



(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)



1-0023380

(51)⁷ G10L 25/78

(13) B

(21) 1-2015-00441

(22) 30/08/2013

(86) PCT/SE2013/051020 30/08/2013

(87) WO2014/035328 06/03/2014

(30) 61/695,623 31/08/2012 US

(45) 27/04/2020 385

(43) 25/05/2015 326A

(73) TELEFONAKTIEBOLAGET L M ERICSSON (PUBL) (SE)

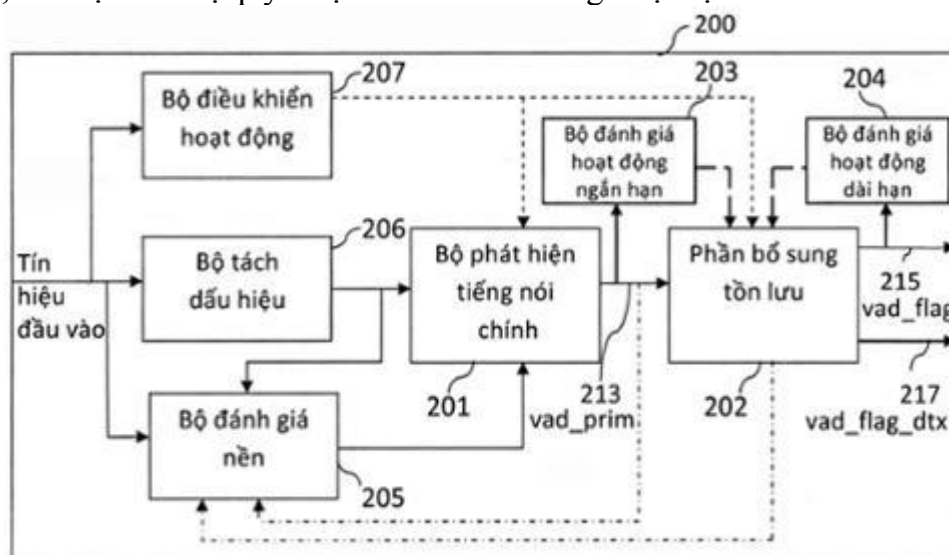
SE-164 83 Stockholm, Sweden

(72) SEHLSTEDT, Martin (SE)

(74) Công ty Luật TNHH T&G (TGVN)

(54) PHƯƠNG PHÁP VÀ THIẾT BỊ ĐỂ PHÁT HIỆN HOẠT ĐỘNG TIẾNG NÓI

(57) Theo một phương án làm ví dụ, sáng chế đề xuất phương pháp và thiết bị để phát hiện hoạt động tiếng nói (voice activity detection - VAD). VAD bao gồm bước tạo ra tín hiệu chỉ thị quyết định VAD chính và xác định việc bổ sung tồn lưu. Bước xác định bổ sung tồn lưu được tiến hành phụ thuộc vào số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn. Sau đó, tín hiệu chỉ thị quyết định VAD cuối cùng được tạo ra.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập chung đến phương pháp và thiết bị để phát hiện hoạt động tiếng nói (voice activity detection - VAD).

Tình trạng kỹ thuật của sáng chế

Trong các hệ thống mã hóa câu nói được dùng cho câu nói hội thoại, người ta thường sử dụng truyền gián đoạn (discontinuous transmission - DTX) để tăng hiệu suất mã hóa. Nguyên nhân là câu nói hội thoại chứa lượng đoạn ngắt lớn lồng vào trong câu nói, ví dụ, trong khi một người đang nói thì người kia nghe. Do đó với DTX bộ mã hóa câu nói chỉ hoạt động trung bình trong khoảng 50 phần trăm thời gian và phần còn lại có thể được mã hóa bằng cách sử dụng tiếng ồn tiện lợi. Một số bộ mã hóa-giải mã (codec) làm ví dụ có chức năng này là Băng hẹp đa tốc độ thích ứng (Adaptive Multi-Rate Narrow Band - AMR NB) và Bộ mã hóa-giải mã tốc độ khả biến nâng cao (Enhanced Variable Rate Codec - EVRC). AMR NB sử dụng DTX và EVRC sử dụng tốc độ bit khả biến (variable bit rate - VBR), trong đó Thuật toán xác định tốc độ (Rate Determination Algorithm - RDA) quyết định tốc độ dữ liệu nào được sử dụng cho mỗi khung, dựa trên quyết định VAD. Trong hoạt động DTX các khung hoạt động của câu nói được mã hóa bằng cách sử dụng bộ mã hóa-giải mã trong khi các khung nằm giữa các vùng hoạt động được thay thế bằng tiếng ồn tiện lợi. Các thông số về tiếng ồn tiện lợi được đánh giá trong bộ mã hóa và được gửi tới bộ giải mã bằng cách sử dụng tốc độ khung giảm và tốc độ bit thấp hơn so với tốc độ được sử dụng cho câu nói có hoạt động.

Đối với hoạt động DTX chất lượng cao, nghĩa là không suy giảm chất lượng câu nói, điều quan trọng là phát hiện được các khoảng thời gian của câu nói trong tín hiệu đầu vào. Điều này thường được thực hiện bởi Bộ phát hiện hoạt động tiếng nói (Voice Activity Detector - VAD) (bộ này được sử dụng cho cả DTX và RDA). Fig.1 là sơ đồ khối chung về một ví dụ của VAD tổng quát 100, bộ này lấy tín hiệu đầu vào 111, thường được chia thành các khung dữ liệu nằm trong khoảng từ 5 đến 30 mili giây tùy

thuộc vào việc thực thi, làm đầu vào và tạo ra các quyết định VAD làm đầu ra, thường là một quyết định cho mỗi khung. Nghĩa là, quyết định VAD là quyết định cho mỗi khung về việc liệu khung chứa câu nói hay tiếng ồn.

Quyết định sơ bộ, vad_prim 113, theo ví dụ này được đưa ra bởi bộ phát hiện tiếng nói chính 101 và, theo ví dụ này, về cơ bản chỉ là sự so sánh các dấu hiệu cho khung hiện thời và các dấu hiệu nền (thường được đánh giá từ các khung đầu vào trước), trong đó độ chênh lệch lớn hơn ngưỡng dẫn đến quyết định chính hoạt động. Theo các ví dụ khác, có thể thu được quyết định sơ bộ theo các cách khác, một số trong các cách này được đề cập ngắn gọn tiếp sau đây. Các chi tiết về hoạt động bên trong của bộ phát hiện tiếng nói chính không có tầm quan trọng thiết yếu cho sáng chế và bộ phát hiện tiếng nói chính bất kỳ mà tạo ra quyết định sơ bộ sẽ hữu dụng trong tình huống này. Khối bổ sung tồn lưu 102, theo ví dụ này, được sử dụng để mở rộng quyết định chính dựa trên các quyết định chính trong quá khứ để tạo thành quyết định cuối cùng, vad_flag 115. Nguyên nhân sử dụng tồn lưu chủ yếu là để giảm/loại bỏ rủi ro về việc cắt đầu sau và giữa câu nói của các nhóm câu nói. Tuy nhiên, tồn lưu cũng có thể được sử dụng để tránh sự cắt trong các đoạn nhạc.

Cũng có thể bổ sung tồn lưu bổ sung cho mục đích của DTX. Trên Fig.1 điều này được thể hiện bởi đầu ra tùy chọn vad_flag_dtx 117. Cần chú ý rằng không có gì bất thường khi chỉ có một đầu ra vad_flag tuy nhiên logic tồn lưu sử dụng các thiết lập khác khi đầu ra cần được sử dụng cho DTX. Trong bản mô tả này, hai đầu ra quyết định cuối cùng vad_flag 115 và vad_flag_dtx 117 sẽ được tách riêng trong hầu hết các phương án, để đơn giản hóa việc mô tả. Tuy nhiên, cũng có thể áp dụng các giải pháp dựa trên các thiết lập tồn lưu luân phiên và một đầu ra đơn.

Có hai nguyên nhân chính khi sử dụng các đầu ra quyết định khác nhau hoặc thiết lập tồn lưu tùy thuộc vào liệu quyết định VAD có được sử dụng cho DTX hay không. Trước tiên, từ quan điểm chất lượng câu nói, có các yêu cầu cao hơn về VAD khi được sử dụng cho DTX. Do đó, có mong muốn đảm bảo là câu nói đã kết thúc trước khi chuyển sang tiếng ồn tiện lợi. Động cơ thứ hai là tồn lưu bổ sung có thể được sử dụng để đánh giá các đặc tính của tiếng ồn nền. Ví dụ, trong AMR NB, việc đánh

giá tiếng ồn tiện lợi thứ nhất được thực hiện trong bộ giải mã dựa trên tồn lưu DTX riêng được sử dụng.

Như được đề cập ở trên, có một số dấu hiệu khác nhau có thể được sử dụng để phát hiện VAD. Một dấu hiệu có khả năng là chỉ cần xem xét năng lượng khung và so sánh nó với ngưỡng để quyết định liệu khung có chứa câu nói hay không. Phương pháp này hoạt động khá tốt với các điều kiện trong đó Tỷ lệ tín hiệu trên tiếng ồn (Signal-to-Noise Ratio - SNR) tốt nhưng lại không dành cho các trường hợp SNR thấp. Trong SNR thấp, tốt hơn nếu sử dụng các metric khác, ví dụ, so sánh các đặc tính của câu nói và các tín hiệu tiếng ồn. Đối với những sự thực thi trong thời gian thực, yêu cầu bổ sung về chức năng của VAD là độ phức tạp tính toán, nó được phản ánh trong biểu diễn thường xuyên của các VAD SNR bằng con trong các bộ mã hóa-giải mã tiêu chuẩn. VAD bằng con thường kết hợp các SNR của các băng con khác nhau thành metric chung, metric này được so sánh với ngưỡng để cho quyết định chính.

VAD 100 bao gồm bộ tách dấu hiệu 106 đưa ra năng lượng băng con của dấu hiệu, và bộ đánh giá nền 105, bộ này đưa ra các đánh giá năng lượng băng con. Đối với mỗi khung, VAD 100 tính toán các dấu hiệu. Để nhận dạng các khung hoạt động, (các) dấu hiệu cho khung hiện thời được so sánh với đánh giá về việc dấu hiệu "trông" ra sao đối với tín hiệu nền.

Khối bổ sung tồn lưu 102 được sử dụng để mở rộng quyết định VAD từ VAD chính dựa trên các quyết định chính trong quá khứ để tạo thành quyết định VAD cuối cùng, "vad_flag", nghĩa là các quyết định VAD trước cũng được tính đến. Như được đề cập ở trên, nguyên nhân sử dụng tồn lưu chủ yếu là để giảm/loại bỏ rủi ro về việc cắt đầu sau và giữa câu nói của các nhóm câu nói. Tuy nhiên, tồn lưu còn có thể được sử dụng để tránh cắt trong các đoạn nhạc. Bộ điều khiển hoạt động 107 có thể điều chỉnh (các) ngưỡng cho bộ phát hiện chính và chiều dài của việc bổ sung tồn lưu theo các đặc tính của tín hiệu đầu vào.

Còn có các giải pháp đã biết trong đó nhiều dấu hiệu với các đặc tính khác nhau được sử dụng cho quyết định chính. Đối với các VAD dựa trên yếu tố cơ bản của SNR băng con, thấy được rằng việc đưa sự phi tuyến tính vào trong tính toán SNR băng con,

đôi khi được gọi là các ngưỡng quan trọng, có thể cải thiện hoạt động VAD đối với các điều kiện có tiếng ồn không ổn định, ví dụ, tiếng ồn văn phòng hoặc nhiễu âm. Tuy nhiên, trong các trường hợp này thường có một quyết định chính được sử dụng để bổ sung tồn lưu, quyết định này có thể thích ứng với các điều kiện tín hiệu đầu vào, để tạo thành quyết định cuối cùng. Ngoài ra, nhiều VAD có ngưỡng năng lượng đầu vào để phát hiện sự im lặng, nghĩa là, đối với các mức đầu vào đủ thấp thì quyết định chính bị ép thành trạng thái không hoạt động.

Một ví dụ trong đó các ngưỡng quan trọng được sử dụng để tạo ra giải pháp VAD kép được mô tả trong Công bố đơn patent quốc tế số WO2008/143569 A1. Trong trường hợp này, các VAD kép được sử dụng để cải thiện việc cập nhật tiếng ồn nền và phát hiện nhạc. Tuy nhiên, chỉ có VAD chính tích cực được sử dụng cho quyết định vad_flag cuối cùng.

Trong WO2008/143569 A1, metric dựa trên hoạt động ngắn hạn lọc thông thấp được sử dụng để phát hiện sự có mặt của nhạc. Metric lọc thông thấp này đưa ra số lượng thay đổi chậm, thích hợp để tìm kiếm các loại âm thanh ít nhiều liên tục, thường là cho, ví dụ, âm nhạc. Sau đó, quyết định vad_music bổ sung có thể được đưa ra cho phần bổ sung tồn lưu, làm cho nó có thể xử lý âm thanh có nhạc theo cách thức cụ thể.

Có một số cách để tạo ra nhiều quyết định VAD chính. Cách cơ bản nhất là sử dụng cùng các dấu hiệu như VAD ban đầu nhưng thu được quyết định chính thứ hai nhờ sử dụng ngưỡng thứ hai. Một tùy chọn khác là chuyển VAD theo các điều kiện SNR được đánh giá, ví dụ, bằng cách sử dụng năng lượng cho các điều kiện SNR cao và chuyển sang hoạt động SNR bằng con cho các điều kiện SNR trung bình và thấp.

Trong Công bố đơn patent quốc tế số WO2011/049516 A1, bộ phát hiện hoạt động tiếng nói và phương pháp dùng cho bộ này được đề xuất. Bộ phát hiện hoạt động tiếng nói được tạo kết cấu để phát hiện hoạt động tiếng nói trong tín hiệu đầu vào nhận được. VAD bao gồm các logic kết hợp được tạo kết cấu để nhận tín hiệu từ bộ phát hiện tiếng nói chính của VAD chỉ thị quyết định VAD chính. Các logic kết hợp còn nhận ít nhất một tín hiệu từ VAD bên ngoài chỉ thị quyết định hoạt động tiếng nói từ VAD bên ngoài. Bộ xử lý kết hợp các quyết định hoạt động tiếng nói được chỉ thị

trong các tín hiệu nhận được để tạo ra quyết định VAD chính biến đổi. Quyết định VAD biến đổi được gửi tới bộ phận bổ sung tồn lưu.

Một vấn đề với tồn lưu là quyết định sử dụng khi nào và sử dụng bao nhiêu. Từ quan điểm chất lượng câu nói, việc bổ sung tồn lưu về cơ bản là tích cực. Tuy nhiên, không mong muốn bổ sung quá nhiều tồn lưu do tồn lưu bổ sung bất kỳ sẽ làm giảm hiệu quả của giải pháp DTX. Do không mong muốn bổ sung tồn lưu vào mọi nhóm hoạt động ngắn, nên thường yêu cầu có số lượng khung hoạt động tối thiểu từ bộ phát hiện chính vad_prim trước khi tính đến việc bổ sung một số tồn lưu để tạo ra quyết định cuối cùng vad_flag. Tuy nhiên, để tránh cắt câu nói, mong muốn giữ số lượng khung hoạt động theo yêu cầu này là thấp nhất có thể.

Đối với tiếng ồn không ổn định số lượng thấp của các khung hoạt động được yêu cầu có thể cho phép chính tiếng ồn tạo ra các sự kiện VAD đủ dài, các sự kiện này sẽ kích hoạt việc bổ sung tồn lưu. Do đó, để tránh hoạt động quá mức, giải pháp này thường không cho phép các tồn lưu dài.

Một vấn đề khác đối với số lượng khung hoạt động được yêu cầu trước khi bổ sung tồn lưu cho VAD hiệu suất cao là khả năng của nó để phát hiện các đoạn ngắt ngắn trong lời phát biểu. Trong trường hợp này, có lời phát biểu đã được phát hiện đúng, nhưng người nói tạo ra đoạn ngắt nhẹ trước khi tiếp tục. Điều này làm cho VAD phát hiện đoạn ngắt và một lần nữa yêu cầu khoảng thời gian mới của các khung chính hoạt động trước khi tồn lưu bất kỳ nào được bổ sung. Điều này có thể tạo ra các thành phần lạ gây khó chịu do cắt đầu sau của các đoạn câu nói kế tiếp như kết thúc lời phát biểu bằng các tiếng nổ bị làm mất tiếng nói.

Bản chất kỹ thuật của sáng chế

Một mục tiêu của các phương án theo sáng chế là nhằm giải quyết ít nhất một trong các vấn đề nêu trên, và mục tiêu này đạt được nhờ các phương pháp và thiết bị theo các điểm yêu cầu bảo hộ độc lập kèm theo, và nhờ các phương án theo các điểm yêu cầu bảo hộ phụ thuộc.

Theo một khía cạnh, sáng chế đề xuất phương pháp để phát hiện hoạt động tiếng nói (voice activity detection - VAD) bao gồm bước tạo ra tín hiệu chỉ thị quyết định VAD chính, và xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện. Việc xác định bổ sung tồn lưu được tiến hành phụ thuộc vào số đo hoạt động ngắn hạn và số đo hoạt động dài hạn. Sau đó, tín hiệu chỉ thị quyết định VAD cuối cùng được tạo ra phụ thuộc ít nhất vào việc xác định bổ sung tồn lưu.

Theo một phương án, số đo hoạt động ngắn hạn được suy ra từ N_{st} quyết định VAD chính gần nhất.

Theo một phương án, số đo hoạt động dài hạn được suy ra từ N_{lt} quyết định VAD cuối cùng gần nhất hoặc từ N_{lt} quyết định VAD chính gần nhất.

Theo một phương án, hai phiên bản của các quyết định cuối cùng, quyết định VAD cuối cùng thứ nhất và quyết định VAD cuối cùng thứ hai được tạo ra. Quyết định VAD cuối cùng thứ hai có thể được đưa ra mà không sử dụng số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn, và số đo hoạt động dài hạn có thể được suy ra từ N_{lt} quyết định VAD cuối cùng thứ hai gần nhất.

Theo một phương án, quyết định VAD cuối cùng bằng với quyết định VAD chính nếu xác định được việc bổ sung tồn lưu không được thực hiện. Trong trường hợp xác định được việc bổ sung tồn lưu được thực hiện, quyết định VAD cuối cùng bằng với quyết định hoạt động tiếng nói, chỉ thị khung hoạt động.

Theo một khía cạnh khác, sáng chế đề xuất thiết bị để phát hiện hoạt động tiếng nói. Thiết bị này bao gồm phần đầu vào, phần bố trí bộ phát hiện tiếng nói chính và bộ phận bổ sung tồn lưu. Phần đầu vào được tạo kết cấu để nhận tín hiệu đầu vào. Phần bố trí bộ phát hiện tiếng nói chính được nối với phần đầu vào. Phần bố trí bộ phát hiện tiếng nói chính được tạo kết cấu để phát hiện hoạt động tiếng nói trong tín hiệu đầu vào nhận được và để tạo ra tín hiệu chỉ thị quyết định VAD chính được kết hợp với tín hiệu đầu vào nhận được. Bộ phận bổ sung tồn lưu được nối với phần bố trí bộ phát hiện tiếng nói chính. Bộ phận bổ sung tồn lưu được tạo kết cấu để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện, và để tạo ra tín hiệu chỉ thị

quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu. Thiết bị này còn bao gồm bộ đánh giá hoạt động ngắn hạn và bộ đánh giá hoạt động dài hạn. Bộ đánh giá hoạt động ngắn hạn được nối với đầu vào của bộ phận bổ sung tồn lưu. Bộ đánh giá hoạt động dài hạn được nối với đầu ra của bộ phận bổ sung tồn lưu. Bộ phận bổ sung tồn lưu được nối với đầu ra của bộ đánh giá hoạt động ngắn hạn và bộ đánh giá hoạt động dài hạn. Bộ phận bổ sung tồn lưu còn được tạo kết cấu để thực hiện việc xác định tồn lưu phụ thuộc vào số đo hoạt động ngắn hạn và số đo hoạt động dài hạn.

Theo một phương án, bộ đánh giá hoạt động ngắn hạn được tạo kết cấu để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất.

Theo một phương án, bộ đánh giá hoạt động dài hạn được tạo kết cấu để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng gần nhất hoặc từ N_{lt} quyết định VAD chính gần nhất.

Theo một phương án, sáng chế đề xuất thiết bị. Phương án này dựa trên bộ xử lý, ví dụ bộ vi xử lý, bộ này chạy thành phần phần mềm để tạo ra tín hiệu chỉ thị quyết định VAD chính, thành phần phần mềm để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện, và thành phần phần mềm để tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu. Theo phương án này bộ xử lý chạy thành phần phần mềm để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất và/hoặc thành phần phần mềm để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng gần nhất. Các thành phần phần mềm này được lưu trữ trong bộ nhớ.

Theo một khía cạnh khác, sáng chế đề xuất chương trình máy tính. Chương trình máy tính này bao gồm các đơn vị mã đọc được bởi máy tính, các đơn vị mã này khi chạy trên thiết bị làm cho thiết bị tạo ra tín hiệu chỉ thị quyết định VAD chính, xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện dựa trên số đo hoạt động ngắn hạn và số đo hoạt động dài hạn, và tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu.

Theo một khía cạnh khác, sáng chế đề xuất sản phẩm chương trình máy tính. Sản phẩm chương trình máy tính này bao gồm phương tiện đọc được bởi máy tính và chương trình máy tính để tạo ra tín hiệu chỉ thị quyết định VAD chính, xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện dựa trên số đo hoạt động ngắn hạn và số đo hoạt động dài hạn, và tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu, được lưu trữ trên phương tiện đọc được bởi máy tính.

Mô tả vắn tắt các hình vẽ

Để hiểu đầy đủ hơn về các phương án làm ví dụ của sáng chế, cần tham khảo phần mô tả sau có dựa vào các hình vẽ kèm theo, trong đó:

Fig.1 là hình vẽ thể hiện một ví dụ về VAD chung có đánh giá nền.

Fig.2 là hình vẽ thể hiện VAD theo một phương án làm ví dụ của sáng chế.

Fig.3 là sơ đồ tiến trình thể hiện phương pháp VAD làm ví dụ theo một phương án của sáng chế.

Fig.4A là hình vẽ thể hiện VAD theo một phương án làm ví dụ của sáng chế.

Fig.4B là hình vẽ thể hiện VAD theo một phương án làm ví dụ khác của sáng chế.

Fig.4C là hình vẽ thể hiện VAD theo một phương án làm ví dụ khác nữa của sáng chế.

Fig.5 là hình vẽ thể hiện VAD theo một phương án làm ví dụ khác của sáng chế.

Fig.6 là hình vẽ thể hiện VAD theo một phương án có tồn lưu.

Fig.7 là hình vẽ thể hiện VAD bổ sung theo một phương án.

Mô tả chi tiết sáng chế

Một cách thức để giảm các vấn đề này đã được tìm ra là sử dụng các đặc tính thời gian của các metric bộ phát hiện chính và các metric quyết định cuối cùng. Chúng được thấy là rất thích hợp để điều chỉnh tồn lưu bổ sung. Tốt hơn nếu ít nhất một quyết định trong số quyết định chính được nhập vào phần bổ sung tồn lưu và quyết định cuối cùng được đưa ra từ phần bổ sung tồn lưu được sử dụng để tác động vào việc bổ sung tồn lưu, và tốt nhất nếu sử dụng cả hai quyết định. Quyết định chính được nhập vào phần bổ sung tồn lưu có thể là quyết định chính ban đầu thu được từ bộ phát hiện tiếng nói chính, hoặc nó có thể là phiên bản biến đổi của quyết định chính ban đầu này. Sự biến đổi này có thể được thực hiện dựa trên các đầu ra từ các VAD khác.

Một phương án của loại VAD 200 chung có sử dụng quyết định chính được nhập vào phần bổ sung tồn lưu 202 và quyết định cuối cùng được đưa ra từ phần bổ sung tồn lưu 202 được thể hiện trên Fig.2.

Bộ tách dấu hiệu 206 đưa ra năng lượng băng con của dấu hiệu, bộ đánh giá nền 205 đưa ra các đánh giá năng lượng băng con, bộ điều khiển hoạt động 207 có thể điều chỉnh (các) ngưỡng cho bộ phát hiện chính và chiều dài của việc bổ sung tồn lưu theo các đặc tính của tín hiệu đầu vào, và bộ phát hiện tiếng nói chính 201 đưa ra quyết định sơ bộ vad_prim 213 như được mô tả liên quan đến Fig.1.

Theo phương án này, bộ phát hiện hoạt động tiếng nói 200 còn bao gồm bộ đánh giá hoạt động ngắn hạn 203 và/hoặc bộ đánh giá hoạt động dài hạn 204. Các đặc tính thời gian được thu lại bằng cách sử dụng hoạt động ngắn hạn theo các dấu hiệu của quyết định chính, vad_prim 213, và hoạt động dài hạn của quyết định cuối cùng, vad_flag 215. Sau đó, các metric này được sử dụng để điều chỉnh việc bổ sung tồn lưu để cải thiện hoạt động của VAD để dùng trong DTX nhờ tạo ra quyết định cuối cùng luân phiên, vad_flag_dtx 217.

Ở đây, trong trường hợp này, hoạt động ngắn hạn được đo nhờ đếm số lượng khung hoạt động trong bộ nhớ của N_st quyết định chính cuối cùng vad_prim 213. Tương tự, hoạt động dài hạn được đo nhờ đếm số lượng khung hoạt động trong quyết định cuối cùng vad_flag 215 trong N_lt khung cuối cùng. N_lt lớn hơn N_st, tốt hơn nếu lớn hơn đáng kể. Sau đó, các metric này được sử dụng để tạo ra quyết định cuối

cùng luân phiên vad_flag_dtx 217. Ưu điểm của việc sử dụng các metric này là nó đơn giản hóa việc điều chỉnh tồn lưu do dễ dàng hơn khi bổ sung tồn lưu ngay tại thời điểm khi hoạt động đã cao.

Hoạt động ngắn hạn cao biểu thị phần đầu, phần giữa hoặc phần cuối của nhóm hoạt động. Khi xem xét qua, metric này có thể xuất hiện tương tự như cách thức thường được sử dụng là chỉ yêu cầu số lượng khung hoạt động liên tiếp như được đề cập trước. Tuy nhiên, khác biệt chính là hoạt động ngắn hạn không được thiết lập lại khi quyết định không hoạt động xuất hiện. Thay vào đó, nó có bộ nhớ, bộ nhớ này nhớ khung hoạt động cho tới N_st khung trước khi nó sau cùng rời khỏi bộ nhớ. Theo đó, khung không hoạt động chỉ giảm hoạt động ngắn hạn trung bình đến một mức độ nhất định. Đối với hoạt động ngắn hạn đủ cao, sẽ an toàn khi bổ sung một vài khung của tồn lưu, do hoạt động ngắn hạn đã cao nên tồn lưu bổ sung sẽ chỉ có tác động nhỏ lên toàn bộ hoạt động. Các khung không hoạt động phân tán sẽ không giảm hoạt động ngắn hạn đủ để ngắt hoạt động tồn lưu này.

Các khung không hoạt động phân tán có thể tương ứng với các đoạn ngắt ngắt trong phần giữa của lời phát biểu hoặc có thể là phát hiện không hoạt động nhằm, ví dụ, gây ra do các chuỗi ngắn của câu nói bị làm mất tiếng. Nhờ sử dụng hoạt động ngắn hạn theo cách thức được chỉ ra như trên, việc bổ sung tồn lưu có thể được duy trì trong các tình huống này.

Tương tự, hoạt động dài hạn cao biểu thị rằng nhóm câu nói đã hoạt động trong một thời gian. Nếu hoạt động dài hạn cao thì xác suất lớn là nó có khả năng bổ sung một số khung tồn lưu bổ sung và vẫn chỉ có tác động nhỏ lên toàn bộ hoạt động.

Theo một phương án, hoạt động ngắn hạn và hoạt động dài hạn lần lượt được so sánh với ngưỡng định trước tương ứng. Nếu đạt đến ngưỡng tương ứng, thì số lượng khung tồn lưu định trước tương ứng được bổ sung.

Do hoạt động dài hạn phản ứng tương đối chậm phụ thuộc vào phần cuối thực tế của hoạt động câu nói, nên có rủi ro là số lượng lớn khung tồn lưu được bổ sung được sử dụng trong thời gian tương đối lâu sau phần cuối của nhóm câu nói. Về vấn đề này,

có thể sử dụng hoạt động ngắn hạn thấp làm chỉ thị cho phần cuối của nhóm câu nói. Do đó, theo một phương án có thể mong muốn giới hạn lượng tồn lưu bổ sung nếu hoạt động ngắn hạn xuống dưới ngưỡng định trước. Nói cách khác, hoạt động ngắn hạn đủ thấp có thể hủy việc bổ sung các khung tồn lưu như được chỉ thị bởi hoạt động dài hạn cao đồng thời.

Sau đây, các phương án trên, trong hầu hết các trường hợp, được mô tả là các phương án biến đổi của các giải pháp đã có trong đó mức tăng lên về độ phức tạp là nhỏ. Tuy nhiên, cũng có thể thiết kế VAD hoàn toàn mới, VAD này sử dụng các metric trên để đưa ra quyết định VAD tin cậy hơn.

Theo một phương án, được thể hiện giản lược trên Fig.3, phương pháp trong bộ phát hiện hoạt động tiếng nói để phát hiện hoạt động tiếng nói trong tín hiệu đầu vào nhận được bao gồm bước tạo ra 310 tín hiệu chỉ thị quyết định VAD chính được kết hợp với tín hiệu đầu vào nhận được, tốt hơn nếu nhờ phân tích các đặc tính của tín hiệu đầu vào nhận được. Xác định 320 liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện hay không. Tín hiệu chỉ thị quyết định VAD cuối cùng được tạo ra 330. Quyết định VAD cuối cùng bằng với quyết định VAD chính nếu xác định được việc bổ sung tồn lưu không được thực hiện. Quyết định VAD cuối cùng bằng với quyết định hoạt động tiếng nói nếu xác định được việc bổ sung tồn lưu được thực hiện. Do tồn lưu được bổ sung, nên quyết định hoạt động tiếng nói được thiết lập để chỉ thị khung hoạt động, nghĩa là khung chứa câu nói chứ không phải tiếng ồn. Số đo hoạt động ngắn hạn được suy ra 340 từ N_{st} quyết định VAD chính gần nhất và/hoặc số đo hoạt động dài hạn được suy ra 342 từ N_{lt} quyết định VAD cuối cùng gần nhất. Việc xác định liệu việc bổ sung tồn lưu có được thực hiện hay không được tiến hành phụ thuộc vào số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn. Mặc dù Fig.3 được thể hiện như một tiến trình đơn gồm các sự kiện, nhưng hệ thống thực tế sẽ xử lý khung này sau khung khác. Các mũi tên bằng nét đứt biểu thị rằng sự phụ thuộc của số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn là hợp lệ cho khung tiếp theo.

Cần hiểu rằng Fig.3 không thể hiện dòng tín hiệu mà thể hiện các bước của phương pháp cần được thực hiện theo một phương án của sáng chế. Nghĩa là, bước tạo

ra quyết định VAD cuối cùng 330 có thể bao gồm tạo ra quyết định cuối cùng luân phiên (ví dụ vad_flag_dtx 217) dựa trên số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn. Tuy nhiên, quyết định cuối cùng luân phiên không được sử dụng làm đầu vào cho bộ đánh giá hoạt động dài hạn 204 do nó sẽ đưa vào vòng lặp hồi tiếp của hoạt động (do sự biến đổi của dấu hiệu cần được đo với bổ sung tồn lưu điều chỉnh). Do đó, bước tạo ra quyết định VAD cuối cùng 330 còn có thể bao gồm tạo ra quyết định cuối cùng (ví dụ vad_flag 215) dựa trên kỹ thuật tồn lưu thông thường và/hoặc các số đo hoạt động ngắn hạn chứ không phải các số đo hoạt động dài hạn, sau đó nó được sử dụng làm đầu vào cho bộ đánh giá hoạt động dài hạn 204, như được thể hiện trên Fig.2.

Theo một phương án, được thể hiện giản lược trên Fig.4A, bộ phát hiện hoạt động tiếng nói 400 bao gồm phần đầu vào 412, phần bố trí bộ phát hiện tiếng nói chính 401 và bộ phận bổ sung tồn lưu 402. Phần đầu vào được tạo kết cấu để nhận tín hiệu đầu vào. Phần bố trí bộ phát hiện tiếng nói chính 401 được nối với phần đầu vào 412. Phần bố trí bộ phát hiện tiếng nói chính 401 được tạo kết cấu để phát hiện hoạt động tiếng nói trong tín hiệu đầu vào nhận được và để tạo ra tín hiệu chỉ thị quyết định VAD chính được kết hợp với tín hiệu đầu vào nhận được. Bộ phận bổ sung tồn lưu 402 được nối với phần bố trí bộ phát hiện tiếng nói chính 401. Bộ phận bổ sung tồn lưu 402 được tạo kết cấu để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện hay không và để tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng. Quyết định VAD cuối cùng bằng với quyết định VAD chính nếu xác định được việc bổ sung tồn lưu không được thực hiện. Quyết định VAD cuối cùng bằng với quyết định hoạt động tiếng nói nếu xác định được việc bổ sung tồn lưu được thực hiện. Bộ phát hiện hoạt động tiếng nói 400 còn bao gồm bộ đánh giá hoạt động ngắn hạn 403 và/hoặc bộ đánh giá hoạt động dài hạn 404. Bộ đánh giá hoạt động ngắn hạn 403 được nối với đầu vào của bộ phận bổ sung tồn lưu 402. Bộ đánh giá hoạt động ngắn hạn 403 được tạo kết cấu để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất. Bộ đánh giá hoạt động dài hạn 404 được nối với đầu ra của bộ phận bổ sung tồn lưu 402. Bộ đánh giá hoạt động dài hạn 404 được tạo kết cấu để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng gần nhất. Bộ phận bổ sung tồn lưu 402 được

nổi với đầu ra của bộ đánh giá hoạt động ngắn hạn 403 và/hoặc bộ đánh giá hoạt động dài hạn 404. Bộ phận bổ sung tồn lưu 402 còn được tạo kết cấu để thực hiện việc xác định tồn lưu phụ thuộc vào số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn. Sau đó, việc xác định tồn lưu phụ thuộc vào số đo hoạt động ngắn hạn và/hoặc số đo hoạt động dài hạn có thể được sử dụng để điều chỉnh việc bổ sung tồn lưu nhằm cải thiện hoạt động của VAD để sử dụng trong DTX nhờ tạo ra quyết định cuối cùng luân phiên.

Bộ phát hiện hoạt động tiếng nói thường được bố trí trong bộ mã hóa-giải mã âm thanh hoặc tiếng nói. Các bộ mã hóa-giải mã này thường được bố trí trong các thiết bị cuối khác nhau, ví dụ trong các mạng viễn thông. Các ví dụ không giới hạn là điện thoại, máy tính, v.v. trong đó thực hiện việc phát hiện hoặc ghi lại âm thanh.

Theo một phương án, quyết định VAD cuối cùng được đưa ra như cờ bổ sung 410, ngoài ra quyết định VAD cuối cùng được đưa ra mà không sử dụng các số đo hoạt động ngắn hạn hoặc các số đo hoạt động dài hạn, thông thường như quyết định VAD cuối cùng để dùng cho DTX, như được thể hiện trên Fig.4B. Sau đó, hai phiên bản của các quyết định cuối cùng có thể được sử dụng song song bởi các bộ phận hoặc chức năng khác nhau. Theo một phương án khác, việc sử dụng các số đo hoạt động ngắn hạn hoặc các số đo hoạt động dài hạn có thể được bật và tắt phụ thuộc vào ngữ cảnh trong đó quyết định VAD sẽ được sử dụng.

Theo một phương án khác, trong đó quyết định VAD cuối cùng không có sẵn hoặc không thích hợp để tiến hành phân tích hoạt động dài hạn bất kỳ, thay vào đó phân tích hoạt động dài hạn có thể được thực hiện trên quyết định VAD chính. Theo phương án này, bộ đánh giá hoạt động dài hạn 404, thay vào đó, được nối với đầu vào của bộ phận bổ sung tồn lưu 402, như được thể hiện trên Fig.4C, và số đo hoạt động dài hạn được suy ra từ N_{lt} quyết định VAD chính gần nhất.

Theo một phương án khác nữa, các đánh giá của hoạt động ngắn hạn và dài hạn có thể được thực hiện trên quyết định VAD cuối cùng và/hoặc chính khác với quyết định VAD cuối cùng và/hoặc chính mà trên đó việc điều chỉnh bổ sung tồn lưu là để được thực hiện. Một khả năng là có VAD đơn giản tạo ra quyết định VAD chính và bộ

phận tồn lưu đơn giản biến đổi nó thành quyết định VAD cuối cùng. Sau đó, động thái hoạt động ngắn hạn và dài hạn của các quyết định VAD cuối cùng và/hoặc chính này có thể được phân tích. Tuy nhiên, sau đó có thể sử dụng việc thiết lập VAD khác, ví dụ việc thiết lập tinh vi hơn, để đưa ra quyết định VAD chính được quan tâm để điều chỉnh việc bổ sung tồn lưu. Sau đó, các hoạt động được phân tích từ hệ thống đơn có thể được sử dụng để điều khiển hoạt động của bộ phận bổ sung tồn lưu 402 của hệ thống VAD phức tạp hơn, đưa ra quyết định VAD cuối cùng tin cậy.

Trong phần sau đây, ví dụ về một phương án của bộ phát hiện hoạt động tiếng nói 500 sẽ được mô tả có dựa vào Fig.5. Phương án này dựa trên bộ xử lý 510, ví dụ bộ vi xử lý, bộ này chạy thành phần phần mềm 501 để tạo ra tín hiệu chỉ thị quyết định VAD chính, thành phần phần mềm 502 để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện, và thành phần phần mềm 503 để tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng. Theo phương án này bộ xử lý 510 chạy thành phần phần mềm 504 để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất và/hoặc thành phần phần mềm 505 để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng gần nhất. Các thành phần phần mềm này được lưu trữ trong bộ nhớ 520. Bộ xử lý 510 truyền thông với bộ nhớ 520 qua buýt hệ thống 515. Tín hiệu audio được nhận bởi bộ điều khiển đầu vào/đầu ra (input/output - I/O) 530 điều khiển buýt I/O 516, mà bộ xử lý 510 và bộ nhớ 520 được nối với nó. Theo phương án này, các tín hiệu nhận được bởi bộ điều khiển I/O 530 được lưu trữ trong bộ nhớ 520, trong đó chúng được xử lý bởi các thành phần phần mềm. Thành phần phần mềm 501 có thể thực thi chức năng của bước 310 theo phương án được mô tả có dựa vào Fig.3 ở trên. Thành phần phần mềm 502 có thể thực thi chức năng của bước 320 theo phương án được mô tả có dựa vào Fig.3 ở trên. Thành phần phần mềm 503 có thể thực thi chức năng của bước 330 theo phương án được mô tả có dựa vào Fig.3 ở trên. Thành phần phần mềm 504 có thể thực thi chức năng của bước 340 theo phương án được mô tả có dựa vào Fig.3 ở trên. Thành phần phần mềm 505 có thể thực thi chức năng của bước 342 theo phương án được mô tả có dựa vào Fig.3 ở trên.

Bộ phận I/O 530 có thể được nối kết với bộ xử lý 510 và/hoặc bộ nhớ 520 qua buýt I/O 516 để cho phép nhập vào và/hoặc đưa ra dữ liệu có liên quan như các tín hiệu đầu vào và các quyết định VAD cuối cùng.

Theo một phương án, các bộ đếm của các khung hoạt động trong bộ nhớ của các quyết định chính và các quyết định cuối cùng được sử dụng như mô tả ở trên. Theo các phương án khác, cũng có thể sử dụng trọng số mà phụ thuộc vào tuổi của khung hoạt động trong bộ nhớ. Điều này có thể được cho cả hoạt động chính ngắn hạn và hoạt động quyết định cuối cùng dài hạn. Theo các phương án khác, có thể có khả năng sử dụng các tồn lưu bổ sung khác nhau phụ thuộc vào các đặc tính tín hiệu đầu vào khác, như mức câu nói được đánh giá, mức tiếng ồn, và/hoặc SNR.

Theo các phương án khác, có thể quan tâm đến việc sử dụng nhiều hơn hai đặc tính thời gian để xác định vị trí tốt hơn đối với phần đầu, giữa, hoặc cuối của nhóm câu nói hoạt động.

Theo các phương án khác, các yếu tố cơ bản của các quyết định tồn lưu được mô tả ở trên còn có thể được kết hợp với các giải pháp cải tiến VAD khác như các yếu tố cơ bản của bộ kết hợp đa VAD được thể hiện trong WO2011/049516. Trong trường hợp này có thể sử dụng quyết định VAD chính biến đổi làm đầu vào bộ đánh giá hoạt động ngắn hạn và khối bổ sung tồn lưu. Sau đó, bộ kết hợp đa VAD có thể được coi là một phần của phân bố trí bộ phát hiện tiếng nói chính.

Tương tự, các phương án bổ sung khác để đánh giá nên có thể được tích hợp một cách thuận lợi và dễ dàng với sáng chế.

Bộ mã hóa-giải mã G.718 theo các chuẩn 3GPP2 được sử dụng làm cơ sở cho một phương án được thể hiện sau đây. Mô tả chi tiết về các phần liên quan có thể tìm thấy trong, ví dụ, Công bố đơn patent quốc tế số WO2009/000073 A1.

Fig.6 thể hiện sơ đồ khối của hệ thống truyền thông âm thanh của WO2009/000073 A1 bao gồm bộ tiền xử lý 601, bộ phân tích phổ 602, bộ phát hiện hoạt động âm thanh 603, bộ đánh giá tiếng ồn 604, bộ giảm tiếng ồn tùy chọn 605, bộ phân tích LP và bộ theo dõi bước 606, môđun cập nhật đánh giá về năng lượng của

tiếng ồn 607, bộ phân loại tiếng ồn 608 và bộ mã hóa âm thanh 609. Việc phát hiện hoạt động âm thanh (giai đoạn thứ nhất của phân loại tín hiệu) được thực hiện trong bộ phát hiện hoạt động âm thanh 603 sử dụng các đánh giá năng lượng của tiếng ồn được tính toán trong khung trước. Đầu ra của bộ phát hiện hoạt động âm thanh 603 là biến nhị phân, biến này còn được sử dụng bởi bộ mã hóa 609 và nó xác định liệu khung hiện thời được mã hóa thành hoạt động hay không hoạt động.

Môđun “SAD dựa trên SNR” 603 là môđun trong đó các phương án của sáng chế có thể được thực thi. Hiện nay, phương án này chỉ bao trùm chuỗi tín hiệu băng rộng, được lấy mẫu ở 16kHz, nhưng biến đổi tương tự cũng có lợi đối với chuỗi tín hiệu băng hẹp, được lấy mẫu ở 8 kHz, hoặc các tốc độ lấy mẫu bất kỳ khác.

Theo một phương án, dựa trên các yếu tố cơ bản được thể hiện trong WO2011/049516 A1, VAD ban đầu từ WO2009/000073 A1 (VAD 1) được sử dụng như VAD thứ nhất, tạo ra các tín hiệu localVAD và vad_flag. Theo sáng chế, localVAD này được sử dụng làm VAD_prim 213 trên đó tiến hành đánh giá hoạt động ngắn hạn.

VAD bổ sung (VAD 2) cũng dựa trên WO2009/000073 A1 nhưng thu được nhờ sử dụng các biến đổi cho đánh giá tiếng ồn nền và SAD dựa trên SNR. Fig.7 thể hiện sơ đồ khối cho VAD thứ hai. Sơ đồ khối thể hiện bộ tiền xử lý 701, bộ phân tích phổ 702, môđun “SAD dựa trên SNR” 703, bộ đánh giá tiếng ồn 704, bộ giảm tiếng ồn tùy chọn 705, bộ phân tích LP và bộ theo dõi bước 706, môđun cập nhật đánh giá về năng lượng của tiếng ồn 707, bộ phân loại tiếng ồn 708 và bộ mã hóa âm thanh 709.

Sơ đồ khối còn thể hiện quyết định VAD chính và quyết định VAD cuối cùng cho VAD 2, lần lượt là, localVAD_he 710 và vad_flag_he 711. localVAD_he 710 và vad_flag_he 711 được sử dụng trong bộ phát hiện tiếng nói chính của VAD1 để tạo ra localVAD.

Đối với phương án này các biến sau được thêm vào trạng thái mã hóa (Encoder_State):

```

long long vad_flag_reg;    /* bộ nhớ của vad_flag cũ */
long long vad_prim_reg;    /* bộ nhớ của localVAD cũ */
short vad_flag_cnt_50;     /* bộ đếm của vad_flag khung hoạt động */
short vad_prim_cnt_16;     /* bộ đếm của các khung hoạt động chính */

short hangover_cnt_dtx;    /* bộ đếm của các khung tồn lưu cho DTX */

```

Tất cả các trạng thái này sẽ được thiết lập thành zerô trong khi khởi tạo, ví dụ điều này có thể được thực hiện trong thường trình `wb_vad_init()`.

Ngoài ra, hoạt động ngắn hạn và dài hạn của các dấu hiệu được cập nhật, hoạt động này được thực hiện ở cuối quá trình xử lý cho mỗi khung. Điều này có thể được thực hiện nhờ thêm mã sau vào trong tệp nguồn thích hợp:

```

if ((st->vad_flag_reg & (long long) 0x01LL << 49) != 0)
{
    st->vad_flag_cnt_50 = st->vad_flag_cnt_50 - 1;
}
st->vad_flag_reg = (st->vad_flag_reg & (long long) 0x3fffffffffffffffLL) << 1;
if (vad_flag)
{
    st->vad_flag_reg = st->vad_flag_reg | 0x01L;
    st->vad_flag_cnt_50 = st->vad_flag_cnt_50 + 1;
}

if ((st->vad_prim_reg & (long long) 1LL << 15) != 0)
{
    st->vad_prim_cnt_16 = st->vad_prim_cnt_16 - 1;
}
st->vad_prim_reg = (st->vad_prim_reg & (long long) 0x3fffffffffffffffLL) << 1;
if (localVAD)
{
    st->vad_prim_reg = st->vad_prim_reg | 0x01L;
    st->vad_prim_cnt_16 = st->vad_prim_cnt_16 + 1;
}

```

}

Ở đây, biến `st` chỉ biến `Encoder_State` được cấp phát trong bộ mã hóa. Theo đó, đối với khung tiếp theo các biến trạng thái `st->vad_flag_cnt_50` sẽ chứa hoạt động quyết định cuối cùng dài hạn dưới dạng số lượng khung mà hoạt động trong 50 khung gần nhất và biến trạng thái `st->vad_prim_cnt_16` sẽ chứa hoạt động chính ngắn hạn dưới dạng số lượng khung hoạt động chính trong 16 khung gần nhất. Chiều dài bộ nhớ của hoạt động ngắn hạn, 16 khung, và chiều dài bộ nhớ của hoạt động dài hạn, 50 khung, là các trị số được sử dụng theo phương án cụ thể này. Các con số này là các trị số điển hình có thể được sử dụng trong quá trình thực thi hoạt động được, nhưng các trị số tuyệt đối không có tính quyết định. Theo đó, các số này có thể được làm thích ứng theo các loại thực thi khác nhau, ví dụ, như sự điều chỉnh các đặc tính tồn lưu. Nói chung, chiều dài bộ nhớ của hoạt động dài hạn dài hơn chiều dài bộ nhớ của hoạt động ngắn hạn, và tốt hơn nếu dài hơn đáng kể, như trong ví dụ được thể hiện ở trên. Theo một phương án điển hình, tỷ lệ giữa chiều dài bộ nhớ của hoạt động dài hạn và chiều dài bộ nhớ của hoạt động ngắn hạn nằm trong khoảng từ 2,5 đến 5. Ngoài ra, tỷ lệ này có thể được làm thích ứng cho các loại thực thi khác nhau trong đó các loại âm thanh khác nhau được mong đợi có mặt thường xuyên.

Mã để quyết định bao nhiêu tồn lưu, `hangover_short`, sẽ được bổ sung có thể được thực thi bằng cách sử dụng sự biến đổi mã sau, trong đó:

`lp_snr` là đánh giá SNR được lọc thông thấp

`th_clean` Ngưỡng SNR sử dụng để quyết định nếu đầu vào là câu nói trong

`thr1` ngưỡng được tính toán cho bộ phát hiện chính

```
if( lp_snr < th_clean )
{
    thr1 = nk * lp_snr + nc; /* Hàm tuyến tính cho câu nói có tiếng ồn */
}
```

```

if( st->Opt_SC_VBR )
{
    hangover_short = 1;
}
else
{
    hangover_short = 4;
}
}
else
{
    thr1 = sk * lp_snr + sc; /* Hàm tuyến tính cho câu nói trong */
    hangover_short = 1;
}

```

Tiếp đó thêm mã sau đây mà cần thiết để làm thích ứng tồn lưu được sử dụng cho DTX hangover_short_dtx.

```

if( lp_snr < th_clean )
{
    thr1 = nk * lp_snr + nc; /* Hàm tuyến tính cho câu nói có tiếng ồn */

    if( st->Opt_SC_VBR )
    {
        hangover_short = 1;
    }
    else
    {
        hangover_short = 4;
    }
}
else
{
    thr1 = sk * lp_snr + sc; /* Hàm tuyến tính cho câu nói trong */

```

```

    hangover_short = 1;
}

hangover_short_dtx = hangover_short;    /* bắt đầu với cùng tồn lưu cho DTX */
if (st->Opt_DTX_ON)
{
    if (st->vad_prim_cnt_16 > 12 ) /* 12 yêu cầu gần đúng > 80% hoạt động chính */
    {
        hangover_short_dtx = hangover_short_dtx + 1;
    }

    if (st->vad_flag_cnt_50 > 40 ) /* 40 yêu cầu gần đúng > 80% hoạt động cò */
    {
        hangover_short_dtx = hangover_short_dtx + 3;
    }

    /* Giữ hangover_short thấp hơn số đếm tồn lưu tối đa */
    if (hangover_short_dtx > HANGOVER_LONG-1)
    {
        hangover_short_dtx=HANGOVER_LONG-1;
    }

    /* Chỉ cho phép HO ngắn nếu không đủ khung hoạt động */
    if ( st->vad_prim_cnt_16 < 7 && hangover_short_dtx > 4 )
    {
        hangover_short_dtx=4;
    }
}

```

Ngoài ra, trong bản mô tả này, có một số con số quy định, các con số này được coi như các biến thiết kế. Theo đó, các số này cũng có thể được làm thích ứng theo các loại thực thi khác nhau, ví dụ như sự điều chỉnh các đặc tính tồn lưu.

Mã để thực thi tồn lưu thực tế có thể được thực hiện với sự biến đổi sau đây:

flag	Quyết định VAD cuối cùng bao gồm tồn lưu
localVAD	quyết định chính
snr_sum	Dấu hiệu VAD dưới dạng đánh giá SNR bằng con
st->nb_active_frames	Số lượng khung hoạt động liên tiếp (các quyết định chính)
st->hangover_cnt	Bộ đếm cho các khung tồn lưu được sử dụng

```

flag = 0;
*localVAD = 0;
if ( snr_sum > thr1 && ( st->Opt_HE_SAD_ON == 0 || (flag_he == 1 && flag_he1 == 1) ) )
/* Có câu nói */
{
    flag = 1;
    if ( snr_sum > thr1 )
    {
        *localVAD = 1;      /* VAD không có tồn lưu */
    }

    st->nb_active_frames++; /* Bộ đếm của các khung câu nói hoạt động liên tiếp */
    if ( st->nb_active_frames >= ACTIVE_FRAMES )
    {
        st->nb_active_frames = ACTIVE_FRAMES;
        st->hangover_cnt = 0; /* Thiết lập lại bộ đếm của các khung tồn lưu sau ít nhất
"active_frames" khung câu nói */
    }

    /* trong khoảng thời gian HO */
    if( st->hangover_cnt < HANGOVER_LONG && st->hangover_cnt != 0 )
    {
        st->hangover_cnt++;
    }
}

```

```

}
else
{ /* Thiết lập lại bộ đếm của các khung câu nói cần thiết để bắt đầu thuật toán tồn lưu */
  st->nb_active_frames = 0;
  if( st->hangover_cnt < HANGOVER_LONG ) /* trong khoảng thời gian HO */
  {
    st->hangover_cnt++;
  }
  if( st->hangover_cnt <= hangover_short ) /* tồn lưu "cứng" */
  {
    flag = 1 ;
  }
}

```

Đoạn này được biến đổi như sau để bao gồm quyết định VAD mới để sử dụng cho DTX, vad_flag_dtx. Sử dụng việc làm thích ứng tồn lưu DTX được xác định ở trên, hangover_short_dtx. Nó thêm các biến sau:

flag_dtx	Quyết định VAD cuối cùng, quyết định này còn bao gồm tồn lưu riêng DTX
st->hangover_cnt_dtx	Bộ đếm cho số lượng khung tồn lưu được sử dụng cho DTX

```

flag = 0;
flag_dtx = 0;
*localVAD = 0;
if ( snr_sum > thr1 && ( st->Opt_HE_SAD_ON == 0 || (flag_he == 1 && flag_he1 == 1) ) )
/* Có câu nói */
{
  flag = 1;
  flag_dtx=1;
  if ( snr_sum > thr1 )

```

```

{
    *localVAD = 1;      /* VAD không có tồn lưu */
}

st->nb_active_frames++; /* Bộ đếm của các khung hoạt động câu nói hoạt động liên tiếp */
if ( st->nb_active_frames >= ACTIVE_FRAMES )
{
    st->nb_active_frames = ACTIVE_FRAMES;
    st->hangover_cnt = 0; /* Thiết lập lại bộ đếm của các khung tồn lưu sau ít nhất
"active_frames" khung câu nói */
}

if (st->Opt_DTX_ON)
{
    if (st->vad_flag_cnt_50 > 45 ) /* 45 yêu cầu gần đúng > 90% hoạt động cò */
    {
        /* Nếu đủ hoạt động trong lần thứ hai thì bổ sung tồn lưu
        mà không có yêu cầu cho các khung hoạt động
        */
        st->hangover_cnt_dtx=0;
    }
}

/* trong khoảng thời gian HO */
if( st->hangover_cnt < HANGOVER_LONG && st->hangover_cnt != 0 )
{
    st->hangover_cnt++;
}

if( st->hangover_cnt_dtx < HANGOVER_LONG && st->hangover_cnt_dtx != 0 )
{
    st->hangover_cnt_dtx++;
}
}

else
{ /* Thiết lập lại bộ đếm của các khung câu nói cần thiết để bắt đầu thuật toán tồn lưu */

```

```

st->nb_active_frames = 0;
if( st->hangover_cnt < HANGOVER_LONG ) /* trong khoảng thời gian HO */
{
    st->hangover_cnt++;
}

if( st->hangover_cnt <= hangover_short ) /* tồn lưu "cứng" */
{
    flag = 1 ;
    flag_dtx = 1 ;
}

if( st->hangover_cnt_dtx < HANGOVER_LONG ) /* inside HO period */
{
    st->hangover_cnt_dtx++;
}

if( st->hangover_cnt_dtx <= hangover_short_dtx ) /* tồn lưu "cứng" */
{
    flag_dtx = 1;
}

```

Với việc sử dụng hoạt động ngắn hạn theo các dấu hiệu của quyết định chính và hoạt động dài hạn của quyết định cuối cùng có thể bổ sung tồn lưu phụ một cách cụ thể hơn vào trong các nhóm câu nói và ở cuối nhóm câu nói, và nhờ đó giảm lượng cắt câu nói, cụ thể đối với các VAD hiệu suất cao.

Hoạt động dài hạn của quyết định cuối cùng còn làm cho nó có thể bổ sung tồn lưu vào các nhóm ngắn sau các lời phát biểu dài, điều này giảm rủi ro về việc cắt đầu sau của các tiếng nổ bị làm mất tiếng nói.

Với việc sử dụng các dấu hiệu hoạt động, có thể mở rộng tồn lưu trên các đoạn có hoạt động câu nói đã cao. Điều này cho phép mở rộng dài hơn mà không gặp rủi ro là toàn bộ hoạt động sẽ tăng đột ngột.

Với các dấu hiệu bổ sung, như được thể hiện rõ hơn ở trên, việc lọc thêm là có thể, việc lọc thêm này làm cho việc mở rộng tồn lưu có thể thực hiện kể cả trong các điều kiện giới hạn hơn, như mức câu nói thấp.

Với SAD tích cực hơn, có thể dễ dàng hơn khi loại bỏ việc cắt câu nói bất kỳ nhờ bổ sung một số tồn lưu mở rộng, cụ thể nếu nó có thể được thực hiện một cách cụ thể hơn cho các đoạn hoạt động đã cao. Giải pháp này có thể điều chỉnh dễ dàng hơn so với thử điều chỉnh lại giải pháp dựa trên hoạt động song song của một số SAD.

Các phương án được mô tả ở trên cần được hiểu là một số ví dụ minh họa của sáng chế. Những người có hiểu biết trung bình trong lĩnh vực sẽ hiểu rằng các phương án biến đổi, kết hợp và thay đổi khác nhau có thể được thực hiện đối với các phương án mà không lệch khỏi phạm vi bảo hộ chung của các phương án này. Cụ thể, các giải pháp bộ phận khác nhau theo các phương án khác nhau có thể được kết hợp trong các kết cấu khác, trong đó có thể thực hiện về mặt kỹ thuật.

YÊU CẦU BẢO HỘ

1. Phương pháp để phát hiện hoạt động tiếng nói (voice activity detection - VAD), phương pháp này bao gồm các bước:

tạo ra (310) tín hiệu chỉ thị quyết định VAD chính;

xác định (320) liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện;

tạo ra (330) tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu;

trong đó việc xác định bổ sung tồn lưu dựa trên số đo hoạt động ngắn hạn và số đo hoạt động dài hạn.

2. Phương pháp theo điểm 1, trong đó số đo hoạt động ngắn hạn được suy ra từ N_{st} quyết định VAD chính gần nhất.

3. Phương pháp theo điểm 1 hoặc 2, trong đó số đo hoạt động dài hạn được suy ra từ N_{lt} quyết định VAD chính gần nhất hoặc từ N_{lt} quyết định VAD cuối cùng gần nhất.

4. Phương pháp theo các điểm 2 và 3, trong đó N_{lt} lớn hơn N_{st} .

5. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó việc tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng bao gồm tạo ra hai phiên bản của các quyết định cuối cùng, quyết định VAD cuối cùng thứ nhất và quyết định VAD cuối cùng thứ hai.

6. Phương pháp theo điểm 5, trong đó quyết định VAD cuối cùng thứ hai được đưa ra mà không sử dụng số đo hoạt động ngắn hạn hoặc số đo hoạt động dài hạn.

7. Phương pháp theo điểm 5 hoặc 6, trong đó số đo hoạt động dài hạn được suy ra từ N_{lt} quyết định VAD cuối cùng thứ hai gần nhất.

8. Phương pháp theo điểm bất kỳ trong số các điểm từ 5 đến 7, trong đó quyết định VAD cuối cùng thứ nhất tương ứng với vad_flag_dtx và quyết định VAD cuối cùng thứ hai tương ứng với vad_flag.
9. Phương pháp theo điểm 2, trong đó số đo hoạt động ngắn hạn dựa trên số lượng khung hoạt động trong bộ nhớ của các quyết định VAD chính gần nhất.
10. Phương pháp theo điểm 3, trong đó số đo hoạt động dài hạn dựa trên số lượng khung hoạt động trong bộ nhớ của các quyết định VAD cuối cùng gần nhất hoặc trong bộ nhớ của các quyết định VAD chính gần nhất.
11. Phương pháp theo điểm 9 hoặc 10, trong đó các khung hoạt động được đánh trọng số phụ thuộc vào tuổi của khung hoạt động trong bộ nhớ của các quyết định VAD gần nhất.
12. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, phương pháp này bao gồm bước bổ sung số lượng khung tồn lưu định trước nếu số đo hoạt động ngắn hạn đạt đến ngưỡng định trước thứ nhất và số đo hoạt động dài hạn đạt đến ngưỡng định trước thứ hai.
13. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó quyết định VAD cuối cùng bằng với quyết định hoạt động tiếng nói nếu xác định được việc bổ sung tồn lưu được thực hiện.
14. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó quyết định VAD cuối cùng bằng với quyết định VAD chính nếu xác định được việc bổ sung tồn lưu không được thực hiện.
15. Thiết bị để phát hiện hoạt động tiếng nói (voice activity detection - VAD), thiết bị này bao gồm:

phần đầu vào (412) để nhận tín hiệu đầu vào;

phần bố trí bộ phát hiện tiếng nói chính (401), được nối với phần đầu vào (412), được tạo kết cấu để phát hiện hoạt động tiếng nói trong tín hiệu đầu vào nhận được và

để tạo ra tín hiệu chỉ thị quyết định VAD chính được kết hợp với tín hiệu đầu vào nhận được;

bộ phận bổ sung tồn lưu (402), được nối với phần bố trí bộ phát hiện tiếng nói chính (401), được tạo kết cấu để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện, và để tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu; và

ít nhất một trong số:

bộ đánh giá hoạt động ngắn hạn (403) được nối với đầu vào của bộ phận bổ sung tồn lưu (402), và

bộ đánh giá hoạt động dài hạn (404) được nối với đầu ra của bộ phận bổ sung tồn lưu (402);

trong đó bộ phận bổ sung tồn lưu (402) còn được nối với đầu ra của bộ đánh giá hoạt động ngắn hạn (403) và bộ đánh giá hoạt động dài hạn (404), và được tạo kết cấu để thực hiện việc xác định tồn lưu phụ thuộc vào số đo hoạt động ngắn hạn và số đo hoạt động dài hạn.

16. Thiết bị theo điểm 15, trong đó bộ đánh giá hoạt động ngắn hạn (403) được tạo kết cấu để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất.

17. Thiết bị theo điểm 15 hoặc 16, trong đó bộ đánh giá hoạt động dài hạn (404) được tạo kết cấu để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD chính gần nhất hoặc từ N_{lt} quyết định VAD cuối cùng gần nhất.

18. Thiết bị theo điểm bất kỳ trong số các điểm từ 15 đến 17, trong đó bộ phận bổ sung tồn lưu (402) được tạo kết cấu để tạo ra hai phiên bản của các quyết định cuối cùng, quyết định VAD cuối cùng thứ nhất và quyết định VAD cuối cùng thứ hai.

19. Thiết bị theo điểm 18, trong đó quyết định VAD cuối cùng thứ hai được đưa ra mà không sử dụng số đo hoạt động ngắn hạn hoặc số đo hoạt động dài hạn.

20. Thiết bị theo điểm 18 hoặc 19, trong đó bộ đánh giá hoạt động dài hạn (404) được tạo kết cấu để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng thứ hai gần nhất.

21. Thiết bị theo điểm bất kỳ trong số các điểm từ 15 đến 20, thiết bị này bao gồm bộ nhớ của các quyết định VAD chính và các quyết định VAD cuối cùng, thiết bị còn bao gồm các bộ đếm của các khung hoạt động trong bộ nhớ này của các quyết định VAD chính và các quyết định VAD cuối cùng.

22. Thiết bị theo điểm 21, trong đó ít nhất một trong số đo hoạt động ngắn hạn và số đo hoạt động dài hạn dựa trên số lượng khung hoạt động trong bộ nhớ này của các quyết định VAD chính và các quyết định VAD cuối cùng.

23. Thiết bị theo điểm bất kỳ trong số các điểm từ 15 đến 22, trong đó bộ phận bổ sung tồn lưu (402) còn được tạo kết cấu để bổ sung số lượng khung tồn lưu định trước nếu số đo hoạt động ngắn hạn đạt đến ngưỡng định trước thứ nhất và số đo hoạt động dài hạn đạt đến ngưỡng định trước thứ hai.

24. Thiết bị theo điểm bất kỳ trong số các điểm từ 15 đến 23, trong đó quyết định VAD cuối cùng bằng với quyết định hoạt động tiếng nói nếu xác định được việc bổ sung tồn lưu được thực hiện và quyết định VAD cuối cùng bằng với quyết định VAD chính nếu xác định được việc bổ sung tồn lưu không được thực hiện

25. Bộ mã hóa-giải mã (codec) để mã hóa tiếng nói hoặc âm thanh, bộ mã hóa-giải mã này bao gồm thiết bị theo ít nhất một trong số các điểm từ 15 đến 24.

26. Phương tiện đọc được bởi máy tính có chương trình máy tính được lưu trữ trên đó, chương trình máy tính này bao gồm các đơn vị mã đọc được bởi máy tính, các đơn vị mã này khi chạy trên thiết bị làm cho thiết bị:

tạo ra (310) tín hiệu chỉ thị quyết định VAD chính;

xác định (320) liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện;

tạo ra (330) tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu;

trong đó việc xác định bổ sung tồn lưu dựa trên số đo hoạt động ngắn hạn và số đo hoạt động dài hạn.

27. Thiết bị (500) để phát hiện hoạt động tiếng nói (voice activity detection - VAD), thiết bị này bao gồm:

bộ xử lý (510); và

bộ nhớ (520) lưu trữ các thành phần phần mềm (501, 502, 503, 504, 505), trong đó bộ xử lý (510) được tạo kết cấu để chạy:

thành phần phần mềm (501) để tạo ra tín hiệu chỉ thị quyết định VAD chính;

thành phần phần mềm (502) để xác định liệu việc bổ sung tồn lưu của quyết định VAD chính có được thực hiện;

thành phần phần mềm (503) để tạo ra tín hiệu chỉ thị quyết định VAD cuối cùng ít nhất là một phần dựa trên việc xác định bổ sung tồn lưu;

thành phần phần mềm (504) để suy ra số đo hoạt động ngắn hạn từ N_{st} quyết định VAD chính gần nhất và thành phần phần mềm (505) để suy ra số đo hoạt động dài hạn từ N_{lt} quyết định VAD cuối cùng gần nhất;

trong đó việc bổ sung tồn lưu dựa trên số đo hoạt động ngắn hạn và số đo hoạt động dài hạn.

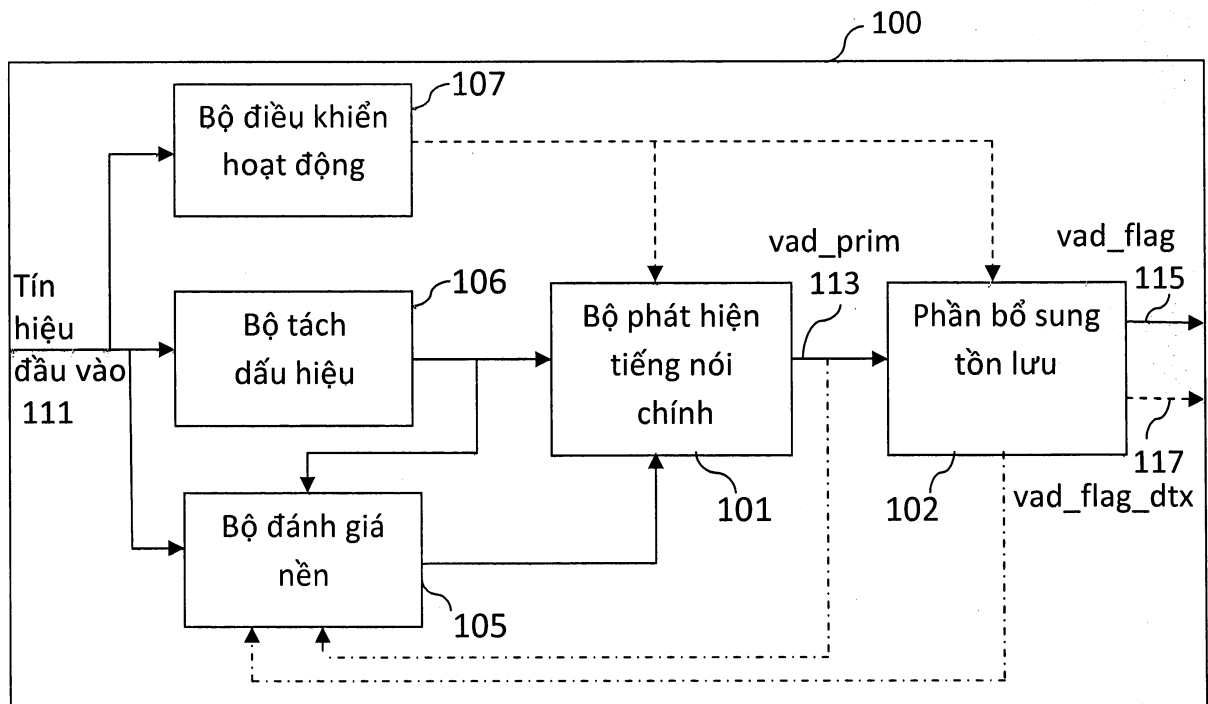


FIG.1

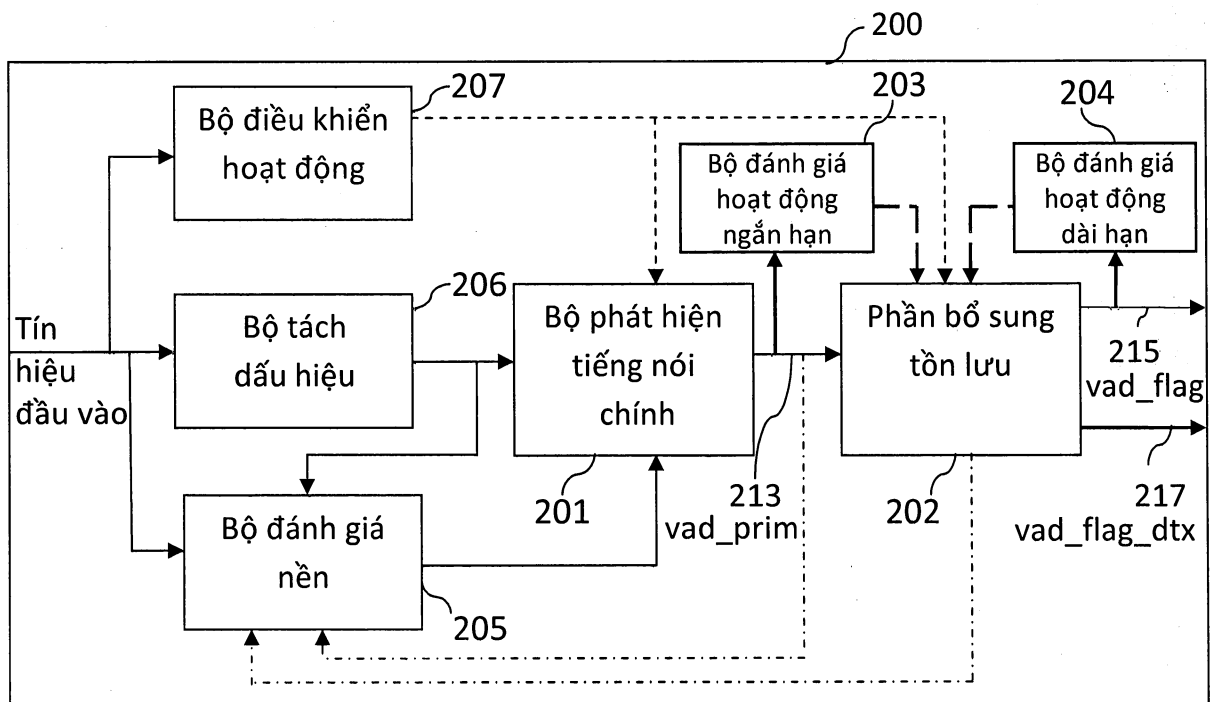


FIG.2

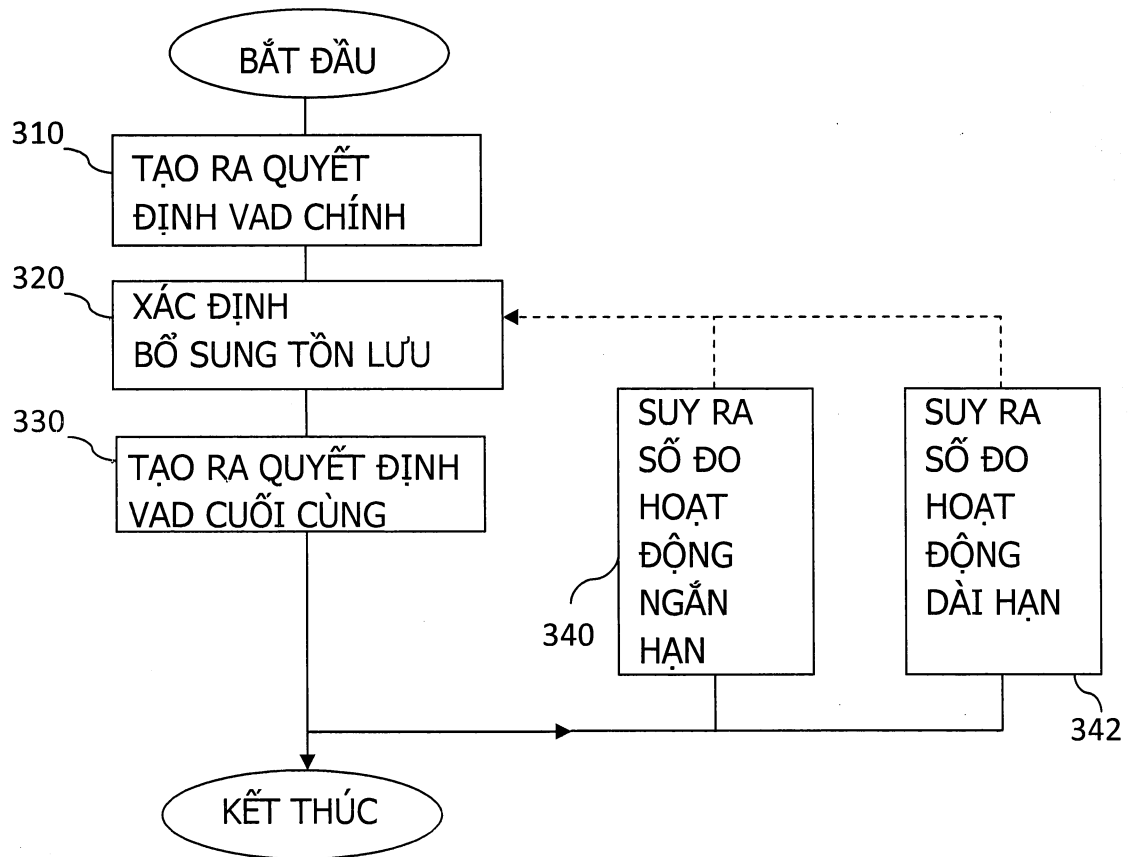


FIG.3

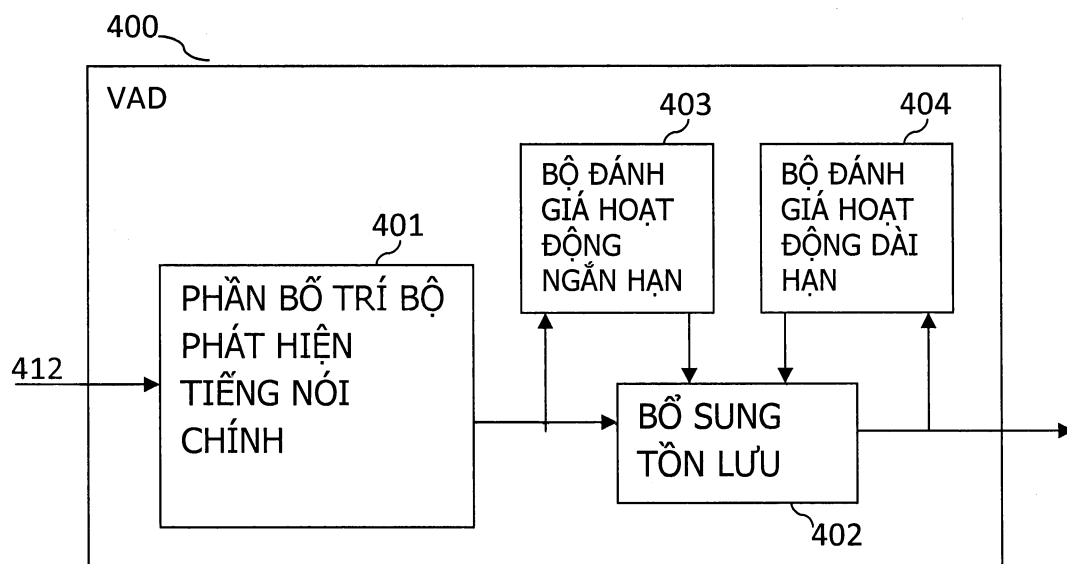


FIG.4A

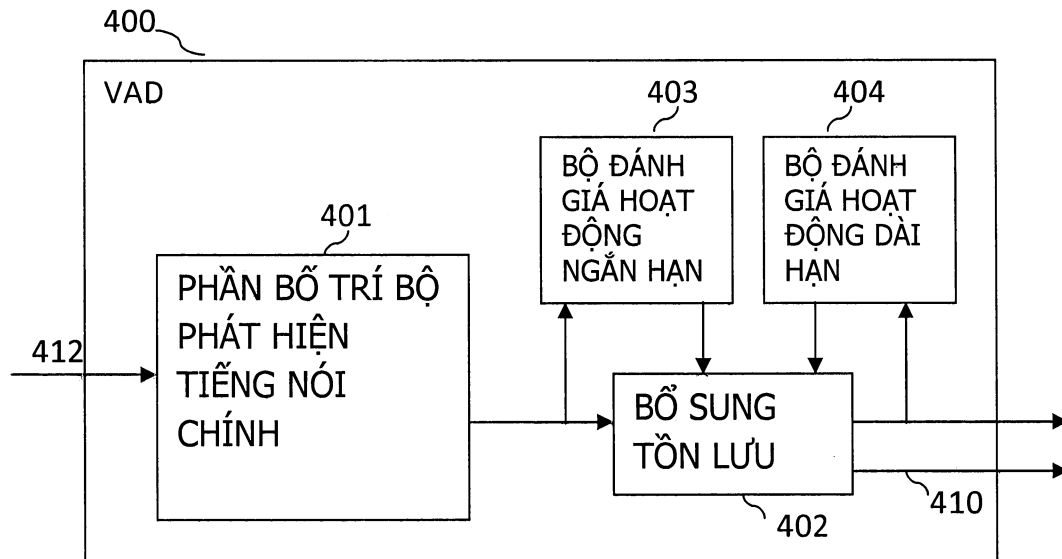


FIG.4B

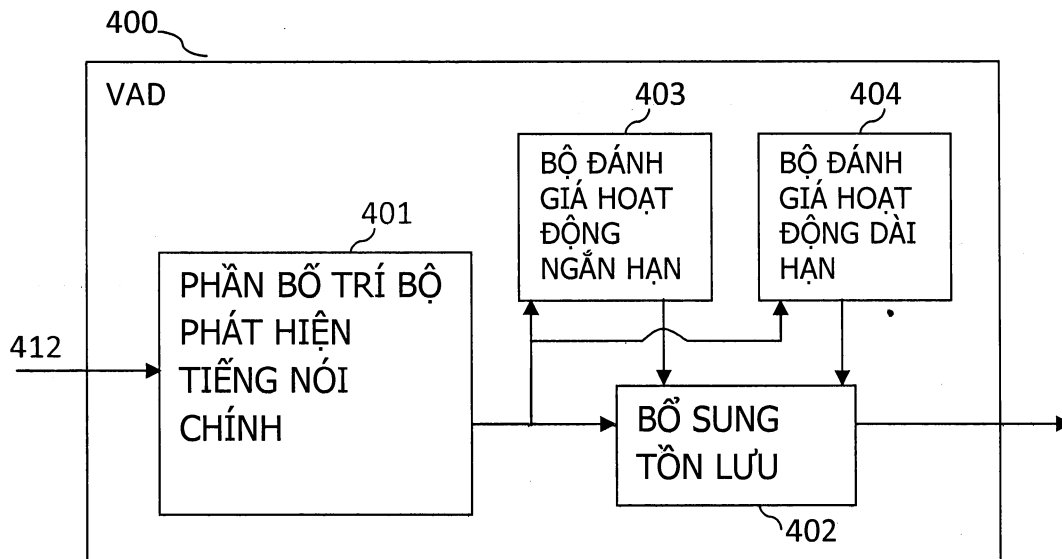


FIG.4C

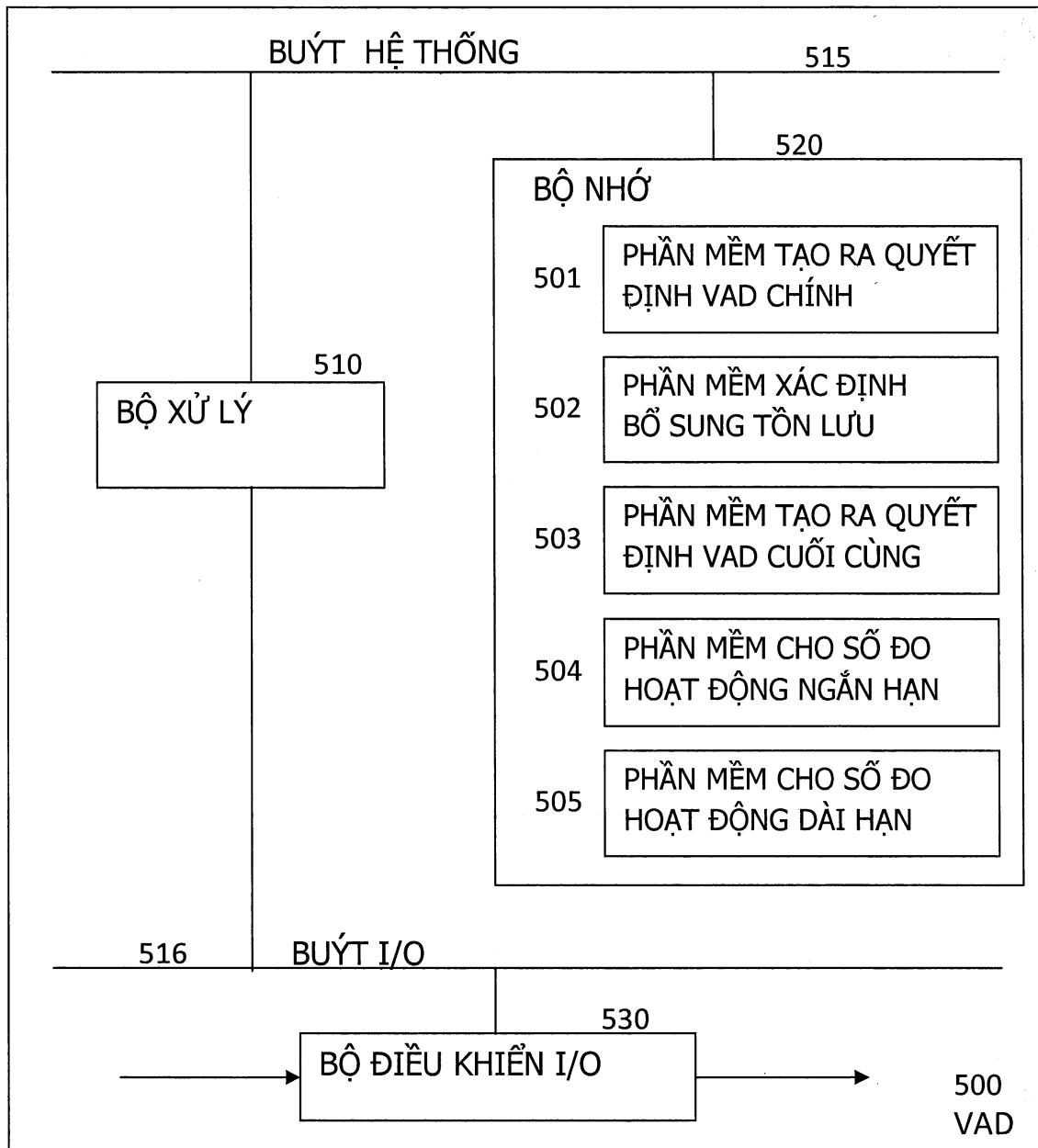


FIG. 5

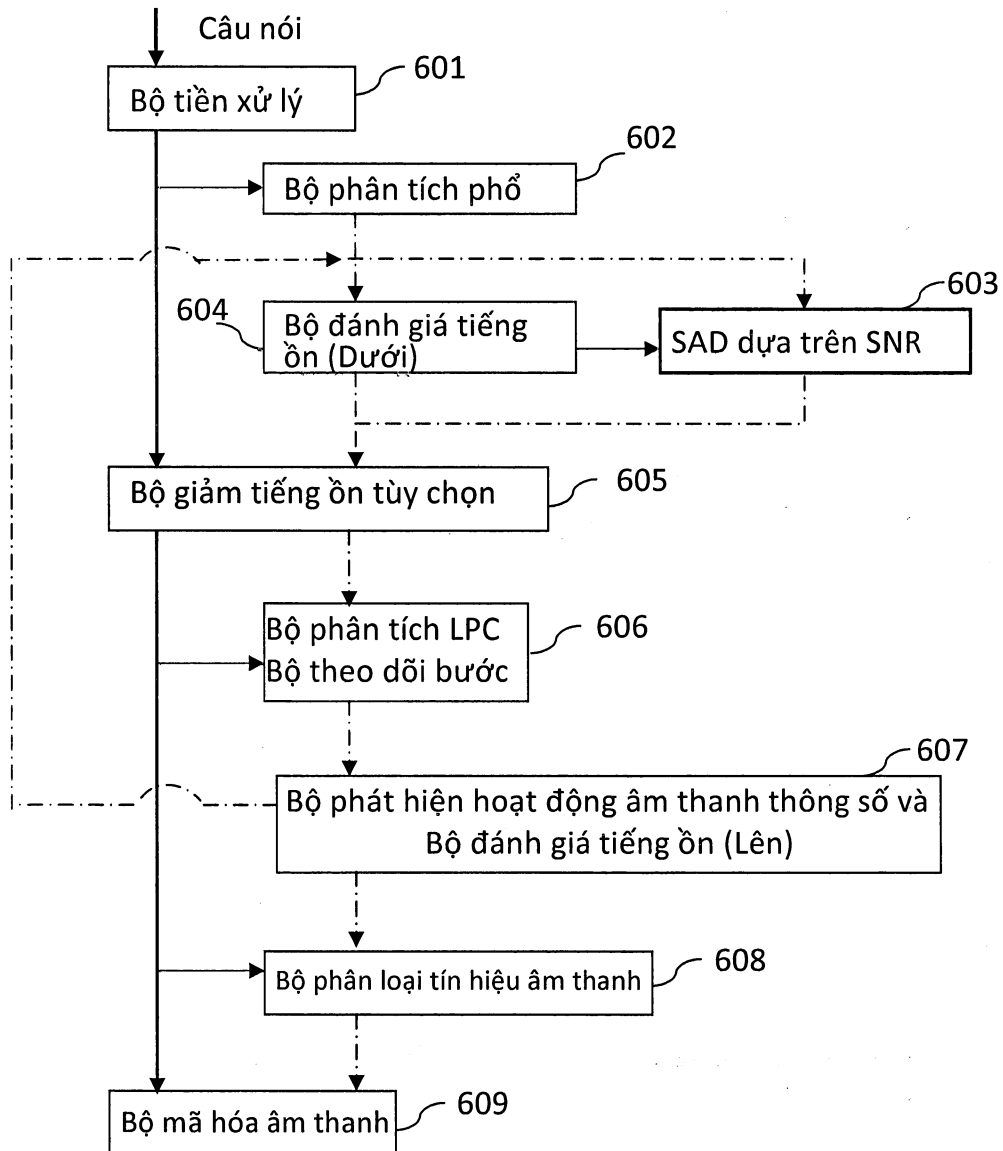


FIG. 6

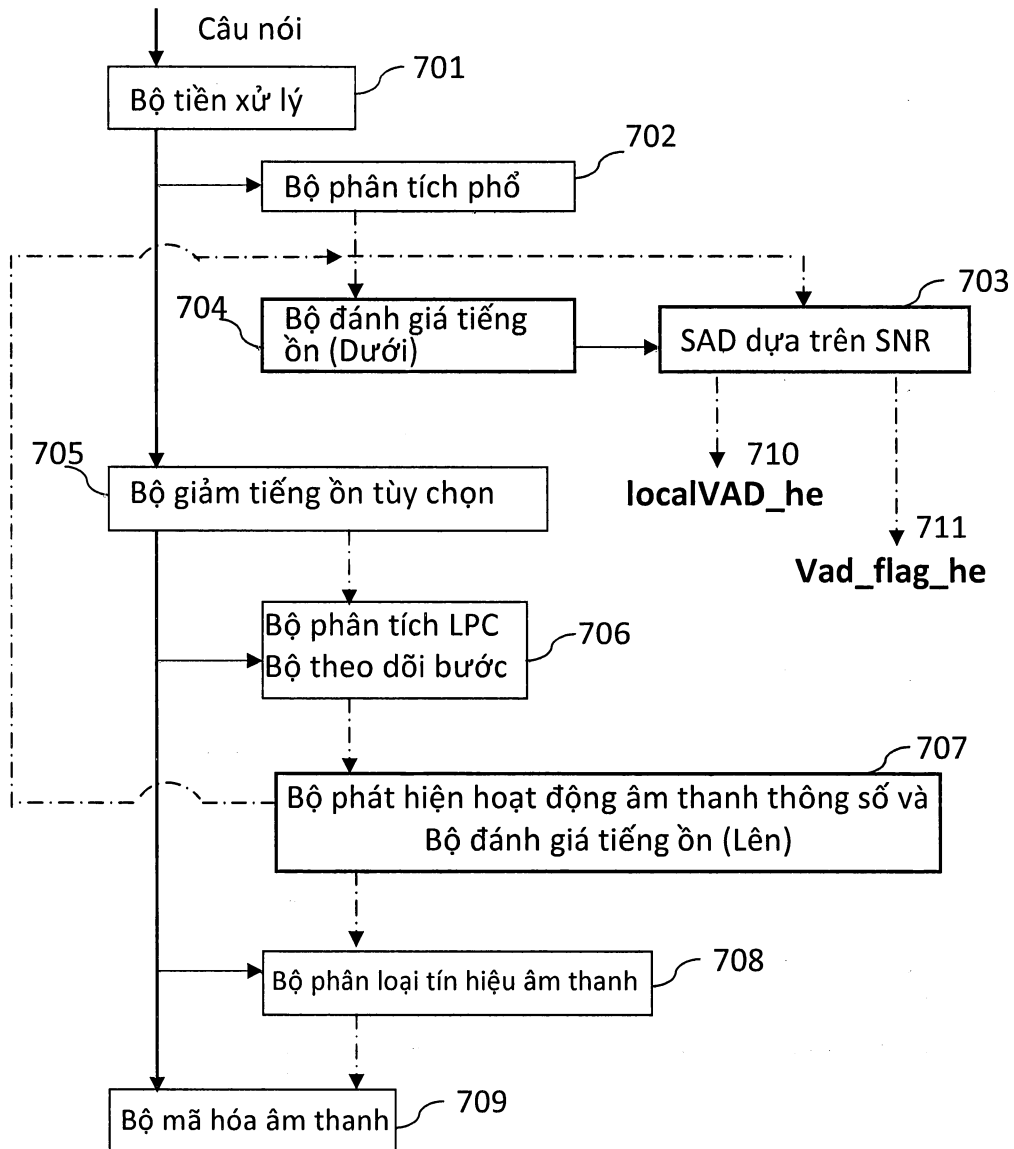


FIG. 7