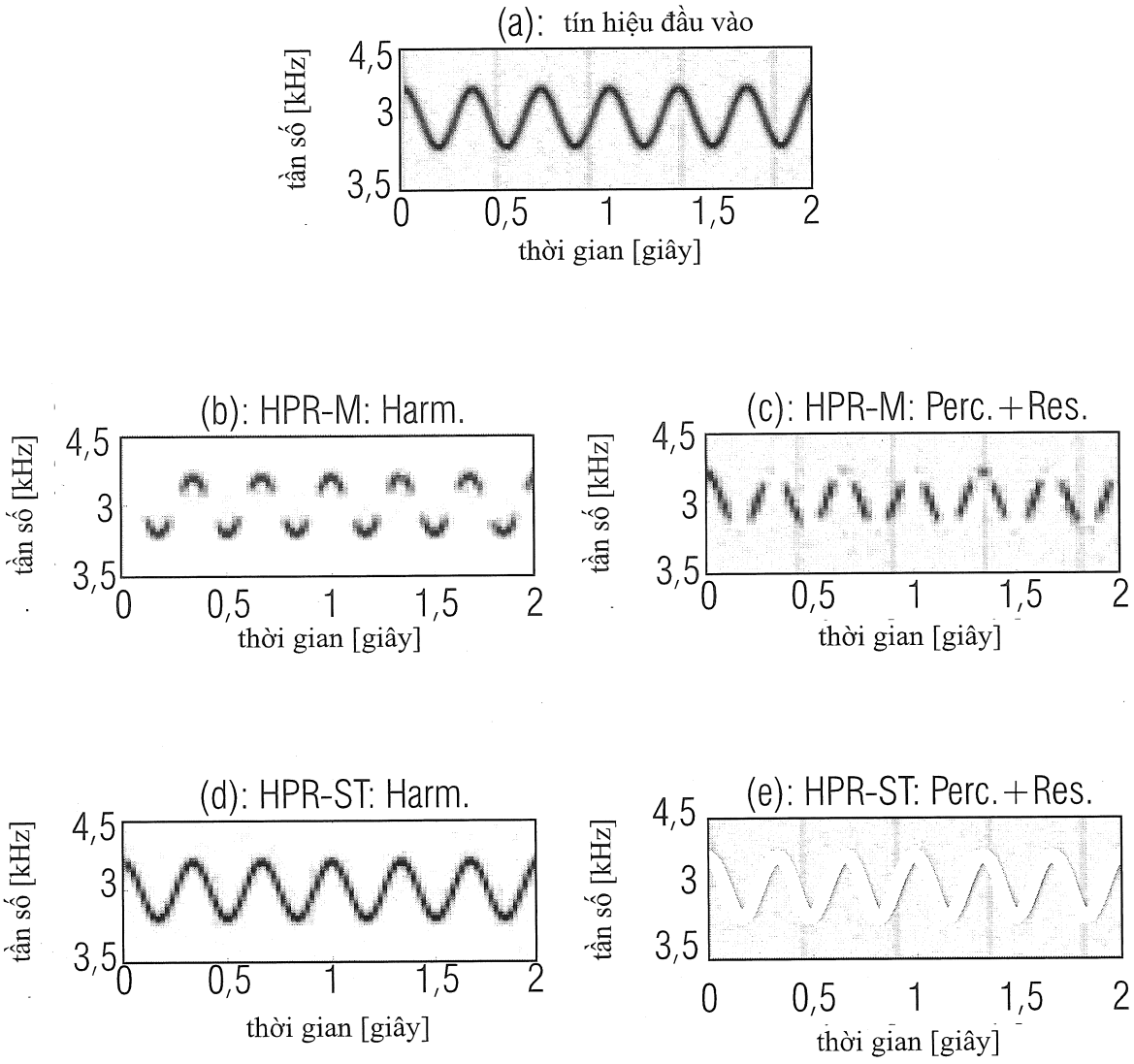

(a): tín hiệu đầu vào
tần số [kHz]
4,5
3
3,5
0
0,5
1
1,5
2
thời gian [giây]
(b): HPR-M: Harm.
(c): HPR-M: Perc.+Res.
(d): HPR-ST: Harm.
(e): HPR-ST: Perc.+Res.

Fig. 4

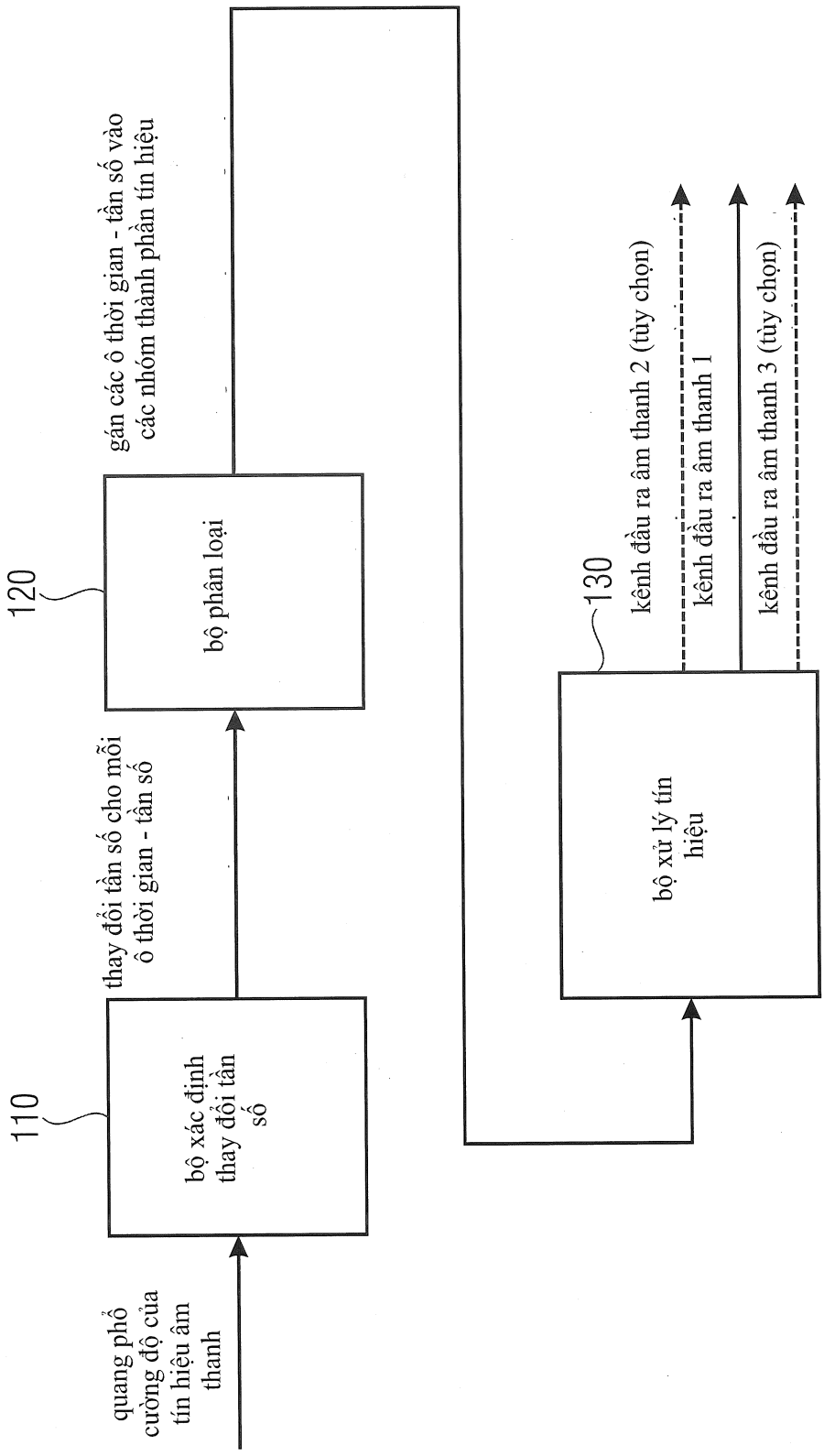

quang phổ cường độ của tín hiệu âm thanh
110
bộ xác định thay đổi tần số
thay đổi tần số cho mỗi ô thời gian - tần số
120
bộ phân loại
gán các ô thời gian - tần số vào các nhóm thành phần tín hiệu
130
bộ xử lý tín hiệu
kênh đầu ra âm thanh 2 (tùy chọn)
kênh đầu ra âm thanh 1
kênh đầu ra âm thanh 3 (tùy chọn)

Fig. 5

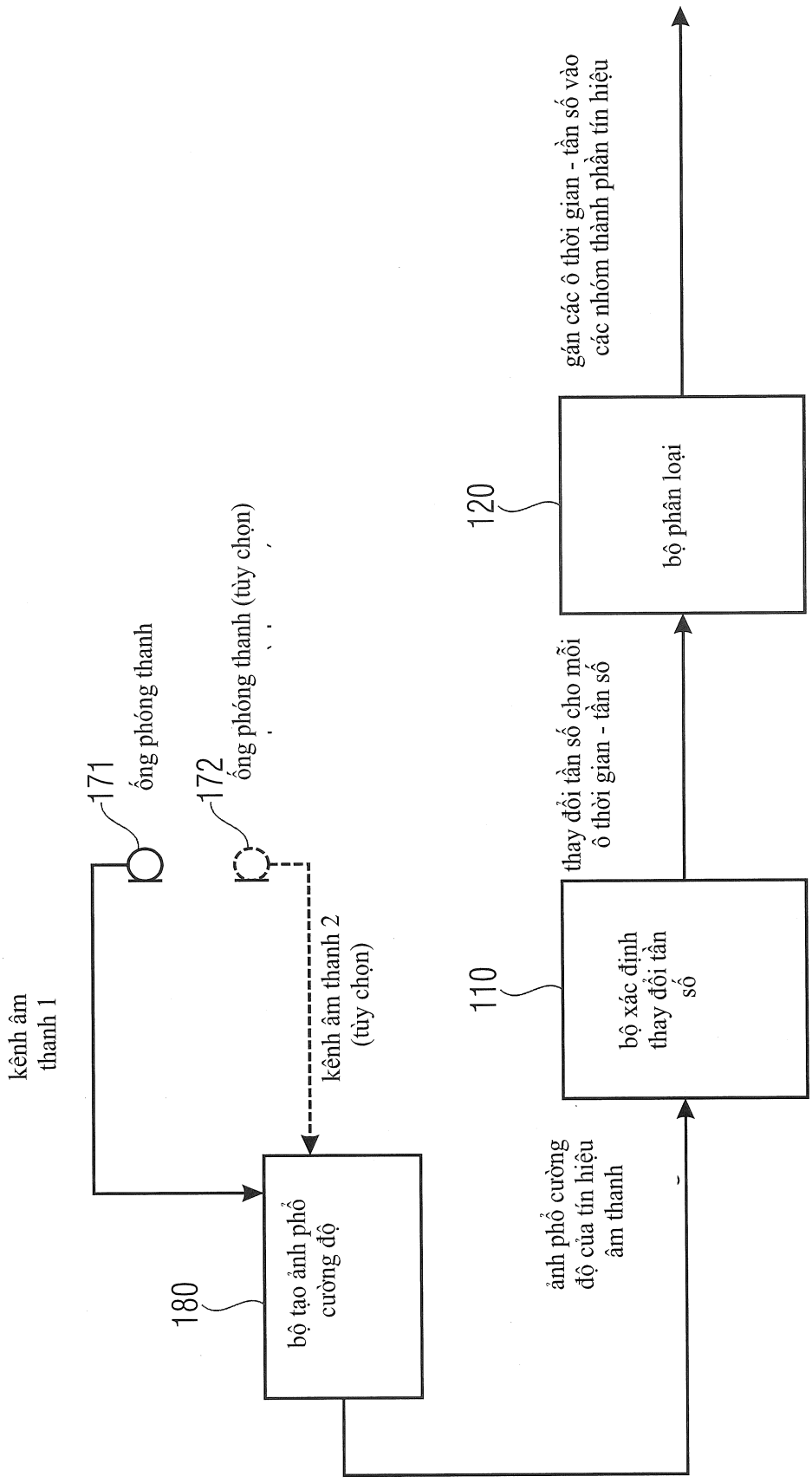

kênh âm thanh 1
171
ống phóng thanh
172
ống phóng thanh (tùy chọn)
kênh âm thanh 2 (tùy chọn)
180
bộ tạo ảnh phổ cường độ
ảnh phổ cường độ của tín hiệu âm thanh
110
bộ xác định thay đổi tần số
thay đổi tần số cho mỗi ô thời gian - tần số
120
bộ phân loại
gán các ô thời gian - tần số vào các nhóm thành phần tín hiệu

Fig. 6