



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

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

(11)



1-0023451

(51)<sup>7</sup> H04S 3/00

(13) B

(21) 1-2013-04145

(22) 27/06/2012

(86) PCT/US2012/044388 27/06/2012

(87) WO2013/006338A2 10/01/2013

(30) 61/504,005 01/07/2011 US; 61/636,429 20/04/2012 US

(45) 27/04/2020 385

(43) 25/08/2014 317A

(73) DOLBY LABORATORIES LICENSING CORPORATION (US)

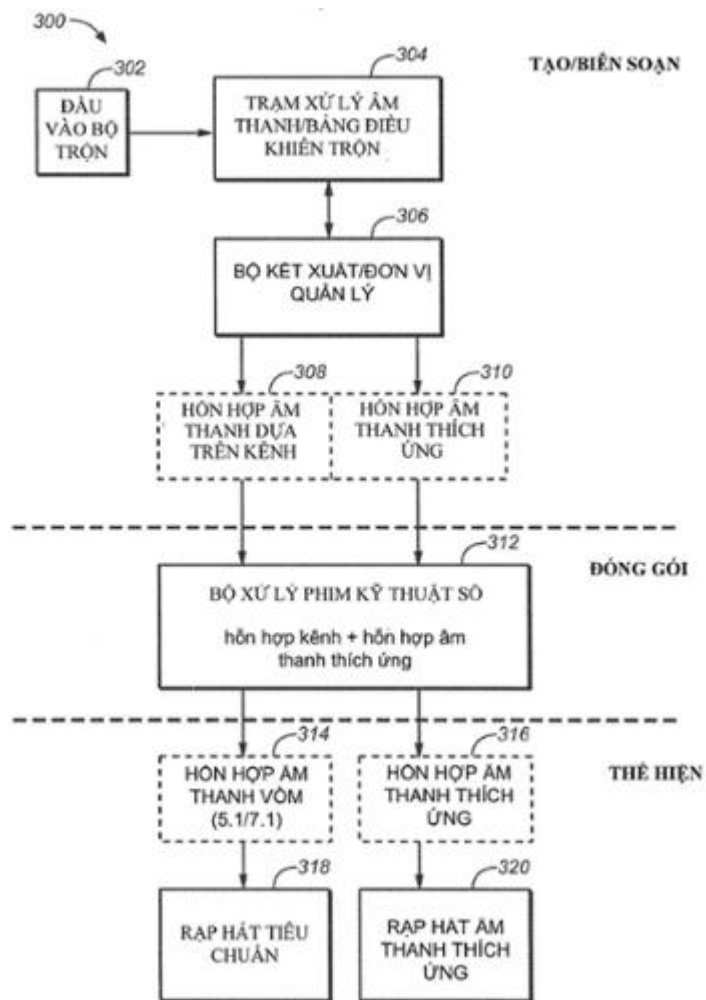
1275 Market Street, San Francisco, California 94103, United States of America.

(72) ROBINSON, Charles Q. (US); TSINGOS, Nicolas R. (FR); CHABANNE, Christophe (FR)

(74) Công ty TNHH Quốc tế D & N (D&N INTERNATIONAL CO.,LTD.)

(54) HỆ THỐNG XỬ LÝ TÍN HIỆU ÂM THANH, PHƯƠNG PHÁP BIÊN SOẠN NỘI DUNG ÂM THANH VÀ KẾT XUẤT TÍN HIỆU ÂM THANH

(57) Sáng chế đề cập đến hệ thống xử lý tín hiệu âm thanh, phương pháp biên soạn, kết xuất tín hiệu âm thanh và phương pháp tạo nội dung âm thanh. Các phương án được mô tả dành cho hệ thống âm thanh thích ứng xử lý dữ liệu âm thanh bao gồm một số dòng âm thanh đơn âm độc lập. Một hoặc nhiều trong số các dòng kết hợp với siêu dữ liệu của nó xác định xem dòng này là dòng dựa trên kênh hay dựa trên đối tượng. Các dòng dựa trên kênh có thông tin kết xuất được mã hóa bởi tên kênh; và các dòng dựa trên đối tượng có thông tin về vị trí được mã hóa thông qua các biểu thị vị trí được mã hóa trong siêu dữ liệu liên kết. Bộ mã hóa-giải mã đóng gói các dòng âm thanh độc lập thành dòng bit tuần tự đơn nhất chứa tất cả dữ liệu âm thanh, cấu hình này cho phép âm thanh được kết xuất theo khung tham chiếu lấy đối tượng làm trung tâm, trong đó vị trí kết xuất của âm thanh dựa trên đặc điểm của môi trường phát lại (ví dụ, kích thước phòng, hình dáng, v.v.) để phù hợp với mục đích của bộ trộn. Siêu dữ liệu vị trí theo đối tượng chứa thông tin khung tham chiếu lấy đối tượng làm trung tâm thích hợp cần thiết để phát âm thanh một cách chính xác bằng cách sử dụng các vị trí loa sẵn có trong phòng được thiết lập để phát nội dung âm thanh thích ứng.



### **Lĩnh vực kỹ thuật được đề cập**

Nói chung, sáng chế đề cập đến việc xử lý tín hiệu âm thanh, và cụ thể hơn đến việc xử lý âm thanh dựa trên kênh và đối tượng lai để sử dụng trong rạp phim, tại nhà, và các môi trường khác. Cụ thể hơn, sáng chế đề cập đến hệ thống xử lý tín hiệu âm thanh, phương pháp biên soạn, kết xuất tín hiệu âm thanh và phương pháp tạo ra nội dung âm thanh.

### **Tình trạng kỹ thuật của sáng chế**

Đối tượng được mô tả trong phần tình trạng kỹ thuật của sáng chế không nên được xem là giải pháp kỹ thuật đã biết chỉ vì nó được nhắc đến trong phần này. Tương tự, vấn đề được đề cập trong phần tình trạng kỹ thuật hoặc có liên quan đến đối tượng trong phần tình trạng kỹ thuật không nên được xem là đã được ghi nhận trước đó trong giải pháp kỹ thuật đã biết. Đối tượng trong phần tình trạng kỹ thuật của sáng chế chỉ thể hiện các phương pháp khác nhau, mà theo các phương pháp này bản thân chúng cũng có thể là sáng chế.

Kể từ khi đưa âm thanh vào phim ảnh, đã có sự phát triển ổn định về công nghệ dùng để nắm bắt ý đồ nghệ thuật của người sáng tác đối với rãnh âm thanh ảnh động và để tái tạo nó một cách chính xác trong môi trường rạp phim. Vai trò cơ bản của âm thanh trong phim là hỗ trợ cho câu chuyện đang được chiếu trên màn hình. Rãnh âm thanh phim thông thường bao gồm nhiều phần tử âm thanh khác nhau tương ứng với các phần tử và hình ảnh trên màn hình, lời thoại, tiếng ồn, và các hiệu ứng âm thanh bắt nguồn từ các phần tử khác nhau trên màn hình và kết hợp với nhạc nền và các hiệu ứng xung quanh nhằm tạo ra trải nghiệm tổng thể cho khán giả. Mục đích nghệ thuật của người sáng tác và nhà sản xuất thể hiện mong muốn tái tạo những âm thanh này theo cách mà tương ứng gần nhất có thể với những gì được chiếu trên màn hình liên quan đến vị trí nguồn âm thanh, cường độ, chuyển động và các tham số tương tự khác.

Việc biên soạn, phân phối và phát lại phim hiện nay phải chịu những hạn chế kìm hãm việc tạo ra âm thanh thực sự sống động và giống như thật. Các hệ thống âm

thanh dựa trên kênh truyền thống gửi nội dung âm thanh dưới dạng tín hiệu cấp cho loa đến từng loa trong môi trường phát lại, chẳng hạn như hệ thống âm thanh lập thể và hệ thống 5.1. Sự ra đời của rạp phim kỹ thuật số đã tạo ra các chuẩn mực mới cho âm thanh trong phim, chẳng hạn như kết hợp tối đa 16 kênh âm thanh để cho phép nhà sáng tạo nội dung có khả năng sáng tạo hơn, và khán giả có trải nghiệm thính giác bao quát và chân thực hơn. Sự ra đời của hệ thống âm thanh vòm 7.1 đã cung cấp một định dạng mới làm tăng số lượng kênh vòm bằng cách chia các kênh vòm trái và phải hiện có thành bốn vùng, do đó tăng phạm vi cho các bộ thiết kế và các bộ trộn âm thanh để kiểm soát việc định vị các phần tử âm thanh trong rạp.

Để cải thiện thêm trải nghiệm cho người nghe, việc phát lại âm thanh trong môi trường ba chiều ảo đã trở thành một lĩnh vực nghiên cứu và phát triển tăng cường. Định nghĩa âm thanh trong không gian sử dụng các đối tượng âm thanh, các đối tượng này là tín hiệu âm thanh có sự mô tả nguồn tham số liên kết về vị trí nguồn bên ngoài (ví dụ, các tọa độ 3D), độ rộng nguồn bên ngoài, và các tham số khác. Âm thanh dựa trên đối tượng ngày càng được sử dụng trong nhiều ứng dụng đa phương tiện hiện nay, ví dụ như phim kỹ thuật số, trò chơi điện tử, các chương trình mô phỏng, và video 3D.

Việc mở rộng vượt quá các tín hiệu cấp cho loa thông thường và âm thanh dựa trên kênh dưới dạng phương tiện phân bố âm thanh không gian là rất quan trọng, và đã có sự quan tâm đáng kể về việc mô tả âm thanh dựa trên mô hình cho phép người nghe/chủ rạp phim tự do lựa chọn cấu hình phát lại phù hợp với các nhu cầu cá nhân hoặc ngân sách của họ, với âm thanh được kết xuất một cách riêng biệt cho cấu hình đã chọn của họ. Ở mức cao, hiện tại có bốn định dạng mô tả âm thanh không gian chính: tín hiệu cấp cho loa trong đó âm thanh được mô tả dưới dạng các tín hiệu dành cho loa ở các vị trí loa danh định; tín hiệu cấp cho micrô trong đó âm thanh được mô tả dưới dạng các tín hiệu được thu bởi micrô thật và ảo trong mảng xác định trước; mô tả dựa theo mô hình trong đó âm thanh được mô tả về trình tự của các sự kiện âm thanh ở các vị trí được mô tả; và âm thanh nổi trong đó âm thanh được mô tả bởi các tín hiệu đi đến tai người nghe. Bốn định dạng mô tả này thường được kết hợp với một hoặc nhiều công nghệ kết xuất để chuyển đổi tín hiệu âm thanh thành tín hiệu cấp cho loa. Các công nghệ kết xuất hiện tại bao gồm kỹ thuật quét, trong đó dòng âm thanh được chuyển đổi thành dạng đầu vào của loa bằng cách sử dụng tập hợp quy tắc quét và các

vị trí loa đã biết hoặc giả định (thường được kết xuất trước khi phân bố); định dạng âm thanh ambisonic, trong đó các tín hiệu micrô được chuyển đổi thành các tín hiệu cấp cho mảng loa có thể thay đổi tỷ lệ (thường được kết xuất sau khi phân bố); kỹ thuật tổng hợp trường sóng (wave field synthesis - WFS) trong đó các sự kiện âm thanh được chuyển đổi thành tín hiệu loa thích hợp để tổng hợp trường âm thanh (thường được kết xuất sau khi phân bố); và âm thanh nổi, trong đó các tín hiệu âm thanh nổi trái/phải (L/R - left/right) được phân bố đến tai L/R, thường sử dụng tai nghe, nhưng cũng sử dụng các loa và công nghệ triệt tiêu xuyên âm (được kết xuất trước hoặc sau quá trình phân bố). Trong số các định dạng này, định dạng đầu vào của loa là phổ biến nhất vì nó đơn giản và hiệu quả. Kết quả âm thanh tốt nhất (chính xác nhất, đáng tin cậy nhất) đạt được bằng cách trộn/giám sát và phân bố cho các dạng đầu vào của loa một cách trực tiếp do không có bước xử lý giữa nhà sáng tạo nội dung và người nghe. Nếu hệ thống phát lại đã được biết trước, mô tả đầu vào của loa thường có độ trung thực cao nhất.

Tuy nhiên, trong nhiều ứng dụng thực tiễn, hệ thống phát lại chưa được biết. Mô tả dựa theo mô hình được coi là có thể thích ứng nhất vì nó không đặt ra giả thuyết về công nghệ kết xuất và do đó nó được ứng dụng một cách dễ dàng nhất vào công nghệ kết xuất bất kỳ. Mặc dù mô tả dựa theo mô hình ghi lại một cách hiệu quả thông tin không gian, nhưng nó trở nên rất kém hiệu quả vì số lượng nguồn âm thanh tăng.

Trong nhiều năm qua, các hệ thống rạp phim có các kênh màn hình riêng đặc trưng ở dạng các kênh trái, giữa, phải và thỉnh thoảng là các kênh “trái bên trong” và “phải bên trong”. Các nguồn riêng này thường có đáp ứng tần số và xử lý công suất đầy đủ để cho phép âm thanh được đặt chính xác vào các khu vực khác nhau của màn hình, và cho phép so khớp âm sắc khi các âm được chuyển dịch hoặc được quét giữa các vị trí. Sự phát triển gần đây trong việc cải thiện trải nghiệm của người nghe đã cố gắng tái tạo một cách chính xác vị trí âm thanh so với người nghe. Trong thiết lập 5.1, “khu vực” vòm bao gồm mảng loa, tất cả các loa này mang cùng thông tin âm thanh bên trong mỗi khu vực vòm trái hoặc vòm phải. Các mảng như vậy có thể có hiệu quả với các hiệu ứng vòm “bao quanh” hoặc phân tán, tuy nhiên, trong cuộc sống hàng ngày nhiều hiệu ứng âm thanh bắt nguồn từ các nguồn có điểm đặt ngẫu nhiên. Ví dụ, trong nhà hàng, âm nhạc xung quanh có thể được phát ra từ mọi nơi xung quanh, còn các âm tĩnh tế nhưng rời rạc phát ra từ các điểm cụ thể: tiếng người nói chuyện từ một điểm;

tiếng lách cách của dao trên đĩa từ một điểm khác. Khả năng bố trí các âm thanh như vậy riêng rẽ quanh thính phòng có thể tăng thêm sự nhận biết chân thực mà không nhận thấy một cách rõ ràng. Các âm thanh trên đầu cũng là một thành phần quan trọng trong định nghĩa về âm thanh vòm. Trong thế giới thực, âm thanh bắt nguồn từ mọi hướng, và không phải lúc nào cũng từ một mặt phẳng nằm ngang duy nhất. Có thể đạt được sự nhận biết nâng cao về độ chân thực nếu âm thanh có thể nghe được từ trên cao, nói cách khác là từ “bán cầu trên”. Tuy nhiên các hệ thống hiện nay không mang đến sự tái tạo âm thanh thật chính xác cho các loại âm thanh khác nhau trong nhiều môi trường phát lại khác nhau. Cần rất nhiều bước xử lý, hiểu biết, và cấu hình của các môi trường phát lại thực tế nhờ sử dụng các hệ thống hiện có để cố gắng thể hiện chính xác âm thanh đặc trưng theo vị trí, do đó khiến cho các hệ thống hiện thời trở nên không thực tiễn đối với hầu hết các ứng dụng.

Điều cần thiết là hệ thống hỗ trợ nhiều kênh màn hình, giúp tăng độ nét và tăng sự kết hợp nghe-nhìn cho các âm thanh và hội thoại trên màn hình, và khả năng định vị chính xác các nguồn ở bất kỳ đâu trong khu vực vòm để cải thiện sự chuyển tiếp nghe nhìn từ màn hình đến phòng. Ví dụ, nếu một nhân vật trên màn hình nhìn vào trong phòng hướng về nguồn âm thanh, thì kỹ sư âm thanh (“bộ trộn”) cần có khả năng định vị âm thanh một cách chính xác sao cho nó khớp với đường nhìn thấy của nhân vật và hiệu ứng sẽ đồng nhất với khán giả. Tuy nhiên, trong quy trình trộn âm thanh vòm 5.1 hoặc 7.1 thông thường, hiệu ứng phụ thuộc nhiều vào vị trí ngồi của người nghe, điều này bất lợi cho hầu hết môi trường nghe quy mô lớn. Độ phân giải âm thanh vòm tăng tạo ra cơ hội mới để sử dụng âm thanh theo cách tập trung vào phòng trái ngược với cách tiếp cận truyền thống, trong đó nội dung được tạo ra giả sử một người nghe ở “điểm hiệu quả”.

Ngoài vấn đề về không gian ra, trạng thái đa kênh hiện tại của các hệ thống theo giải pháp đã biết gặp vấn đề liên quan đến âm sắc. Ví dụ, chất lượng âm sắc của một số âm thanh, chẳng hạn như tiếng hơi nước xì ra khỏi ống nước bị vỡ, có thể bị ảnh hưởng khi được tái tạo bởi mảng loa. Khả năng hướng âm thanh cụ thể đến một loa duy nhất tạo cho bộ trộn cơ hội loại bỏ các nhiễu lạ của quá trình tái tạo mảng loa và mang lại trải nghiệm chân thực hơn cho khán giả. Thông thường, các loa vòm không hỗ trợ cùng một dải tần số âm thanh đầy đủ và cùng một mức âm thanh mà các kênh màn ảnh rộng

hỗ trợ. Trong quá khứ, điều này đã tạo ra các vấn đề cho bộ trộn, làm giảm khả năng di chuyển tự do các âm thanh toàn dải từ màn hình tới phòng của các bộ trộn này. Do đó, các chủ rạp không cảm thấy bắt buộc phải nâng cấp cấu hình kênh vòm của họ, việc này ngăn chặn sự chấp nhận rộng rãi của các hệ thống chất lượng cao hơn.

### **Bản chất kỹ thuật của sáng chế**

Mục đích của sáng chế là đề xuất hệ thống xử lý tín hiệu âm thanh. Hệ thống này có thể bao gồm bộ phận biên soạn được tạo cấu hình để nhận nhiều tín hiệu âm thanh, và tạo ra hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và một hoặc nhiều tập hợp siêu dữ liệu liên quan đến mỗi trong số nhiều dòng âm thanh đơn âm và định rõ vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và các dòng khác trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của âm thanh dựa trên kênh bao gồm các ký hiệu định rõ loa của các loa trong mảng loa, và trong đó vị trí phát lại của âm thanh dựa trên đối tượng bao gồm vị trí trong không gian ba chiều; và ngoài ra trong đó tập hợp siêu dữ liệu thứ nhất được áp dụng theo mặc định vào một hoặc nhiều dòng trong số nhiều dòng âm thanh đơn âm, và tập hợp siêu dữ liệu thứ hai có liên quan đến điều kiện cụ thể của môi trường phát lại và được áp dụng vào một hoặc nhiều trong số các dòng âm thanh đơn âm thay vì tập hợp siêu dữ liệu thứ nhất nếu điều kiện của môi trường phát lại khớp với điều kiện cụ thể của môi trường phát lại. Hệ thống theo sáng chế có thể còn bao gồm hệ thống kết xuất được nối với bộ phận biên soạn và được tạo cấu hình để nhận dòng bit đóng gói nhiều dòng âm thanh đơn âm và các tập hợp siêu dữ liệu, và để kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại phù hợp với các tập hợp siêu dữ liệu dựa trên điều kiện của môi trường phát lại.

Theo một số phương án của sáng chế, mỗi tập hợp siêu dữ liệu có thể bao gồm các phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng, phần tử siêu dữ liệu cho mỗi dòng dựa trên đối tượng xác định các tham số không gian kiểm soát việc phát lại âm thanh dựa trên đối tượng tương ứng, và bao gồm một hoặc nhiều trong số các tham số: vị trí âm thanh, khoảng rộng âm thanh, và vận tốc âm thanh; và ngoài ra trong đó mỗi tập hợp siêu dữ liệu bao gồm phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên

kênh, và mảng loa bao gồm các loa được bố trí theo cấu hình âm thanh vòm xác định, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh bao gồm ký hiệu định rõ các kênh âm thanh vòm của các loa trong mảng loa tương ứng với chuẩn âm thanh vòm xác định.

Theo một số phương án của sáng chế, mảng loa có thể bao gồm các loa bổ sung để phát lại các dòng dựa trên đối tượng được định vị trong môi trường phát lại theo các lệnh thiết lập từ người dùng dựa trên điều kiện của môi trường phát lại, và trong đó điều kiện phát lại phụ thuộc vào các biến số bao gồm: kích thước và hình dạng phòng của môi trường phát lại, độ chiếm giữ, thành phần vật liệu, và tiếng ồn xung quanh; và ngoài ra trong đó hệ thống nhận tập tin thiết lập từ người dùng bao gồm ít nhất một danh sách ký hiệu định rõ loa và ánh xạ kênh đến từng loa của mảng loa, thông tin về việc nhóm loa, và ánh xạ dựa trên vị trí tương đối của loa so với môi trường phát lại.

Theo một số phương án của sáng chế, bộ phận biên soạn có thể bao gồm bảng điều khiển trộn có các bộ điều chỉnh vận hành bởi người dùng để xác định các mức phát lại của các dòng âm thanh đơn âm chứa nội dung dạng âm thanh gốc, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng tương ứng được tạo ra tự động khi người dùng nhập vào các lệnh điều chỉnh của bảng điều khiển trộn.

Theo một số phương án của sáng chế, các tập hợp siêu dữ liệu có thể bao gồm siêu dữ liệu để cho phép tăng trộn hoặc giảm trộn ít nhất một trong số các dòng âm thanh đơn âm dựa trên kênh và các dòng âm thanh đơn âm dựa trên đối tượng phù hợp với sự thay đổi từ cấu hình thứ nhất của mảng loa sang cấu hình thứ hai của mảng loa.

Theo một số phương án của sáng chế, tập hợp siêu dữ liệu có thể bao gồm siêu dữ liệu biểu thị loại nội dung của dòng âm thanh đơn âm, trong đó loại nội dung được chọn từ nhóm bao gồm: lời thoại, âm nhạc, và các hiệu ứng, và mỗi loại nội dung được chứa trong tập hợp tương ứng của dòng dựa trên kênh hoặc dòng dựa trên đối tượng, và ngoài ra trong đó các thành phần âm thanh của mỗi loại nội dung được truyền đến nhóm loa xác định của một hoặc nhiều nhóm loa được gán trong mảng loa.

Theo một số phương án của sáng chế, các loa của mảng loa có thể được đặt ở các vị trí cụ thể trong môi trường phát lại, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng tương ứng xác định rằng một hoặc nhiều thành phần âm thanh



được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh, như được chỉ báo bởi siêu dữ liệu về vị trí.

Theo một số phương án của sáng chế, vị trí phát lại có thể bao gồm vị trí không gian liên quan đến màn hình trong môi trường phát lại, hoặc bề mặt bao quanh môi trường phát lại, và trong đó bề mặt này bao gồm mặt trước, mặt sau, mặt trái, mặt phải, mặt trên, và mặt dưới.

Theo một số phương án của sáng chế, hệ thống có thể còn bao gồm bộ mã hóa-giải mã được nối với bộ phận biên soạn và/hoặc với bộ phận kết xuất và được tạo cấu hình để nhận nhiều dòng âm thanh đơn âm và siêu dữ liệu và để tạo ra một dòng bit số chứa nhiều dòng âm thanh đơn âm theo cách được sắp xếp.

Theo một số phương án của sáng chế, bộ phận kết xuất có thể còn bao gồm các phương tiện chọn thuật toán kết xuất được sử dụng bởi bộ phận kết xuất, thuật toán kết xuất được chọn từ nhóm bao gồm: âm thanh nổi, lưỡng cực âm lập thể, định dạng âm thanh ambisonic, tổng hợp trường sóng (Wave Field Synthesis - WFS), quét đa kênh, các dạng âm thanh stems thô với siêu dữ liệu vị trí, cân bằng kép, và quét biên độ dựa trên vector.

Theo một số phương án của sáng chế, vị trí phát lại cho mỗi trong số nhiều dòng âm thanh đơn âm có thể được xác định một cách độc lập đối với khung tham chiếu lấy chủ thể làm trung tâm hoặc khung tham chiếu lấy đối tượng làm trung tâm, trong đó khung tham chiếu lấy chủ thể làm trung tâm được áp dụng liên quan đến người nghe trong môi trường phát lại, và trong đó khung tham chiếu lấy đối tượng làm trung tâm được áp dụng đối với đặc điểm của môi trường phát lại.

Mục đích khác của sáng chế là đề xuất hệ thống xử lý tín hiệu âm thanh. Hệ thống này có thể bao gồm bộ phận biên soạn được tạo cấu hình để nhận nhiều tín hiệu âm thanh và để tạo ra hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi dòng trong số các dòng âm thanh và xác định vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh hoặc các dòng khác trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của âm thanh dựa trên kênh bao gồm kỹ hiệu định rõ loa của các loa trong mảng loa, và vị trí phát lại của âm thanh dựa trên đối tượng bao gồm vị trí

trong không gian ba chiều, và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa cụ thể của mảng loa. Hệ thống theo sáng chế còn có thể bao gồm hệ thống kết xuất được nối với bộ phận biên soạn và được tạo cấu hình để nhận dòng bit đóng gói nhiều dòng âm thanh đơn âm và siêu dữ liệu, và để kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại, trong đó các loa của mảng loa được đặt ở các vị trí cụ thể trong môi trường phát lại, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng xác định rằng một hoặc nhiều thành phần âm thanh được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh, sao cho dòng âm thanh đơn âm dựa trên đối tượng được kết xuất một cách hiệu quả bởi loa gần nhất với vị trí phát lại dự tính.

Theo một số phương án của sáng chế, siêu dữ liệu có thể bao gồm hai hoặc nhiều tập hợp siêu dữ liệu, và hệ thống kết xuất kết xuất các dòng âm thanh đơn âm tương ứng với một trong số hai hoặc nhiều tập hợp siêu dữ liệu dựa trên điều kiện của môi trường phát lại, trong đó tập hợp siêu dữ liệu thứ nhất được áp dụng cho một hoặc nhiều dòng trong số các dòng âm thanh đơn âm cho điều kiện thứ nhất của môi trường phát lại, và tập hợp siêu dữ liệu thứ hai được áp dụng cho một hoặc nhiều dòng âm thanh cho điều kiện thứ hai của môi trường phát lại; và trong đó mỗi tập hợp siêu dữ liệu bao gồm phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng, phần tử siêu dữ liệu cho mỗi dòng dựa trên đối tượng này xác định các tham số không gian kiểm soát việc phát lại của âm thanh dựa trên đối tượng tương ứng, và bao gồm một hoặc nhiều tham số: vị trí âm thanh, khoảng rộng âm thanh, và vận tốc âm thanh; và trong đó mỗi tập hợp siêu dữ liệu còn bao gồm phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh, và mảng loa bao gồm các loa được bố trí trong cấu hình âm thanh vòm xác định, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh bao gồm các ký hiệu định rõ của các kênh âm thanh vòm của các loa trong mảng loa theo chuẩn âm thanh vòm xác định.

Theo một số phương án của sáng chế, mảng loa có thể bao gồm các loa bổ sung để phát lại các dòng dựa trên đối tượng được định vị trong môi trường phát lại theo các lệnh thiết lập từ người dùng dựa trên điều kiện của môi trường phát lại, và trong đó điều kiện phát lại phụ thuộc vào các biến số bao gồm: kích thước và hình dạng phòng của

môi trường phát lại, độ chiếm giữ, thành phần vật liệu, và tiếng ồn xung quanh; và ngoài ra trong đó hệ thống nhận tập tin thiết lập từ người sử dụng bao gồm ít nhất một danh sách ký hiệu định rõ loa và ánh xạ kênh đến từng loa của mảng loa, thông tin về nhóm loa, và ánh xạ chạy thực dựa trên vị trí tương đối của loa so với môi trường phát lại, và trong đó dòng dựa trên đối tượng, được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh, sẽ bắt vào một loa trong số các loa bổ sung.

Theo một số phương án của sáng chế, vị trí phát lại dự tính có thể bao gồm vị trí không gian tương ứng với màn hình trong môi trường phát lại hoặc bề mặt bao quanh môi trường phát lại, và trong đó bề mặt này bao gồm mặt trước, mặt sau, mặt trái, mặt trên, và mặt sàn.

Mục đích khác nữa của sáng chế là đề xuất hệ thống xử lý tín hiệu âm thanh. Hệ thống này có thể bao gồm bộ phận biên soạn được tạo cấu hình để nhận nhiều tín hiệu âm thanh và để tạo ra hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi dòng trong số nhiều dòng âm thanh đơn âm và xác định vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số dòng trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và trong đó các dòng khác trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của âm thanh dựa trên kênh bao gồm ký hiệu định rõ các ký hiệu định rõ loa của các loa trong mảng loa, và vị trí phát lại của âm thanh dựa trên đối tượng bao gồm vị trí trong không gian ba chiều tương đối với môi trường phát lại chứa mảng loa, và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa cụ thể của mảng loa. Hệ thống theo sáng chế còn có thể bao gồm hệ thống kết xuất được nối với bộ phận biên soạn và được tạo cấu hình để nhận ánh xạ thứ nhất của loa đến kênh âm thanh chứa danh sách các loa và vị trí tương ứng của chúng trong môi trường phát lại và dòng bit đóng gói nhiều dòng âm thanh đơn âm và siêu dữ liệu, và để kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại theo ánh xạ dựa trên vị trí tương đối của các loa so với môi trường phát lại và điều kiện của môi trường phát lại.

Theo một số phương án của sáng chế, điều kiện của môi trường phát lại có thể phụ thuộc vào các biến số bao gồm: kích thước và hình dạng phòng của môi trường phát lại, độ chiếm giữ, thành phần vật liệu, và tiếng ồn xung quanh.

Theo một số phương án của sáng chế, ánh xạ thứ nhất có thể được xác định trong tập tin thiết lập bao gồm ít nhất một danh sách ký hiệu định rõ loa và ánh xạ kênh đến từng loa của mảng loa, và thông tin về việc nhóm loa.

Theo một số phương án của sáng chế, vị trí phát lại dự tính có thể bao gồm vị trí trong không gian liên quan đến màn hình bên trong môi trường phát lại hoặc bề mặt của vỏ chứa môi trường phát lại, và trong đó bề mặt bao gồm một trong số: mặt trước, mặt sau, mặt bên, mặt trên, và mặt sàn của vỏ.

Theo một số phương án của sáng chế, mảng loa có thể bao gồm các loa được sắp xếp theo cấu hình âm thanh vòm xác định, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh chứa các ký hiệu định rõ của các kênh âm thanh vòm của các loa trong mảng loa theo chuẩn âm thanh vòm xác định, và trong đó các dòng dựa trên đối tượng nhất định còn được phát qua các loa bổ sung của mảng loa, và trong đó ánh xạ chạy thực xác định động các loa riêng lẻ nào của mảng loa phát lại dòng dựa trên đối tượng tương ứng trong suốt quá trình phát lại.

Mục đích khác nữa của sáng chế là đề xuất các phương pháp biên soạn tín hiệu âm thanh để kết xuất tương ứng với các bộ phận biên soạn nêu trên, và các phương pháp kết xuất tín hiệu âm thanh tương ứng với các hệ thống kết xuất nêu trên.

Theo một số phương án của sáng chế, mỗi tập hợp siêu dữ liệu có thể bao gồm phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng, phần tử siêu dữ liệu cho mỗi dòng dựa trên đối tượng xác định các tham số không gian kiểm soát sự phát lại của các âm thanh dựa trên đối tượng tương ứng, và bao gồm một hoặc nhiều yếu tố: vị trí âm thanh, khoảng rộng âm thanh, và vận tốc âm thanh; và trong đó mỗi tập hợp siêu dữ liệu còn bao gồm phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh, và mảng loa bao gồm các loa được bố trí trong cấu hình âm thanh vòm xác định, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên kênh bao gồm các ký hiệu định rõ của các kênh âm thanh vòm của các loa trong mảng loa theo chuẩn âm thanh vòm xác định.

Theo một số phương án của sáng chế, mảng loa có thể bao gồm các loa bổ sung để phát lại các dòng dựa trên đối tượng được định vị trong môi trường phát lại, phương pháp còn bao gồm bước nhận lệnh thiết lập từ người dùng dựa trên điều kiện của môi trường phát lại, và trong đó điều kiện phát lại phụ thuộc vào các tham số bao gồm: kích thước và hình dạng của phòng của môi trường phát lại, độ chiếm giữ, thành phần vật liệu, và tiếng ồn xung quanh; các lệnh thiết lập còn bao gồm ít nhất danh sách ký hiệu định rõ loa và ánh xạ kênh đến từng loa của mảng loa, thông tin về việc nhóm loa, và ánh xạ dựa trên vị trí tương đối của loa so với môi trường phát lại.

Theo một số phương án của sáng chế, phương pháp biên soạn tín hiệu âm thanh để kết xuất còn có thể bao gồm các bước: nhận, từ bảng điều khiển trộn có bộ điều chỉnh được vận hành bởi người dùng để xác định mức phát lại của nhiều dòng âm thanh đơn âm chứa nội dung âm thanh gốc; và tự động tạo ra phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng tương ứng được tạo ra khi nhận đầu vào từ người dùng.

Mục đích khác nữa của sáng chế là đề xuất phương pháp tạo ra nội dung âm thanh chứa nhiều dòng âm thanh đơn âm được xử lý trong bộ phận biên soạn, trong đó nhiều dòng âm thanh đơn âm chứa ít nhất một dòng âm thanh dựa trên kênh và ít nhất một dòng âm thanh dựa trên đối tượng. Phương pháp này có thể bao gồm các bước: chỉ rõ cho mỗi dòng âm thanh đơn âm trong số nhiều dòng âm thanh đơn âm xem nó là là dòng dựa trên kênh hay dòng dựa trên đối tượng; liên kết phần tử siêu dữ liệu xác định vị trí kênh với mỗi dòng dựa trên kênh để kết xuất dòng dựa trên kênh tương ứng đến một hoặc nhiều loa trong môi trường phát lại; liên kết một hoặc nhiều phần tử siêu dữ liệu xác định vị trí dựa trên đối tượng với mỗi dòng dựa trên đối tượng để kết xuất dòng dựa trên đối tượng tương ứng đến một hoặc nhiều loa trong môi trường phát lại đối với khung tham chiếu lấy đối tượng làm trung tâm được xác định theo kích thước và các chiều của môi trường phát lại; và ghép nhiều dòng đơn âm và siêu dữ liệu liên kết thành tín hiệu.

Theo một số phương án của sáng chế, môi trường phát lại có thể bao gồm mảng loa được đặt ở các vị trí và hướng xác định tương đối với điểm chuẩn của vỏ bao gồm môi trường phát lại.

Theo một số phương án của sáng chế, tập hợp loa thứ nhất của mảng loa có thể bao gồm các loa được bố trí theo hệ thống âm thanh vòm xác định, và trong đó tập hợp loa thứ hai của mảng loa bao gồm các loa được bố trí theo sơ đồ âm thanh thích ứng.

Theo một số phương án của sáng chế, phương pháp có thể còn bao gồm các bước: xác định loại âm thanh cho tập hợp nhiều dòng âm thanh đơn âm, trong đó loại âm thanh này được chọn từ nhóm bao gồm lời thoại, âm nhạc, và các hiệu ứng; và truyền tập hợp dòng âm thanh đơn âm đến các tập hợp loa cụ thể dựa trên loại âm thanh dành cho tập hợp dòng âm thanh đơn âm tương ứng.

Theo một số phương án của sáng chế, phương pháp có thể còn bao gồm bước tự động tạo ra phần tử siêu dữ liệu qua bộ phận biên soạn được thực hiện trong bảng điều khiển trộn có bộ điều chỉnh vận hành bởi người dùng để xác định mức phát lại của các dòng âm thanh đơn âm.

Theo một số phương án của sáng chế, phương pháp có thể còn bao gồm bước đóng gói nhiều dòng âm thanh đơn âm và phần tử siêu dữ liệu liên kết thành dòng bit số trong bộ mã hóa.

Mục đích khác nữa của sáng chế là đề xuất phương pháp tạo ra nội dung âm thanh. Phương pháp này có thể bao gồm các bước: xác định các giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ nhất kết hợp với việc lập trình nội dung âm thanh để xử lý trong hệ thống âm thanh lai xử lý cả nội dung âm thanh dựa trên kênh và dựa trên đối tượng; xác định giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ hai kết hợp với các đặc tính lưu trữ và kết xuất nội dung âm thanh trong hệ thống âm thanh lai; và xác định giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ ba được kết hợp với vị trí nguồn âm thanh và thông tin điều khiển để kết xuất nội dung âm thanh dựa trên kênh và dựa trên đối tượng.

Theo một số phương án của sáng chế, vị trí nguồn âm thanh để kết xuất nội dung âm thanh dựa trên kênh có thể bao gồm tên gắn với loa trong hệ thống loa âm thanh vòm, trong đó tên này xác định vị trí của loa tương ứng tương đối với một hoặc nhiều vị trí chuẩn trong môi trường phát lại.

Theo một số phương án của sáng chế, thông tin điều khiển để kết xuất nội dung âm thanh dựa trên kênh có thể bao gồm thông tin tăng trộn và giảm trộn để kết xuất nội dung âm thanh trong các cấu hình âm thanh vòm khác nhau, và trong đó siêu dữ liệu bao gồm siêu dữ liệu cho phép hoặc không cho phép chức năng tăng trộn và/hoặc giảm trộn.

Theo một số phương án của sáng chế, vị trí nguồn âm thanh để kết xuất nội dung âm thanh dựa trên đối tượng có thể bao gồm các trị số gắn với một hoặc nhiều hàm toán học xác định vị trí phát lại dự tính để phát lại thành phần âm thanh của nội dung âm thanh dựa trên đối tượng.

Theo một số phương án của sáng chế, hàm toán học có thể được chọn từ nhóm bao gồm: các tọa độ ba chiều được xác định dưới dạng các trị số tọa độ  $x$ ,  $y$ ,  $z$ , định nghĩa bề mặt kèm theo tập hợp tọa độ hai chiều, và định nghĩa đường cong kèm theo tọa độ vị trí tuyến tính một chiều, và vị trí vô hướng trên màn hình trong môi trường phát lại.

Theo một số phương án của sáng chế, thông tin điều khiển để kết xuất nội dung âm thanh dựa trên đối tượng có thể bao gồm các trị số xác định các loa hoặc nhóm loa riêng lẻ trong môi trường phát lại mà thành phần âm thanh được phát qua đó.

Theo một số phương án của sáng chế, thông tin điều khiển để kết xuất nội dung âm thanh dựa trên đối tượng còn có thể bao gồm trị số nhị phân xác định nguồn âm thanh cần được bắt vào loa gần nhất hoặc nhóm loa gần nhất trong môi trường phát lại.

Mục đích nữa của sáng chế là đề xuất phương pháp xác định giao thức truyền âm thanh. Phương pháp này có thể bao gồm các bước: xác định giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ nhất kết hợp với việc lập trình nội dung âm thanh để xử lý trong hệ thống âm thanh lai xử lý cả nội dung âm thanh dựa trên kênh và dựa trên đối tượng; xác định giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ hai kết hợp với các đặc tính lưu trữ và kết xuất của nội dung âm thanh trong hệ thống âm thanh lai; và xác định giá trị của một hoặc nhiều phần tử siêu dữ liệu trong nhóm siêu dữ liệu thứ ba kết hợp với vị trí nguồn âm thanh và thông tin điều khiển để kết xuất nội dung âm thanh dựa trên kênh và dựa trên đối tượng.

Các hệ thống và phương pháp được mô tả cho hệ thống định dạng và xử lý âm thanh rạp phim bao gồm bố cục loa mới (cấu hình kênh) và định dạng mô tả không gian kết hợp. Hệ thống và định dạng âm thanh thích ứng được xác định để hỗ trợ nhiều công nghệ kết xuất. Các dòng âm thanh được truyền cùng với siêu dữ liệu mô tả “mục đích của bộ trộn” bao gồm vị trí mong muốn của dòng âm thanh. Vị trí có thể được biểu diễn dưới dạng kênh có tên (từ trong cấu hình kênh định trước) hoặc dưới dạng thông tin vị trí ba chiều. Định dạng các kênh cộng các đối tượng này kết hợp các phương pháp mô tả cảnh âm thanh dựa trên kênh và dựa trên mô hình tối ưu. Dữ liệu âm thanh cho hệ thống âm thanh thích ứng bao gồm một số dòng âm thanh đơn âm độc lập. Mỗi dòng được kết hợp với siêu dữ liệu của nó để xác định xem dòng là dòng dựa trên kênh hay dựa trên đối tượng. Các dòng dựa trên kênh có thông tin kết xuất được mã hóa bằng tên kênh; và các dòng dựa trên đối tượng có thông tin vị trí được mã hóa thông qua các biểu thức toán học được mã hóa trong siêu dữ liệu gắn kèm khác. Các dòng âm thanh độc lập ban đầu được đóng gói dưới dạng một dòng bit nối tiếp chứa tất cả dữ liệu âm thanh. Cấu hình này cho phép âm thanh được kết xuất theo khung tham chiếu phân bố, trong đó vị trí kết xuất của âm thanh được dựa trên các đặc điểm của môi trường phát lại (ví dụ, kích thước, hình dạng phòng, v.v.) để tương ứng với mục đích của bộ trộn. Siêu dữ liệu vị trí đối tượng chứa thông tin khung tham chiếu phân bố thích hợp cần thiết để phát âm thanh chính xác bằng cách sử dụng các vị trí loa có sẵn trong phòng mà được thiết lập để phát nội dung âm thanh thích ứng. Việc này cho phép âm thanh được trộn tối ưu cho một môi trường phát lại cụ thể mà có thể khác với môi trường trộn do kỹ sư âm thanh trải nghiệm.

Hệ thống âm thanh thích ứng cải thiện chất lượng âm thanh ở các phòng khác nhau thông qua các lợi ích như quy trình cân bằng trong phòng và quản lý âm trầm xung quanh được cải thiện, vì vậy các loa (dù trên màn hình hay ngoài màn hình) có thể được định địa chỉ tự do bởi bộ trộn mà không phải suy nghĩ về so khớp âm. Hệ thống âm thanh thích ứng bổ sung tính linh hoạt và công suất của các đối tượng âm thanh động vào luồng công việc dựa trên kênh truyền thống. Các đối tượng âm thanh này cho phép các nhà sáng tạo kiểm soát các thành phần âm thanh riêng biệt bất kể cấu hình loa phát lại cụ thể bất kỳ, bao gồm các loa trên cao. Hệ thống cũng giới thiệu hiệu suất mới với quy trình sau sản xuất, cho phép các kỹ sư âm thanh nắm bắt hiệu quả tất cả ý định



của họ và sau đó kiểm soát theo thời gian thực, hoặc tự động tạo ra, các phiên bản âm thanh vòm 7.1 và 5.1.

Hệ thống âm thanh thích ứng đơn giản hóa việc phân phối bằng cách gói bản chất âm thanh và mục đích nghệ thuật trong một tệp rãnh duy nhất trong bộ xử lý rạp phim kỹ thuật số, mà có thể được phát lại một cách trung thực trong một loạt các cấu hình rạp. Hệ thống cung cấp khả năng tái tạo tối ưu mục đích nghệ thuật khi trộn và kết xuất sử dụng cùng một cấu hình kênh và một bản tóm tắt duy nhất có sự thích ứng giảm dần cho cấu hình kết xuất, ví dụ, giảm trộn.

Các ưu điểm này và các ưu điểm khác được đưa ra qua các phương án định hướng cho nền tảng âm thanh rạp phim, giải quyết những hạn chế của hệ thống hiện tại và mang lại trải nghiệm âm thanh vượt trội so với các hệ thống hiện có.

### **Mô tả vắn tắt các hình vẽ**

Trong các hình vẽ sau, các số chỉ dẫn giống nhau được dùng để chỉ các phần tử giống nhau. Mặc dù các hình vẽ sau mô tả các ví dụ khác nhau, nhưng một hoặc nhiều phương án thực hiện không bị giới hạn ở các ví dụ được mô tả trên các hình vẽ này.

Fig.1 là sơ đồ mức cao nhất minh họa quy trình tạo ra âm thanh và môi trường phát lại sử dụng hệ thống âm thanh thích ứng, theo một phương án.

Fig.2 minh họa sự kết hợp dữ liệu dựa trên kênh và dựa trên đối tượng để tạo ra hỗn hợp âm thanh thích ứng, theo một phương án.

Fig.3 là sơ đồ khối minh họa tiến trình tạo, đóng gói và kết xuất nội dung âm thanh thích ứng, theo một phương án.

Fig.4 là sơ đồ khối của giai đoạn kết xuất của hệ thống âm thanh thích ứng, theo một phương án.

Fig.5 là bảng liệt kê các loại siêu dữ liệu và phần tử siêu dữ liệu liên kết cho hệ thống âm thanh thích ứng, theo một phương án.

Fig.6 là sơ đồ minh họa quy trình sau sản xuất và hoàn chỉnh hệ thống âm thanh thích ứng, theo một phương án.

Fig.7 là sơ đồ minh họa tiến trình làm ví dụ đối với quy trình đóng gói phim kỹ thuật số sử dụng các tập tin âm thanh thích ứng, theo một phương án.

Fig.8 là hình chiếu nhìn từ trên cao sơ đồ các vị trí loa được đề xuất làm ví dụ dùng với hệ thống âm thanh thích ứng trong thính phòng thông thường.

Fig.9 là hình chiếu nhìn từ trước sơ đồ các vị trí loa được đề xuất làm ví dụ tại màn hình dùng trong thính phòng thông thường.

Fig.10 là hình chiếu bên sơ đồ các vị trí loa được đề xuất làm ví dụ dùng với hệ thống âm thanh thích ứng trong thính phòng thông thường.

Fig.11 là một ví dụ về định vị các loa vòm bên trên và các loa vòm bên cạnh tương ứng với điểm chuẩn, theo một phương án.

### **Mô tả chi tiết sáng chế**

Các hệ thống và phương pháp được mô tả cho hệ thống âm thanh thích ứng và tín hiệu âm thanh đi kèm và định dạng dữ liệu hỗ trợ nhiều công nghệ kết xuất. Các khía cạnh của một hoặc nhiều phương án được mô tả trong sáng chế này có thể được thực hiện trong hệ thống âm thanh hoặc hệ thống nghe nhìn để xử lý thông tin âm thanh nguồn trong hệ thống trộn, kết xuất và phát lại bao gồm một hoặc nhiều máy tính hoặc các thiết bị xử lý thực thi các lệnh phần mềm. Phương án bất kỳ trong số các phương án được mô tả có thể được sử dụng riêng hoặc kết hợp với phương án khác. Mặc dù các phương án khác nhau có thể được thúc đẩy bởi các thiếu sót khác nhau của giải pháp kỹ thuật đã biết, mà có thể được mô tả hoặc ngụ ý ở một hoặc nhiều chỗ trong bản mô tả, nhưng các phương án này không nhất thiết giải quyết bất kỳ thiếu sót nào trong số các thiếu sót này. Nói cách khác, các phương án khác nhau có thể giải quyết các thiếu sót khác nhau mà có thể đã được mô tả trong sáng chế. Một số phương án có thể chỉ giải quyết một phần các thiếu sót hoặc chỉ một thiếu sót mà có thể được mô tả trong sáng chế, và một số phương án có thể không giải quyết bất kỳ thiếu sót nào.

Nhằm mục đích mô tả sáng chế, các thuật ngữ sau có ý nghĩa liên quan:

Kênh hoặc kênh âm thanh: tín hiệu âm thanh đơn âm hoặc dòng âm thanh kèm theo siêu dữ liệu trong đó vị trí được mã hóa làm thông tin nhận dạng (ID - Identifier) kênh, ví dụ, Vòm Trái Trước hoặc Phải Trên. Đối tượng kênh có thể điều khiển nhiều loa, ví dụ, các kênh Vòm Trái (Ls) sẽ cấp phát cho tất cả loa trong mảng Ls.

Cấu hình Kênh: tập hợp vùng loa định trước với các vị trí danh định đi kèm, ví dụ 5.1, 7.1, và v.v; hệ thống 5.1 là hệ thống âm thanh vòm 6 kênh có các kênh trái và

phải phía trước, kênh trung tâm, hai kênh vòm, và kênh siêu trầm; hệ thống 7.1 là hệ thống âm thanh vòm 8 kênh bổ sung 2 kênh vòm khác vào hệ thống 5.1. Các ví dụ về cấu hình hệ thống 5.1 và 7.1 bao gồm hệ thống âm thanh vòm Dolby®.

Loa: bộ chuyển đổi âm thanh hoặc tập hợp bộ chuyển đổi kết xuất tín hiệu âm thanh.

Vùng Loa: mảng gồm một hoặc nhiều loa có thể được tham chiếu một lần và nhận tín hiệu âm thanh đơn nhất, ví dụ Vòm Trái thường thấy ở rạp phim, và cụ thể là để loại trừ hoặc bao hàm cho bước kết xuất đối tượng.

Kênh Loa hoặc Kênh Cấp phát loa: kênh âm thanh được kết hợp với loa hoặc vùng loa có tên trong cấu hình loa xác định. Kênh loa được kết xuất một cách danh định bằng cách sử dụng vùng loa kết hợp.

Nhóm Kênh Loa: tập hợp một hoặc nhiều kênh loa tương ứng với cấu hình kênh (ví dụ, rãnh âm lập thể, rãnh đơn âm, v.v.)

Đối tượng hoặc Kênh Đối tượng: một hoặc nhiều kênh âm thanh có mô tả nguồn tham số, như vị trí nguồn rõ ràng (ví dụ, tọa độ 3D), độ rộng nguồn rõ ràng, v.v. Dòng âm thanh kèm theo siêu dữ liệu trong đó vị trí được mã hóa như là vị trí 3D trong không gian.

Chương trình Âm thanh: tập hợp hoàn chỉnh bao gồm các kênh loa và/hoặc các kênh đối tượng và siêu dữ liệu liên kết mô tả dạng biểu diễn âm thanh không gian mong muốn.

Tham chiếu lấy đối tượng làm trung tâm (Allocentric reference): tham chiếu không gian trong đó các đối tượng âm thanh được định nghĩa tương quan với các đặc điểm trong môi trường kết xuất như tường và góc phòng, vị trí loa tiêu chuẩn, và vị trí màn hình (ví dụ, góc trên bên trái của phòng).

Tham chiếu lấy chủ thể làm trung tâm (Egocentric reference): tham chiếu không gian trong đó các đối tượng âm thanh được định nghĩa tương quan với góc nhìn của (khán giả) người nghe và thường được định rõ đối với các góc tương quan với người nghe (ví dụ, 30 độ về bên phải của người nghe).

Khung: khung là các đoạn ngắn có thể giải mã độc lập mà toàn bộ chương trình âm thanh được chia từ đó. Tốc độ và biên khung âm thanh thường được căn chỉnh với các khung video.

Âm thanh thích ứng: tín hiệu âm thanh dựa trên kênh và/hoặc dựa trên đối tượng kèm siêu dữ liệu kết xuất tín hiệu âm thanh dựa trên môi trường phát lại.

Hệ thống định dạng và xử lý âm thanh rạp phim được mô tả ở đây, còn được gọi là "hệ thống âm thanh thích ứng," sử dụng mô tả âm thanh không gian mới và công nghệ kết xuất để cho phép khán giả đắm chìm hơn, làm chủ về mặt nghệ thuật hơn, hệ thống linh hoạt và có khả năng mở rộng, và dễ dàng lắp đặt và bảo trì. Các phương án của nền âm thanh rạp phim bao gồm một số bộ phận rời rạc bao gồm các công cụ trộn, bộ đóng gói/bộ mã hóa, bộ mở gói/bộ giải mã, bộ phận trộn và kết xuất sau cùng trong rạp, thiết kế loa mới, và bộ khuếch đại nổi mạng. Hệ thống bao gồm các đề xuất dành cho cấu hình kênh mới cần được dùng bởi nhà sáng tạo nội dung và các chủ rạp phim. Hệ thống sử dụng mô tả dựa theo mô hình để hỗ trợ một số đặc tính như: một bản tóm tắt với sự thích ứng dưới và trên với cấu hình kết xuất, ví dụ, trì hoãn việc kết xuất và cho phép sử dụng tối ưu các loa sẵn có; sự đóng gói âm thanh được cải thiện, bao gồm việc giảm trộn được tối ưu hóa để tránh sự tương quan giữa các kênh; độ phân giải không gian tăng thêm nhờ các mảng dẫn hướng (ví dụ, đối tượng âm thanh được gán động cho một hoặc nhiều loa trong mảng vòm); và hỗ trợ các phương pháp kết xuất thay thế.

Fig.1 là sơ đồ mức cao nhất minh họa quy trình tạo ra âm thanh và môi trường phát lại sử dụng hệ thống âm thanh thích ứng theo một phương án. Như được thể hiện trên Fig.1, môi trường toàn diện giữa các đầu 100 bao gồm các bộ phận tạo nội dung, đóng gói, phân phối và phát lại/kết xuất qua rất nhiều thiết bị điểm cuối và các trường hợp sử dụng. Toàn bộ hệ thống 100 bắt đầu với nội dung được bắt giữ từ và cho một số trường hợp sử dụng khác nhau bao gồm các trải nghiệm người dùng khác nhau 112. Phần tử bắt giữ nội dung 102 bao gồm, ví dụ, rạp phim, TV, phát sóng trực tiếp, nội dung do người dùng tạo ra, nội dung được ghi âm, trò chơi, âm nhạc, và nội dung tương tự, và có thể bao gồm nội dung âm thanh đơn thuần hoặc nội dung nghe/nhìn. Nội dung, khi đi qua hệ thống 100 từ giai đoạn bắt giữ 102 đến trải nghiệm người dùng cuối cùng 112, vượt qua một số bước xử lý then chốt thông qua các bộ phận hệ thống riêng biệt.

Các bước xử lý này bao gồm tiền xử lý âm thanh 104, các công cụ và quy trình biên soạn 106, mã hóa bởi bộ mã hóa-giải mã âm thanh 108 để thu nạp, ví dụ, dữ liệu âm thanh, siêu dữ liệu bổ sung và thông tin tái tạo, và các kênh đối tượng. Các hiệu ứng xử lý khác nhau, chẳng hạn như nén (mất hoặc không mất), mã hóa, và tương tự có thể được áp dụng cho các kênh đối tượng để phân phối hiệu quả và an toàn qua một số phương tiện khác nhau. Các quy trình giải mã và kết xuất điểm cuối cụ thể thích hợp 110 sau đó được áp dụng để tái tạo và vận chuyển trải nghiệm âm thanh thích ứng cụ thể của người dùng 112. Trải nghiệm âm thanh 112 là sự phát lại nội dung âm thanh hoặc nội dung nghe/nhìn qua các loa và thiết bị phát lại thích hợp, và có thể là môi trường bất kỳ trong đó người nghe được trải nghiệm sự phát lại nội dung đã bắt giữ, như rạp phim, phòng hòa nhạc, rạp ngoài trời, trong nhà hoặc phòng, buồng nghe, ô tô, bảng điều khiển trò chơi, tai nghe hoặc bộ tai nghe và micrô, hệ thống phát thanh công cộng (public address - PA), hoặc môi trường phát lại khác bất kỳ.

Phương án của hệ thống 100 bao gồm bộ mã hóa-giải mã âm thanh 108 có khả năng phân bổ và lưu trữ hiệu quả các chương trình âm thanh đa kênh, và do đó có thể được gọi là bộ mã hóa-giải mã “lai”. Bộ mã hóa-giải mã 108 kết hợp dữ liệu âm thanh dựa trên kênh truyền thống với siêu dữ liệu liên kết để tạo ra các đối tượng âm thanh hỗ trợ việc tạo ra và phân phối âm thanh mà được làm thích ứng và được tối ưu hóa để kết xuất và phát lại trong các môi trường có thể khác với môi trường trộn. Điều này cho phép kỹ sư âm thanh mã hóa ý muốn của mình về âm thanh cuối cùng nên được nghe như thế nào bởi người nghe, dựa trên môi trường nghe thực tế của người nghe.

Bộ mã hóa-giải mã âm thanh dựa trên kênh truyền thống hoạt động với giả thiết rằng chương trình âm thanh sẽ được tái tạo bởi mảng gồm các loa ở các vị trí định trước tương quan với người nghe. Để tạo ra chương trình âm thanh đa kênh hoàn chỉnh, các kỹ sư âm thanh thường trộn lượng lớn các dòng âm thanh tách rời (ví dụ, lời thoại, âm nhạc, hiệu ứng) để tạo ra ấn tượng tổng thể mong muốn. Các quyết định trộn âm thanh thường được đưa ra bằng cách nghe chương trình âm thanh như được tái tạo bởi mảng gồm các loa ở các vị trí định trước, ví dụ, hệ thống 5.1 hoặc 7.1 cụ thể ở rạp cụ thể. Tín hiệu trộn cuối cùng đóng vai trò làm đầu vào cho bộ mã hóa-giải mã âm thanh. Để tái tạo, các trường âm thanh có độ chính xác trong không gian được thu nhận chỉ khi các loa được bố trí ở các vị trí định trước.

Một dạng mã hóa âm thanh mới gọi là mã hóa đối tượng âm thanh cung cấp các nguồn âm thanh riêng (các đối tượng âm thanh) làm đầu vào cho bộ mã hóa dưới dạng các dòng âm thanh tách rời.

Các ví dụ về đối tượng âm thanh bao gồm các rãnh hội thoại, các dụng cụ đơn nhất, các hiệu ứng âm thanh riêng lẻ, và các nguồn điểm khác. Mỗi đối tượng âm thanh được kết hợp với các tham số không gian, các tham số này có thể bao gồm, nhưng không bị giới hạn ở, vị trí âm thanh, độ rộng âm thanh, và thông tin vận tốc. Các đối tượng âm thanh và các tham số được liên kết sau đó được mã hóa để phân phối và lưu trữ. Quy trình trộn và kết xuất đối tượng âm thanh cuối cùng được thực hiện ở đầu thu của chuỗi phân phối âm thanh, như là một phần của quy trình phát lại chương trình âm thanh. Bước này có thể dựa trên hiểu biết về vị trí loa thực tế nhờ đó kết quả là hệ thống phân phối âm thanh có thể tùy chỉnh được theo các điều kiện nghe cụ thể của người dùng. Hai dạng mã hóa, dựa trên kênh và dựa trên đối tượng, thực hiện một cách tối ưu cho các điều kiện tín hiệu đầu vào khác nhau. Bộ mã hóa âm thanh dựa trên kênh thường hiệu quả hơn trong việc mã hóa các tín hiệu đầu vào chứa hỗn hợp dày đặc gồm các nguồn âm thanh khác nhau và trong việc khuếch tán âm thanh. Ngược lại, bộ mã hóa đối tượng âm thanh hiệu quả hơn trong việc mã hóa một số ít các nguồn âm thanh định hướng cao.

Theo một phương án, các phương pháp và thành phần của hệ thống 100 bao gồm hệ thống mã hóa, phân bổ, và giải mã âm thanh được tạo cấu hình để tạo ra một hoặc nhiều dòng bit chứa cả phần tử âm thanh dựa trên kênh và thành phần mã hóa đối tượng âm thanh truyền thống. Phương pháp kết hợp như vậy mang lại hiệu quả mã hóa và độ linh hoạt kết xuất lớn hơn so với phương pháp dựa trên kênh hoặc dựa trên đối tượng được thực hiện tách rời.

Các khía cạnh khác của các phương án được mô tả bao gồm việc mở rộng bộ mã hóa-giải mã âm thanh dựa trên kênh định trước theo cách thức tương thích ngược để bao gồm các phần tử mã hóa đối tượng âm thanh. “Lớp mở rộng” mới chứa các phần tử mã hóa đối tượng âm thanh được định nghĩa và bổ sung vào lớp “nền” hoặc “tương thích ngược” của dòng bit mã hóa-giải mã âm thanh dựa trên kênh. Phương pháp này cho phép một hoặc nhiều dòng bit, chứa lớp mở rộng cần được xử lý bởi các bộ giải mã kế thừa, đồng thời mang lại trải nghiệm người nghe được nâng cao cho người sử dụng

có bộ giải mã mới. Một ví dụ về trải nghiệm người dùng được nâng cao bao gồm kiểm soát việc kết xuất đối tượng âm thanh. Một ưu điểm nữa của phương pháp này là ở chỗ các đối tượng âm thanh có thể được bổ sung hoặc được biến đổi ở bất kỳ đâu trong chuỗi phân phối mà không cần giải mã/trộn/mã hóa lại âm thanh đa kênh được mã hóa bằng bộ mã hóa-giải mã âm thanh dựa trên kênh.

Đối với khung tham chiếu, các hiệu ứng không gian của các tín hiệu âm thanh là rất quan trọng trong việc mang lại trải nghiệm chìm đắm cho người nghe. Các âm thanh mà được hiểu là phát ra từ khu vực cụ thể của màn hình hoặc phòng chiếu nên được phát qua (các) loa đặt ở cùng vị trí tương đối. Do đó, siêu dữ liệu âm thanh chủ yếu của sự kiện âm thanh trong mô tả dựa theo mô hình là vị trí, mặc dù các tham số khác chẳng hạn như kích thước, định hướng, vận tốc và sự tán âm cũng có thể được mô tả. Để truyền tải vị trí, mô tả không gian âm thanh 3D dựa trên mô hình cần có hệ tọa độ 3D. Hệ tọa độ (Oclit, hình cầu, v.v) dùng để truyền thường được chọn vì thuận tiện hoặc tính nhỏ gọn, tuy nhiên, các hệ tọa độ khác có thể được dùng cho quy trình kết xuất. Ngoài hệ tọa độ ra, cần có khung tham chiếu để thể hiện vị trí của các đối tượng trong không gian. Để các hệ thống này tái tạo một cách chính xác âm thanh dựa trên vị trí trong các môi trường khác nhau, việc lựa chọn khung tham chiếu thích hợp có thể là một yếu tố quan trọng. Đối với khung tham chiếu lấy đối tượng làm trung tâm, vị trí nguồn âm thanh được định nghĩa tương đối với các đặc điểm bên trong môi trường kết xuất chẳng hạn như tường và góc phòng, vị trí loa tiêu chuẩn, và vị trí màn hình. Trong khung tham chiếu lấy chủ thể làm trung tâm, các vị trí được thể hiện đối với quan điểm của người nghe, chẳng hạn như "phía trước tôi, hơi chếch về bên trái," v.v. Các nghiên cứu khoa học về cảm giác không gian (âm thanh và âm thanh khác), đã cho thấy rằng phối cảnh lấy chủ thể làm trung tâm được dùng ở hầu hết mọi nơi. Tuy nhiên đối với rạp phim, khung tham chiếu lấy đối tượng làm trung tâm thường phù hợp hơn vì một số lý do. Ví dụ, vị trí chính xác của đối tượng âm thanh là quan trọng nhất khi có đối tượng được liên kết trên màn hình. Nhờ sử dụng khung tham chiếu lấy đối tượng làm trung tâm, với mỗi vị trí nghe, và đối với kích thước màn hình bất kỳ, âm thanh sẽ định vị tại cùng một vị trí tương đối trên màn hình, ví dụ, chếch về bên trái một phần ba so với chính giữa màn hình. Một lý do khác là bộ trộn có xu hướng nghĩ và trộn trong các điều kiện lấy đối tượng làm trung tâmkhung tham chiếu lấy đối tượng làm trung tâm, và các công cụ quét được bố trí với khung tham chiếu lấy đối tượng làm trung tâm (tường

trong phòng), và bộ trộn muốn chúng được kết xuất theo cách đó, ví dụ, âm thanh này nên ở trên màn hình, âm thanh kia nên ở ngoài màn hình, hoặc từ tường bên trái, v.v.

Mặc dù sử dụng khung tham chiếu lấy đối tượng làm trung tâm trong môi trường rạp phim, có một số trường hợp mà ở đó khung tham chiếu lấy chủ thể làm trung tâm có thể hữu ích, và thích hợp hơn. Các trường hợp này bao gồm các âm thanh không phải lời dẫn, ví dụ, các âm thanh không có trong “không gian câu chuyện”, chẳng hạn như, âm nhạc thể hiện tâm trạng, mà dạng thể hiện đồng đều lấy chủ thể làm trung tâm cho âm nhạc này có thể được mong muốn. Một trường hợp khác là các hiệu ứng trường gần (ví dụ, tiếng muỗi vo ve trong tai trái của người nghe) đòi hỏi dạng thể hiện lấy chủ thể làm trung tâm. Hiện nay vẫn chưa có cách nào để kết xuất trường âm thanh này mà không phải sử dụng tai nghe hoặc loa trường rất gần. Ngoài ra, các nguồn âm thanh xa vô tận (và các sóng phẳng tạo thành) xuất hiện từ vị trí lấy chủ thể làm trung tâm không đổi (ví dụ, 30 độ về bên trái), và các âm như vậy dễ mô tả trong các điều kiện lấy chủ thể làm trung tâm hơn là trong các điều kiện lấy đối tượng làm trung tâm.

Trong một số trường hợp, có thể sử dụng khung tham chiếu lấy đối tượng làm trung tâm miễn là vị trí nghe danh định được định nghĩa, trong khi một số ví dụ đòi hỏi dạng biểu diễn lấy chủ thể làm trung tâm mà vẫn chưa thể kết xuất. Mặc dù tham chiếu lấy đối tượng làm trung tâm có thể hữu ích và thích hợp hơn, nhưng dạng biểu diễn âm thanh nên được mở rộng, do nhiều đặc tính mới, bao gồm dạng biểu diễn lấy chủ thể làm trung tâm có thể mong muốn hơn trong các ứng dụng và môi trường nghe nhất định. Các phương án của hệ thống âm thanh thích ứng bao gồm phương pháp mô tả không gian lai chứa cấu hình kênh được đề xuất cho độ chính xác tối ưu và để kết xuất các nguồn đa điểm phân tán hoặc phức hợp (ví dụ, đám đông trong sân vận động, môi trường xung quanh) bằng cách sử dụng tham chiếu lấy chủ thể làm trung tâm, kèm theo mô tả âm thanh dựa trên mô hình lấy đối tượng làm trung tâm để cho phép tăng một cách hiệu quả độ phân giải không gian và khả năng mở rộng.

#### Thành phần hệ thống

Như thể hiện trên Fig.1, dữ liệu nội dung âm thanh gốc 102 trước hết được xử lý trong khối tiền xử lý 104. Khối tiền xử lý 104 của hệ thống 100 bao gồm thành phần lọc kênh đối tượng. Trong nhiều trường hợp, các đối tượng âm thanh chứa các nguồn âm thanh riêng lẻ để cho phép quét độc lập các âm thanh. Trong một số trường hợp, chẳng



hạn như khi tạo ra các chương trình âm thanh bằng cách sử dụng âm thanh tự nhiên hoặc “nhân tạo”, có thể cần trích xuất các đối tượng âm thanh riêng lẻ từ bản ghi âm chứa nhiều nguồn âm thanh. Các phương án bao gồm phương pháp tách tín hiệu nguồn độc lập từ tín hiệu phức tạp hơn. Các phần tử không mong muốn cần tách ra từ các tín hiệu nguồn độc lập có thể bao gồm, nhưng không bị giới hạn ở, nguồn âm thanh độc lập hoặc nhiều nền khác. Ngoài ra, buồng vang có thể được loại bỏ để phục hồi nguồn âm thanh “khô”.

Bộ tiền xử lý 104 cũng bao gồm chức năng tách nguồn và phát hiện loại nội dung. Hệ thống cung cấp sự tạo ra siêu dữ liệu tự động qua quy trình phân tích âm thanh đầu vào. Siêu dữ liệu về vị trí được suy ra từ bản ghi đa kênh qua quy trình phân tích các mức tương ứng của đầu vào tương ứng giữa các cặp kênh. Việc phát hiện loại nội dung, chẳng hạn như “lời nói” hoặc “âm nhạc”, có thể được thực hiện, ví dụ, bằng cách trích và phân loại đặc tính.

#### Công cụ Biên soạn

Khối công cụ biên soạn 106 bao gồm các đặc tính để cải thiện việc biên soạn các chương trình âm thanh bằng cách tối ưu hóa đầu vào và việc mã hóa ý định sáng tạo của kỹ sư âm thanh cho phép anh ta tạo ra hỗn hợp âm thanh cuối cùng một khi nó được tối ưu hóa để phát lại trong môi trường phát lại thực tế bất kỳ. Việc này được thực hiện nhờ việc sử dụng các đối tượng âm thanh và dữ liệu vị trí được liên kết và mã hóa với nội dung âm thanh gốc. Để bố trí âm thanh một cách chính xác xung quanh thính phòng, kỹ sư âm thanh cần kiểm soát về việc các âm thanh sẽ được kết xuất cuối cùng như thế nào dựa trên các điều kiện hạn chế và đặc điểm thực tế của môi trường phát lại. Hệ thống âm thanh thích ứng cung cấp quyền kiểm soát này bằng cách cho phép kỹ sư âm thanh thay đổi cách thức nội dung âm thanh được thiết kế và được trộn thông qua sử dụng các đối tượng âm thanh và dữ liệu vị trí.

Các đối tượng âm thanh có thể được xem là các nhóm phần tử âm thanh có thể được cảm nhận phát ra từ một vị trí hoặc các vị trí vật lý cụ thể trong phòng. Các đối tượng như vậy có thể tĩnh, hoặc chúng có thể chuyển động. Trong hệ thống âm thanh thích ứng 100, các đối tượng âm thanh được kiểm soát bởi siêu dữ liệu, mà trong số những cái khác, mô tả chi tiết vị trí của âm thanh ở một thời điểm cho trước. Khi các đối tượng được giám sát hoặc được phát lại trong rạp, chúng được kết xuất theo siêu dữ

liệu về vị trí bằng cách sử dụng các loa có mặt, thay vì nhất thiết phải được xuất ra kênh vật lý. Rãnh trong một phiên có thể là đối tượng âm thanh, và dữ liệu quét tiêu chuẩn tương tự với siêu dữ liệu vị trí. Theo cách này, nội dung được bố trí trên màn hình có thể quét một cách hiệu quả theo cách thức giống như với nội dung dựa trên kênh, nhưng nội dung được bố trí trong các vòm có thể được kết xuất cho loa riêng lẻ nếu muốn. Mặc dù việc sử dụng các đối tượng âm thanh tạo ra sự kiểm soát mong muốn đối với các hiệu ứng rời rạc, nhưng các khía cạnh khác của rãnh âm thanh trong phim lại làm việc hiệu quả trong môi trường dựa trên kênh. Ví dụ, nhiều hiệu ứng xung quanh hoặc âm vang thực sự có lợi khi được cấp phát cho các mảng loa. Mặc dù các mảng loa này có thể được xem như các đối tượng có chiều rộng đủ để lấp đầy mảng, nhưng vẫn có lợi nếu giữ lại một số chức năng dựa trên kênh.

Theo một phương án, hệ thống âm thanh thích ứng hỗ trợ dạng âm thanh 'beds' ngoài các đối tượng âm thanh, trong đó dạng âm thanh beds là hỗn hợp âm thanh phụ hoặc âm thanh stems dựa trên kênh. Hỗn hợp âm thanh này có thể được phân phối để phát lại lần cuối (kết xuất) riêng lẻ, hoặc được kết hợp thành một dạng âm thanh bed, phụ thuộc vào ý định của nhà sáng tạo nội dung. Các dạng âm thanh beds này có thể được tạo ra trong các cấu hình dựa trên kênh khác nhau, chẳng hạn như 5.1, 7.1, và có thể được mở rộng thành các định dạng rộng hơn chẳng hạn như 9.1, và mảng chứa các loa trên đầu.

Fig.2 minh họa sự kết hợp của kênh và dữ liệu dựa trên đối tượng để tạo ra hỗn hợp âm thanh thích ứng theo một phương án. Như được thể hiện trên quy trình 200, dữ liệu dựa trên kênh 202, mà ví dụ có thể là dữ liệu âm thanh vòm 5.1 hoặc 7.1 cung cấp dưới dạng dữ liệu điều biến mã xung (pulse-code modulated - PCM), được kết hợp với dữ liệu đối tượng âm thanh 204 để tạo ra hỗn hợp âm thanh thích ứng 208. Dữ liệu đối tượng âm thanh 204 được tạo ra bằng cách kết hợp các phần tử của dữ liệu dựa trên kênh gốc với siêu dữ liệu liên kết để định rõ một số tham số nhất định liên quan đến vị trí của các đối tượng âm thanh.

Như được thể hiện về mặt khái niệm trên Fig.2, các công cụ biên soạn có khả năng tạo ra các chương trình âm thanh chứa đồng thời tổ hợp của các nhóm kênh loa và các kênh đối tượng. Ví dụ, chương trình âm thanh có thể chứa một hoặc nhiều kênh loa được sắp xếp một cách tùy ý vào các nhóm (hoặc các rãnh, ví dụ, rãnh âm lập thể hoặc

5.1), siêu dữ liệu mô tả cho một hoặc nhiều kênh loa, một hoặc nhiều kênh đối tượng, và siêu dữ liệu mô tả cho một hoặc nhiều kênh đối tượng. Trong một chương trình âm thanh, mỗi nhóm kênh loa, và mỗi kênh đối tượng có thể được thể hiện bằng cách sử dụng một hoặc nhiều các tốc độ mẫu khác nhau. Ví dụ, các ứng dụng Phim Kỹ thuật số (D-Cinema) hỗ trợ các tốc độ mẫu 48 kHz và 96 kHz, nhưng các tốc độ mẫu khác cũng có thể được hỗ trợ. Ngoài ra, sự tiêu thụ, lưu trữ và biên soạn kênh với các tốc độ mẫu khác nhau cũng có thể được hỗ trợ.

Quy trình tạo ra chương trình âm thanh cần bước thiết kế âm thanh, bước này bao gồm bước kết hợp các phần tử âm thanh làm tổng của các phần tử âm thanh hợp thành mức điều chỉnh để tạo ra hiệu ứng âm thanh mong muốn mới. Công cụ biên soạn của hệ thống âm thanh thích ứng cho phép tạo ra các hiệu ứng âm thanh dưới dạng tập hợp gồm các đối tượng âm thanh với vị trí tương đối bằng cách sử dụng giao diện người dùng đồ họa thiết kế âm thanh không gian-trực quan. Ví dụ, dạng biểu diễn trực quan của đối tượng tạo ra âm thanh (ví dụ, ô tô) có thể được dùng làm mẫu để thu thập các phần tử âm thanh (tiếng nổ ở cuối pô xe, tiếng ồn của lốp xe, tiếng máy xe)) là các kênh đối tượng chứa âm thanh và vị trí không gian thích hợp (ở ống xả, lốp xe, mũi xe). Các kênh đối tượng riêng lẻ khi đó có thể được kết hợp và sử dụng như một nhóm. Công cụ biên soạn 106 bao gồm một số phần tử giao diện người dùng cho phép kỹ sư âm thanh nhập thông tin điều khiển và xem các tham số trộn, và cải thiện chức năng của hệ thống. Quy trình biên soạn và thiết kế âm thanh cũng được cải thiện bằng cách cho phép các kênh đối tượng và các kênh loa được liên kết và sử dụng làm một nhóm. Ví dụ, kết hợp kênh đối tượng có nguồn âm thanh khô, rời rạc với tập hợp kênh loa chứa tín hiệu vang liên quan.

Công cụ biên soạn âm thanh 106 hỗ trợ khả năng kết hợp nhiều kênh âm thanh, thường gọi là trộn. Nhiều phương pháp trộn được hỗ trợ, và có thể bao gồm trộn dựa trên mức truyền thống và trộn dựa trên âm lượng. Theo phương pháp trộn dựa trên mức, việc định tỷ lệ dải rộng được áp dụng cho các kênh âm thanh, và các kênh âm thanh được định tỷ lệ sau đó được cộng với nhau. Các hệ số định tỷ lệ dải rộng của mỗi kênh được lựa chọn để kiểm soát mức tuyệt đối của tín hiệu được trộn thu được, và cả các mức tương đối của các kênh được trộn bên trong tín hiệu được trộn. Theo phương pháp trộn dựa trên âm lượng, một hoặc nhiều tín hiệu đầu vào được điều biến bằng cách sử

dụng phép định tỷ lệ biên độ phụ thuộc tần số, trong đó biên độ phụ thuộc tần số được lựa chọn để tạo ra âm lượng tuyệt đối và tương đối nhận thấy được mong muốn, trong khi bảo toàn âm sắc nhận thấy được của âm thanh đầu vào.

Các công cụ biên soạn cho phép tạo ra các kênh loa và nhóm kênh loa. Việc này cho phép siêu dữ liệu được kết hợp với mỗi nhóm kênh loa. Mỗi nhóm kênh loa có thể được gắn thẻ theo loại nội dung. Loại nội dung này có thể mở rộng được qua sự mô tả bằng văn bản. Các loại nội dung có thể bao gồm, nhưng không bị giới hạn ở, lời thoại, âm nhạc, và các hiệu ứng. Mỗi nhóm kênh loa có thể được gắn các lệnh duy nhất về cách thức tăng trộn từ cấu hình kênh này đến một cấu hình kênh khác, trong đó việc tăng trộn được định nghĩa là việc tạo ra M kênh âm thanh từ N kênh trong đó  $M > N$ . Các lệnh tăng trộn có thể bao gồm, nhưng không bị giới hạn ở, các lệnh sau: cờ cho phép/cấm biểu thị liệu tăng trộn có được phép không; ma trận tăng trộn để kiểm soát ánh xạ giữa mỗi kênh đầu vào và đầu ra; và các thiết lập kích hoạt mặc định và các thiết lập của ma trận có thể được gắn dựa trên loại nội dung, ví dụ, chỉ cho phép tăng trộn đối với âm nhạc. Mỗi nhóm kênh loa cũng có thể được gắn các lệnh duy nhất về cách thức giảm trộn từ cấu hình kênh này đến một cấu hình kênh khác, trong đó việc giảm trộn được định nghĩa là việc tạo ra Y kênh âm thanh từ X kênh trong đó  $Y < X$ . Các lệnh giảm trộn có thể bao gồm, nhưng không bị giới hạn ở các lệnh sau: ma trận để kiểm soát ánh xạ giữa mỗi kênh đầu vào và đầu ra; và thiết lập ma trận mặc định có thể được gắn dựa trên loại nội dung, ví dụ, lời thoại sẽ giảm trộn trên màn hình; các hiệu ứng sẽ giảm trộn ra khỏi màn hình. Mỗi kênh loa cũng có thể được kết hợp với cờ siêu dữ liệu để không cho phép điều chỉnh âm trầm trong khi kết xuất.

Các phương án chứa dấu hiệu cho phép tạo ra nhiều kênh đối tượng và nhiều nhóm kênh đối tượng. Sáng chế cho phép siêu dữ liệu được kết hợp với mỗi nhóm kênh đối tượng. Mỗi nhóm kênh đối tượng có thể được đánh dấu theo loại nội dung. Loại nội dung có thể mở rộng được qua sự mô tả bằng văn bản, trong đó loại nội dung có thể bao gồm, nhưng không bị giới hạn ở, lời thoại, âm nhạc, và các hiệu ứng. Mỗi nhóm kênh đối tượng có thể được gắn siêu dữ liệu để mô tả cách (các) đối tượng nên được kết xuất.

Thông tin vị trí được cung cấp để biểu thị vị trí nguồn rõ ràng mong muốn. Vị trí này có thể được biểu thị bằng cách sử dụng khung tham chiếu lấy chủ thể làm trung tâm hoặc lấy đối tượng làm trung tâm. Khung tham chiếu lấy chủ thể làm trung tâm thích

hợp khi vị trí nguồn được tham chiếu với người nghe. Đối với vị trí có khung tham chiếu lấy chủ thể làm trung tâm, các tọa độ hình cầu là có ích cho việc mô tả vị trí. Tham chiếu lấy đối tượng làm trung tâm là khung tham chiếu thông thường cho rạp phim hoặc các dạng biểu diễn âm thanh/hình ảnh khác trong đó vị trí nguồn được tham chiếu tương ứng với đối tượng trong môi trường biểu diễn như màn hình trực quan hoặc các cạnh trong phòng. Thông tin quỹ đạo ba chiều (3D - Three-dimensional) được cung cấp để cho phép nội suy vị trí hoặc để sử dụng các quyết định kết xuất khác như kích hoạt "chế độ bắt nhanh - snap to mode." Thông tin kích thước được cung cấp để biểu thị kích thước nguồn âm thanh nhận thức được rõ ràng mong muốn.

Quy trình lượng tử hóa không gian được thực hiện nhờ điều khiển “bắt giữ cho loa gần nhất- snap to closest speaker” biểu thị ý muốn của kỹ sư hoặc bộ trộn âm thanh để có đối tượng được kết xuất bằng chỉ một loa (có thể kém chính xác về không gian). Giới hạn ở biến dạng không gian cho phép có thể được biểu thị qua các ngưỡng dung sai độ cao và góc phương vị sao cho nếu vượt quá ngưỡng này, thì chức năng "bắt" sẽ không xuất hiện. Ngoài các ngưỡng về khoảng cách, tham số tốc độ làm mờ âm có thể được biểu thị để kiểm soát nhanh cách thức đối tượng chuyển động sẽ chuyển tiếp hoặc nhảy từ loa này đến loa khác khi vị trí mong muốn cắt ngang giữa các loa.

Theo một phương án, siêu dữ liệu không gian phụ thuộc được dùng cho siêu dữ liệu vị trí nhất định. Ví dụ, siêu dữ liệu có thể được tạo ra một cách tự động cho đối tượng “phụ - slave” bằng cách kết hợp nó với đối tượng “chính - master” mà đối tượng phụ này theo sau. Độ trễ thời gian hoặc tốc độ tương đối có thể được gán cho đối tượng phụ. Các cơ chế cũng có thể được cung cấp để cho phép định nghĩa trọng tâm âm thanh cho các tập hợp hoặc các nhóm đối tượng, nhờ đó đối tượng có thể được kết xuất sao cho nhận thức được là nó chuyển động xung quanh đối tượng khác. Trong trường hợp như vậy, một hoặc nhiều đối tượng có thể quay quanh đối tượng hoặc khu vực xác định, chẳng hạn như điểm cao, hoặc khu vực khô trong phòng. Trọng tâm âm thanh sau đó sẽ được dùng trong giai đoạn kết xuất để giúp xác định thông tin vị trí cho mỗi âm thanh dựa trên đối tượng thích hợp, mặc dù thông tin vị trí cuối cùng sẽ được thể hiện là vị trí tương đối với phòng, ngược với vị trí tương đối với đối tượng khác.

Khi đối tượng được kết xuất nó được gán cho một hoặc nhiều loa theo siêu dữ liệu vị trí, và vị trí của các loa phát lại. Siêu dữ liệu bổ sung có thể được kết hợp với đối

tượng để giới hạn số loa sẽ được sử dụng. Việc sử dụng các giới hạn này có thể cấm sử dụng các loa gắn hoặc đơn thuần là hạn chế các loa gắn (cho phép ít năng lượng hơn vào loa hoặc các loa so với áp dụng theo phương án khác). Tập hợp loa bị giới hạn có thể bao gồm, nhưng không bị giới hạn ở, bất kỳ trong số các loa hoặc vùng loa có tên (ví dụ, L, C, R, v.v.), hoặc các khu vực loa, chẳng hạn như: tường trước, tường sau, tường trái, tường phải, trần nhà, sàn, loa trong phòng, và v.v. Tương tự, trong quá trình xác định hỗn hợp mong muốn gồm nhiều phần tử âm thanh, có thể khiến cho một hoặc nhiều phần tử âm thanh trở nên không nghe được hoặc "bị chặn" do sự có mặt của các phần tử âm thanh "chặn" khác. Ví dụ, khi dò được các phần tử bị chặn, chúng có thể được nhận diện với người sử dụng qua hiển thị đồ họa.

Như được mô tả ở chỗ khác, phần mô tả chương trình âm thanh có thể được làm thích ứng để kết xuất trên rất nhiều hệ cài đặt loa và cấu hình kênh. Khi chương trình âm thanh được biên soạn, điều quan trọng là giám sát hiệu ứng kết xuất chương trình trong các cấu hình phát lại dự kiến để xác minh rằng đã đạt được kết quả mong muốn. Sáng chế bao gồm khả năng lựa chọn các cấu hình phát lại đích và kiểm tra kết quả. Ngoài ra, hệ thống có thể tự động giám sát các mức tín hiệu trong trường hợp xấu nhất (tức là, mức cao nhất) mà sẽ được tạo ra trong mỗi cấu hình phát lại dự kiến, và cung cấp chỉ báo nếu xảy ra tình trạng cắt hoặc giới hạn.

Fig.3 là sơ đồ khối minh họa tiến trình tạo, đóng gói và kết xuất nội dung âm thanh thích ứng theo một phương án. Tiến trình 300 trên Fig.3 được chia thành ba nhóm tác vụ riêng được đánh dấu là tạo/biên soạn, đóng gói, và thể hiện. Nói chung, mô hình lai của dạng âm thanh beds và đối tượng thể hiện trên Fig.2 cho phép hầu hết các bước thiết kế, biên soạn, trộn trước, và trộn cuối cùng được thực hiện theo cách giống như cách của ngày nay mà không phát sinh chi phí trong các quy trình hiện tại. Theo một phương án, chức năng âm thanh thích ứng được cung cấp dưới dạng phần mềm, phần sụn hoặc mạch được dùng với thiết bị tạo và xử lý âm thanh, trong đó thiết bị này có thể là hệ thống phân cứng mới hoặc là bản nâng cấp so với hệ thống đã có. Ví dụ, các ứng dụng cảm có thể được cung cấp cho các trạm xử lý âm thanh kỹ thuật số để cho phép các kỹ thuật quét hiện có trong quá trình thiết kế và biên soạn âm thanh không thay đổi. Theo cách này, có thể bố trí cả dạng âm thanh beds và đối tượng trong máy trạm trong các phòng biên soạn trang bị âm thanh vòm 5.1 hoặc tương tự. Đối tượng âm thanh và

siêu dữ liệu được ghi trong một phiên để chuẩn bị cho các giai đoạn trộn trước và trộn cuối cùng trong phòng lồng tiếng.

Như được thể hiện trên Fig.3, tác vụ tạo hoặc biên soạn bao gồm bước nhập các lệnh điều khiển trộn 302 bởi người dùng, ví dụ, kỹ sư âm thanh trong ví dụ sau đây, vào bảng điều khiển trộn hoặc trạm xử lý âm thanh 304. Theo một phương án, siêu dữ liệu được tích hợp vào bề mặt bảng điều khiển trộn, cho phép giám âm trên các dải kênh, quét và xử lý âm thanh để làm việc với cả các dạng âm thanh beds hoặc stems và các đối tượng âm thanh. Siêu dữ liệu có thể được biên soạn bằng cách sử dụng cả bề mặt bảng điều khiển hoặc giao diện người dùng của trạm xử lý, và âm thanh được giám sát nhờ sử dụng đơn vị kết xuất và hoàn chỉnh (rendering and mastering unit - RMU) 306. Dữ liệu âm thanh bed và âm thanh đối tượng và siêu dữ liệu liên kết được ghi lại trong phiên hoàn chỉnh để tạo ra “bản in gốc – print master” mà bao gồm hỗn hợp âm thanh thích ứng 310 và mọi giao phẩm được kết xuất khác (chẳng hạn như hỗn hợp âm thanh vòm trong rạp 7.1 hoặc 5.1) 308.

Các công cụ biên soạn hiện có (ví dụ, trạm xử lý âm thanh kỹ thuật số chẳng hạn như công cụ Pro Tools) có thể được dùng để cho phép các kỹ sư âm thanh đánh dấu các rãnh âm thanh riêng trong phiên trộn. Các phương án mở rộng khái niệm này bằng cách cho phép người dùng đánh dấu các đoạn con riêng trong rãnh để hỗ trợ tìm kiếm hoặc nhận dạng nhanh các phần tử âm thanh. Giao diện người dùng đối với bảng điều khiển trộn mà cho phép định nghĩa và tạo ra siêu dữ liệu có thể được cài đặt qua các phần tử giao diện người dùng đồ họa, các bộ phận điều khiển vật lý (ví dụ, thanh trượt và nút bấm), hoặc kết hợp bất kỳ của các bộ phận này.

Trong giai đoạn đóng gói, tệp bản in gốc được gói bằng cách sử dụng các thủ tục gói MXF tiêu chuẩn công nghiệp, được tạo hàm băm và được mã hóa tùy ý để đảm bảo sự nguyên vẹn của nội dung âm thanh để phân phối cho bộ phận đóng gói phim kỹ thuật số. Bước này có thể được thực hiện bởi bộ xử lý rạp phim kỹ thuật số (digital cinema processor - DCP) 312 hoặc bộ xử lý âm thanh thích hợp bất kỳ tùy thuộc vào môi trường phát lại cuối cùng, chẳng hạn như rạp được trang bị âm thanh vòm chuẩn 318, rạp cho phép âm thanh thích ứng 320, hoặc môi trường phát lại khác bất kỳ. Như được thể hiện trên Fig.3, bộ xử lý 312 xuất tín hiệu âm thanh thích hợp 314 và 316 tùy thuộc vào môi trường thể hiện.

Theo một phương án, bản in gốc âm thanh thích ứng chứa hỗn hợp âm thanh thích ứng, cùng với hỗn hợp Điều biến Mã Xung (PCM - Pulse Code Modulated) phù hợp với chuẩn DCI. Hỗn hợp PCM có thể được kết xuất bởi đơn vị kết xuất và hoàn chỉnh trong phòng lồng tiếng, hoặc được tạo ra bởi đường trộn riêng nếu muốn. Âm thanh PCM tạo ra tập tin rãnh âm thanh chuẩn chính bên trong bộ xử lý phim kỹ thuật số 312, và âm thanh thích ứng tạo ra tập tin rãnh bổ sung. Tập tin rãnh như vậy có thể phù hợp với các tiêu chuẩn công nghiệp hiện tại, và các máy chủ tương ứng với chuẩn DCI bỏ qua mà không thể sử dụng nó.

Trong môi trường phát lại rạp phim làm ví dụ, DCP chứa tập tin rãnh âm thanh thích ứng được máy chủ nhận biết là gói hợp lệ, và được nạp vào máy chủ và sau đó được truyền theo dòng cho bộ xử lý âm thanh thích ứng trong phim. Nếu hệ thống có cả tập tin âm thanh thích ứng và PCM tuyến tính, thì hệ thống đó có thể chuyển giữa hai loại này nếu cần thiết. Để phân phối cho giai đoạn thể hiện, sơ đồ đóng gói âm thanh thích ứng cho phép phân phối một loại gói cần phân phối cho phim. Gói DCP chứa cả các tập tin PCM và âm thanh thích ứng. Việc sử dụng các khóa bảo mật, chẳng hạn như thông báo phân phối khóa (key delivery message - KDM) có thể được hợp nhất để cho phép phân phối an toàn nội dung phim, hoặc nội dung tương tự khác.

Như được thể hiện trên Fig.3, phương pháp luận về âm thanh thích ứng được thực hiện bằng cách cho phép kỹ sư âm thanh biểu diễn ý định của mình liên quan đến việc kết xuất và phát lại nội dung âm thanh qua trạm xử lý âm thanh 304. Bằng cách kiểm soát các bộ phận điều khiển đầu vào, kỹ sư có thể định rõ vị trí và cách thức các đối tượng âm thanh và các phần tử âm thanh được phát lại tùy thuộc vào môi trường nghe. Siêu dữ liệu được tạo ra trong trạm âm thanh 304 đáp lại đầu vào trộn của kỹ sư 302 để cung cấp hàng đợi kết xuất kiểm soát các tham số không gian (ví dụ, vị trí, vận tốc, cường độ, âm sắc, v.v.) và định rõ (các) loa hoặc nhóm loa nào trong môi trường nghe phát ra âm thanh tương ứng trong quá trình chiếu. Siêu dữ liệu này được kết hợp với dữ liệu âm thanh tương ứng trong trạm xử lý âm thanh 304 hoặc RMU 306 để DCP 312 đóng gói và vận chuyển.

Giao diện người dùng đồ họa và các công cụ phần mềm được kỹ sư cung cấp quyền kiểm soát trạm 304 bao gồm ít nhất một phần trong số các công cụ biên soạn 106 của bộ mã hóa-giải mã âm thanh lai trên Fig.1.



Như được thể hiện trên Fig.1, hệ thống 100 bao gồm bộ mã hóa-giải mã âm thanh lai 108. Bộ phận này chứa hệ thống mã hóa, phân phối và giải mã âm thanh được tạo cấu hình để tạo ra một dòng bit chứa cả các phần tử âm thanh dựa trên kênh và các phần tử mã hóa đối tượng âm thanh thông thường. Hệ thống mã hóa âm thanh lai được xây dựng xung quanh hệ thống mã hóa dựa trên kênh đã được tạo cấu hình để tạo ra một dòng bit (hợp nhất) tương thích đồng bộ với (ví dụ, có thể giải mã bởi) bộ giải mã thứ nhất được tạo cấu hình để giải mã dữ liệu âm thanh được mã hóa theo giao thức mã hóa thứ nhất (dựa trên kênh) và một hoặc nhiều bộ giải mã thứ hai được tạo cấu hình để giải mã dữ liệu âm thanh được mã hóa theo một hoặc nhiều giao thức mã hóa thứ hai (dựa trên đối tượng). Dòng bit có thể bao gồm cả dữ liệu được mã hóa (dưới dạng các khối dữ liệu) có thể giải mã được bởi bộ giải mã thứ nhất (và được bỏ qua bởi bộ giải mã thứ hai bất kỳ) và dữ liệu được mã hóa (ví dụ, các khối dữ liệu khác) có thể giải mã bởi một hoặc nhiều bộ giải mã thứ hai (và bị bỏ qua bởi bộ giải mã thứ nhất). Âm thanh được giải mã và thông tin được kết hợp (siêu dữ liệu) từ bộ giải mã thứ nhất và một hoặc nhiều bộ giải mã thứ hai sau đó có thể được kết hợp theo cách thức sao cho cả thông tin dựa trên kênh và dựa trên đối tượng được kết xuất đồng thời để tái tạo bản sao của môi trường, kênh, thông tin không gian, và các đối tượng được thể hiện với hệ thống mã hóa lai (ví dụ, trong không gian 3D hoặc môi trường nghe).

Bộ mã hóa-giải mã 108 tạo ra dòng bit chứa thông tin âm thanh được mã hóa và thông tin liên quan đến nhiều tập hợp vị trí kênh (các loa). Theo một phương án, một tập hợp gồm các vị trí kênh được cố định và được dùng cho giao thức mã hóa dựa trên kênh, còn một tập hợp khác gồm các vị trí kênh là thích ứng và được dùng cho giao thức mã hóa dựa trên đối tượng âm thanh, sao cho cấu hình kênh của đối tượng âm thanh có thể thay đổi là hàm thời gian (tùy thuộc vào nơi đối tượng được bố trí trong trường âm thanh). Do đó, hệ thống mã hóa âm thanh lai có thể mang thông tin về hai tập hợp vị trí loa để phát lại, trong đó một tập hợp có thể được cố định và là tập hợp con của tập hợp còn lại. Các thiết bị hỗ trợ thông tin âm thanh mã hóa kế thừa sẽ giải mã và kết xuất thông tin âm thanh từ tập hợp con cố định, còn thiết bị có khả năng hỗ trợ tập hợp lớn hơn có thể giải mã và kết xuất thông tin âm thanh được mã hóa bổ sung mà có thể được gán theo cách thay đổi theo thời gian cho các loa khác nhau từ tập hợp lớn hơn. Ngoài ra, hệ thống không phụ thuộc vào bộ giải mã thứ nhất và một hoặc nhiều bộ giải mã thứ hai có mặt đồng thời bên trong hệ thống và/hoặc thiết bị. Do đó, thiết bị/hệ

thống kế thừa và/hoặc hiện có chỉ chứa bộ giải mã hỗ trợ giao thức thứ nhất sẽ đạt được trường âm thanh thích ứng hoàn toàn cần được kết xuất qua hệ thống tái tạo dựa trên kênh truyền thống. Trong trường hợp này, (các) phần không biết hoặc không được hỗ trợ của giao thức dòng bit-lai (tức là, thông tin âm thanh được biểu diễn bởi giao thức mã hóa thứ hai) sẽ được bỏ qua bởi hệ thống hoặc bộ giải mã thiết bị hỗ trợ giao thức mã hóa lai thứ nhất.

Theo một phương án khác, bộ mã hóa-giải mã 108 được tạo cấu hình để vận hành theo chế độ trong đó hệ thống con mã hóa thứ nhất (hỗ trợ giao thức thứ nhất) bao gồm dạng biểu diễn kết hợp của tất cả thông tin trường âm thanh (kênh và đối tượng) được biểu diễn trong cả hệ thống con mã hóa thứ nhất và một hoặc nhiều hệ thống con mã hóa thứ hai có mặt trong bộ mã hóa lai. Điều này đảm bảo rằng dòng bit lai có khả năng tương thích ngược với bộ giải mã chỉ hỗ trợ giao thức của hệ thống con mã hóa thứ nhất bằng cách cho phép các đối tượng âm thanh (thường được mang trong một hoặc nhiều giao thức mã hóa thứ hai) được biểu diễn và kết xuất trong bộ giải mã chỉ hỗ trợ giao thức thứ nhất.

Theo một phương án khác nữa, bộ mã hóa-giải mã 108 bao gồm hai hoặc nhiều hệ thống con mã hóa, trong đó mỗi hệ thống con trong số các hệ thống con này được tạo cấu hình để mã hóa dữ liệu âm thanh theo giao thức khác, và được tạo cấu hình để kết hợp các đầu ra của các hệ thống con để tạo ra dòng bit có định dạng lai (đồng nhất).

Một trong số các lợi ích của các phương án là khả năng dòng bit âm thanh mã hóa lai được mang qua rất nhiều hệ thống phân phối nội dung, trong đó mỗi trong số các hệ thống phân phối thường chỉ hỗ trợ dữ liệu được mã hóa theo giao thức mã hóa thứ nhất. Điều này loại bỏ nhu cầu sửa đổi/thay đổi hệ thống và/hoặc giao thức mức vận chuyển (transport level protocol) bất kỳ để hỗ trợ riêng hệ thống mã hóa lai.

Hệ thống mã hóa âm thanh thường sử dụng các phần tử dòng bit chuẩn hóa để cho phép truyền dữ liệu (tùy ý) bổ sung trong chính dòng bit đó. Dữ liệu (tùy ý) bổ sung này thường được nhảy (tức là, bỏ qua) trong khi giải mã âm thanh mã hóa bao gồm trong dòng bit, nhưng có thể được dùng cho mục đích khác ngoài giải mã. Các tiêu chuẩn mã hóa âm thanh khác nhau thể hiện các trường dữ liệu bổ sung này sử dụng một thuật ngữ duy nhất. Các phần tử dòng bit thuộc loại chung này có thể bao gồm, nhưng không bị giới hạn ở, dữ liệu bổ sung, các trường bỏ qua, các phần tử dòng dữ liệu, các

phần tử nạp, dữ liệu phụ, và các phần tử dòng phụ. Trừ khi có quy định khác, việc sử dụng cách diễn đạt “dữ liệu bổ sung” trong tài liệu này không chỉ loại hoặc định dạng dữ liệu bổ sung cụ thể, mà nên được hiểu như cách diễn đạt chung bao hàm bất kỳ hoặc tất cả các ví dụ liên quan đến sáng chế.

Kênh dữ liệu được cho phép thông qua các phần tử dòng bit “bổ sung” của giao thức mã hóa thứ nhất trong dòng bit hệ thống mã hóa lai kết hợp có thể mang một hoặc nhiều dòng bit âm thanh (độc lập hoặc phụ thuộc) thứ hai (được mã hóa theo một hoặc nhiều giao thức mã hóa thứ hai). Một hoặc nhiều dòng bit âm thanh thứ hai có thể được chia thành các khối N mẫu và được dồn kênh thành các trường “dữ liệu bổ sung” của dòng bit thứ nhất. Dòng bit thứ nhất có thể giải mã bởi bộ giải mã (bổ sung) thích hợp. Ngoài ra, dữ liệu bổ sung của dòng bit thứ nhất có thể được trích xuất, tái kết hợp thành một hoặc nhiều dòng bit âm thanh thứ hai, được giải mã bởi bộ xử lý hỗ trợ cú pháp của một hoặc nhiều dòng bit thứ hai, và sau đó được kết hợp và kết xuất cùng nhau hoặc độc lập.

Ngoài ra, cũng có thể đảo ngược vai trò của dòng bit thứ nhất và thứ hai, nhờ đó các khối dữ liệu của dòng bit thứ nhất được dồn kênh vào dữ liệu bổ sung của dòng bit thứ hai.

Các phần tử dòng bit kết hợp với giao thức mã hóa thứ hai cũng mang và vận chuyển các đặc điểm thông tin (siêu dữ liệu) của âm thanh nền, các đặc điểm này có thể bao gồm, nhưng không bị giới hạn ở, vị trí nguồn âm thanh mong muốn, vận tốc, và kích thước. Siêu dữ liệu này được dùng trong quá trình giải mã và kết xuất để tái tạo vị trí thích hợp (ví dụ, góc) cho đối tượng âm thanh kết hợp được mang trong dòng bit có thể áp dụng được. Cũng có thể mang siêu dữ liệu được mô tả ở trên, siêu dữ liệu này có thể ứng dụng cho các đối tượng âm thanh được chứa trong một hoặc nhiều dòng bit thứ hai có mặt trong dòng lai, trong các phần tử dòng bit được kết hợp với giao thức mã hóa thứ nhất.

Các phần tử dòng bit liên quan đến một hoặc cả giao thức mã hóa thứ nhất và thứ hai của hệ thống mã hóa lai mang/vận chuyển siêu dữ liệu ngữ cảnh xác định các tham số không gian (tức là, bản chất của chính các thuộc tính tín hiệu) và thông tin khác mô tả dạng bản chất của âm thanh nền dưới dạng các lớp âm thanh cụ thể được mang trong dòng bit âm thanh được mã hóa lai. Siêu dữ liệu như vậy có thể biểu thị, ví dụ, sự có

mặt của lời thoại được nói, âm nhạc, lời thoại trên nền nhạc, tiếng vỗ tay, giọng hát, v.v., và có thể được dùng để thay đổi một cách thích ứng hành vi của các mô-đun xử lý trước hoặc sau được liên kết đứng trước hoặc đứng sau hệ thống mã hóa lai.

Theo một phương án, bộ mã hóa-giải mã 108 được tạo cấu hình để hoạt động với nhóm bit chung hoặc nhóm bit dùng chung trong đó các bit có sẵn để mã hóa được "dùng chung" giữa tất cả hoặc một phần của các hệ thống con mã hóa hỗ trợ một hoặc nhiều giao thức. Bộ mã hóa-giải mã như vậy có thể phân phối các bit có sẵn (từ nhóm bit "dùng chung" chung) giữa các hệ thống con mã hóa để tối ưu hóa toàn bộ chất lượng âm thanh của dòng bit được hợp nhất. Ví dụ, trong khoảng thời gian thứ nhất, bộ mã hóa-giải mã có thể gán nhiều bit có sẵn hơn cho hệ thống con mã hóa thứ nhất, và ít bit có sẵn hơn cho các hệ thống con còn lại, còn trong khoảng thời gian thứ hai, bộ mã hóa-giải mã có thể gán ít bit có sẵn hơn cho hệ thống con mã hóa thứ nhất, và nhiều bit có sẵn hơn cho các hệ thống con còn lại. Quyết định về cách gán các bit giữa các hệ thống con mã hóa có thể phụ thuộc, ví dụ, vào các kết quả phân tích thống kê của nhóm bit được dùng chung, và/hoặc phân tích nội dung âm thanh được mã hóa bởi mỗi hệ thống con. Bộ mã hóa-giải mã có thể phân bổ các bit từ nhóm bit dùng chung theo cách sao cho dòng bit hợp nhất được xây dựng bằng cách ghép các đầu ra của các hệ thống con mã hóa vẫn duy trì độ dài khung/tốc độ truyền bit không đổi trong một khoảng thời gian cụ thể. Trong một số trường hợp, đối với độ dài khung/tốc độ bit của dòng bit hợp nhất cũng có thể thay đổi trong một khoảng thời gian cụ thể.

Theo một phương án khác, bộ mã hóa-giải mã 108 tạo ra dòng bit hợp nhất chứa dữ liệu được mã hóa theo giao thức mã hóa thứ nhất được tạo cấu hình và được truyền như dòng phụ độc lập của dòng dữ liệu mã hóa (trong đó bộ giải mã hỗ trợ giao thức mã hóa thứ nhất sẽ giải mã), và dữ liệu được mã hóa theo giao thức thứ hai được gửi đi như dòng bit phụ độc lập hoặc phụ thuộc của dòng dữ liệu mã hóa (trong đó bộ giải mã hỗ trợ giao thức thứ nhất sẽ bỏ qua). Khái quát hơn, trong một lớp các phương án, bộ mã hóa-giải mã tạo ra dòng bit hợp nhất bao gồm hai hoặc nhiều dòng phụ độc lập hoặc phụ thuộc (trong đó mỗi dòng phụ chứa dữ liệu được mã hóa theo giao thức mã hóa khác nhau hoặc giống nhau).

Theo một phương án khác nữa, bộ mã hóa-giải mã 108 tạo ra dòng bit hợp nhất chứa dữ liệu được mã hóa theo giao thức mã hóa thứ nhất được tạo cấu hình và được

truyền với bộ nhận dạng dòng bit độc nhất (mà bộ giải mã hỗ trợ giao thức mã hóa thứ nhất kết hợp với bộ nhận dạng dòng bit độc nhất sẽ giải mã), và dữ liệu được mã hóa theo giao thức thứ hai được tạo cấu hình và truyền với bộ nhận biết dòng bit độc nhất, mà bộ giải mã hỗ trợ giao thức thứ nhất sẽ bỏ qua. Khái quát hơn, trong lớp của các phương án bộ mã hóa-giải mã tạo ra dòng bit hợp nhất bao gồm hai hoặc nhiều dòng phụ (trong đó mỗi dòng phụ chứa dữ liệu được mã hóa theo giao thức mã hóa khác nhau hoặc giống nhau và mỗi dòng mang bộ nhận biết dòng bit độc nhất). Các phương pháp và hệ thống để tạo ra dòng bit hợp nhất mô tả ở trên có khả năng báo hiệu một cách không rõ ràng (cho bộ giải mã) sự xen kẽ và/hoặc giao thức nào đã được sử dụng trong dòng bit lai (ví dụ, để báo hiệu xem dữ liệu AUX, SKIP, DSE hay phương pháp dòng phụ đã được mô tả trong sáng chế có được sử dụng hay không).

Hệ thống mã hóa lai được tạo cấu hình để hỗ trợ phương pháp khử đan xen/khử dồn kênh và tái đan xen/tái dồn kênh dòng bit hỗ trợ một hoặc nhiều giao thức thứ hai thành dòng bit thứ nhất (hỗ trợ giao thức thứ nhất) ở điểm xử lý bất kỳ tìm thấy nhờ hệ thống phân phối phương tiện. Bộ mã hóa-giải mã lai cũng được tạo cấu hình để có khả năng mã hóa dòng đầu vào âm thanh với các tốc độ mẫu khác nhau thành một dòng bit. Việc này mang lại phương tiện mã hóa và phân phối một cách hiệu quả nguồn âm thanh chứa tín hiệu có các dải thông cố hữu khác nhau. Ví dụ, các rãnh hội thoại thường có dải thông cố hữu thấp hơn so với các rãnh âm nhạc và các rãnh hiệu ứng.

### Kết xuất

Theo một phương án, hệ thống âm thanh thích ứng cho phép nhiều (ví dụ, tối đa 128) rãnh được đóng gói, thường dưới dạng kết hợp của dạng âm thanh beds và đối tượng. Định dạng cơ bản của dữ liệu âm thanh đối với hệ thống âm thanh thích ứng bao gồm một số dòng âm thanh đơn âm độc lập. Mỗi dòng kết hợp với siêu dữ liệu của nó để xác định liệu dòng là dòng dựa trên kênh hay dòng dựa trên đối tượng. Các dòng dựa trên kênh có thông tin kết xuất được mã hóa bởi tên hoặc nhãn kênh; và các dòng dựa trên đối tượng có thông tin vị trí được mã hóa qua các biểu thức toán học còn được mã hóa trong siêu dữ liệu liên kết khác. Các dòng âm thanh độc lập gốc sau đó được đóng gói là một dòng bit tuần tự chứa tất cả dữ liệu âm thanh theo cách được sắp xếp. Cấu hình dữ liệu thích ứng này cho phép âm thanh được kết xuất theo khung tham chiếu lấy đối tượng làm trung tâm, trong đó vị trí kết xuất cuối cùng của âm thanh được dựa trên

môi trường phát lại phù hợp với mục đích của bộ trộn. Do đó, âm thanh có thể được định rõ để bắt đầu từ khung tham chiếu của phòng phát lại (ví dụ, giữa của tường trái), thay vì từ loa hay nhóm loa được đánh dấu cụ thể (ví dụ, âm thanh vòm trái). Siêu dữ liệu vị trí đối tượng chứa thông tin khung tham chiếu lấy đối tượng làm trung tâm thích hợp cần thiết để phát âm thanh một cách chính xác sử dụng các vị trí loa sẵn có trong phòng được thiết lập để phát nội dung âm thanh thích ứng.

Bộ kết xuất lấy dòng bit mã hóa các rãnh âm thanh, và xử lý nội dung theo loại tín hiệu. Dạng âm thanh beds được nạp vào các mảng, các mảng này có thể cần các độ trễ khác nhau và xử lý cân bằng so với các đối tượng riêng. Quy trình này hỗ trợ việc kết xuất các dạng âm thanh bed và các đối tượng này đối với nhiều (tối đa là 64) đầu ra loa. Fig.4 là sơ đồ khối của giai đoạn kết xuất của hệ thống âm thanh thích ứng, theo một phương án. Như được thể hiện trên hệ thống 400 trên Fig.4, số lượng tín hiệu đầu vào, chẳng hạn tối đa là 128 rãnh âm thanh chứa tín hiệu âm thanh thích ứng 402 được cung cấp bởi một số bộ phận thuộc các giai đoạn tạo, biên soạn và đóng gói của hệ thống 300, chẳng hạn như RMU 306 và bộ xử lý 312. Các tín hiệu này chứa các dạng âm thanh bed và đối tượng dựa trên kênh được dùng bởi bộ kết xuất 404. Âm thanh dựa trên kênh (dạng âm thanh bed) và các đối tượng được nhập vào bộ quản lý mức 406 cung cấp quyền điều khiển các mức hoặc biên độ đầu ra của các thành phần âm thanh khác nhau. Một số thành phần âm thanh có thể được xử lý bởi bộ phận hiệu chỉnh mảng 408. Tín hiệu âm thanh thích ứng sau đó được chuyển qua bộ phận xử lý chuỗi B 410, bộ phận này tạo ra số lượng (ví dụ, tối đa 64) tín hiệu đầu ra cấp cho loa. Nói chung, tín hiệu nạp chuỗi B chỉ các tín hiệu được xử lý bởi các bộ khuếch đại công suất, bộ xuyên âm và các loa, trái với nội dung chuỗi A thiết lập rãnh âm thanh trên kho phim.

Theo một phương án, bộ kết xuất 404 chạy thuật toán kết xuất sử dụng một cách thông minh các loa vòm trong rạp với khả năng tốt nhất của chúng. Bằng cách cải thiện công suất tối đa và đáp ứng tần số của các loa vòm này, và giữ cùng mức chuẩn giám sát cho mỗi kênh hoặc loa đầu ra trong rạp, các đối tượng được quét giữa màn hình và loa vòm có thể duy trì mức áp suất âm thanh của chúng và có âm sắc phù hợp hơn, mà quan trọng là không tăng mức áp suất âm thanh tổng thể trong rạp. Mảng các loa vòm được xác định một cách phù hợp thường sẽ có đủ khoảng trống để tái tạo khoảng động tối đa khả dụng trong rãnh âm thanh vòm 7.1 hoặc 5.1 (tức là, 20dB cao hơn mức

chuẩn), tuy nhiên không chắc là một loa vòm sẽ có cùng khoảng trống của loa màn hình lớn đa chiều. Kết quả là, có thể có các trường hợp khi đối tượng được đặt trong trường âm thanh sẽ cần có áp suất âm thanh lớn hơn so với áp suất có thể đạt được nhờ sử dụng một loa vòm. Trong các trường hợp này, bộ kết xuất sẽ dàn trải âm thanh trên số lượng loa thích hợp để đạt được mức áp suất âm thanh yêu cầu. Hệ thống âm thanh thích ứng cải thiện chất lượng và công suất tối đa của các loa vòm mang lại sự cải thiện độ trung thực của việc kết xuất. Sáng chế hỗ trợ quản lý âm trầm của các loa vòm nhờ sử dụng tùy ý các loa siêu trầm ở phía sau mà cho phép mỗi loa vòm đạt được công suất tối đa cải thiện, và có thể sử dụng đồng thời các thùng loa nhỏ hơn. Sáng chế còn cho phép bổ sung các loa vòm bên cạnh gần màn hình hơn so với thực tế áp dụng hiện nay để đảm bảo rằng các đối tượng có thể chuyển tiếp một cách thuận lợi từ màn hình ra loa vòm.

Nhờ sử dụng siêu dữ liệu để xác định thông tin vị trí của các đối tượng âm thanh cùng với một số quy trình kết xuất, hệ thống 400 cung cấp phương pháp tổng thể, linh hoạt cho các nhà sáng tạo nội dung để loại bỏ các hạn chế của hệ thống hiện tại. Như đã đề cập trên đây, các hệ thống hiện thời tạo ra và phân phối âm thanh cố định theo vị trí loa cụ thể với sự hiểu biết giới hạn về loại nội dung được truyền dưới dạng bản chất âm thanh (một phần của âm thanh được phát lại). Hệ thống âm thanh thích ứng 100 cung cấp phương pháp lai mới bao gồm sự lựa chọn cho cả âm thanh định rõ vị trí loa (kênh trái, kênh phải, v.v.) và các phần tử âm thanh định hướng đối tượng được tổng quát hóa thông tin không gian, thông tin này có thể bao gồm, nhưng không bị giới hạn ở vị trí, kích thước và vận tốc. Phương pháp lai này là phương pháp cân bằng để kết xuất chính xác (được tạo ra bởi các vị trí loa cố định) và linh hoạt (các đối tượng âm thanh tổng quát hóa). Hệ thống còn cung cấp thông tin hữu dụng bổ sung về nội dung âm thanh được ghép với bản chất âm thanh bởi nhà sáng tạo nội dung tại thời điểm tạo ra nội dung. Thông tin này là thông tin có hiệu lực, chi tiết về các thuộc tính của âm thanh mà có thể được dùng theo cách rất hiệu quả trong suốt quy trình kết xuất. Các thuộc tính này có thể bao gồm, nhưng không bị giới hạn ở, loại nội dung (lời thoại, âm nhạc, hiệu ứng, hiệu ứng Foley, nhạc nền/tiếng ồn xung quanh, v.v.), các thuộc tính không gian (vị trí 3D, kích thước 3D, vận tốc), và thông tin kết xuất (bắt giữ vị trí loa, trọng lượng kênh, độ khuếch đại, thông tin quản lý âm trầm, v.v.).

Hệ thống âm thanh thích ứng được mô tả ở đây cung cấp thông tin mạnh mẽ có thể được dùng để kết xuất bằng rất nhiều điểm cuối khác nhau. Trong nhiều trường hợp kỹ thuật kết xuất tối ưu được áp dụng phụ thuộc nhiều vào thiết bị điểm cuối. Ví dụ, hệ thống rạp hát tại nhà và các loa dạng thanh có thể có 2, 3, 5, 7 hoặc thậm chí 9 loa riêng. Nhiều kiểu hệ thống khác, chẳng hạn như ti vi, máy vi tính, và các kệ âm nhạc chỉ có hai loa, và gần như tất cả các thiết bị thường dùng chung có đầu ra cắm tai nghe âm thanh nổi (máy tính cá nhân, máy tính xách tay, máy tính bảng, điện thoại di động, máy nghe nhạc, v.v.). Tuy nhiên, đối với âm thanh truyền thống ngày nay (các kênh đơn sắc, lập thể, 5.1, 7.1), các thiết bị điểm cuối thường cần đưa ra các quyết định và thỏa hiệp đơn giản để kết xuất và tái tạo âm thanh mà hiện được phân phối dưới dạng kênh/loa riêng. Ngoài ra có ít hoặc không có thông tin được truyền về nội dung thực tế được phân bổ (lời thoại, âm nhạc, tiếng ồn xung quanh, v.v.) và có ít hoặc không có thông tin về ý định của nhà sáng tạo nội dung để tái tạo âm thanh. Tuy nhiên, hệ thống âm thanh thích ứng 100 cung cấp thông tin này và, có thể truy cập vào các đối tượng âm thanh, các đối tượng này có thể được dùng để tạo ra trải nghiệm người dùng thế hệ tiếp theo hấp dẫn.

Hệ thống 100 cho phép nhà sáng tạo nội dung nhúng ý định không gian của hỗn hợp trong dòng bit bằng cách sử dụng siêu dữ liệu như vị trí, kích thước, vận tốc, và v.v, thông qua định dạng truyền âm thanh thích ứng và siêu dữ liệu duy nhất và mạnh mẽ. Điều này cho phép rất linh hoạt trong việc tái tạo âm thanh trong không gian. Từ quan điểm kết xuất không gian, âm thanh thích ứng cho phép điều chỉnh hỗn hợp với vị trí chính xác của các loa trong phòng riêng để tránh sự biến dạng không gian xảy ra khi dạng hình học của hệ thống phát lại không đồng nhất với hệ thống biên soạn. Trong các hệ thống tái tạo âm thanh hiện tại trong đó chỉ âm thanh dành cho kênh loa được gửi đi, mục đích của nhà sáng tạo nội dung là không xác định. Hệ thống 100 sử dụng siêu dữ liệu được truyền qua đường ống tạo và phân phối. Hệ thống tái tạo nhận biết âm thanh thích ứng có thể sử dụng thông tin siêu dữ liệu này để tái tạo nội dung theo cách thức trùng khớp với mục đích ban đầu của nhà sáng tạo nội dung. Tương tự, hỗn hợp này có thể được làm thích ứng với cấu hình phần cứng chính xác của hệ thống tái tạo. Hiện nay, có nhiều cấu hình và loại loa có thể có khác nhau trong thiết bị kết xuất chẳng hạn như ti vi, rạp hát tại nhà, các loa dạng thanh, các ổ cắm máy chơi nhạc di động, v.v. Khi các hệ thống này gửi thông tin âm thanh định rõ kênh hôm nay (tức là, âm thanh kênh



phải và kênh trái hoặc âm thanh đa kênh), hệ thống phải xử lý âm thanh để phù hợp với các khả năng của thiết bị kết xuất. Ví dụ như âm thanh lập thể tiêu chuẩn được gửi cho loa dạng thanh có nhiều hơn hai loa. Trong quy trình tái tạo âm thanh hiện tại mà chỉ có âm thanh của kênh loa được gửi đi, mục đích của nhà sáng tạo nội dung là không xác định. Thông qua việc sử dụng siêu dữ liệu được truyền qua đường ống tạo ra và phân phối, hệ thống tái tạo nhận biết âm thanh thích ứng có thể sử dụng thông tin này để tái tạo nội dung theo cách thức trùng khớp với ý định ban đầu của nhà sáng tạo nội dung. Ví dụ, một số loa dạng thanh có các loa hướng ra bên để tạo ra cảm giác bao quanh. Với âm thanh thích ứng, thông tin không gian và loại nội dung (như các hiệu ứng xung quanh) có thể được dùng bởi loa dạng thanh để chỉ gửi âm thanh thích hợp cho các loa hướng ra bên này.

Hệ thống âm thanh thích ứng cho phép nội suy không giới hạn các loa trong hệ thống trên tất cả các kích thước trước/sau, trái/phải, trên/dưới, gần/xa. Trong các hệ thống tái tạo âm thanh hiện tại, không có thông tin nào chỉ cách thức xử lý âm thanh khi có mong muốn định vị âm thanh sao cho người nghe nhận thức được là ở giữa hai loa. Hiện nay, với âm thanh chỉ được gán cho một loa cụ thể, hệ số lượng tử hóa không gian được đưa vào. Với âm thanh thích ứng, sự định vị không gian của âm thanh có thể được biết một cách chính xác và do đó được tái tạo sao cho phù hợp trên hệ thống tái tạo âm thanh.

Đối với việc kết xuất tai nghe, ý định của người tạo nội dung được hiện thực hóa bằng cách so khớp các Hàm Truyền Liên quan đến Đầu (Head Related Transfer Functions - HRTF) với vị trí không gian. Khi âm thanh được tái tạo qua tai nghe, quy trình ảo hóa không gian có thể đạt được bằng cách áp dụng hàm HRTF, hàm này xử lý âm thanh, bổ sung tín hiệu về cảm giác để tạo ra sự cảm nhận về âm thanh được phát trong không gian 3D và không phải qua tai nghe. Độ chính xác của sự tái tạo trong không gian phụ thuộc vào sự lựa chọn hàm HRTF thích hợp, hàm này có thể thay đổi dựa trên một số yếu tố bao gồm vị trí không gian. Việc sử dụng thông tin không gian được cung cấp bởi hệ thống Âm thanh Thích ứng có thể dẫn đến việc lựa chọn một hàm HRTF hoặc nhiều hàm HRTF thay đổi liên tục để cải thiện đáng kể trải nghiệm tái tạo.

Thông tin không gian được truyền bởi hệ thống âm thanh thích ứng có thể không những được dùng bởi nhà sáng tạo nội dung để tạo ra trải nghiệm giải trí thú vị (phim,

truyền hình, âm nhạc, v.v.), mà thông tin không gian còn có thể biểu thị nơi mà người nghe được định vị tương quan với các đối tượng vật lý chẳng hạn như các tòa nhà hoặc các điểm địa lý được chú ý. Điều này sẽ cho phép người dùng tương tác với trải nghiệm âm thanh ảo hóa có liên quan đến thế giới thực, tức là, tăng thực tế.

Các phương án cũng cho phép tăng trộn không gian, bằng cách thực hiện tăng trộn nâng cao nhờ chỉ đọc siêu dữ liệu nếu các đối tượng dữ liệu âm thanh không có sẵn. Việc biết vị trí của tất cả đối tượng và loại của chúng cho phép bộ tăng trộn phân biệt tốt hơn các phần tử trong các rãnh dựa trên kênh. Các thuật toán tăng trộn hiện có phải suy ra thông tin chẳng hạn như loại nội dung âm thanh (lời nói, âm nhạc, các hiệu ứng xung quanh) cũng như vị trí của các phần tử khác nhau trong dòng âm thanh để tạo ra sự tăng trộn chất lượng cao với thành phần lạ nhỏ nhất hoặc không nghe thấy được. Nhiều lần thông tin được suy ra có thể không chính xác hoặc không thích hợp. Với âm thanh thích ứng, thông tin bổ sung có sẵn từ siêu dữ liệu liên quan đến, ví dụ, loại nội dung âm thanh, vị trí không gian, vận tốc, kích thước đối tượng âm thanh, v.v., có thể được dùng bởi thuật toán tăng trộn để tạo ra kết quả tái tạo chất lượng cao. Hệ thống này cũng so khớp về mặt không gian âm thanh với dữ liệu video bằng cách định vị chính xác đối tượng âm thanh của màn hình với các phần tử trực quan. Trong trường hợp này, có thể trải nghiệm tái tạo âm thanh/video hấp dẫn, đặc biệt với kích thước màn hình rộng hơn, nếu vị trí không gian được tái tạo của các phần tử âm thanh khớp với các phần tử hình ảnh trên màn hình. Một ví dụ là lời thoại trong phim hoặc chương trình truyền hình trùng khớp về mặt không gian với người hoặc nhân vật đang nói trên màn hình. Với âm thanh dựa trên kênh loa thông thường, không có phương pháp dễ dàng để xác định vị trí không gian của lời thoại để khớp với vị trí của người hoặc nhân vật trên màn hình. Đối với thông tin âm thanh có sẵn có âm thanh thích ứng, có thể đạt được sự điều chỉnh âm thanh/hình ảnh như vậy. Sự điều chỉnh vị trí trực quan và không gian âm thanh cũng có thể được dùng cho các đối tượng không phải nhân vật/lời thoại như ô tô, xe tải, hoạt hình, v.v.

Việc xử lý mặt nạ không gian được hệ thống 100 tạo điều kiện thuận lợi, vì hiểu biết về ý muốn không gian của hỗn hợp thông qua siêu dữ liệu âm thanh thích ứng có nghĩa là hỗn hợp có thể được làm thích ứng với mọi cấu hình loa. Tuy nhiên, có nguy cơ giảm trộn các đối tượng ở vị trí giống hoặc gần giống do các hạn chế của hệ thống

phát lại. Ví dụ, đối tượng cần được quét ở phía sau bên trái có thể được giảm trộn so với phía trước bên trái nếu không có các kênh vòm, nhưng nếu phần tử to hơn xuất hiện đồng thời ở phía trước bên trái, thì đối tượng được giảm trộn sẽ bị che và biến mất khỏi hỗn hợp. Nhờ sử dụng siêu dữ liệu âm thanh thích ứng, việc chấn không gian có thể được bộ kết xuất biết trước, và các tham số không gian và/hoặc giảm trộn âm lượng của mỗi đối tượng có thể được điều chỉnh vì vậy tất cả các phần tử âm thanh của hỗn hợp duy trì khả năng nhận thấy được như trong hỗn hợp gốc. Do bộ kết xuất hiểu được quan hệ không gian giữa hỗn hợp và hệ thống phát lại, nên nó có khả năng "bắt" đối tượng vào các loa gần nhất thay vì tạo ra hình ảnh ảo giữa hai hoặc nhiều loa. Trong khi việc này có thể làm biến dạng nhẹ dạng biểu diễn không gian của hỗn hợp, nó cũng cho phép bộ kết xuất tránh được hình ảnh ảo không mong muốn. Ví dụ, nếu vị trí góc của loa trái của giai đoạn trộn không tương ứng với vị trí góc của loa trái của hệ thống phát lại, thì việc sử dụng chức năng bắt loa gần nhất có thể tránh được việc hệ thống phát lại tái tạo hình ảnh ảo không đối của kênh trái của giai đoạn trộn.

Đối với xử lý nội dung, hệ thống âm thanh thích ứng 100 cho phép nhà sáng tạo nội dung tạo ra các đối tượng âm thanh riêng và bổ sung thông tin về nội dung mà có thể được truyền cho hệ thống tái tạo. Điều này cho phép tính linh hoạt cao trong quá trình xử lý âm thanh trước khi tái tạo. Từ quan điểm xử lý và kết xuất nội dung, hệ thống âm thanh thích ứng cho phép quy trình xử lý được làm thích ứng với loại đối tượng. Ví dụ, việc nâng cao thoại chỉ có thể được áp dụng cho đối tượng thoại. Việc nâng cao thoại đề cập đến phương pháp xử lý âm thanh chứa thoại sao cho khả năng nghe và/hoặc độ nghe rõ của thoại được tăng lên và hoặc được cải thiện. Trong nhiều trường hợp, việc xử lý âm thanh mà được áp dụng cho thoại là không thích hợp với nội dung âm thanh không phải lời thoại (tức là, âm nhạc, các hiệu ứng xung quanh, v.v.) và có thể dẫn đến các thành phần lạ nghe thấy khó chịu. Đối với âm thanh thích ứng, một đối tượng âm thanh chỉ có thể chứa lời thoại trong một phần nội dung và nó có thể được đánh dấu sao cho phù hợp để giải pháp kết xuất có thể áp dụng một cách chọn lọc việc nâng cao thoại chỉ cho nội dung lời thoại. Ngoài ra, nếu đối tượng âm thanh chỉ là lời thoại (và không phải là hỗn hợp của lời thoại và nội dung khác thường xảy ra), thì quá trình xử lý nâng cao lời thoại có thể xử lý hoàn toàn lời thoại (do đó hạn chế việc xử lý bất kỳ được thực hiện trên nội dung khác bất kỳ). Tương tự, việc quản lý âm trầm (lọc, làm suy giảm, khuếch đại) có thể được nhắm vào các đối tượng cụ thể dựa trên loại của

chúng. Việc quản lý âm trầm đề cập đến việc tách có lựa chọn và chỉ xử lý các tần số âm trầm (hoặc thấp hơn) trong một phần nội dung cụ thể. Với các hệ thống âm thanh và cơ chế phân phối hiện tại, đây là một quy trình “mù” được áp dụng cho tất cả âm thanh. Với âm thanh thích ứng, các đối tượng âm thanh cụ thể mà việc quản lý âm trầm phù hợp có thể được siêu dữ liệu nhận ra, và quá trình kết xuất có thể được áp dụng một cách thích hợp.

Hệ thống âm thanh thích ứng 100 còn cung cấp khả năng nén dải động dựa trên đối tượng và tăng trộn có lựa chọn. Các rãnh âm thanh truyền thống có khoảng thời gian giống với chính nội dung của nó, còn đối tượng âm thanh có thể xuất hiện chỉ trong một khoảng thời gian giới hạn trong nội dung. Siêu dữ liệu kết hợp với đối tượng có thể chứa thông tin về biên độ tín hiệu đỉnh và trung bình của nó, cũng như thời gian bắt đầu hoặc can thiệp của nó (đặc biệt là đối với vật liệu chuyển tiếp). Thông tin này sẽ cho phép bộ nén làm thích ứng tốt hơn các hằng số thời gian và nén của nó (can thiệp, giải phóng, v.v.) để phù hợp hơn với nội dung. Đối với việc tăng trộn có lựa chọn, các nhà sáng tạo nội dung có thể lựa chọn để biểu thị trong dòng bit âm thanh thích ứng xem đối tượng có nên được tăng trộn hay không. Thông tin này cho phép Bộ kết xuất Âm thanh Thích ứng và bộ tăng trộn phân biệt phần tử âm thanh nào có thể được tăng trộn một cách an toàn, trong khi vẫn tôn trọng ý muốn của người sáng tạo nội dung.

Các phương án này cũng cho phép hệ thống âm thanh thích ứng lựa chọn thuật toán kết xuất ưu tiên từ nhiều thuật toán kết xuất và/hoặc định dạng âm thanh vòm sẵn có. Các ví dụ về thuật toán kết xuất sẵn có bao gồm: âm thanh nổi, lưỡng cực lập thể, định dạng âm thanh ambisonic, tổng hợp trường sóng (Wave Field Synthesis - WFS), quét đa kênh, các dạng âm stem thô với siêu dữ liệu vị trí. Các thuật toán khác bao gồm cân bằng kép, và quét biên độ dựa trên vector.

Định dạng phân phối âm thanh nổi sử dụng dạng biểu diễn hai kênh của trường âm thanh theo tín hiệu có ở tai trái và tai phải. Thông tin âm thanh nổi có thể được tạo ra bằng cách ghi âm trong tai hoặc được tổng hợp bằng cách sử dụng các mẫu HRTF. Việc phát lại dạng biểu diễn âm thanh nổi này thường được thực hiện qua tai nghe, hoặc bằng cách sử dụng kỹ thuật triệt tiêu xuyên âm. Việc phát lại qua thiết lập loa tùy ý sẽ đòi hỏi phân tích tín hiệu để xác định trường âm thanh và/hoặc (các) nguồn tín hiệu liên quan.

Phương pháp kết xuất lưỡng cực lập thể là quy trình triệt tiêu xuyên âm để làm cho các tín hiệu âm thanh nổi có thể phát được trên các loa lập thể (ví dụ, cách trung tâm + và - 10 độ).

Định dạng âm thanh ambisonic (định dạng phân phối và phương pháp kết xuất) được mã hóa dưới dạng bốn kênh gọi là định dạng-B. Kênh thứ nhất, W, là tín hiệu áp suất vô hướng; kênh thứ hai, X, là gradient áp suất định hướng chứa thông tin trước và sau; kênh thứ ba, Y, chứa thông tin trái và phải, và kênh Z chứa thông tin trên và dưới. Các kênh này định nghĩa mẫu lệnh thứ nhất của trường âm thanh hoàn chỉnh tại một điểm. Âm thanh ambisonic sử dụng tất cả các loa sẵn có để tái tạo trường âm thanh lấy mẫu (hoặc tổng hợp) trong mảng loa sao cho khi một số loa được đầy, các loa khác kéo.

WFS là phương pháp kết xuất để tái tạo âm thanh, dựa trên cấu trúc chính xác của trường sóng mong muốn bởi các nguồn thứ hai. WFS dựa trên nguyên lý của Huygen, và được triển khai dưới dạng các mảng loa (hàng chục hoặc hàng trăm) bao quanh không gian nghe và hoạt động theo kiểu được định pha, phối hợp để tái tạo mỗi sóng âm thanh riêng.

Kỹ thuật quét đa kênh là định dạng phân phối và/hoặc phương pháp kết xuất, và có thể được gọi là âm thanh dựa trên kênh. Trong trường hợp này, âm thanh được thể hiện là số lượng nguồn rời rạc được phát lại qua số lượng loa bằng nhau ở các góc xác định trước so với người nghe. Nhà sáng tạo/bộ trộn nội dung có thể tạo ra các hình ảnh ảo bằng cách quét các tín hiệu giữa các kênh liên kế để cung cấp các tín hiệu dẫn hướng; phản xạ sớm, hồi âm, v.v., có thể được trộn thành nhiều kênh để tạo ra các tín hiệu dẫn hướng và môi trường.

Các dạng âm thanh stem thô với siêu dữ liệu vị trí là định dạng phân phối, và cũng có thể được gọi là âm thanh dựa trên đối tượng. Trong định dạng này, nguồn âm thanh “micro đóng” riêng được biểu diễn cùng với siêu dữ liệu vị trí và môi trường. Nguồn ảo được kết xuất dựa trên siêu dữ liệu và thiết bị phát lại và môi trường nghe.

Định dạng âm thanh thích ứng là lai của định dạng quét đa kênh và định dạng âm thanh stem thô. Phương pháp kết xuất theo phương án này là quét đa kênh. Đối với các kênh âm thanh, bước kết xuất (quét) xảy ra ở thời điểm biên soạn, còn đối với các đối tượng thì bước kết xuất (quét) xảy ra ở thời điểm phát lại.

### Định dạng truyền siêu dữ liệu và âm thanh thích ứng

Như đã nêu ở trên, siêu dữ liệu được tạo ra trong giai đoạn tạo để mã hóa thông tin vị trí nhất định cho các đối tượng âm thanh và đi kèm với chương trình âm thanh để hỗ trợ kết xuất chương trình âm thanh, và cụ thể là, để mô tả chương trình âm thanh theo cách cho phép kết xuất chương trình âm thanh trên nhiều thiết bị phát lại và môi trường phát lại. Siêu dữ liệu được tạo ra cho một chương trình cho trước và các bộ biên soạn và bộ trộn để tạo ra, thu thập, biên soạn và điều khiển âm thanh trong giai đoạn sau sản xuất. Đặc điểm quan trọng của định dạng âm thanh thích ứng là khả năng kiểm soát cách thức âm thanh sẽ dịch sang hệ thống và môi trường phát lại khác với môi trường trộn. Cụ thể, rạp phim cho trước có thể có ít khả năng hơn so với môi trường trộn.

Bộ kết xuất âm thanh thích ứng được thiết kế để sử dụng tốt nhất thiết bị sẵn có để tái tạo ý định của bộ trộn. Ngoài ra, các công cụ biên soạn âm thanh thích ứng cho phép bộ trộn xem trước và điều chỉnh cách thức hỗn hợp sẽ được kết xuất trên nhiều cấu hình phát lại. Tất cả các giá trị siêu dữ liệu có thể được quy định trong môi trường phát lại và cấu hình loa. Ví dụ, mức trộn khác nhau đối với phần tử âm thanh cho trước có thể được xác định dựa trên cấu hình hoặc chế độ phát lại. Theo một phương án, danh sách các chế độ phát lại được quy định có thể mở rộng và bao gồm các chế độ sau: (1) chỉ phát lại dựa trên kênh: 5.1, 7.1, 7.1 (chiều cao), 9.1; và (2) phát lại qua loa rời rạc: 3D, 2D (không có chiều cao).

Theo một phương án, siêu dữ liệu kiểm soát hoặc đặc tả các khía cạnh khác nhau của nội dung âm thanh thích ứng và được sắp xếp dựa trên các loại khác nhau bao gồm: siêu dữ liệu chương trình, siêu dữ liệu âm thanh, và siêu dữ liệu kết xuất (cho kênh và đối tượng). Mỗi loại siêu dữ liệu bao gồm một hoặc nhiều mục siêu dữ liệu cung cấp các giá trị cho các đặc điểm được tham chiếu bởi định danh (identifier - ID). Fig.5 là bảng liệt kê các loại siêu dữ liệu và các phần tử siêu dữ liệu liên kết cho hệ thống âm thanh thích ứng, theo một phương án.

Như được thể hiện trong bảng 500 trên Fig.5, loại siêu dữ liệu đầu tiên là siêu dữ liệu chương trình, bao gồm các phần tử siêu dữ liệu xác định tốc độ khung, đếm rãnh, mô tả kênh có thể mở rộng, và mô tả giai đoạn trộn. Phần tử siêu dữ liệu tốc độ khung xác định tốc độ của các khung nội dung âm thanh theo các đơn vị khung trên giây (fps -

frames per second). Định dạng âm thanh thô không cần bao gồm việc tạo khung âm thanh hoặc siêu dữ liệu do âm thanh được cung cấp như các rãnh đầy đủ (khoảng thời gian của cuộn phim hoặc toàn bộ đặc điểm) thay vì các đoạn âm thanh (khoảng thời gian của đối tượng). Định dạng thô cần phải mang tất cả thông tin cần thiết để cho phép bộ mã hóa âm thanh thích ứng tạo khung âm thanh và siêu dữ liệu, bao gồm tốc độ khung thực tế. Bảng 1 thể hiện ID, các giá trị làm ví dụ và mô tả của phần tử siêu dữ liệu tốc độ khung.

BẢNG 1

| ID        | Giá trị                                                           | Mô tả 2                                                                                                                                                                                                          |
|-----------|-------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| FrameRate | 24, 25, 30, 48, 50, 60, 96, 100, 120, có thể mở rộng (khung/giây) | Biểu thị tốc độ khung dự tính cho toàn bộ chương trình. Trường này sẽ tạo ra sự mã hóa hiệu quả các tốc độ chung, cũng như khả năng mở rộng đến trường điểm di động có thể thay đổi tỷ lệ với độ phân giải 0,01. |

Phần tử siêu dữ liệu đếm rãnh biểu thị số lượng rãnh âm thanh trong khung. Bộ giải mã/bộ xử lý âm thanh thích ứng làm ví dụ có thể hỗ trợ tối đa đồng thời 128 rãnh âm thanh, trong khi định dạng âm thanh thích ứng sẽ hỗ trợ số lượng rãnh âm thanh bất kỳ. Bảng 2 thể hiện ID, các giá trị làm ví dụ và mô tả của phần tử siêu dữ liệu đếm rãnh.

BẢNG 2

| ID      | Giá trị                               | Mô tả 2                                      |
|---------|---------------------------------------|----------------------------------------------|
| nTracks | Số nguyên dương, khoảng mở rộng được. | Biểu thị số lượng rãnh âm thanh trong khung. |

Âm thanh dựa trên kênh có thể được gán cho các kênh không tham chiếu và phần tử siêu dữ liệu mô tả kênh mở rộng được cho phép các hỗn hợp sử dụng các vị trí kênh mới. Với mỗi kênh mở rộng, siêu dữ liệu sau đây sẽ được cung cấp như được thể hiện trên Bảng 3:

BẢNG 3

| ID              | Giá trị         | Mô tả 2 |
|-----------------|-----------------|---------|
| ExtChanPosition | Tọa độ x, y, z. | Vị trí  |

|                 |                 |         |
|-----------------|-----------------|---------|
| ExtChanPosition | Tọa độ x, y, z. | Độ rộng |
|-----------------|-----------------|---------|

Phần tử siêu dữ liệu mô tả giai đoạn trộn xác định tần số mà một loa cụ thể sinh ra tại đó một nửa công suất của dải thông. Bảng 4 thể hiện ID, các giá trị làm ví dụ và mô tả của phần tử siêu dữ liệu mô tả giai đoạn trộn, trong đó LF = Tần số thấp; HF = Tần số cao; điểm 3dB = mép dải thông loa.

BẢNG 4

| ID             | Giá trị                                                                                                                                                                             | Mô tả                                                                                                                                                                                                                                                                                                                                                                                                            |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| nMixSpeakers   | Số nguyên dương                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                  |
| MixSpeakerPos  | Tọa độ x, y, z của mỗi loa.                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                  |
| MixSpeakersTyp | {FR, LLF, Sub}, của mỗi loa.                                                                                                                                                        | Toàn dải, Đáp ứng LF bị giới hạn, Loa siêu trầm                                                                                                                                                                                                                                                                                                                                                                  |
| MixSpeaker3db  | Số nguyên dương (Hz), của mỗi loa.                                                                                                                                                  | Điểm LF 3dB cho các loa FR và LLF, điểm HF 3dB cho các kiểu loa Sub. Có thể sử dụng để so khớp khả năng tái tạo phổ của thiết bị giai đoạn trộn.                                                                                                                                                                                                                                                                 |
| MixChannel     | {L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, không, khác}, của mỗi loa                                                                                                           | Loa -> ánh xạ kênh.<br>Sử dụng “không” cho các loa không được liên kết                                                                                                                                                                                                                                                                                                                                           |
| MixSpeakerSub  | Danh sách cặp (độ khuếch đại, số loa). Độ khuếch đại là giá trị thực: $0 \leq \text{độ khuếch đại} \leq 1,0$ .<br>Số loa là số nguyên. $0 < \text{số loa} < n \text{ loa trộn} - 1$ | Loa -> ánh xạ loa siêu trầm. Được dùng để biểu thị loa siêu trầm đích để quản lý âm trầm của mỗi loa. Mỗi loa có thể được quản lý âm trầm nhiều hơn loa siêu trầm. Độ khuếch đại biểu thị phân tín hiệu âm trầm mà nên đi đến mỗi loa siêu trầm. Độ khuếch đại=0 biểu thị cuối danh sách, và số loa không đi theo. Nếu loa không được quản lý âm trầm, thì giá trị độ khuếch đại thứ nhất được thiết lập bằng 0. |
| MixPos         | Tọa độ x, y, z cho vị trí trộn                                                                                                                                                      | Vị trí trộn danh định                                                                                                                                                                                                                                                                                                                                                                                            |
| MixRoomDim     | Tọa độ x, y, z cho các chiều không gian (mét)                                                                                                                                       | Các chiều trong giai đoạn trộn danh định                                                                                                                                                                                                                                                                                                                                                                         |
| MixRoomRT60    | Giá trị thực <20                                                                                                                                                                    | Giai đoạn trộn danh định RT60                                                                                                                                                                                                                                                                                                                                                                                    |
| MixScreenDim   | Tọa độ x, y, z cho các chiều màn hình (mét)                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                  |



|              |                                             |  |
|--------------|---------------------------------------------|--|
| MixScreenPos | Tọa độ x, y, z cho trung tâm màn hình (mét) |  |
|--------------|---------------------------------------------|--|

Như được thể hiện trên Fig.5, loại siêu dữ liệu thứ hai là siêu dữ liệu âm thanh. Mỗi phần tử âm thanh dựa trên kênh hoặc dựa trên đối tượng bao gồm dạng bản chất âm thanh và siêu dữ liệu. Dạng bản chất âm thanh là dòng âm thanh đơn âm được mang trong một trong số nhiều rãnh âm thanh. Siêu dữ liệu kết hợp mô tả cách dạng bản chất âm thanh được lưu trữ (siêu dữ liệu âm thanh, ví dụ, tốc độ mẫu) hoặc cách nó được kết xuất (siêu dữ liệu kết xuất, ví dụ, vị trí nguồn âm thanh mong muốn). Nói chung, các rãnh âm thanh kế tiếp nhau qua khoảng thời gian của chương trình âm thanh. Người biên soạn chương trình hoặc bộ trộn chịu trách nhiệm gán các phần tử âm thanh cho các rãnh. Việc sử dụng rãnh được dự kiến là thừa thớt, tức là số rãnh sử dụng đồng thời trung bình chỉ có thể nằm trong khoảng từ 16 đến 32. Trong phương án thực hiện điển hình, âm thanh sẽ được truyền một cách hiệu quả bằng cách sử dụng bộ mã hóa không tổn thất. Tuy nhiên, có thể sử dụng các phương án thực hiện khác, ví dụ truyền dữ liệu âm thanh chưa mã hóa hoặc dữ liệu âm thanh được mã hóa có tổn thất. Trong phương án thực hiện điển hình, định dạng này chứa tối đa 128 rãnh âm thanh trong đó mỗi rãnh có một tốc độ mẫu và một hệ thống mã hóa. Mỗi rãnh kéo dài khoảng thời gian của đặc điểm (không hỗ trợ cuộn rõ ràng). Việc ánh xạ các đối tượng đến các rãnh (dồn kênh phân thời) là trách nhiệm của nhà sáng tạo nội dung (bộ trộn).

Như được thể hiện trên Fig.3, siêu dữ liệu âm thanh bao gồm các phần tử của tốc độ mẫu, độ sâu bit, và các hệ thống mã hóa. Bảng 5 thể hiện ID, các giá trị làm ví dụ và sự mô tả của phần tử siêu dữ liệu tốc độ mẫu.

BẢNG 5

| ID         | Giá trị                                                           | Mô tả                                                                                                                                                      |
|------------|-------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SampleRate | 16, 24, 32, 44.1, 48, 88.2 96, và có thể mở rộng (x1000 mẫu/giây) | Trường SampleRate sẽ tạo ra sự mã hóa hiệu quả tốc độ chung, cũng như khả năng mở rộng đến trường điểm di động có thể thay đổi tỷ lệ với độ phân giải 0,01 |

Bảng 6 thể hiện ID, các giá trị làm ví dụ và sự mô tả của phần tử siêu dữ liệu độ sâu bit (cho PCM và nén không tổn thất).

BẢNG 6

| ID       | Giá trị                      | Mô tả                                                                                                                          |
|----------|------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| BitDepth | Số nguyên dương tối đa là 32 | Biểu thị độ sâu bit mẫu. Các mẫu sẽ được căn chỉnh trái nếu độ sâu bit nhỏ hơn thành phần chứa (tức là, các LSB điền số không) |

Bảng 7 thể hiện ID, các giá trị làm ví dụ và sự mô tả của thành phần siêu dữ liệu hệ thống mã hóa.

BẢNG 7

| ID                | Giá trị                                                                  | Mô tả                                                                                                                                                        |
|-------------------|--------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bộ mã hóa-giải mã | PCM, không tổn thất, có thể mở rộng                                      | Biểu thị định dạng âm thanh. Mỗi rãnh âm thanh có thể được gán loại mã hóa được hỗ trợ bất kỳ                                                                |
| GIAI ĐOẠN 1       |                                                                          | GIAI ĐOẠN 2                                                                                                                                                  |
| GroupNumber       | Số nguyên dương                                                          | Thông tin nhóm đối tượng. Áp dụng cho các đối tượng âm thanh và đối tượng kênh, ví dụ, để biểu thị các dạng âm thanh stem.                                   |
| AudioTyp          | {lời thoại, âm nhạc, hiệu ứng, âm nhạc & hiệu ứng, không xác định, khác} | Loại âm thanh. Danh sách sẽ có thể mở rộng và bao gồm các loại sau: Không xác định, lời thoại, âm nhạc, Hiệu ứng, hiệu ứng Foley, tiếng ồn xung quanh, Khác. |
| AudioTypTxt       | Mô tả bằng văn bản tự do                                                 |                                                                                                                                                              |

Như được thể hiện trên Fig.5, loại siêu dữ liệu thứ ba là siêu dữ liệu kết xuất. Siêu dữ liệu kết xuất xác định các giá trị giúp bộ kết xuất so khớp gần nhất có thể ý định bộ trộn gốc bất kể môi trường phát lại. Tập hợp các phần tử siêu dữ liệu là khác đối với âm thanh dựa trên kênh và âm thanh dựa trên đối tượng. Trường siêu dữ liệu kết xuất thứ nhất lựa chọn giữa hai loại âm thanh - dựa trên kênh hoặc dựa trên đối tượng, như được thể hiện trên Bảng 8.

BẢNG 8

| ID        | Giá trị           | GIAI ĐOẠN 2                                                                                                                          |
|-----------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| ChanOrObj | {Kênh, Đối tượng} | Đối với mỗi phần tử âm thanh, biểu thị liệu xem nó có được mô tả nhờ sử dụng siêu dữ liệu đối tượng hoặc siêu dữ liệu kênh hay không |

Siêu dữ liệu kết xuất cho âm thanh dựa trên kênh bao gồm phần tử siêu dữ liệu vị trí xác định vị trí nguồn âm thanh như một hoặc nhiều vị trí loa. Bảng 9 thể hiện ID và các giá trị cho phần tử siêu dữ liệu vị trí đối với trường hợp dựa trên kênh.

BẢNG 9

| ID         | Giá trị                                                                 | Mô tả                                                                                                                                                                                               |
|------------|-------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ChannelPos | {L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc, Crs, Cts, khác} | Vị trí nguồn âm thanh được biểu thị là một trong số tập hợp vị trí loa có tên. Tập hợp này có thể mở rộng được. Vị trí và sự mở rộng (các) kênh mở rộng được tạo ra bởi ExtChanPos và ExtChanWidth. |

Siêu dữ liệu kết xuất cho âm thanh dựa trên kênh cũng bao gồm phần tử kiểm soát kết xuất định rõ một số đặc điểm liên quan đến việc phát lại âm thanh dựa trên kênh, như được thể hiện trên Bảng 10.

BẢNG 10

| ID              | Giá trị                                                                 | Mô tả                                                                                                                                                                      |
|-----------------|-------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ChanUpmix       | {không, có}                                                             | Không cho phép (mặc định) hoặc cho phép tăng trộn                                                                                                                          |
| ChanUpmixZones  | {L, C, R, Ls, Rs, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc, Crs, Cts, khác} | Biểu thị các miền cho phép tăng trộn kênh trong đó.                                                                                                                        |
| ChanDownmixVect | Giá trị thực dương $\leq 1$                                             | Các ma trận giảm trộn đối tượng kênh tùy chỉnh cho các cấu hình kênh cụ thể. Danh sách cấu hình kênh có thể mở rộng và bao gồm 5.1 và âm thanh vòm Dolby 7.1.              |
| ChanUpmixVect   | Giá trị thực dương $\leq 1$                                             | Các ma trận tăng trộn đối tượng kênh tùy chỉnh cho các cấu hình kênh cụ thể. Danh sách cấu hình kênh có thể mở rộng được và bao gồm 5.1 và âm thanh vòm Dolby 7.1, và 9.1. |
| ChanSSBias      |                                                                         | Biểu thị độ nghiêng màn hình so với vòm. Hữu dụng nhất cho việc điều chỉnh kết xuất mặc định của các chế độ phát lại khác (5.1, 7.1)                                       |

Đối với âm thanh dựa trên đối tượng, siêu dữ liệu bao gồm các phần tử tương tự như đối với âm thanh dựa trên kênh. Bảng 11 cung cấp ID và các giá trị cho phần tử siêu dữ liệu vị trí đối tượng. Vị trí đối tượng được mô tả theo một trong ba cách: tọa độ ba chiều; tọa độ một mặt phẳng và hai chiều; hoặc đường thẳng và tọa độ một chiều. Phương pháp kết xuất có thể thích ứng dựa trên loại thông tin vị trí.

BẢNG 11

| ID           | Giá trị                                                                                                         | Mô tả                                                                                                                                    |
|--------------|-----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| ObjPosFormat | {3D, 2D, 1D}                                                                                                    | Định dạng vị trí                                                                                                                         |
| ObjPos3D     | Tọa độ x, y, z                                                                                                  | Vị trí 3D                                                                                                                                |
| ObjPos2D     | 3 tập hợp tọa độ x, y, z để định nghĩa mặt phẳng, và tập hợp tọa độ x, y để biểu thị vị trí trên mặt phẳng.     | Mặt phẳng + Vị trí 2D                                                                                                                    |
| ObjPos1D     | 2 tập hợp tọa độ x, y, z để định nghĩa đường thẳng, và 1 đại lượng vô hướng để biểu thị vị trí trên đường thẳng | Đường thẳng + Vị trí 1D hoặc Đường cong + Vị trí 1D                                                                                      |
| ObjPosScreen | {có, không}                                                                                                     | Sử dụng màn hình làm chuẩn. Thông tin vị trí nên được đo và dịch chuyển dựa trên hỗn hợp so với kích thước và vị trí màn hình biểu hiện. |

ID và các giá trị cho các phần tử siêu dữ liệu kiểm soát kết xuất đối tượng được thể hiện trên Bảng 12. Các giá trị này cung cấp các cách thức bổ sung để kiểm soát hoặc tối ưu hóa việc kết xuất cho âm thanh dựa trên đối tượng.

BẢNG 12

| ID        | Giá trị                            | Mô tả                                                                                                                                                                                                                                                                          |
|-----------|------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ObjSpread | x hoặc (x, y, z), số thực dương <1 | Độ rộng của hàm trải. Giá trị > 0 biểu thị nhiều hơn 1 loa nên được sử dụng. Khi giá trị càng tăng thì càng nhiều loa được sử dụng đến mức lớn hơn. Độ trải được biểu thị là một giá trị, hoặc giá trị độc lập cho mỗi chiều. Có thể sử dụng để nhấn việc quét, hoặc để tạo ra |

|                  |                                               |                                                                                                                                                                                                                                                                                                     |
|------------------|-----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                  |                                               | sự không rõ ràng về vị trí.                                                                                                                                                                                                                                                                         |
| ObjASW           | x hoặc (x, y, z), Số thực dương <1            | Độ rộng nguồn bên ngoài. Các giá trị lớn hơn biểu thị độ rộng nguồn lớn hơn. Có thể được thực thi nhờ việc sử dụng phương pháp khử tương quan.                                                                                                                                                      |
| ObjSnap          | {có, không}                                   | Bắt vào loa gần nhất. Hữu dụng khi âm sắc điểm nguồn quan trọng hơn độ chính xác về không gian.                                                                                                                                                                                                     |
| ObjSnapSmoothing | Giá trị thực dương < 10 (giây)                | Hằng số thời gian làm nhẵn không gian cho chế độ “Bắt vào”. Làm cho nhiều hơn “trượt vào”.                                                                                                                                                                                                          |
| ObjSnapTol       | Giá trị thực dương < 10                       | Bắt vào dung sai: Chấp nhận bao nhiêu lỗi không gian (theo khoảng cách được chuẩn hóa, chiều rộng phòng =1) trước khi chuyển lại hình ảnh ảo.                                                                                                                                                       |
| ObjRendAlg       | {def, dualBallance, vbap, dbap, 2D, 1D, khác} | Def: lựa chọn của bộ kết xuất<br>dualBallance: phương pháp Dolby<br>vbap: quét biên độ dựa trên vector<br>dbap: quét biên độ dựa trên khoảng cách<br>2D: kết hợp với ObjPos2D, sử dụng vbap chỉ với 3 vị trí nguồn (ảo)<br>ID: kết hợp với ObjPos1D, sử dụng quét theo cặp giữa 2 vị trí nguồn (ảo) |
| ObjZones         | Các giá trị thực dương <=1                    | Mức phân phối của miền loa có tên bất kỳ. Các miền loa được hỗ trợ bao gồm: L, C, R, Lss, Rss, Lrs, Rrs, Lts, Rts, Lc, Rc. Danh sách miền loa có thể được mở rộng để hỗ trợ các miền loa tương lai.                                                                                                 |
| ObjLevel         | Giá trị thực dương <=2                        | Mức đối tượng âm thanh khác của các cấu hình kênh cụ thể. Danh sách cấu hình kênh có thể mở rộng và bao gồm 5.1 và âm thanh vòm Dolby 7.1. Đối tượng có thể                                                                                                                                         |

|           |  |                                                                                                                                                                                                                                                                                                                                                |
|-----------|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |  | được làm suy giảm hoặc loại bỏ hoàn toàn khi kết xuất cho các cấu hình kênh nhỏ hơn.                                                                                                                                                                                                                                                           |
| ObjSSBias |  | <p>Biểu thị độ nghiêng màn hình so với phòng. Hữu dụng nhất cho việc điều chỉnh sự kết xuất mặc định của các chế độ phát lại thay thế (5.1, 7.1).</p> <p>Được xem là “tùy ý” bởi vì đặc điểm này có thể không yêu cầu siêu dữ liệu bổ sung – dữ liệu kết xuất khác có thể được sửa đổi trực tiếp (ví dụ, quỹ đạo quét, ma trận giảm trộn).</p> |

Theo một phương án, siêu dữ liệu được mô tả ở trên và được minh họa trên Fig.5 được tạo ra và lưu trữ như là một hoặc nhiều tập tin mà được liên kết hoặc được chú thích với nội dung âm thanh tương ứng nhờ đó các dòng âm thanh được xử lý bởi hệ thống âm thanh thích ứng thông dịch siêu dữ liệu được tạo ra bởi bộ trộn. Cần lưu ý rằng siêu dữ liệu được mô tả ở trên là ví dụ tập hợp gồm ID, giá trị, và định nghĩa, và phần tử khác hoặc các phần tử siêu dữ liệu bổ sung có thể được bao gồm để sử dụng trong hệ thống âm thanh thích ứng.

Theo một phương án, hai (hoặc nhiều) tập hợp phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh dựa trên kênh và dựa trên đối tượng. Tập hợp siêu dữ liệu thứ nhất được áp dụng cho nhiều dòng âm thanh đối với điều kiện thứ nhất của môi trường phát lại, và tập hợp siêu dữ liệu thứ hai được áp dụng cho nhiều dòng âm thanh đối với điều kiện thứ hai của môi trường phát lại. Tập hợp phần tử siêu dữ liệu thứ hai hoặc tiếp theo thay thế tập hợp phần tử siêu dữ liệu thứ nhất đối với dòng âm thanh cho trước dựa trên điều kiện của môi trường phát lại. Điều kiện này có thể bao gồm các yếu tố như kích thước phòng, hình dáng, thành phần vật liệu bên trong phòng, sự chiếm giữ hiện thời và mật độ người trong phòng, đặc điểm tiếng ồn xung quanh, đặc điểm ánh sáng xung quanh, và yếu tố khác bất kỳ có thể ảnh hưởng đến âm thanh hoặc thậm chí là tình trạng của môi trường phát lại.

Sau sản xuất và hoàn chỉnh

Giai đoạn kết xuất 110 của hệ thống xử lý âm thanh thích ứng 100 có thể bao gồm các bước sau sản xuất âm thanh dẫn đến việc tạo ra hỗn hợp cuối cùng. Trong ứng

dụng rạp phim, ba nhóm âm thanh chính được dùng trong hỗn hợp phim là lời thoại, âm nhạc, và hiệu ứng.

Các hiệu ứng bao gồm âm thanh không phải lời thoại hoặc âm nhạc (ví dụ, tiếng ồn xung quanh, tiếng ồn nền/cảnh). Hiệu ứng âm thanh có thể được ghi lại hoặc được tổng hợp bởi bộ thiết kế âm thanh hoặc chúng có thể được lấy ra từ thư viện hiệu ứng. Nhóm phụ gồm các hiệu ứng mà bao gồm các nguồn nhiễu cụ thể (ví dụ, bước chân, cửa, v.v.) được gọi là hiệu ứng Foley và được thực hiện bởi các diễn viên Foley. Các loại âm thanh khác nhau được đánh dấu và quét cho phù hợp bởi các kỹ sư ghi âm.

Fig.6 mô tả tiến trình công việc làm ví dụ của quy trình sau sản xuất trong hệ thống âm thanh thích ứng, theo một phương án. Như được thể hiện trên sơ đồ 600, tất cả các thành phần âm thanh riêng lẻ bao gồm âm nhạc, lời thoại, hiệu ứng Foley, và các hiệu ứng được ghép lại với nhau trong phòng lồng tiếng trong công đoạn trộn cuối cùng 606, và (các) bộ trộn ghi âm lại 604 sử dụng các bản phối âm (premixes) (còn được gọi là 'trộn âm-mix minus') cùng với các đối tượng âm thanh riêng lẻ và dữ liệu vị trí để tạo ra các dạng âm thanh stem như là một cách để nhóm, ví dụ, lời thoại, âm nhạc, hiệu ứng, hiệu ứng Foley và âm thanh nền. Ngoài việc tạo ra hỗn hợp cuối cùng 606, âm nhạc và tất cả các hiệu ứng gốc có thể được sử dụng như là cơ sở để tạo ra các phiên bản ngôn ngữ lồng tiếng cho phim. Mỗi dạng âm thanh stem chứa dạng âm thanh bed dựa trên kênh và nhiều đối tượng âm thanh với siêu dữ liệu. Các dạng âm thanh stem kết hợp để tạo ra hỗn hợp cuối cùng. Nhờ sử dụng thông tin quét đối tượng từ cả trạm xử lý âm thanh và bảng điều khiển trộn, bộ kết xuất và hoàn chỉnh 608 kết xuất âm thanh đến các vị trí loa trong phòng lồng tiếng. Sự kết xuất này cho phép bộ trộn nghe cách thức các dạng âm thanh bed dựa trên kênh và các đối tượng âm thanh kết hợp, và cũng mang lại khả năng kết xuất cho các cấu hình khác nhau. Bộ trộn có thể sử dụng siêu dữ liệu có điều kiện, siêu dữ liệu này mặc định cho các cấu hình tương ứng, để kiểm soát cách nội dung được kết xuất cho các kênh vòm. Bằng cách này, bộ trộn giữ lại toàn bộ sự kiểm soát cách thức phim phát lại trong tất cả các môi trường có thể mở rộng được. Bước giám sát có thể được bao gồm sau một hoặc cả hai bước ghi âm lại 604 và bước trộn cuối cùng 606 để cho phép bộ trộn nghe và đánh giá nội dung trung gian được sinh ra trong suốt mỗi giai đoạn này.

Trong phiên hoàn chỉnh, các dạng âm thanh stem, đối tượng, và siêu dữ liệu được ghép lại với nhau trong gói âm thanh thích ứng 614, gói này được tạo bởi bản in gốc 610. Gói này cũng chứa hỗn hợp cho rạp âm thanh vòm tương thích ngược (hệ thống âm thanh kế thừa 5.1 hoặc 7.1) 612. Bộ kết xuất/hoàn chỉnh (RMU - rendering/mastering unit) 608 có thể kết xuất đầu ra này nếu muốn; nhờ đó loại bỏ nhu cầu về các bước tiến trình bổ sung bất kỳ trong việc tạo ra các giao phẩm dựa trên kênh hiện tại. Theo một phương án, các tập tin âm thanh được đóng gói bằng cách sử dụng chuẩn đóng gói định dạng trao đổi nguyên liệu (Material Exchange Format - MXF). Tập tin chủ hỗn hợp âm thanh thích ứng cũng có thể được sử dụng để tạo ra các giao phẩm khác, chẳng hạn như hỗn hợp âm lập thể hoặc đa kênh người tiêu dùng. Các cấu hình thông minh và siêu dữ liệu có điều kiện cho phép kết xuất có kiểm soát mà có thể giảm đáng kể thời gian cần thiết để tạo ra các hỗn hợp này.

Theo một phương án, hệ thống đóng gói có thể được dùng để tạo ra gói phim kỹ thuật số cho các giao phẩm bao gồm hỗn hợp âm thanh thích ứng. Các tập tin rãnh âm thanh có thể được khóa với nhau để giúp ngăn ngừa các lỗi đồng bộ hóa với các tập tin rãnh âm thanh thích ứng.

Một số vùng cần bổ sung các tập tin rãnh trong giai đoạn đóng gói, ví dụ, bổ sung các rãnh khiếm khuyết thính giác (Hearing Impaired - HI) hoặc các rãnh tường thuật khiếm khuyết thị giác (Visually Impaired Narration - VI-N) vào tập tin rãnh âm thanh chính.

Theo một phương án, mảng loa trong môi trường phát lại có thể bao gồm số lượng bất kỳ các loa âm thanh vòm được bố trí và ký hiệu theo tiêu chuẩn âm thanh vòm đã thiết lập. Số lượng loa bổ sung bất kỳ để kết xuất chính xác nội dung âm thanh dựa trên đối tượng cũng có thể được bố trí dựa trên điều kiện của môi trường phát lại. Các loa bổ sung này có thể được thiết lập bởi kỹ sư âm thanh, và sự thiết lập này được thực hiện với hệ thống dưới dạng tập tin thiết lập được sử dụng bởi hệ thống để kết xuất các thành phần dựa trên đối tượng của âm thanh thích ứng cho loa hoặc các loa cụ thể bên trong toàn bộ mảng loa. Tập tin cài đặt bao gồm ít nhất một danh sách ký hiệu định rõ loa và ánh xạ kênh đến từng loa, thông tin về việc nhóm loa, và ánh xạ thời gian chạy dựa trên vị trí tương đối của loa so với môi trường phát lại. Ánh xạ thời gian chạy được sử dụng bởi đặc tính bất của hệ thống kết xuất nội dung âm thanh dựa trên đối tượng



nguồn điểm vào loa cụ thể gần nhất với vị trí nhận biết được của âm thanh như được dự định bởi kỹ sư âm thanh.

Fig.7 là sơ đồ của tiến trình công việc ví dụ của quy trình đóng gói phim kỹ thuật số sử dụng các tập tin âm thanh thích ứng, theo một phương án. Như được thể hiện trên sơ đồ 700, các tập tin âm thanh chứa cả các tập tin âm thanh thích ứng và các tập tin âm thanh vòm 5.1 hoặc 7.1 được nhập vào khối đóng gói/mã hóa 704. Theo một phương án, nhờ vào việc tạo ra gói phim kỹ thuật số trong khối 706, tập tin PCM MXF (với các rãnh bổ sung thích hợp được nối thêm) được mã hóa bằng cách sử dụng ngôn ngữ đặc tả SMPTE tương ứng với thực tế hiện tại. Âm thanh thích ứng MXF được đóng gói làm tập tin rãnh bổ trợ, và được mã hóa tùy ý bằng cách sử dụng khóa nội dung đối xứng trên mỗi ngôn ngữ đặc tả SMPTE. Một DCP 708 sau đó có thể được phân phối cho máy chủ phù hợp hệ thống chiếu phim kỹ thuật số (Digital Cinema Initiatives - DCI) bất kỳ. Nói chung, mọi cài đặt không được trang bị phù hợp sẽ chỉ cần bỏ qua tập tin rãnh bổ sung chứa rãnh âm thanh thích ứng và sẽ sử dụng tập tin rãnh âm thanh chính đã có cho việc phát lại chuẩn. Các hệ thống cài đặt được trang bị các bộ xử lý âm thanh thích ứng thích hợp sẽ có thể nhận và phát lại rãnh âm thanh thích ứng trong trường hợp ứng dụng được, phục hồi về rãnh âm thanh chuẩn nếu cần thiết. Bộ phận đóng gói/mã hóa 704 cũng có thể cung cấp đầu vào trực tiếp cho khối KDM phân phối 710 để tạo ra mã bảo mật thích hợp để sử dụng trong máy chủ chiếu phim kỹ thuật số. Các thành phần phim hoặc tập tin khác, chẳng hạn như phụ đề 714 và hình ảnh 716 có thể được đóng gói và mã hóa cùng với các tập tin âm thanh 702. Trong trường hợp này, có thể bao gồm các bước xử lý nhất định, chẳng hạn như nén 712 trong trường hợp các tập tin hình ảnh 716.

Đối với việc quản lý nội dung, hệ thống âm thanh thích ứng 100 cho phép nhà sáng tạo nội dung tạo ra các đối tượng âm thanh riêng và bổ sung thông tin về nội dung có thể được vận chuyển đến hệ thống tái tạo. Điều này cho phép tính linh hoạt cao trong quản lý nội dung âm thanh. Từ quan điểm quản lý nội dung, phương pháp âm thanh thích ứng cho phép nhiều đặc tính khác nhau. Các đặc tính này bao gồm thay đổi ngôn ngữ của nội dung bằng cách chỉ thay thế đối tượng lời thoại để tiết kiệm không gian, hiệu quả tải xuống, thích ứng phát lại địa lý, v.v. Phim, truyền hình và các chương trình giải trí khác thường được phân phối trên toàn thế giới. Điều này thường đòi hỏi ngôn ngữ trong đoạn nội dung cần được thay đổi tùy thuộc vào nơi mà nó sẽ được tái sản

xuất (tiếng Pháp đối với phim được chiếu ở Pháp, tiếng Đức đối với chương trình truyền hình được chiếu ở Đức, v.v.). Ngày nay điều này thường đòi hỏi rãnh âm thanh độc lập hoàn toàn được tạo ra, đóng gói và phân phối. Với âm thanh thích ứng và khái niệm vốn có của nó về các đối tượng âm thanh, lời thoại cho phần nội dung có thể là đối tượng âm thanh độc lập. Điều này cho phép ngôn ngữ của nội dung được thay đổi dễ dàng mà không cần cập nhật hoặc thay đổi các yếu tố khác của rãnh âm thanh chẳng hạn như âm nhạc, hiệu ứng, v.v. Điều này không chỉ áp dụng cho tiếng nước ngoài mà còn cả ngôn ngữ không phù hợp cho khán giả nhất định (ví dụ, chương trình truyền hình cho trẻ em, phim trên máy bay, v.v.), quảng cáo nhắm mục tiêu, v.v.

#### Cân nhắc lắp đặt và thiết bị

Định dạng tập tin âm thanh thích ứng và các bộ xử lý kết hợp cho phép thay đổi cách lắp đặt, hiệu chỉnh và bảo trì thiết bị rạp. Với sự ra đời của nhiều đầu ra loa tiềm năng hơn, mỗi đầu ra được cân bằng và cân chỉnh riêng, có nhu cầu về sự cân bằng trong phòng tự động thông minh và hiệu quả thời gian, việc này có thể được thực hiện nhờ khả năng điều chỉnh thủ công sự cân bằng tự động trong phòng bất kỳ. Theo một phương án, hệ thống âm thanh thích ứng sử dụng máy cân bằng dải octa 1/12\* tối ưu hóa. Tối đa 64 đầu vào có thể được xử lý để cân bằng chính xác hơn âm thanh trong rạp. Hệ thống này cũng cho phép giám sát theo lịch trình các đầu ra loa riêng, từ đầu ra của bộ xử lý phim đến âm thanh được tái tạo trong khán phòng. Cảnh báo cục bộ hoặc cảnh báo mạng có thể được tạo ra để đảm bảo rằng hành động thích hợp được tiến hành. Hệ thống kết xuất linh hoạt có thể loại bỏ tự động loa hoặc bộ khuếch đại bị hỏng ra khỏi chuỗi phát lại và kết xuất quanh nó, do đó cho phép chương trình tiếp tục.

Bộ xử lý rạp phim có thể được nối với máy chủ rạp phim kỹ thuật số có các kết nối âm thanh chính 8xAES hiện có, và kết nối Ethernet để truyền dữ liệu âm thanh thích ứng. Sự phát lại nội dung âm thanh vòm 5.1 hoặc 7.1 sử dụng kết nối PCM hiện có. Dữ liệu âm thanh thích ứng được truyền theo dòng qua kết nối Ethernet đến bộ xử lý phim để giải mã và kết xuất, và kết nối giữa máy chủ và bộ xử lý phim cho phép âm thanh được xác định và đồng bộ hóa. Trong trường hợp có vấn đề bất kỳ với việc phát lại rãnh âm thanh thích ứng, âm thanh được phục hồi thành âm thanh vòm Dolby 7.1 hoặc âm thanh PCM 5.1.

Mặc dù các phương án được mô tả liên quan đến hệ thống âm thanh vòm 5.1 và 7.1, nhưng cần lưu ý rằng nhiều cấu hình âm thanh vòm khác hiện tại và tương lai khác có thể được sử dụng cùng với các phương án bao gồm 9.1, 11.1 và 13.1 và hơn nữa.

Hệ thống âm thanh thích ứng được thiết kế để cho phép cả các nhà sáng tạo nội dung và các chủ rạp phim quyết định cách nội dung âm thanh được kết xuất trong các cấu hình loa phát lại khác nhau. Số lượng kênh đầu ra loa lý tưởng được sử dụng sẽ khác nhau tùy theo kích thước phòng. Do đó, việc sắp đặt loa được đề xuất phụ thuộc vào nhiều yếu tố, chẳng hạn như kích thước, thành phần, sắp xếp chỗ ngồi, môi trường, số lượng khán giả trung bình, v.v.

Cấu hình và sơ đồ loa làm ví dụ hoặc làm đại diện được đề cập ở đây chỉ nhằm mục đích minh họa, và không nhằm giới hạn phạm vi của các phương án yêu cầu bảo hộ bất kỳ.

Sơ đồ loa được đề xuất cho hệ thống âm thanh thích ứng vẫn tương thích với các hệ thống chiếu phim hiện tại, điều này rất quan trọng vì không ảnh hưởng đến việc phát lại các định dạng dựa trên kênh 5.1 và 7.1 hiện tại. Để duy trì ý định của kỹ sư âm thanh thích ứng và ý định của các bộ trộn nội dung 7.1 và 5.1, các vị trí của các kênh màn hình hiện tại không nên bị thay đổi quá nhiều trong nỗ lực nâng cao hoặc làm nổi bật vị trí loa mới. Trái với việc sử dụng tất cả 64 kênh đầu ra có sẵn, định dạng âm thanh thích ứng có khả năng được kết xuất chính xác trong rạp phim thành các cấu hình loa chẳng hạn như 7.1, do đó còn cho phép định dạng (và các lợi ích liên quan) được sử dụng trong rạp hiện tại mà không thay đổi các bộ khuếch đại hoặc loa.

Các vị trí loa khác nhau có thể có hiệu quả khác nhau tùy thuộc vào thiết kế rạp, do đó hiện nay không có số lượng hoặc sự bố trí kênh lý tưởng được định rõ trong ngành công nghiệp. Âm thanh thích ứng được thiết kế để thực sự thích ứng và có khả năng phát lại chính xác trong nhiều thính phòng khác nhau, cho dù chúng có số lượng kênh phát lại hạn chế hoặc nhiều kênh có các cấu hình rất linh hoạt.

Fig.8 là hình chiếu nhìn từ trên cao 800 của sơ đồ ví dụ về các vị trí loa được đề xuất để sử dụng với hệ thống âm thanh thích ứng trong thính phòng điển hình, và Fig.9 là hình vẽ nhìn từ đằng trước 900 của bố cục làm ví dụ về các vị trí loa được đề xuất tại màn hình của thính phòng. Vị trí chuẩn được đề cập sau đây tương ứng với vị trí 2/3 khoảng cách từ màn hình đến tường phía sau, trên đường trung tâm của màn hình. Loa

màn hình tiêu chuẩn 801 được thể hiện ở các vị trí thông thường so với màn hình. Các nghiên cứu về cảm nhận độ cao trên bề mặt màn hình đã cho thấy rằng các loa bổ sung 804 sau màn hình, chẳng hạn như các loa màn hình giữa trái (Lc) và giữa phải (Rc) (ở các vị trí của kênh ngoài cùng trái và ngoài cùng phải trong định dạng phim 70mm), có thể có lợi trong việc thực hiện quét đều đặn hơn trên màn hình. Do đó các loa tùy chọn như vậy, đặc biệt trong các thính phòng có màn hình rộng hơn 12m (40ft.) được khuyến dùng. Tất cả các loa màn hình nên được đặt ở góc sao cho chúng hướng về vị trí chuẩn. Vị trí được đề xuất cho loa siêu trầm 810 đằng sau màn hình không nên thay đổi, bao gồm việc duy trì vị trí buồng bất đối xứng, liên quan đến trung tâm của phòng, để ngăn chặn sự kích thích của sóng đứng. Các loa siêu trầm bổ sung 816 có thể được đặt ở đằng sau rạp.

Các loa vòm 802 phải được nối riêng ngược với rãnh của bộ khuếch đại, và có thể được khuếch đại riêng nếu có thể với kênh khuếch đại công suất chuyên dụng để phù hợp với việc quản lý công suất của loa theo thông số kỹ thuật của nhà sản xuất. Lý tưởng là, các loa vòm nên được xác định để xử lý SPL gia tăng cho mỗi loa riêng, và cũng đáp ứng tần số rộng hơn nếu có thể. Theo quy tắc ngón tay cái đối với rạp có kích thước trung bình, khoảng cách giữa các loa vòm phải nằm trong khoảng từ 2 đến 3m (từ 6'6" đến 9'9"), với các loa vòm trái và phải được bố trí đối xứng.

Tuy nhiên, khoảng cách của các loa vòm được xem là hiệu quả nhất khi các góc đối diện với người nghe cho trước giữa các loa liền kề, trái ngược với việc sử dụng khoảng cách tuyệt đối giữa các loa. Để phát lại tối ưu trong toàn bộ thính phòng, khoảng cách góc giữa các loa liền kề nên từ 30 độ hoặc ít hơn, được tham chiếu từ mỗi góc trong bốn góc của khu vực nghe cơ bản. Có thể đạt được kết quả tốt với khoảng cách lên đến 50 độ. Đối với mỗi vùng loa vòm, các loa nên duy trì khoảng cách tuyến tính bằng nhau cạnh khu vực ghế ngồi nếu có thể. Khoảng cách tuyến tính ngoài khu vực nghe, ví dụ, giữa hàng trước và màn hình, có thể lớn hơn một chút. Fig.11 là ví dụ về sự định vị các loa vòm trên 808 và các loa vòm bên cạnh 806 so với vị trí chuẩn, theo một phương án.

Các loa vòm bên phụ 806 nên được gắn gần với màn hình hơn so với thực tế hiện được khuyến nghị là bắt đầu khoảng một-phần-ba khoảng cách đến mặt sau thính phòng. Các loa này không được sử dụng làm các loa vòm bên trong khi phát lại rãnh âm

thanh vòm Dolby 7.1 hoặc 5.1, nhưng sẽ cho phép chuyển đổi mượt mà và sự so khớp âm sắc cải thiện khi quét các đối tượng từ các loa màn hình đến các vùng loa vòm. Để tối đa hóa sự ấn tượng về không gian, các mảng loa vòm phải được bố trí thấp ở mức như thực tế, chịu sự ràng buộc sau: vị trí dọc của loa vòm ở đằng trước mảng phải gần với chiều cao của trung tâm âm thanh loa màn hình và đủ cao để duy trì vùng bao phủ tốt trên toàn bộ khu vực ghế ngồi theo sự định hướng của loa. Vị trí dọc của loa vòm phải được bố trí sao cho chúng tạo thành đường thẳng từ trước tới sau, và (thường) nghiêng hướng lên vì vậy độ cao tương đối của các loa vòm phía trên người nghe được duy trì về phía sau rạp phim do độ cao của ghế ngồi gia tăng, như được thể hiện trên Fig.10, đây là hình chiếu bên của sơ đồ làm ví dụ về các vị trí loa được đề xuất để sử dụng với hệ thống âm thanh thích ứng trong thính phòng điển hình. Trên thực tế, điều này có thể đạt được đơn giản nhất bằng cách lựa chọn độ cao cho các loa vòm bên trước nhất và sau nhất, và đặt các loa còn lại trên đường thẳng giữa các điểm này.

Để tạo ra vùng bao phủ tối ưu cho mỗi loa trên khu vực ngồi, loa vòm bên 806 và các loa sau 816 và các loa vòm trên 808 phải hướng đến vị trí chuẩn trong rạp, theo các chỉ dẫn xác định về khoảng cách, vị trí, góc, v.v.

Các phương án của hệ thống và định dạng rạp phim có âm thanh thích ứng đạt được các mức đắm chìm và thu hút khán giả được cải thiện cho các hệ thống hiện tại bằng cách cung cấp các công cụ biên soạn mới mạnh mẽ cho bộ trộn và bộ xử lý phim mới có công cụ kết xuất linh hoạt giúp tối ưu hóa chất lượng âm thanh và hiệu ứng âm thanh vòm của rãnh âm thanh cho sơ đồ và đặc điểm loa của mỗi phòng. Ngoài ra, hệ thống duy trì khả năng tương thích ngược và giảm thiểu tác động lên các tiến trình sản xuất và phân phối hiện tại.

Mặc dù các phương án được mô tả liên quan đến các ví dụ và các phương án thực hiện trong môi trường rạp phim trong đó nội dung âm thanh thích ứng được kết hợp với nội dung phim để sử dụng trong các hệ thống xử lý phim kỹ thuật số, nhưng cần lưu ý rằng các phương án này có thể cũng được thực hiện trong các môi trường không phải rạp phim. Nội dung âm thanh thích ứng bao gồm âm thanh dựa trên đối tượng và âm thanh dựa trên kênh có thể được sử dụng cùng với nội dung liên quan bất kỳ (âm thanh, hình ảnh, đồ họa được kết hợp, v.v.), hoặc nó có thể tạo thành nội dung âm thanh riêng. Môi trường phát lại có thể là môi trường nghe thích hợp bất kỳ từ tai

nghe hoặc màn hình trường gần đến các phòng nhỏ hoặc lớn, ô tô, sân khấu ngoài trời, phòng hòa nhạc, v.v..

Các khía cạnh của hệ thống 100 có thể được thực hiện trong môi trường mạng xử lý âm thanh bằng máy tính thích hợp để xử lý các tập tin âm thanh kỹ thuật số hoặc được số hóa. Các phần của hệ thống âm thanh thích ứng có thể bao gồm một hoặc nhiều mạng bao gồm số lượng máy riêng lẻ mong muốn bất kỳ, bao gồm một hoặc nhiều bộ định tuyến (không được thể hiện trên hình vẽ) đóng vai trò làm bộ đệm và định tuyến dữ liệu được truyền giữa các máy tính với nhau. Mạng như vậy có thể được xây dựng trên nhiều giao thức mạng khác nhau, và có thể là mạng Internet, mạng diện rộng (Wide Area Network - WAN), mạng cục bộ (Local Area Network - LAN), hoặc kết hợp bất kỳ của các mạng trên. Theo một phương án trong đó mạng này bao gồm mạng Internet, một hoặc nhiều máy có thể được tạo cấu hình để truy cập mạng Internet thông qua các chương trình trình duyệt web.

Một hoặc nhiều bộ phận, khối, quy trình hoặc các bộ phận chức năng khác có thể được thực hiện qua chương trình máy tính để kiểm soát sự thực thi của thiết bị điện toán dựa trên bộ xử lý của hệ thống. Cũng nên lưu ý rằng các chức năng khác nhau được bộc lộ ở đây có thể được mô tả bằng cách sử dụng số lượng tổ hợp bất kỳ của phần cứng, phần sụn, và/hoặc dữ liệu và/hoặc các lệnh được thể hiện trong các phương tiện đọc được bằng máy tính hoặc đọc được trên các máy khác nhau, xét về hành vi, truyền bằng thanh ghi, thành phần logic, và/hoặc các đặc điểm khác của chúng. Phương tiện đọc được bằng máy tính trong đó các dữ liệu và/hoặc lệnh được định dạng như vậy có thể được biểu hiện bao gồm, nhưng không bị giới hạn ở, phương tiện lưu trữ vật lý không thay đổi (bất biến) ở các dạng khác nhau, chẳng hạn như phương tiện lưu trữ quang, từ hoặc bán dẫn.

Trừ khi ngữ cảnh quy định khác rõ ràng, trong suốt bản mô tả và yêu cầu bảo hộ, các từ "bao gồm -comprise," "bao gồm-comprising," và tương tự phải được hiểu theo nghĩa bao hàm trái với ý nghĩa loại bỏ hoặc loại trừ; tức là, có nghĩa "bao gồm, nhưng không bị giới hạn ở." Các từ sử dụng dạng số ít hoặc số nhiều cũng bao gồm số ít hoặc số nhiều tương ứng. Ngoài ra, các từ "ở đây," "dưới đây," "trên đây," "sau đây," và các từ có ý nghĩa tương tự để chỉ trong toàn bộ sáng chế này và không chỉ phần cụ thể bất kỳ của đơn này. Khi từ "hoặc" được sử dụng để chỉ một danh sách gồm hai mục trở lên,

từ này bao hàm tất cả các cách hiểu sau: thành phần bất kỳ trong số các thành phần có trong danh sách liệt kê, tất cả các thành phần có trong danh sách liệt kê và kết hợp bất kỳ của các thành phần có trong danh sách liệt kê.

Mặc dù một hoặc nhiều phương án thực hiện được mô tả bằng ví dụ và theo các phương án cụ thể, nhưng phải hiểu rằng một hoặc nhiều phương án thực hiện không bị giới hạn ở các phương án được bộc lộ. Ngược lại, nó được dự định để bao gồm các cải biến khác nhau và các cách sắp xếp tương tự rõ ràng đối với người có trình độ trung bình trong lĩnh vực này. Do đó, phạm vi của các yêu cầu bảo hộ nên được hiểu theo nghĩa rộng nhất để bao gồm tất cả các cải biến như vậy và các sắp xếp tương tự.

## YÊU CẦU BẢO HỘ

1. Hệ thống xử lý tín hiệu âm thanh, hệ thống này bao gồm bộ phận biên soạn được tạo cấu hình để:

nhận nhiều tín hiệu âm thanh;

tạo ra hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi trong số các dòng âm thanh và chỉ báo vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và các dòng âm thanh còn lại trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của dòng âm thanh đơn âm dựa trên kênh bao gồm ký hiệu định rõ loa trong mảng loa, và vị trí phát lại của dòng âm thanh đơn âm dựa trên đối tượng bao gồm vị trí trong không gian ba chiều, và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa riêng của mảng loa; và

đóng gói nhiều dòng âm thanh đơn âm và siêu dữ liệu trong dòng bit để truyền cho hệ thống kết xuất được tạo cấu hình để kết xuất các dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại, trong đó các loa của mảng loa được đặt ở các vị trí riêng trong môi trường phát lại, và trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng chỉ báo xem một hoặc nhiều thành phần âm thanh có được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh hay không, sao cho dòng âm thanh đơn âm dựa trên đối tượng tương ứng được kết xuất hiệu quả bởi loa gần nhất với vị trí phát lại dự tính.

2. Hệ thống theo điểm 1, trong đó thành phần biên soạn bao gồm bảng điều khiển trộn có các bộ điều chỉnh vận hành bởi người sử dụng để chỉ báo các mức phát lại của nhiều dòng âm thanh đơn âm, và trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng tương ứng được tạo ra tự động khi người sử dụng nhập vào các bộ điều chỉnh bảng điều khiển trộn.

3. Hệ thống theo điểm 1, trong đó hệ thống này còn bao gồm bộ mã hóa được nối với bộ phận biên soạn và được tạo cấu hình để nhận nhiều dòng âm thanh đơn âm và siêu



dữ liệu và để tạo ra một dòng bit số chứa nhiều dòng âm thanh đơn âm theo cách có thứ tự.

4. Hệ thống xử lý tín hiệu âm thanh bao gồm hệ thống kết xuất được tạo cấu hình để:

nhận dòng bit đóng gói hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi dòng trong số các dòng âm thanh và chỉ báo vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số trong số nhiều dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và các dòng âm thanh còn lại trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của dòng âm thanh đơn âm dựa trên kênh bao gồm ký hiệu định rõ loa trong mảng loa, và vị trí phát lại của dòng âm thanh đơn âm dựa trên đối tượng bao gồm vị trí trong không gian ba chiều, và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa cụ thể của mảng loa; và

kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại, trong đó các loa của mảng loa được đặt ở các vị trí riêng trong môi trường phát lại, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng chỉ báo xem một hoặc nhiều thành phần âm thanh có được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh hay không, sao cho dòng âm thanh đơn âm dựa trên đối tượng tương ứng được kết xuất hiệu quả bởi loa gần nhất với vị trí phát lại dự tính.

5. Hệ thống theo điểm 4, trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng còn chỉ báo ngưỡng biến dạng không gian, và trong đó phần tử siêu dữ liệu chỉ báo xem thành phần âm thanh tương ứng có được kết xuất bởi loa gần nhất với vị trí phát lại dự tính hay không được bỏ qua nếu sự biến dạng không gian tạo thành từ việc kết xuất thành phần âm thanh tương ứng bằng loa gần nhất với vị trí phát lại dự tính vượt quá ngưỡng biến dạng không gian.

6. Hệ thống theo điểm 5, trong đó ngưỡng biến dạng không gian bao gồm ít nhất một trong số ngưỡng dung sai góc phương vị và ngưỡng dung sai độ cao.

7. Hệ thống theo điểm 4, trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng còn chỉ báo tham số tốc độ tăng giảm chéo (crossfade rate parameter), và trong đó khi loa gần nhất với vị trí phát lại dự tính cho thành phần âm thanh thay đổi từ loa thứ nhất đến loa thứ hai, tốc độ mà thành phần âm thanh thay đổi từ loa thứ nhất đến loa thứ hai được kiểm soát đáp lại tham số tốc độ tăng giảm chéo.

8. Hệ thống theo điểm 4, trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng còn chỉ báo các tham số không gian kiểm soát sự phát lại của thành phần âm thanh tương ứng bao gồm một hoặc nhiều trong số: vị trí âm thanh, độ rộng âm thanh và vận tốc âm thanh.

9. Hệ thống theo điểm 4, trong đó vị trí phát lại của mỗi trong số nhiều dòng âm thanh đơn âm dựa trên đối tượng bao gồm vị trí không gian liên quan đến màn hình trong môi trường phát lại, hoặc bề mặt bao quanh môi trường phát lại, và trong đó bề mặt này bao gồm mặt trước, mặt sau, mặt trái, mặt phải, mặt trên, và mặt dưới.

10. Hệ thống theo điểm 4, trong đó hệ thống kết xuất chọn thuật toán kết xuất được sử dụng bởi hệ thống kết xuất, thuật toán kết xuất này được chọn từ nhóm bao gồm: âm thanh nổi, lưỡng cực lập thể, dạng âm thanh Ambisonics, tổng hợp trường sóng (wave field synthesis - WFS), quét đa kênh, dạng âm thanh stems thô (raw stems) có siêu dữ liệu vị trí, cân bằng kép, và quét biên độ dựa trên vector.

11. Hệ thống theo điểm 4, trong đó vị trí phát lại của mỗi trong số nhiều dòng âm thanh đơn âm dựa trên đối tượng được xác định một cách độc lập đối với khung tham chiếu lấy chủ thể làm trung tâm (egocentric frame of reference) hoặc khung tham chiếu lấy đối tượng làm trung tâm (allocentric frame of reference), trong đó khung tham chiếu lấy chủ thể làm trung tâm được áp dụng liên quan đến người nghe trong môi trường phát lại, và trong đó khung tham chiếu lấy đối tượng làm trung tâm được áp dụng đối với đặc điểm của môi trường phát lại.

12. Phương pháp biên soạn nội dung âm thanh để kết xuất, phương pháp này bao gồm các bước:

nhận nhiều tín hiệu âm thanh;

tạo ra hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi dòng trong số các dòng âm thanh và chỉ báo vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số dòng trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và các dòng còn lại trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của âm thanh dựa trên kênh bao gồm các ký hiệu định rõ loa của các loa trong mảng loa, và vị trí phát lại của âm thanh dựa trên đối tượng bao gồm vị trí trong không gian ba chiều; và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa riêng của mảng loa; và

đóng gói nhiều dòng âm thanh đơn âm và siêu dữ liệu trong dòng bit để truyền cho hệ thống kết xuất được tạo cấu hình để kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại, trong đó các loa trong mảng loa được đặt ở các vị trí riêng trong môi trường phát lại, và trong đó phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng chỉ báo xem một hoặc nhiều thành phần âm thanh có được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh hay không, sao cho dòng âm thanh đơn âm dựa trên đối tượng được kết xuất hiệu quả bởi loa gần nhất với vị trí phát lại dự tính.

13. Phương pháp theo điểm 12 còn bao gồm các bước:

nhận, từ bảng điều khiển trộn có bộ điều chỉnh được vận hành bởi người dùng để chỉ báo các mức phát lại của các dòng âm thanh đơn âm bao gồm nội dung âm thanh; và

tự động tạo ra các phần tử siêu dữ liệu kết hợp với mỗi dòng dựa trên đối tượng tương ứng khi nhận đầu vào từ người dùng.

14. Phương pháp kết xuất tín hiệu âm thanh, phương pháp này bao gồm các bước:

nhận dòng bit đóng gói hỗn hợp âm thanh thích ứng bao gồm nhiều dòng âm thanh đơn âm và siêu dữ liệu kết hợp với mỗi dòng âm thanh trong số các dòng âm thanh và chỉ báo vị trí phát lại của dòng âm thanh đơn âm tương ứng, trong đó ít nhất một số trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên kênh và các dòng âm thanh còn lại trong số các dòng âm thanh đơn âm được xác định là âm thanh dựa trên đối tượng, và trong đó vị trí phát lại của dòng âm thanh đơn âm dựa trên

kênh bao gồm ký hiệu định rõ loa trong mảng loa, và vị trí phát lại của dòng âm thanh đơn âm dựa trên đối tượng bao gồm vị trí trong không gian ba chiều, và trong đó mỗi dòng âm thanh đơn âm dựa trên đối tượng được kết xuất trong ít nhất một loa riêng của mảng loa; và

kết xuất nhiều dòng âm thanh đơn âm thành nhiều tín hiệu cấp cho loa tương ứng với các loa trong môi trường phát lại, trong đó các loa của mảng loa được đặt ở các vị trí riêng trong môi trường phát lại, và trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng chỉ báo xem một hoặc nhiều thành phần âm thanh có được kết xuất thành tín hiệu cấp cho loa để phát lại qua loa gần nhất với vị trí phát lại dự tính của thành phần âm thanh hay không sao cho dòng âm thanh đơn âm dựa trên đối tượng được kết xuất hiệu quả bởi loa gần nhất với vị trí phát lại dự tính.

15. Phương pháp theo điểm 14, trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng còn chỉ báo ngưỡng biến dạng không gian, và trong đó phần tử siêu dữ liệu chỉ báo xem thành phần âm thanh tương ứng có được kết xuất bởi loa gần nhất với vị trí phát lại dự tính hay không được bỏ qua nếu sự biến dạng không gian tạo thành từ việc kết xuất thành phần âm thanh tương ứng bởi loa gần nhất với vị trí phát lại dự tính vượt quá ngưỡng biến dạng không gian.

16. Phương pháp theo điểm 15, trong đó ngưỡng biến dạng không gian bao gồm ít nhất một trong số ngưỡng dung sai góc phương vị và ngưỡng dung sai độ cao.

17. Phương pháp theo điểm 14, trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng tương ứng còn chỉ báo tham số tốc độ tăng giảm chéo, và trong đó khi loa gần nhất với vị trí phát lại dự tính cho thành phần âm thanh thay đổi từ loa thứ nhất đến loa thứ hai, tốc độ mà đối tượng tương ứng chuyển tiếp từ loa thứ nhất đến loa thứ hai được kiểm soát đáp lại tham số tốc độ tăng giảm chéo.

18. Phương pháp theo điểm 14, trong đó các phần tử siêu dữ liệu kết hợp với mỗi dòng âm thanh đơn âm dựa trên đối tượng còn chỉ báo các tham số không gian kiểm soát sự phát lại của thành phần âm thanh tương ứng bao gồm một hoặc nhiều trong số các: vị trí âm thanh, độ rộng âm thanh và vận tốc âm thanh.

19. Phương pháp theo điểm 14, trong đó vị trí phát lại cho mỗi dòng trong số các dòng âm thanh đơn âm dựa trên đối tượng bao gồm vị trí không gian liên quan đến màn hình trong môi trường phát lại, hoặc bề mặt bao quanh môi trường phát lại, và trong đó bề mặt này bao gồm mặt trước, mặt sau, mặt trái, mặt phải, mặt trên, và mặt dưới và/hoặc được xác định độc lập đối với khung tham chiếu lấy chủ thể làm trung tâm hoặc khung tham chiếu lấy đối tượng làm trung tâm, trong đó khung tham chiếu lấy chủ thể làm trung tâm được áp dụng liên quan đến người nghe trong môi trường phát lại, và trong đó khung tham chiếu lấy đối tượng làm trung tâm được áp dụng đối với đặc điểm của môi trường phát lại.

20. Phương pháp theo điểm 14, trong đó hệ thống kết xuất chọn thuật toán kết xuất được sử dụng bởi hệ thống kết xuất, thuật toán kết xuất này được chọn từ nhóm bao gồm: âm thanh nổi, lưỡng cực lập thể, dạng âm thanh Ambisonics, WFS, quét đa kênh, dạng âm thanh stems thô với siêu dữ liệu về vị trí, cân bằng kép, và quét biên độ dựa trên vector.

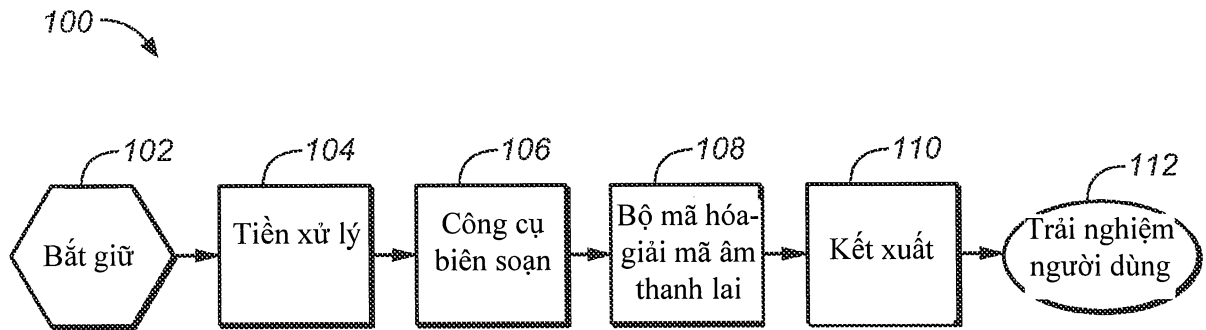


Fig.1

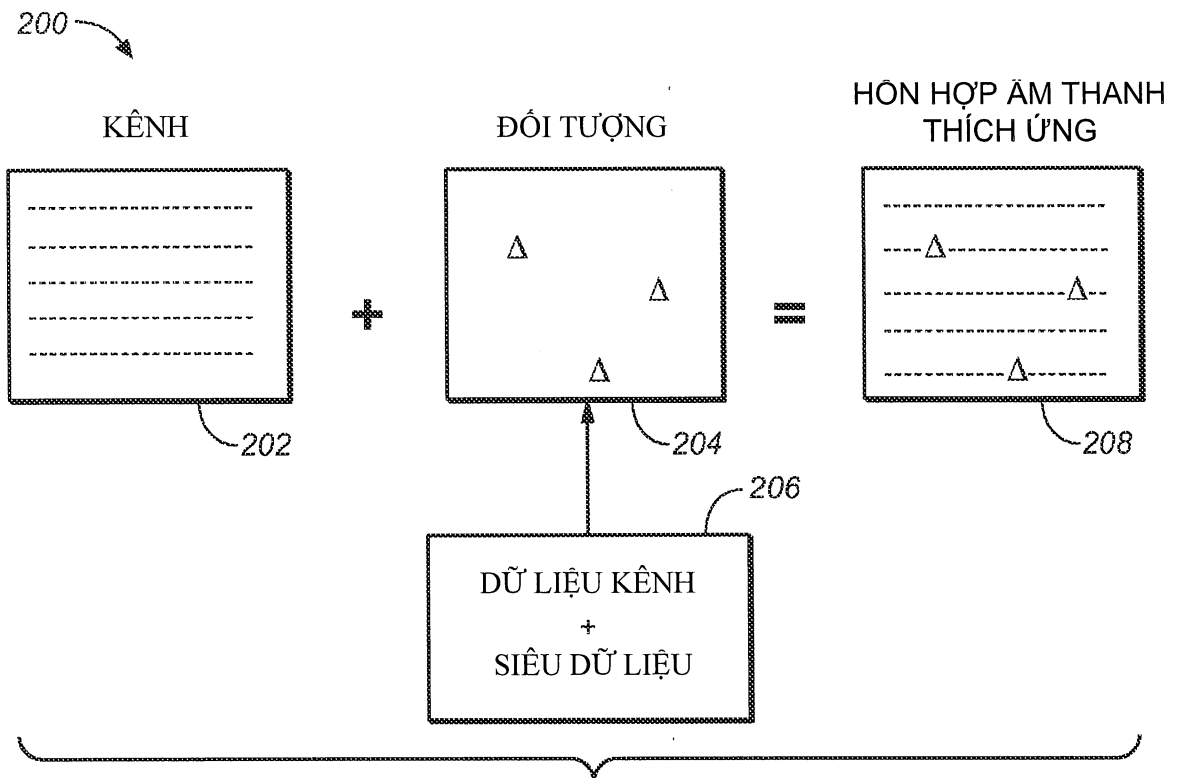


Fig.2

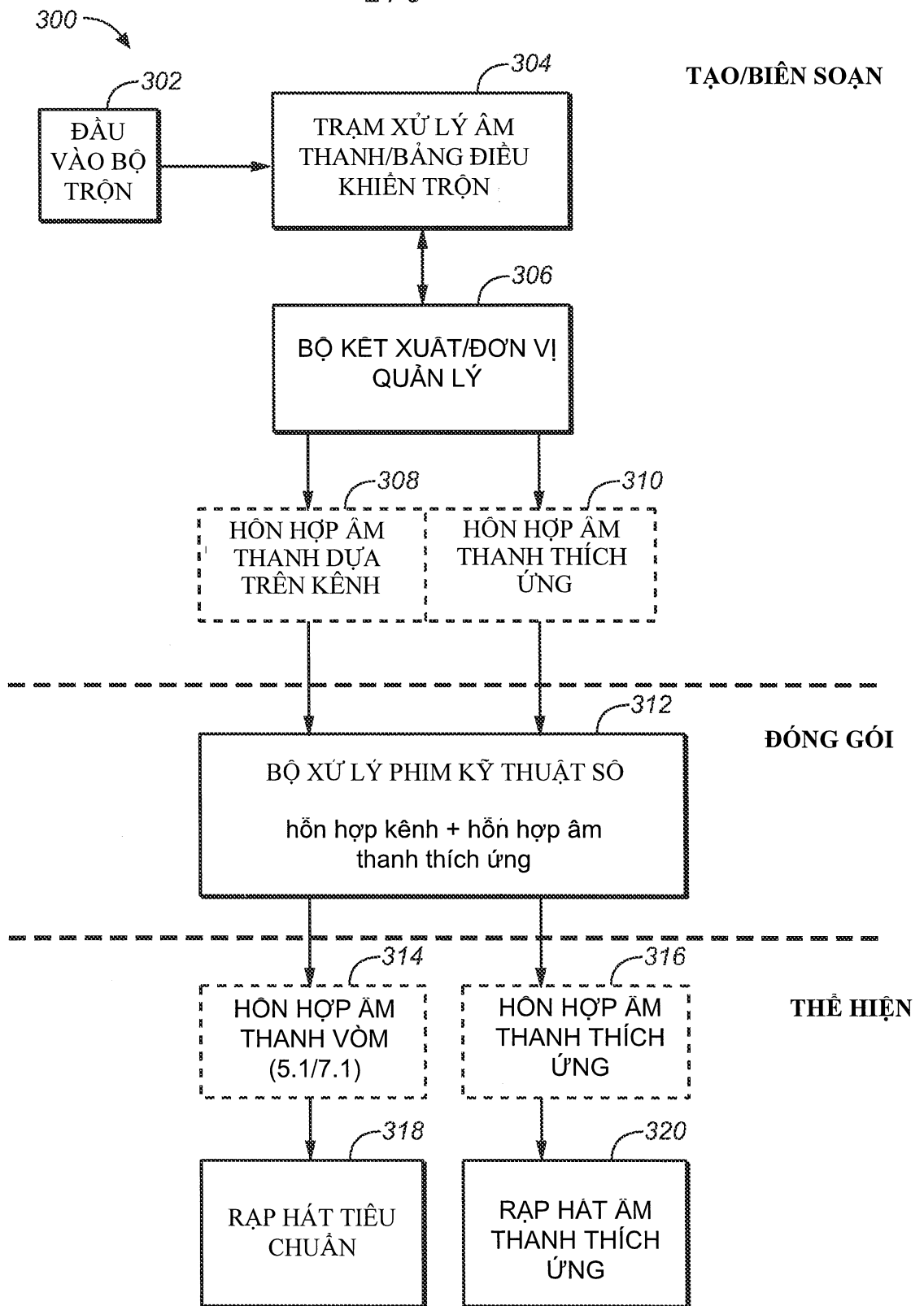


Fig.3

3 / 9

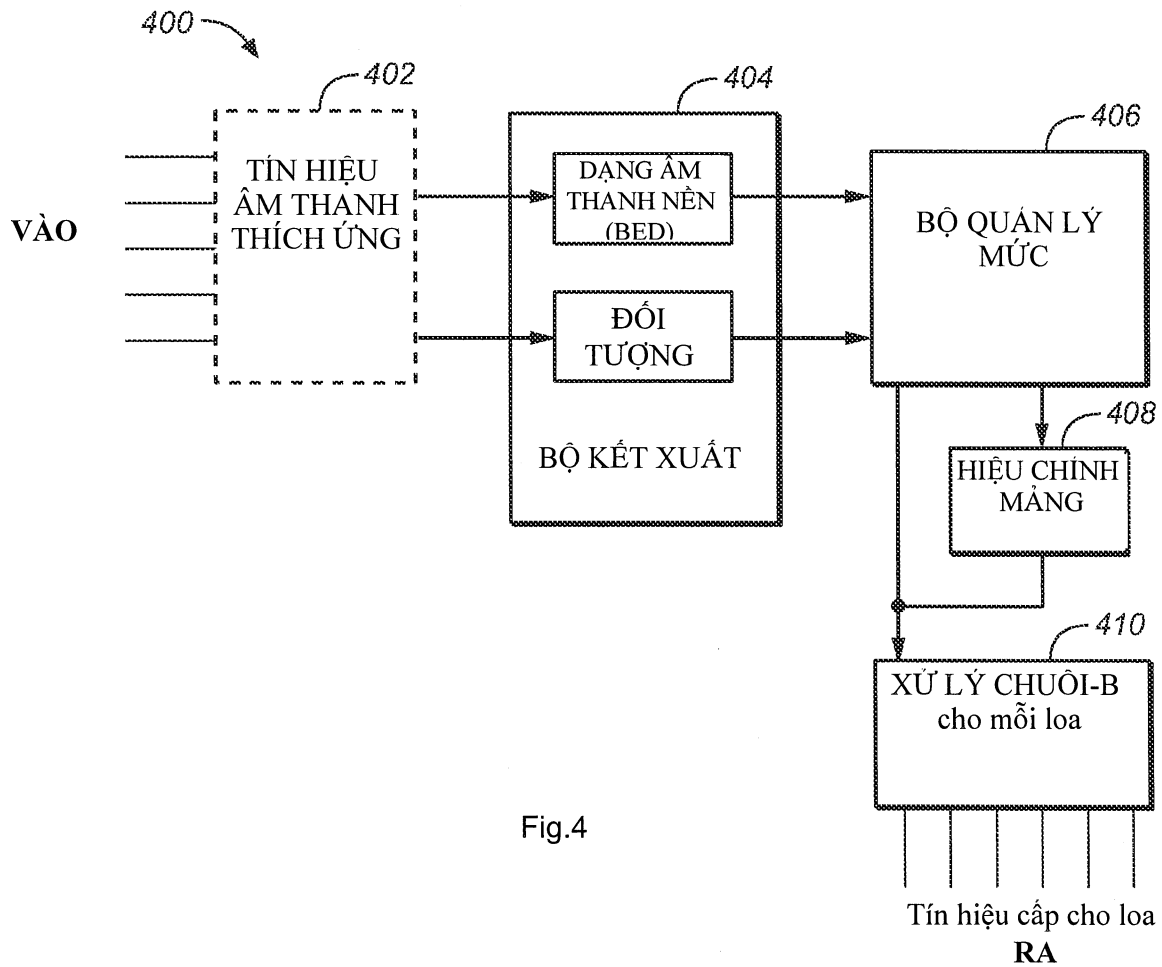


Fig.4

500

| LOẠI SIÊU DỮ LIỆU         | THÀNH PHẦN SIÊU DỮ LIỆU      |
|---------------------------|------------------------------|
| SIÊU DỮ LIỆU CHƯƠNG TRÌNH | Tốc độ Khung                 |
|                           | Đếm Rãnh                     |
|                           | Mô tả Kênh có thể mở rộng    |
|                           | Mô tả giai đoạn trộn         |
| SIÊU DỮ LIỆU ÂM THANH     | Tốc độ Mẫu                   |
|                           | Độ sâu Bit                   |
|                           | Hệ thống mã hóa              |
| SIÊU DỮ LIỆU KẾT XUẤT     | Vị trí Kênh                  |
|                           | Kiểm soát Kết xuất Kênh      |
|                           | Vị trí Đối tượng             |
|                           | Kiểm soát Kết xuất Đối tượng |

Fig.5



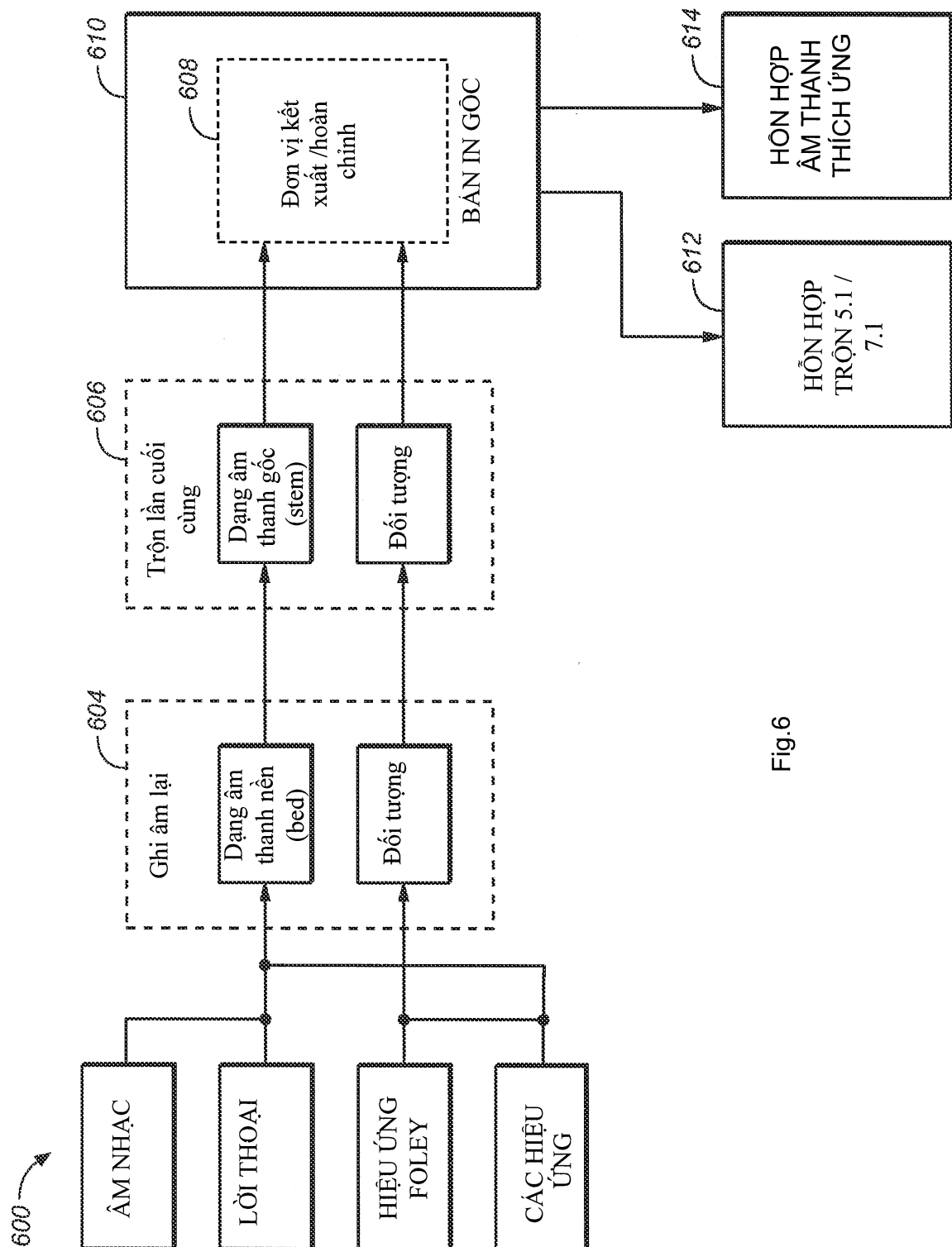


Fig.6

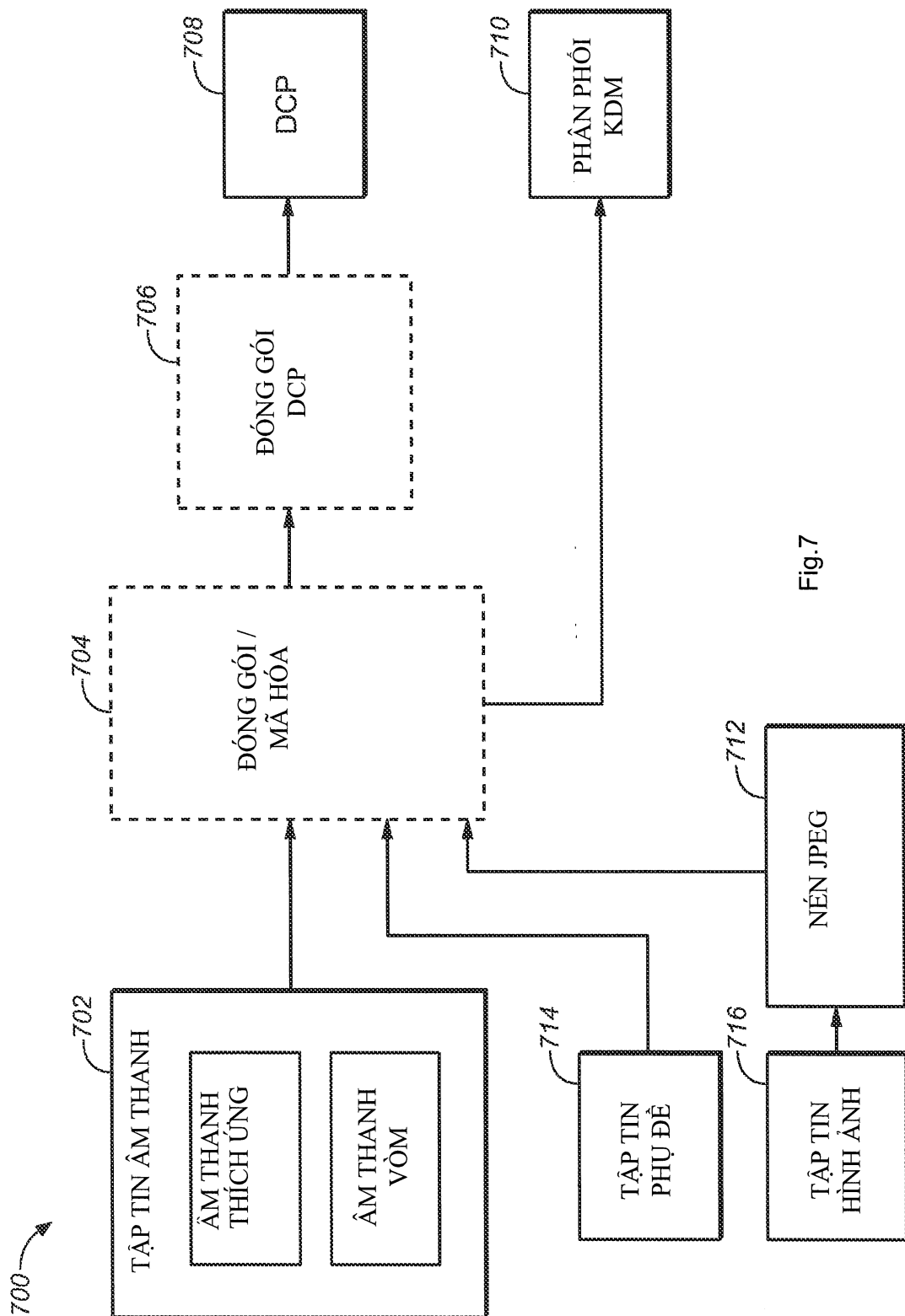


Fig.7

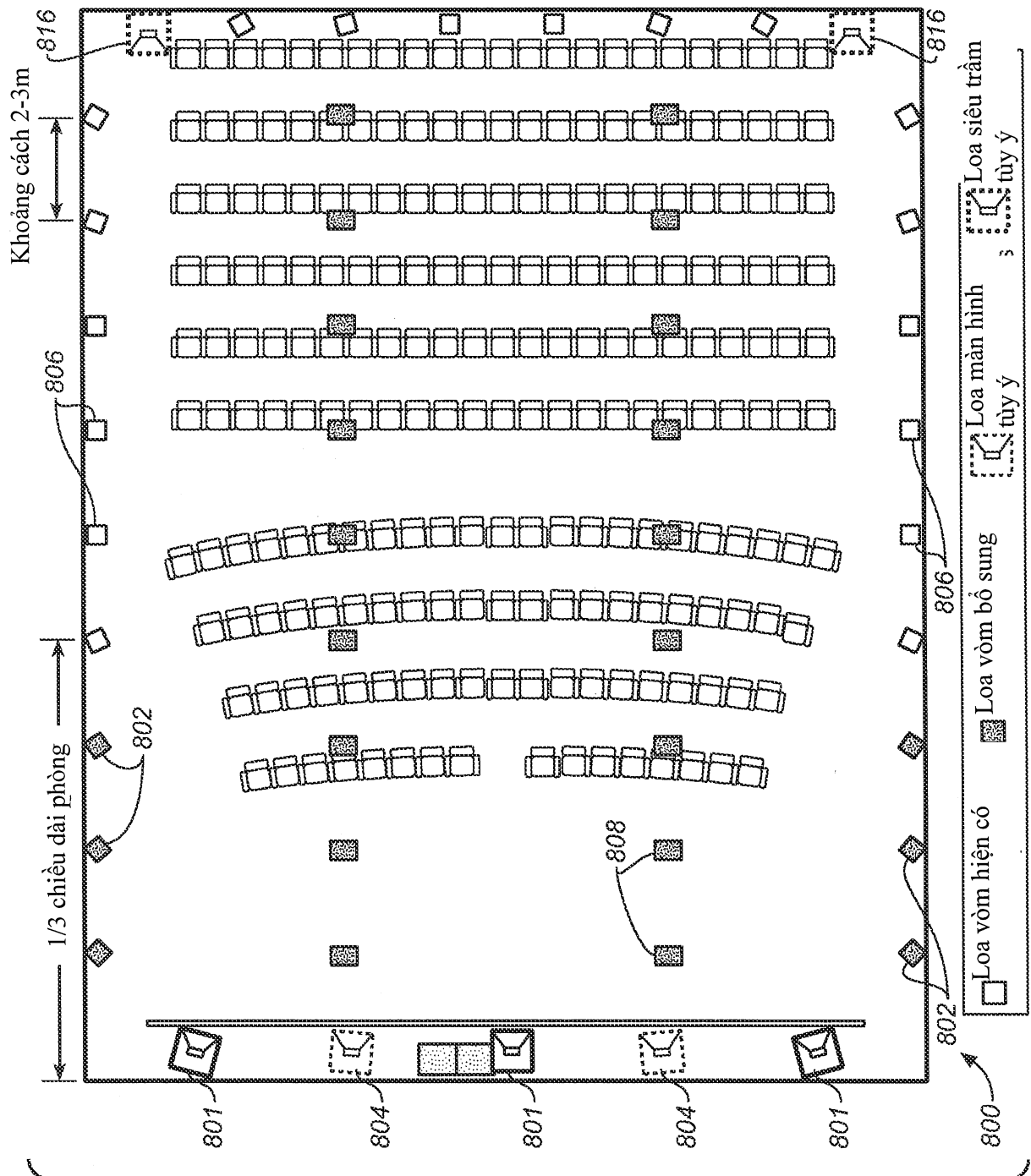


Fig.8

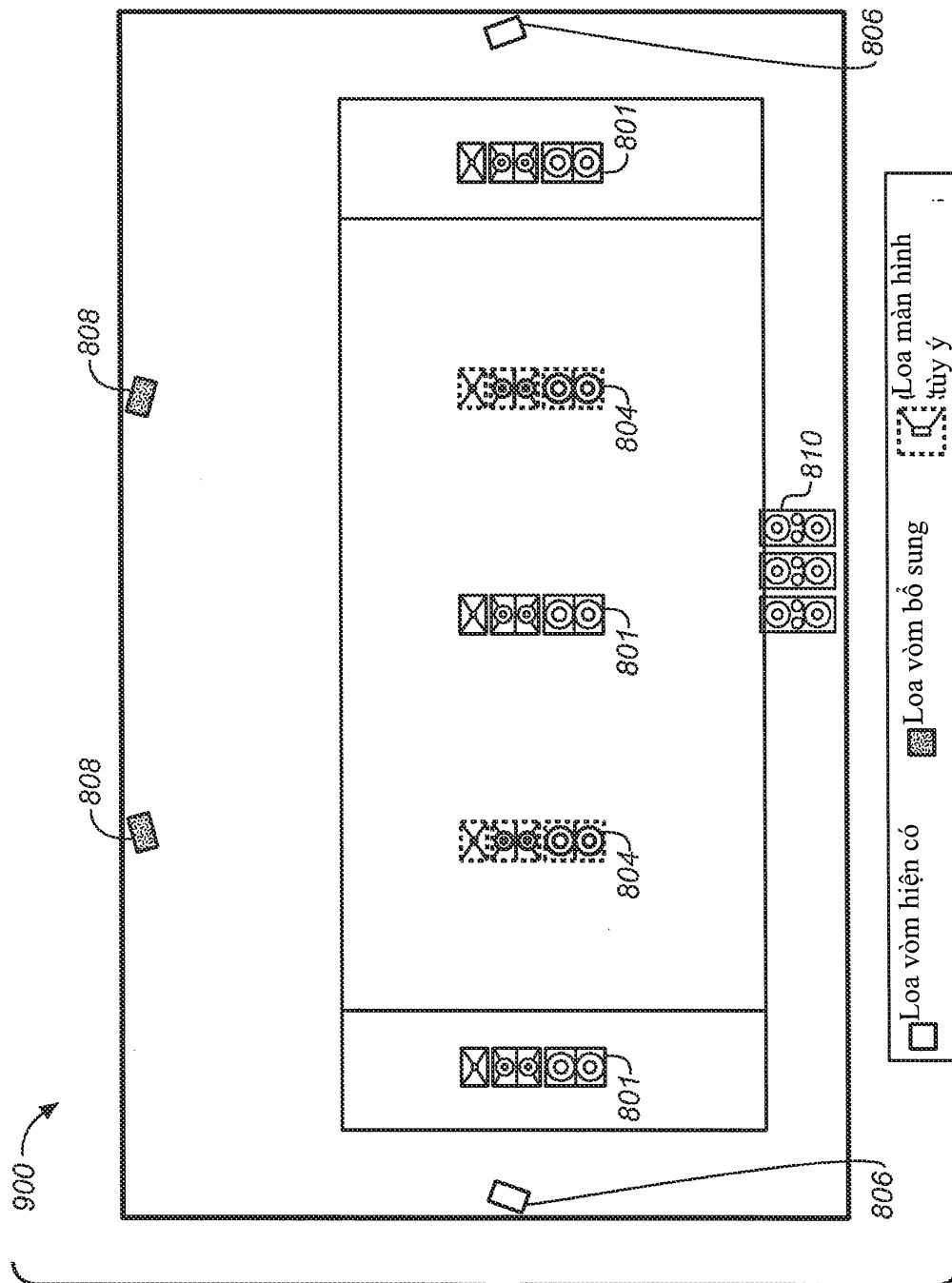


Fig.9

