

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7680571号  
(P7680571)

(45)発行日 令和7年5月20日(2025.5.20)

(24)登録日 令和7年5月12日(2025.5.12)

(51)国際特許分類

F I

G 1 0 L 19/008 (2013.01)

G 1 0 L 19/008

G 1 0 L 19/22 (2013.01)

G 1 0 L 19/22

請求項の数 42 (全73頁)

|                   |                               |          |                                                                                                         |
|-------------------|-------------------------------|----------|---------------------------------------------------------------------------------------------------------|
| (21)出願番号          | 特願2023-573612(P2023-573612)   | (73)特許権者 | 503433420                                                                                               |
| (86)(22)出願日       | 令和4年5月30日(2022.5.30)          |          | 華為技術有限公司                                                                                                |
| (65)公表番号          | 特表2024-521204(P2024-521204 A) |          | HUAWEI TECHNOLOGIES CO., LTD.                                                                           |
| (43)公表日           | 令和6年5月28日(2024.5.28)          |          | 中華人民共和國 5 1 8 1 2 9 広東省深                                                                                |
| (86)国際出願番号        | PCT/CN2022/096025             |          | ▲チェン▼市龍崗区坂田 華為総部▲ベ                                                                                      |
| (87)国際公開番号        | WO2022/253187                 |          | ン▼公樓                                                                                                    |
| (87)国際公開日         | 令和4年12月8日(2022.12.8)          |          | Huawei Administration Building, Bantian, Longgang District, Shenzhen, Guangdong 5 1 8 1 2 9, P.R. China |
| 審査請求日             | 令和5年12月26日(2023.12.26)        |          |                                                                                                         |
| (31)優先権主張番号       | 202110602507.4                | (74)代理人  | 100132481                                                                                               |
| (32)優先日           | 令和3年5月31日(2021.5.31)          |          | 弁理士 赤澤 克豪                                                                                               |
| (33)優先権主張国・地域又は機関 | 中国(CN)                        |          |                                                                                                         |

最終頁に続く

(54)【発明の名称】 三次元音声信号処理方法および装置

(57)【特許請求の範囲】

【請求項1】

三次元音声信号処理方法であって、  
三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得するステップと、  
前記線形分解結果に基づいて、前記現行フレームに対応する音場分類パラメータを取得するステップと、  
前記音場分類パラメータに基づいて、前記現行フレームの音場分類結果を決定するステップと、  
前記音場分類結果に基づいて、前記現行フレームに対応する符号化モード決定するステップと、  
を備え、  
前記音場分類結果に基づいて、前記現行フレームに対応する符号化モードを決定する前記ステップは、  
前記音場分類結果が不均一型音源数を含むか、もしくは前記音場分類結果が前記不均一型音源数および音場種別を含む場合、前記不均一型音源数に基づいて、前記現行フレームに対応する前記符号化モードを決定するステップ、  
前記音場分類結果が前記音場種別を含むか、もしくは前記音場分類結果が前記不均一型音源数および前記音場種別を含む場合、前記音場種別に基づいて、前記現行フレームに対応する前記符号化モードを決定するステップ、または

10

20

前記音場分類結果が前記不均一型音源数および前記音場種別を含む場合、前記不均一型音源数および前記音場種別に基づいて、前記現行フレームに対応する前記符号化モードを決定するステップを含む、  
方法。

【請求項 2】

前記三次元音声信号は、高次アンビソニックス H O A 信号、もしくは一次アンビソニックス F O A 信号を含む、請求項 1 に記載の方法。

【請求項 3】

三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得する前記ステップは、

前記現行フレームに対して特異値分解を実行して、前記現行フレームに対応する特異値を取得するステップであって、前記線形分解結果は前記特異値を含む、ステップ、

前記現行フレームに対して主成分分析を実行して、前記現行フレームに対応する第一の特徴値を取得するステップであって、前記線形分解結果は前記第一の特徴値を含む、ステップ、または

前記現行フレームに対して独立成分分析を実行して、前記現行フレームに対応する第二の特徴値を取得するステップであって、前記線形分解結果は前記第二の特徴値を含む、ステップ

を含む、請求項 1 に記載の方法。

【請求項 4】

複数の線形分解結果が存在し、複数の音場分類パラメータが存在し、

前記線形分解結果に基づいて、前記現行フレームに対応する音場分類パラメータを取得する前記ステップは、

前記現行フレームの  $(i + 1)$  番目の線形解析結果に対する前記現行フレームの  $i$  番目の線形解析結果の比を取得するステップであって、 $i$  は正の整数である、ステップと、

前記比に基づいて、前記現行フレームに対応する  $i$  番目の音場分類パラメータを取得するステップと

を含む、

請求項 1 に記載の方法。

【請求項 5】

複数の音場分類パラメータが存在し、前記音場分類結果が音場種別を含み、

前記音場分類パラメータに基づいて、前記現行フレームの音場分類結果を決定する前記ステップは、

前記複数の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合、前記音場種別は分散型音場であると判定するステップ、または

前記複数の音場分類パラメータの値のうちの少なくとも一つの値が予め設定される不均一型音源判定条件を満たす場合、前記音場種別は不均一型音場であると判定するステップを含む、

請求項 1 に記載の方法。

【請求項 6】

前記分散型音源判定条件は、前記音場分類パラメータの値が予め設定される分散型音源判定閾値未満であることを含むか、または

前記不均一型音源判定条件は、前記音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることを含む、

請求項 5 に記載の方法。

【請求項 7】

複数の音場分類パラメータが存在し、

前記音場分類結果は音場種別を含むか、または前記音場分類結果は不均一型音源数および音場種別を含み、

10

20

30

40

50

前記音場分類パラメータに基づいて、前記現行フレームの音場分類結果を決定する前記ステップは、

前記複数の音場分類パラメータの値に基づいて、前記現行フレームに対応する不均一型音源数を取得するステップと、

前記現行フレームに対応する前記不均一型音源数に基づいて、前記音場種別を決定するステップと

を含む、

請求項 1 に記載の方法。

【請求項 8】

複数の音場分類パラメータが存在し、

10

前記音場分類パラメータは、不均一型音源数を含み、

前記音場分類パラメータに基づいて、前記現行フレームの音場分類結果を決定する前記ステップは、

前記複数の音場分類パラメータの値に基づいて、前記現行フレームに対応する前記不均一型音源数を取得するステップ

を含む、

請求項 1 に記載の方法。

【請求項 9】

前記複数の音場分類パラメータは、 $temp[i]$ 、 $i = 0, 1, \dots, \min(L, K) - 2$  であり、 $L$  は前記現行フレームのチャンネル数を表し、 $K$  は前記現行フレームの各チャンネルに対応する信号点の数を表し、 $\min$  は最小値を選択する演算を表し、

20

前記複数の音場分類パラメータの値に基づいて、前記現行フレームに対応する不均一型音源数を取得する前記ステップは、

$i = 0$  から以下の判定手順、

$temp[i]$  が予め設定される不均一型音源判定閾値を超えるか否かを判定するステップと、

$temp[i]$  が本判定手順における前記不均一型音源判定閾値未満である場合、 $i$  の値を  $i + 1$  に更新し、次の判定手順を実行するステップ、または

$temp[i]$  が本判定手順における前記不均一型音源判定閾値以上である場合、前記判定手順の実行を終了し、本判定手順における  $i$  に 1 を加えた値が前記不均一型音源数に等しいと判定するステップと

30

を順次実行するステップを含む、

請求項 7 に記載の方法。

【請求項 10】

現行フレームに対応する前記不均一型音源数に基づいて、音場種別を決定する前記ステップは、

前記不均一型音源数が第一のプリセット条件を満たす場合、前記音場種別が第一の音場種別であると判定するステップ、または

前記不均一型音源数が第一のプリセット条件を満たさない場合、前記音場種別が第二の音場種別であると判定するステップ

40

を含み、

前記第一の音場種別に対応する不均一型音源数が、前記第二の音場種別に対応する不均一型音源数とは相違する、

請求項 7 に記載の方法。

【請求項 11】

前記第一のプリセット条件は、前記不均一型音源数が第一の閾値を超え、かつ第二の閾値未満であること、および前記第二の閾値が前記第一の閾値を超えることを含むか、または、

前記第一のプリセット条件は、前記不均一型音源数が前記第一の閾値以下であるか、もしくは第二の閾値以上であること、および前記第二の閾値が前記第一の閾値を超えること

50

を含む、

請求項 10 に記載の方法。

【請求項 12】

前記不均一型音源数に基づいて、前記現行フレームに対応する前記符号化モードを決定する前記ステップは、

前記不均一型音源数が第二のプリセット条件を満たす場合、前記符号化モードが第一の符号化モードであると判定するステップ、または

前記不均一型音源数が第二のプリセット条件を満たさない場合、前記符号化モードが第二の符号化モードであると判定するステップ

を含み、

10

前記第一の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、前記第二の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、前記第一の符号化モードおよび前記第二の符号化モードは、相違する符号化モードである、

請求項 1 に記載の方法。

【請求項 13】

前記第二のプリセット条件は、前記不均一型音源数が第一の閾値を超え、かつ第二の閾値未満であること、および前記第二の閾値が前記第一の閾値を超えることを含むか、または、

20

前記第二のプリセット条件は、前記不均一型音源数が前記第一の閾値以下であるか、もしくは前記第二の閾値以上であること、および前記第二の閾値が前記第一の閾値を超えることを含む、

請求項 12 に記載の方法。

【請求項 14】

前記音場種別に基づいて、前記現行フレームに対応する前記符号化モードを決定する前記ステップは、

前記音場種別が不均一型音場である場合、前記符号化モードは仮想スピーカー選択に基づく H O A 符号化モードであると判定するステップ、または

前記音場種別が分散型音場である場合、前記符号化モードは指向性音声コーディングに基づく H O A 符号化モードであると判定するステップ

30

を含む、

請求項 1 に記載の方法。

【請求項 15】

前記音場分類結果に基づいて、前記現行フレームに対応する符号化モードを決定する前記ステップは、

前記現行フレームの音場分類結果に基づいて、前記現行フレームに対応する初期符号化モードを決定するステップと、

前記現行フレームが配置されているハングオーバー時間枠を取得するステップであって、前記ハングオーバー時間枠は、前記現行フレームの前記初期符号化モード、および前記現行フレームより前の  $N - 1$  個のフレームの符号化モードを含み、 $N$  は前記ハングオーバー時間枠の長さである、ステップと、

40

前記現行フレームの前記初期符号化モードと前記ハングオーバー時間枠における  $N - 1$  個のフレームの符号化モードとに基づいて、前記現行フレームの符号化モードを決定するステップと

を含む、

請求項 1 に記載の方法。

【請求項 16】

前記音場分類結果に基づいて、前記現行フレームに対応する符号化パラメータを決定するステップをさらに備える、請求項 1 に記載の方法。

50

## 【請求項 17】

前記符号化パラメータは、仮想スピーカ信号のチャンネル数、残差信号のチャンネル数、仮想スピーカ信号の符号化ビット数、残差信号の符号化ビット数、もしくは最適スピーカを探索するための投票回数のうちの少なくとも一つを含み、

前記仮想スピーカ信号および前記残差信号は、前記三次元音声信号に基づいて生成される、

請求項 16 に記載の方法。

## 【請求項 18】

前記投票回数は、

$$1 - I - d$$

10

の関係を満たし、

I は前記投票回数であり、d は前記音場分類結果に含まれる不均一型音源数である、請求項 17 に記載の方法。

## 【請求項 19】

前記音場分類結果は、不均一型音源数および音場種別を含み、

前記音場種別が不均一型音場である場合、前記仮想スピーカ信号の前記チャンネル数は、

$$F = \min(S, PF)$$

の関係を満たし、

F は前記仮想スピーカ信号の前記チャンネル数であり、S は前記不均一型音源数であり、PF はエンコードによって予め設定される前記仮想スピーカ信号のチャンネル数である

20

、または、

前記音場種別が分散型音場である場合、前記仮想スピーカ信号の前記チャンネル数は、

$$F = 1$$

の関係を満たし、

F は前記仮想スピーカ信号の前記チャンネル数である、

請求項 17 に記載の方法。

## 【請求項 20】

音場種別が分散型音場である場合、前記残差信号の前記チャンネル数は、

$$R = \max(C - 1, PR)$$

の関係を満たし、

30

R は前記残差信号の前記チャンネル数であり、PR はエンコードによって予め設定される前記残差信号のチャンネル数であり、C は前記エンコードによって予め設定される前記残差信号の前記チャンネル数と前記エンコードによって予め設定される前記仮想スピーカ信号のチャンネル数との合計である、または

前記音場種別が不均一型音場である場合、前記残差信号の前記チャンネル数は、

$$R = C - F$$

の関係を満たし、

R は前記残差信号の前記チャンネル数であり、C は前記エンコードによって予め設定される前記残差信号の前記チャンネル数と前記エンコードによって予め設定される前記仮想スピーカ信号の前記チャンネル数との合計であり、F は前記仮想スピーカ信号の前記チャンネル数である、

40

請求項 17 に記載の方法。

## 【請求項 21】

前記音場分類結果は、不均一型音源数を含み、

前記仮想スピーカ信号の前記チャンネル数は、

$$F = \min(S, PF)$$

の関係を満たし、

F は前記仮想スピーカ信号の前記チャンネル数であり、S は前記不均一型音源数であり、PF はエンコードによって予め設定される前記仮想スピーカ信号のチャンネル数である、

請求項 17 に記載の方法。

50

## 【請求項 2 2】

前記残差信号の前記チャンネル数は、

$$F = C - R$$

の関係を満たし、 $R$  は前記残差信号の前記チャンネル数であり、 $C$  はエンコーダによって予め設定される前記残差信号のチャンネル数と前記エンコーダによって予め設定される前記仮想スピーカー信号のチャンネル数との合計であり、 $F$  は前記仮想スピーカー信号の前記チャンネル数である、

請求項 1 7 に記載の方法。

## 【請求項 2 3】

前記音場分類結果は、不均一型音源数を含むか、または、前記音場分類結果は、前記不均一型音源数および音場種別を含み、

10

前記仮想スピーカー信号の前記符号化ビット数は、伝送チャンネルの符号化ビット数に対する前記仮想スピーカー信号の前記符号化ビット数の比に基づいて取得され、

前記残差信号の前記符号化ビット数は、前記伝送チャンネルの前記符号化ビット数に対する前記仮想スピーカー信号の前記符号化ビット数の比によって取得され、

前記伝送チャンネルの前記符号化ビット数は、前記仮想スピーカー信号の前記符号化ビット数、および前記残差信号の前記符号化ビット数を含み、前記不均一型音源数が前記仮想スピーカー信号の前記チャンネル数以下であり、前記伝送チャンネルの前記符号化ビット数に対する前記仮想スピーカー信号の前記符号化ビット数の前記比は、前記伝送チャンネルの前記符号化ビット数に対する前記仮想スピーカー信号の前記符号化ビット数の初期比を増加させることによって取得される、

20

請求項 1 7 に記載の方法。

## 【請求項 2 4】

前記現行フレームおよび前記音場分類結果を符号化するステップと、前記符号化された現行フレームおよび前記音場分類結果をビットストリームに書き込むステップとをさらに備える、請求項 1 に記載の方法。

## 【請求項 2 5】

三次元音声信号処理方法であって

ビットストリームを受信するステップと、

前記ビットストリームを復号化して、現行フレームの音場分類結果を取得するステップと、

30

前記音場分類結果に基づいて、前記復号化された現行フレームの三次元音声信号を取得するステップと

を備え、

前記音場分類結果に基づいて、前記復号化された現行フレームの三次元音声信号を取得する前記ステップは、

前記音場分類結果に基づいて、前記現行フレームの復号化モードを決定するステップと、

前記復号化モードに基づいて、前記復号化された現行フレームの三次元音声信号を取得するステップと

を含み、

40

前記音場分類結果に基づいて、前記現行フレームの復号化モードを決定する前記ステップは、

前記音場分類結果が不均一型音源数を含むか、もしくは前記音場分類結果が不均一型音源数および音場種別を含む場合、前記不均一型音源数に基づいて、前記現行フレームの前記復号化モードを決定するステップ、

前記音場分類結果が音場種別を含むか、もしくは前記音場分類結果が不均一型音源数および音場種別を含む場合、前記音場種別に基づいて、前記現行フレームの前記復号化モードを決定するステップ、または

前記音場分類結果が不均一型音源数および音場種別を含む場合、前記不均一型音源数および前記音場種別に基づいて、前記現行フレームの前記復号化モードを決定するステップ

50

を含む、  
方法。

【請求項 26】

前記不均一型音源数に基づいて、前記現行フレームに対応する前記復号化モードを決定する前記ステップは、

前記不均一型音源数がプリセット条件を満たす場合、前記復号化モードが第一の復号化モードであると判定するステップ、または

前記不均一型音源数がプリセット条件を満たさない場合前記復号化モードが第二の復号化モードであると判定するステップ

を含み、

10

前記第一の復号化モードは、仮想スピーカー選択に基づくH O A復号化モード、もしくは指向性音声コーディングに基づくH O A復号化モードであり、前記第二の復号化モードは、仮想スピーカー選択に基づくH O A復号化モード、もしくは指向性音声コーディングに基づくH O A復号化モードであり、前記第一の復号化モードおよび前記第二の復号化モードは、相違する復号化モードである、

請求項 25 に記載の方法。

【請求項 27】

前記プリセット条件は、前記不均一型音源数が第一の閾値を超え、かつ第二の閾値未満であること、および前記第二の閾値は、前記第一の閾値を超えることを含むか、または、

前記プリセット条件は、前記不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および前記第二の閾値は、前記第一の閾値を超えることを含む、  
請求項 26 に記載の方法。

20

【請求項 28】

前記音場分類結果に基づいて、前記復号化された現行フレームの三次元音声信号を取得する前記ステップは、

前記音場分類結果に基づいて、前記現行フレームの復号化パラメータを決定するステップと、

前記復号化パラメータに基づいて、前記復号化された現行フレームの三次元音声信号を取得するステップと

を含む、請求項 25 に記載の方法。

30

【請求項 29】

前記復号化パラメータは、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の復号化ビット数、もしくは残差信号の復号化ビット数のうちの少なくとも一つを含み、

前記仮想スピーカー信号および前記残差信号は、前記ビットストリームを復号化することによって取得される、

請求項 28 に記載の方法。

【請求項 30】

前記音場分類結果が、不均一型音源数および音場種別を含み、

前記音場種別が不均一型音場である場合、前記仮想スピーカー信号の前記チャンネル数は、  
$$F = \min(S, PF)$$

40

の関係を満たし、F は前記仮想スピーカー信号の前記チャンネル数であり、S は前記不均一型音源数であり、PF はデコードによって予め設定される前記仮想スピーカー信号のチャンネル数であり、または、

前記音場種別が分散型音場である場合、前記仮想スピーカー信号の前記チャンネル数は、  
$$F = 1$$

の関係を満たし、F は前記仮想スピーカー信号の前記チャンネル数である、

請求項 29 に記載の方法。

【請求項 31】

前記音場種別が分散型音場である場合、前記残差信号の前記チャンネル数は、

50

$$R = \max(C - 1, PR)$$

の関係を満たし、

R は前記残差信号の前記チャンネル数であり、PR は前記デコーダによって予め設定される残差信号のチャンネル数であり、C は前記デコーダによって予め設定される前記残差信号の前記チャンネル数と前記デコーダによって予め設定される前記仮想スピーカ信号の前記チャンネル数との合計である、または

前記音場種別が不均一型音場である場合、前記残差信号の前記チャンネル数は、

$$R = C - F$$

の関係を満たし、

R は前記残差信号の前記チャンネル数を表し、C は前記デコーダによって予め設定される前記残差信号の前記チャンネル数と前記デコーダによって予め設定される前記仮想スピーカ信号の前記チャンネル数との合計であり、F は前記仮想スピーカ信号の前記チャンネル数である、

10

請求項 30 に記載の方法。

【請求項 32】

前記音場分類結果は、前記不均一型音源数を含み、

前記仮想スピーカ信号の前記チャンネル数は、

$$F = \min(S, PF)$$

の関係を満たし、

F は前記仮想スピーカ信号の前記チャンネル数であり、S は前記不均一型音源数であり、PF は前記デコーダによって予め設定される前記仮想スピーカ信号の前記チャンネル数である、

20

請求項 30 に記載の方法。

【請求項 33】

前記残差信号の前記チャンネル数は、

$$R = C - F$$

の関係を満たし、R は前記残差信号の前記チャンネル数であり、C はデコーダによって予め設定される前記残差信号のチャンネル数と前記デコーダによって予め設定される前記仮想スピーカ信号のチャンネル数との合計であり、F は前記仮想スピーカ信号の前記チャンネル数である、

30

請求項 29 に記載の方法。

【請求項 34】

前記音場分類結果は前記不均一型音源数を含むか、または前記音場分類結果は前記不均一型音源数および前記音場種別を含み、

前記仮想スピーカ信号の前記復号化ビット数は、伝送チャンネルの復号化ビット数に対する前記仮想スピーカ信号の前記復号化ビット数の比によって取得され、

前記残差信号の前記復号化ビット数は、前記伝送チャンネルの前記復号化ビット数に対する、前記仮想スピーカ信号の前記復号化ビット数の比によって取得され、

前記伝送チャンネルの前記復号化ビット数は、前記仮想スピーカ信号の前記復号化ビット数、および前記残差信号の前記復号化ビット数を含み、前記不均一型音源数が前記仮想スピーカ信号の前記チャンネル数以下である場合、前記伝送チャンネルの前記復号化ビット数に対する前記仮想スピーカ信号の前記復号化ビット数の比は、前記伝送チャンネルの復号化ビット数に対する前記仮想スピーカ信号の前記復号化ビット数の初期比を増加させることによって取得される、

40

請求項 30 に記載の方法。

【請求項 35】

三次元音声信号処理装置であって、

三次元音声信号に対して線形分解を実行して、線形分解結果を取得するように構成される線形解析モジュールと、

前記線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得する

50

ように構成されるパラメータ生成モジュールと、

前記音場分類パラメータに基づいて、前記現行フレームの音場分類結果を決定するように構成される音場分類モジュールと、

前記音場分類結果に基づいて、前記現行フレームに対応する符号化モードを決定するように構成される符号化モード決定モジュールと

備え、

前記符号化モード決定モジュールは、

前記音場分類結果が不均一型音源数を含むか、もしくは前記音場分類結果が前記不均一型音源数および音場種別を含む場合、前記不均一型音源数に基づいて、前記現行フレームに対応する前記符号化モードを決定すること、

前記音場分類結果が前記音場種別を含むか、もしくは前記音場分類結果が前記不均一型音源数および前記音場種別を含む場合、前記音場種別に基づいて、前記現行フレームに対応する前記符号化モードを決定すること、または

前記音場分類結果が前記不均一型音源数および前記音場種別を含む場合、前記不均一型音源数および前記音場種別に基づいて、前記現行フレームに対応する前記符号化モードを決定すること

を行うようにさらに構成される、

三次元音声信号処理装置。

#### 【請求項 36】

三次元音声信号処理装置であって、

ビットストリームを受信するように構成される受信モジュールと、

前記ビットストリームを復号化して、現行フレームの音場分類結果を取得するように構成される復号化モジュールと、

前記音場分類結果に基づいて、前記復号化された現行フレームの三次元音声信号を取得するように構成される信号生成モジュールと

を備え、

前記信号生成モジュールは、

前記音場分類結果に基づいて、前記現行フレームの復号化モードを決定することと、

前記復号化モードに基づいて、前記復号化された現行フレームの三次元音声信号を取得することと

を行うようにさらに構成され、

前記音場分類結果に基づいて、前記現行フレームの復号化モードを決定することは、

前記音場分類結果が不均一型音源数を含むか、もしくは前記音場分類結果が不均一型音源数および音場種別を含む場合、前記不均一型音源数に基づいて、前記現行フレームの前記復号化モードを決定すること、

前記音場分類結果が音場種別を含むか、もしくは前記音場分類結果が不均一型音源数および音場種別を含む場合、前記音場種別に基づいて、前記現行フレームの前記復号化モードを決定すること、または

前記音場分類結果が不均一型音源数および音場種別を含む場合、前記不均一型音源数および前記音場種別に基づいて、前記現行フレームの前記復号化モードを決定すること

を行うようにさらに構成される、

三次元音声信号処理装置。

#### 【請求項 37】

三次元音声信号処理装置であって、前記三次元音声信号処理装置は、少なくとも一つのプロセッサを含み、前記少なくとも一つのプロセッサは、メモリと結合され、前記メモリに格納される命令を読み出し実行して、請求項 1 ないし 24 の何れか一つに記載の方法を実行するように構成される、三次元音声信号処理装置。

#### 【請求項 38】

前記三次元音声信号処理装置は、前記メモリをさらに備える、請求項 37 に記載の三次元音声信号処理装置。

10

20

30

40

50

**【請求項 39】**

三次元音声信号処理装置であって、前記三次元音声信号処理装置は、少なくとも一つのプロセッサを含み、前記少なくとも一つのプロセッサは、メモリと結合され、前記メモリに格納される命令を読み出し実行して、請求項 25 ないし 34 の何れか一つに記載の方法を実装するように構成される、三次元音声信号処理装置。

**【請求項 40】**

前記三次元音声信号処理装置は、前記メモリをさらに備える、請求項 39 に記載の三次元音声信号処理装置。

**【請求項 41】**

コンピュータ可読記憶媒体であって、命令を備え、前記命令がコンピュータ上で実行されると、前記コンピュータは、請求項 1 ないし 24 の何れか一つに記載の方法、または請求項 25 ないし 34 の何れか一つに記載の方法を実行する、コンピュータ可読記憶媒体。

10

**【請求項 42】**

コンピュータ可読記憶媒体であって、請求項 24 に記載の方法に対応する命令がコンピュータ上で実行されると、前記命令の実行結果として前記コンピュータによって生成される前記ビットストリームを含む、コンピュータ可読記憶媒体。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、音声処理技術の分野に関し、特に、三次元音声信号処理方法および装置に関する。

20

**【背景技術】****【0002】**

本出願は、2021年5月31日に中国国家知識産権局に提出された、「三次元音声信号処理方法および装置」と題される中国特許出願第202110602507.4号に対する優先権を主張するものであり、この内容は、その全体が参照により本明細書に組み込まれる。

**【0003】**

三次元音声技術は、無線通信会話、仮想現実／拡張現実、およびメディア音声などに広く利用されている。三次元音声技術は、現実世界における音響イベントおよび三次元音場情報の取得、処理、伝送、レンダリング、および再生を行う音声技術である。三次元音声技術は、強い空間感覚、包み込み感、および没入感を持たせ、非日常的な「没入型」聴覚体験を音に与える。高次アンビソニックス（HOA）技術は、録音、符号化、および再生時におけるスピーカ配置に依存することなく、HOAフォーマットにおけるデータを回転再生する機能を備えている。高次アンビソニックス技術は、三次元音声再生により高い柔軟性を有しており、そのため、より多くの関心および研究がなされている。

30

**【0004】**

撮像装置（例えば、マイクロホンなど）は、大量のデータを取り込んで、三次元音場情報を記録し、三次元音声信号を再生装置（例えば、スピーカもしくはマイクロフォンなど）に送信し、これにより、再生装置は、三次元音声信号を再生する。三次元音場情報のデータ量は大きいと、大きな記憶容量が、そのデータを保存するために必要とされ、また、高い帯域幅が、三次元音声信号を搬送するために必要とされる。前述の課題を解決するために、三次元音声信号を圧縮することがあり、圧縮データを保存もしくは送信し得る。

40

**【0005】**

現在、エンコーダは、事前に構成される複数の仮想スピーカを使用することによって、三次元音声信号を符号化し得る。しかしながら、三次元音声信号を符号化する前、エンコーダは、三次元音声信号を分類することができず、その結果、三次元音声信号を効果的に識別することができない。

**【発明の概要】****【0006】**

50

本出願発明の実施形態は、三次元音声信号の音場分類を実装し、三次元音声信号を正確に識別するための、三次元音声信号処理方法および装置を提供する。

【0007】

前述の技術的課題を解決するために、本出願の実施形態は、以下の技術的解決策を提供する。

【0008】

第一の態様によれば、本出願の一実施形態は、以下を含む三次元音声信号処理方法を提供する。すなわち、三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得するステップ。この線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するステップ。および、この音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップ。前述の解決策では、最初に、三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得する。次いで、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得する。最後に、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定する。本出願の実施形態では、三次元音声信号の現行フレームに対して線形分解を実行して、現行フレームの線形分解結果を取得する。次いで、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得する。そのため、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定し、音場分類結果に基づいて、現行フレームの音場分類を実装することができる。本出願の実施形態では、三次元音声信号に対して音場分類を実行して、三次元音声信号を正確に識別する。

【0009】

可能な実装では、三次元音声信号は、高次アンビソニックスH O A信号、もしくは一次アンビソニックスF O A信号を含む。

【0010】

可能な実装では、三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得するステップは、以下を含む。すなわち、現行フレームに対して特異値分解を実行して、現行フレームに対応する特異値を取得するステップであって、線形分解結果は特異値を含む、ステップ。現行フレームに対して主成分分析を実行して、現行フレームに対応する第一の特徴値を取得するステップであって、線形分解結果は第一の特徴値を含む、ステップ。または、現行フレームに対して独立成分分析を実行して、現行フレームに対応する第二の特徴値を取得するステップであって、線形分解結果は第二の特徴値を含む、ステップ。前述の解決策では、線形分解は特異値分解であり得る。線形分解は、代替的に、特徴値を取得するための主成分分析であり得るか、または線形分解は、代替的に、第二の特徴値を取得するための独立成分分析であり得る。これら三つの方式の何れか一つでは、現行フレームの線形分解が実行されて、後続の音声チャネル決定のための線形解析結果を提供し得る。

【0011】

可能な実装では、複数の線形分解結果が存在し、複数の音場分類パラメータが存在する。線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するステップは、以下を含む。すなわち、現行フレームの $(i + 1)$ 番目の線形解析結果に対する、現行フレームの $i$ 番目の線形解析結果の比を取得するステップであって、 $i$ は正の整数である、ステップ。および、この比に基づいて、現行フレームに対応する $i$ 番目の音場分類パラメータを取得するステップ。

【0012】

さらに、 $i$ 番目の線形解析結果および $(i + 1)$ 番目の線形解析結果は、現行フレームにおける連続する二つの線形解析結果である。

【0013】

前述の解決策では、エンコーダ側は、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得し得る。例えば、現行フレームにおける複数の線形分解結果が存在し、複数の線形解析結果における連続する二つの線形解析結果は、現行フレームの

$i$  番目の線形解析結果および  $(i + 1)$  番目の線形解析結果として表現される。この場合、現行フレームの  $(i + 1)$  番目の線形解析結果に対する、現行フレームの  $i$  番目の線形解析結果の比が算出され得て、 $i$  の具体的な値は特に限定されない。この比が取得された後、現行フレームの  $(i + 1)$  番目の線形解析結果に対する、 $i$  番目の線形解析結果の比に基づいて、現行フレームに対応する  $i$  番目の音場分類パラメータが取得され得る。

#### 【0014】

可能な実装では、複数の音場分類パラメータが存在し、音場分類結果が音場種別を含む。音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下を含む。すなわち、複数の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合、音場種別が分散音であると判定するステップ。または、複数の音場分類パラメータの値のうち少なくとも一つが予め設定される不均一型音源判定条件を満たす場合、音場種別が不均一型音場であると判定するステップ。前述の解決策では、音場種別には、不均一型音場および分散型音場が含まれ得る。本発明の本実施形態では、分散型音源判定条件および不均一型音源判定条件が予め設定される。分散型音源判定条件は、音場種別が分散型音場であるか否かを判定するために使用され、不均一型音源判定条件は、音場種別が不均一型音場であるか否かを判定するために使用される。現行フレームにおける複数の音場分類パラメータが取得された後、複数の音場分類パラメータの値およびブリセット条件に基づいて、判定が実行される。

#### 【0015】

可能な実装では、分散型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値未満であることを含むか、または不均一型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることを含む。前述の解決策では、不均一型音源判定閾値は、予め設定される閾値であり得て、具体的な値は限定されない。分散型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値未満であることを含む。そのため、複数の音場分類パラメータの値が全て予め設定される不均一型音源判定閾値未満である場合、音場種別が分散型音場であると判定される。不均一型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることを含む。そのため、複数の音場分類パラメータの値の少なくとも一つが、予め設定される不均一型音源判定閾値以上である場合、音場種別が不均一型音場であると判定される。

#### 【0016】

可能な実装では、複数の音場分類パラメータが存在し、音場分類結果は、音場種別を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下を含む。すなわち、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップ。および、現行フレームに対応する不均一型音源数に基づいて、音場種別を決定するステップ。前述の解決策では、現行フレームに対応する複数の音場分類パラメータを取得した後、エンコーダ側は、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得し得る。不均一型音源は、異なる位置および/もしくは方向を有する点音源であり、現行フレームに含まれる不均一型音源数は、不均一型音源数と呼ばれる。現行フレームの音場は、不均一型音源数に基づいて分類することができる。現行フレームに対応する不均一型音源数が音場種別を決定するために取得された後、現行フレームに対応する不均一型音源数を分析することによって、現行フレームに対応する音場種別が決定され得る。

#### 【0017】

可能な実装では、複数の音場分類パラメータが存在し、音場分類結果は、不均一型音源数を含む。音場分類パラメータに基づいて、現行フレームの音場分類結果を判定するステップは、以下を含む。すなわち、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップ。前述の解決策では、現行フレームに対応する複数の音場分類パラメータを取得した後、エンコーダ側は、複数の音場分類パラメ

10

20

30

40

50

ータの値に基づいて、現行フレームに対応する不均一型音源数を取得し得る。不均一型音源は、異なる位置および/もしくは方向を有する点音源であり、現行フレームに含まれる不均一型音源数は、不均一型音源数と呼ばれる。

【 0 0 1 8 】

可能な実装では、複数の音場分類パラメータは、 $temp[i]$ 、 $i = 0, 1, \dots, min(L, K) - 2$ であり、 $L$ は現行フレームのチャンネル数を表し、 $K$ は現行フレームの各チャンネルに対応する信号点の数を表し、 $min$ は最小値を選択する動作を表す。複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップは、以下を含む。すなわち、 $i = 0$ から以下の判定手順を順次実行するステップ。 $temp[i]$ が予め設定される不均一型音源判定閾値を超えるか否かを判定するステップ。 $temp[i]$ が本判定手順における不均一型音源判定閾値未満である場合、 $i$ の値を $i + 1$ に更新し、次の判定手順の実行を継続するステップ。または、 $temp[i]$ が本判定手順における不均一型音源判定閾値以上である場合、判定手順の実行を終了し、本判定手順における1を加えた $i$ が不均一型音源数に等しいと判定するステップ。前述の解決策では、判定手順は、複数回の間実行され、その都度、不均一型音源数を取得するように、判定手順の実行を終了するか否かが判定される。

10

【 0 0 1 9 】

可能な実装では、現行フレームに対応する不均一型音源数に基づいて、音場種別を決定するステップは、以下を含む。すなわち、不均一型音源数が第一のプリセット条件を満たす場合、音場種別が第一の音場種別であると判定するステップ。または、不均一型音源数が第一のプリセット条件を満たさない場合、音場種別が第二の音場種別であると判定するステップ。第一の音場種別に対応する不均一型音源数は、第二の音場種別に対応する不均一型音源数とは相違する。前述の解決手段では、不均一型音源数の差異に基づいて、音場種別は、第一の音場種別および第二の音場種別という二種類に分類され得る。エンコーダ側は、プリセット条件を取得する。すなわち、不均一型音源数がプリセット条件を満たすか否かを判定し、不均一型音源数が第一のプリセット条件を満たす場合、音場種別を第一の音場種別と判定し、または不均一型音源数が第一のプリセット条件を満たさない場合、音場種別を第二の音場種別と判定する。本出願の本実施形態では、現行フレームの音場種別の分割を実装し、現行フレームの音場種別が第一の音場種別に属するか、もしくは第二の音場種別に属するかを正確に識別するために、不均一型音源数が第一のプリセット条件を満たすか否かが判定され得る。

20

30

【 0 0 2 0 】

可能な実装では、第一のプリセット条件は、不均一型音源数が第一の閾値を超え、かつ第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または、第一のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値は、第一の閾値を超えることを含む。前述の解決策では、第一の閾値および第二の閾値における具体的な値は制限されないで、用途のシナリオに基づいて具体的に決定され得る。第二の閾値は、第一の閾値を超える。そのため、第一の閾値および第二の閾値は、プリセット範囲を構成し得て、第一のプリセット条件は、不均一型音源数がプリセット範囲内に収まることであってもよいし、または第一のプリセット条件は、不均一型音源数がプリセット範囲を超えることであってもよい。不均一型音源数は、不均一型音源数が第一のプリセット条件を満たすか否かを判定し、現行フレームの音場種別が第一の音場種別もしくは第二の音場種別に属することを正確に識別するように、第一のプリセット条件における第一の閾値および第二の閾値に基づいて決定され得る。

40

【 0 0 2 1 】

可能な実装では、本方法は、以下をさらに含む。すなわち、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップ。前述の解決策では、エンコーダ側は、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定し得る。この符号化モードは、三次元音声信号の現行フレームを符号化する際に使用されるモード

50

である。複数の符号化モードが存在し、現行フレームの異なる音場分類結果に基づいて、異なる符号化モードが使用され得る。本出願の実施形態では、現行フレームの異なる音場分類結果に応じて、適切な符号化モードが選択され、これにより、その符号化モードを使用することによって、現行フレームは符号化される。これは、音声信号の圧縮効率および聴覚品質を改善する。

【 0 0 2 2 】

可能な実装では、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、音場分類結果が不均一型音源数を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームに対応する符号化モードを決定するステップ。音場分類結果が音場種別を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップ。または、音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップ。前述の解決策では、エンコーダ側は、不均一型音源数および/もしくは音場種別に基づいて、現行フレームに対応する符号化モードを決定して、現行フレームの音場分類結果に基づいて、対応する符号化モードを決定し得て、これにより、決定された符号化モードは、三次元音声信号の現行フレームに適用することができる。これは、符号化の効率を改善する。

【 0 0 2 3 】

可能な実装では、不均一型音源数に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、不均一型音源数が第二のプリセット条件を満たす場合、符号化モードが第一の符号化モードであると判定するステップ。または、不均一型音源数が第二のプリセット条件を満たさない場合、符号化モードが第二の符号化モードであると判定するステップ。第一の符号化モードは、仮想スピーカー選択に基づくH O A符号化モード、もしくは指向性音声コーディングに基づくH O A符号化モードであり、第二の符号化モードは、仮想スピーカー選択に基づくH O A符号化モード、もしくは指向性音声コーディングに基づくH O A符号化モードであり、第一の符号化モードおよび第二の符号化モードは、異なる符号化モードである。前述の解決策では、符号化モードは、異なる不均一型音源数に基づいて、第一の符号化モードおよび第二の符号化モードという二種類に分類され得る。エンコーダ側は、第二のプリセット条件を取得する。すなわち、不均一型音源数が第二のプリセット条件を満たすか否かを判定する。および、不均一型音源数が第二のプリセット条件を満たす場合、符号化モードが第一の符号化モードであると判定する。または、不均一型音源数が第二のプリセット条件を満たさない場合、符号化モードが第二の符号化モードであると判定する。本出願の本実施形態では、現行フレームの符号化モードの分割を実装し、現行フレームの符号化モードが第一のプリセット条件に属するか、もしくは第二のプリセット条件に属するかを正確に識別するために、不均一型音源数が第二のプリセット条件を満たすか否かが判定され得る。

【 0 0 2 4 】

可能な実装では、第二のプリセット条件は、不均一型音源数が第一の閾値を超え、かつ第二の閾値未満であること、および第二の閾値が、第一の閾値を超えることを含む。または、第二のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が、第一の閾値を超えることを含む。

【 0 0 2 5 】

可能な実装では、音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、音場種別が不均一型音場である場合、符号化モードが仮想スピーカーに基づくH O A符号化モードであると判定するステップ。または、音場種別が分散型音場である場合、符号化モードが指向性音声コーディングに基づくH O A符号化モードであると判定するステップ。

【 0 0 2 6 】

可能な実装では、音場分類結果に基づいて、現行フレームに対応する符号化モードを決

定するステップは、以下を含む。すなわち、現行フレームの音場分類結果に基づいて、現行フレームに対応する初期符号化モードを決定するステップ。現行フレームが位置するハングオーバー時間枠を取得するステップであって、ハングオーバー時間枠は、現行フレームの初期符号化モードと、現行フレームより前の $N - 1$ 個のフレームの符号化モードとを含み、 $N$ は、ハングオーバー時間枠の長さである、ステップ。および、現行フレームの初期符号化モードと、 $N - 1$ 個のフレームの符号化モードとに基づいて、現行フレームの符号化モードを決定するステップ。前述の解決策では、本出願の本実施形態では、現行フレームの初期符号化モードが、現行フレームの符号化モードを取得するために、ハングオーバー時間枠に基づいて修正される。これは、連続するフレームの符号化モードが頻繁に切り替わらなくなることを確実にして、符号化の効率を改善する。

10

【0027】

可能な実装では、本方法は、以下をさらに含む。すなわち、音場分類結果に基づいて、現行フレームに対応する符号化パラメータを決定するステップ。前述の解決策では、エンコーダ側は、音場分類結果に基づいて、現行フレームに対応する符号化パラメータを決定し得る。この符号化パラメータは、三次元音声信号の現行フレームを符号化する際に使用されるパラメータである。複数の符号化パラメータが存在し、現行フレームの異なる音場分類結果に基づいて、異なる符号化パラメータが使用され得る。本出願の本実施形態では、現行フレームの異なる音場分類結果に対して、適切な符号化パラメータが選択され、これにより、その符号化パラメータに基づいて、現行フレームが符号化される。これは、音声信号の圧縮効率および聴覚品質を改善する。

20

【0028】

可能な実装では、符号化パラメータは、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の符号化ビット数、残差信号の符号化ビット数、もしくは最適スピーカーを探索するための投票回数のうちの少なくとも一つを含む。仮想スピーカー信号および残差信号は、三次元音声信号に基づいて生成される。

【0029】

投票回数は、 $1 \leq I \leq d$  の関係を満たす。 $I$  は投票回数であり、 $d$  は音場分類結果に含まれる不均一型音源数である。前述の解決策では、エンコーダ側は、現行フレームの不均一型音源数に基づいて、最適スピーカーを探索するための投票回数を決定する。投票回数は、現行フレームの不均一型音源数以下であり、これにより、投票回数は、現行フレームの音場分類における実際の状況に適合することができる。これは、現行フレームが符号化される際に、最適スピーカーを探索するための投票回数が決定される必要があるという課題を解決する。

30

【0030】

可能な実装では、音場分類結果には、不均一型音源数および音場種別が含まれる。音場種別が不均一型音場である場合、仮想スピーカー信号のチャンネル数は、 $F = \min(S, PF)$  の関係を満たす。ここで、 $F$  は仮想スピーカー信号のチャンネル数であり、 $S$  は不均一型音源数であり、 $PF$  はエンコーダによって予め設定される仮想スピーカー信号のチャンネル数である。または、音場種別が分散型音場である場合、仮想スピーカー信号のチャンネル数は、 $F = 1$  の関係を満たす。ここで、 $F$  は仮想スピーカー信号のチャンネル数である。前述の解決策では、仮想スピーカー信号のチャンネル数は、仮想スピーカー信号を送信するためのチャンネル数であり、仮想スピーカー信号のチャンネル数は、不均一型音源および音場種別に基づいて決定され得る。前述の計算方式では、音場種別が分散型音場である場合、仮想スピーカー信号のチャンネル数は、現行フレームの符号化効率を改善するために、1であると判定される。音場種別が不均一型音源である場合、 $\min$  は最小値を選択する演算、すなわち仮想スピーカー信号のチャンネル数として、 $S$  および  $PF$  の最小値を選択する演算を表し、これにより、仮想スピーカー信号のチャンネルは、現行フレームの音場分類における実際の状況に適合することができる。これは、現行フレームを符号化する際に、仮想スピーカー信号のチャンネル数が決定される必要があるという課題を解決する。

40

【0031】

50

可能な実装では、音場種別が分散型音場である場合、残差信号のチャンネル数は、 $R = \max(C - 1, PR)$  の関係を満たす。ここで、 $PR$  はエンコーダによって予め設定される残差信号のチャンネル数であり、 $C$  はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカー信号のチャンネル数との合計である。または、音場種別が不均一型音場である場合、残差信号のチャンネル数は、 $R = C - F$  の関係を満たす。ここで、 $R$  は残差信号のチャンネル数であり、 $C$  はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカー信号のチャンネル数との合計であり、 $F$  は仮想スピーカー信号のチャンネル数である。前述の解決策では、仮想スピーカー信号のチャンネル数が取得された後、残差信号のチャンネル数は、残差信号のプリセットチャンネル数と、残差信号のプリセットチャンネル数および仮想スピーカー信号のプリセットチャンネル数の合計とに基づいて計算され得る。 $PR$  の値は、エンコーダ側において予め設定され得て、 $R$  の値は、 $\max(C - 1, PR)$  の計算式に従って取得され得る。残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計は、エンコーダ側において予め設定される。なお、 $C$  は、伝送チャンネルの総数と呼ばれることもある。

10

#### 【0032】

可能な実装では、音場分類結果は、不均一型音源数を含む。仮想スピーカー信号のチャンネル数は、 $F = \min(S, PF)$  の関係を満たす。ここで、 $F$  は仮想スピーカー信号のチャンネル数であり、 $S$  は不均一型音源数であり、 $PF$  はエンコーダによって予め設定される仮想スピーカー信号のチャンネル数である。

20

#### 【0033】

可能な実装では、残差信号のチャンネル数は、 $R = C - F$  の関係を満たす。ここで、 $R$  は残差信号のチャンネル数であり、 $C$  はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカーのチャンネル数との合計であり、 $F$  は仮想スピーカー信号のチャンネル数である。前述の解決策では、仮想スピーカー信号のチャンネル数が取得された後、残差信号のチャンネル数は、仮想スピーカー信号のチャンネル数と、残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計とに基づいて計算され得る。残差信号のプリセットチャンネル数および仮想スピーカー信号のプリセットチャンネル数の合計は、エンコーダ側で予め設定される。なお、 $C$  は伝送チャンネルの総数と呼ばれることもある。

30

#### 【0034】

可能な実装では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。仮想スピーカー信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカー信号の符号化ビット数の比に基づいて取得される。残差信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカー信号の符号化ビット数の比に基づいて取得される。伝送チャンネルの符号化ビット数には、仮想スピーカー信号の符号化ビット数および残差信号の符号化ビット数が含まれ、不均一型音源数が仮想スピーカー信号のチャンネル数以下である場合、伝送チャンネルの符号化ビット数に対する、仮想スピーカー信号の符号化ビット数の比は、伝送チャンネルの符号化ビット数に対して、仮想スピーカー信号の符号化ビット数の初期比を増加させることによって取得される。

40

#### 【0035】

可能な実装では、本方法は、以下をさらに含む。すなわち、現行フレームおよび音場分類結果を符号化するステップ。および、符号化された現行フレームおよび音場分類結果をビットストリームに書き込むステップ。

#### 【0036】

第二の態様によれば、本出願の実施形態は、以下を含む三次元音声信号処理方法をさらに提供する。すなわち、ビットストリームを受信するステップ。ビットストリームを復号化して、現行フレームの音場分類結果を取得するステップ。および、音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するステップ。前述の解決策で

50

は、音場分類結果は、ビットストリームにおける現行フレームを復号化するために使用することができる。そのため、デコーダ側は、現行フレームの音場に適合する復号化方式において復号化を実行して、エンコーダ側から送信された三次元音声信号を取得する。これは、エンコーダ側からデコーダ側への音声信号の伝送を実装する。

【0037】

可能な実装では、音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するステップは、以下を含む。すなわち、音場分類結果に基づいて、現行フレームの復号化モードを決定するステップ。および、復号化モードに基づいて、復号化された現行フレームの三次元音声信号を取得するステップ。

【0038】

可能な実装では、音場分類結果に基づいて、現行フレームの復号化モードを決定するステップは、以下を含む。すなわち、音場分類結果が不均一型音源数を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームの復号化モードを決定するステップ。音場分類結果が音場種別を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームの復号化モードを決定するステップ。または、音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームの復号化モードを決定するステップ。

【0039】

可能な実装では、不均一型音源数に基づいて、現行フレームに対応する復号化モードを決定するステップは、以下を含む。すなわち、不均一型音源数がプリセット条件を満たす場合、復号化モードが第一の復号化モードであると判定するステップ。または、不均一型音源数がプリセット条件を満たさない場合、復号化モードが第二の復号化モードであると判定するステップ。第一の復号化モードは、仮想スピーカー選択に基づくH O A復号化モード、もしくは指向性音声コーディングに基づくH O A復号化モードであり、第二の復号化モードは、仮想スピーカー選択に基づくH O A復号化モード、もしくは指向性音声コーディングに基づくH O A復号化モードであり、第一の復号化モードおよび第二の復号化モードは、相違する復号化モードである。

【0040】

可能な実装では、プリセット条件は、不均一型音源数が第一の閾値を超え、かつ、第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または、プリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

【0041】

可能な実装では、音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するステップは、以下を含む。すなわち、音場分類結果に基づいて、現行フレームの復号化パラメータを決定するステップ。および、復号化パラメータに基づいて、復号化された現行フレームの三次元音声信号を取得するステップ。

【0042】

可能な実装では、復号化パラメータは、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の復号化ビット数、もしくは仮想スピーカー信号の復号化ビット数のうちの少なくとも一つを含む。仮想スピーカー信号および残差信号は、ビットストリームを復号化することによって取得される。

【0043】

可能な実装では、音場分類結果には、不均一型音源数および音場種別が含まれる。音場種別が不均一型音場である場合、仮想スピーカー信号のチャンネル数は、 $F = \min(S, PF)$  の関係を満たす。ここで、 $F$  は仮想スピーカー信号のチャンネル数であり、 $S$  は不均一型音源数であり、 $PF$  はデコーダによって予め設定される仮想スピーカー信号のチャンネル数である。または、音場種別が分散型音場である場合、仮想スピーカー信号のチャンネル数は、 $F = 1$  の関係を満たす。ここで、 $F$  は仮想スピーカー信号のチャンネル数である。

10

20

30

40

50

## 【 0 0 4 4 】

可能な実装では、音場種別が分散型音場である場合、残差信号のチャンネル数は、 $R = \max(C - 1, PR)$  の関係を満たす。ここで、 $PR$  はデコーダによって予め設定される残差信号のチャンネル数であり、 $C$  はデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計である。または、音場種別が不均一型音場である場合、残差信号のチャンネル数は、 $R = C - F$  の関係を満たす。ここで、 $R$  は残差信号のチャンネル数であり、 $C$  はデコーダによって予め設定される残差信号のチャンネル数とデコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$  は仮想スピーカ信号のチャンネル数である。

## 【 0 0 4 5 】

可能な実装では、音場分類結果は、不均一型音源数を含む。仮想スピーカ信号のチャンネル数は、 $F = \min(S, PF)$  の関係を満たす。ここで、 $F$  は仮想スピーカ信号のチャンネル数であり、 $S$  は不均一型音源数であり、 $PF$  はデコーダによって予め設定される仮想スピーカ信号のチャンネル数である。

## 【 0 0 4 6 】

可能な実装では、残差信号のチャンネル数は、 $R = C - F$  の関係を満たす。ここで、 $R$  は残差信号のチャンネル数であり、 $C$  はデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$  は仮想スピーカ信号のチャンネル数である。

## 【 0 0 4 7 】

可能な実装では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。仮想スピーカ信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。残差信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。伝送チャンネルの復号化ビット数は、仮想スピーカ信号の復号化ビット数と、残差信号の復号化ビット数とが含まれ、不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の初期比を増加させることによって取得される。

## 【 0 0 4 8 】

第三の態様によれば、本出願の一実施形態は、以下を含む三次元音声信号処理装置をさらに提供する。すなわち、パラメータ生成モジュールは、三次元音声信号に対して線形分解を実行して、線形分解結果を取得するように構成される線形解析モジュール。線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するように構成されるパラメータ生成モジュール。および、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するように構成される音場分類モジュール。

## 【 0 0 4 9 】

本出願における第三の態様では、三次元音声信号処理装置に含まれるモジュールは、第一の態様および可能な実装において説明されるステップをさらに実行し得る。詳細については、第一の態様および可能な実装の説明を参照されたい。

## 【 0 0 5 0 】

第四の態様によれば、本出願の実施形態は、以下を含む三次元音声信号処理装置をさらに提供する。すなわち、ビットストリームを受信するように構成される受信モジュール。ビットストリームを復号化して、現行フレームの音場分類結果を取得するように構成される復号化モジュール。および、音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するように構成される信号生成モジュール。

## 【 0 0 5 1 】

本出願における第四の態様では、三次元音声信号処理装置に含まれるモジュールは、第二の態様および可能な実装において説明されるステップをさらに実行し得る。詳細につい

10

20

30

40

50

ては、第二の態様および可能な実装の説明を参照されたい。

【0052】

可能な実装では、仮想スピーカ信号の符号化ビット数は、次の関係を満たす。すなわち、

【0053】

【数1】

$$\text{core\_numbit} = \text{round}\left(\text{fac1} * F * \frac{\text{numbit}}{\text{fac1} * F + \text{fac2} * R}\right)$$

10

【0054】

`core_numbit`は、仮想スピーカ信号の符号化ビット数であり、`fac1`は、仮想スピーカ信号の符号化ビットに割り当てられた重み係数であり、`fac2`は、残差信号の符号化ビットに割り当てられた重み係数であり、`round`は、切り捨てを表し、`F`は、仮想スピーカ信号のチャンネル数であり、`R`は、残差信号のチャンネル数を表し、`numbit`は、仮想スピーカ信号の符号化ビット数と残差信号の符号化ビット数との合計である。残差信号の符号化ビット数は、次の関係を満たす。すなわち、

【0055】

【数2】

$$\text{res\_numbit} = \text{numbit} - \text{core\_numbit}$$

20

【0056】

`res_numbit`は、残差信号の符号化ビット数であり、`core_numbit`は、仮想スピーカ信号の符号化ビット数であり、`numbit`は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数の合計である。

【0057】

可能な実装では、

【0058】

【数3】

$$\text{fac1} > \text{fac2}$$

30

【0059】

である。

【0060】

可能な実装では、残差信号の符号化ビット数は、次の関係を満たす。すなわち、

【0061】

【数4】

$$\text{res\_numbit} = \text{round}\left(\text{fac2} * R * \frac{\text{numbit}}{\text{fac1} * F + \text{fac2} * R}\right)$$

40

【0062】

`res_numbit`は、残差信号の符号化ビット数であり、`fac1`は、仮想スピーカ信号の符号化ビットに割り当てられた重み係数であり、`fac2`は、残差信号の符号化ビットに割り当てられた重み係数であり、`round`は、切り捨てを表し、`F`は、仮想スピーカ信号のチャンネル数であり、`R`は残差信号のチャンネル数を表し、`numbit`は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数の合計である。

50

【 0 0 6 3 】

仮想スピーカ信号の符号化ビット数は、次の関係を満たす。すなわち、

【 0 0 6 4 】

【 数 5 】

$$core\_numbit = numbit - res\_numbit$$

【 0 0 6 5 】

$core\_numbit$  は、仮想スピーカ信号の符号化ビット数であり、 $res\_numbit$  は、残差信号の符号化ビット数であり、 $numbit$  は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数の合計である。

10

【 0 0 6 6 】

可能な実装では、各仮想スピーカ信号の符号化ビット数は、次の関係を満たす。すなわち、

【 0 0 6 7 】

【 数 6 】

$$core\_ch\_numbit = round\left(fac1 * \frac{numbit}{fac1 * F + fac2 * R}\right)$$

20

【 0 0 6 8 】

$core\_ch\_numbit$  は、各仮想スピーカ信号の符号化ビット数であり、 $fac1$  は、仮想スピーカ信号の符号化ビットに割り当てられた重み係数であり、 $fac2$  は、残差信号の符号化ビットに割り当てられた重み係数であり、 $round$  は、切り捨てを表し、 $F$  は仮想スピーカ信号のチャンネル数であり、 $R$  は残差信号のチャンネル数を表し、 $numbit$  は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数の合計である。

【 0 0 6 9 】

各残差信号の符号化ビット数は、次の関係を満たす。すなわち、

【 0 0 7 0 】

【 数 7 】

$$res\_ch\_numbit = round\left(fac2 * \frac{numbit}{fac1 * F + fac2 * R}\right)$$

30

【 0 0 7 1 】

$res\_numbit$  は、各残差信号の符号化ビット数であり、 $fac1$  は、仮想スピーカ信号の符号化ビットに割り当てられた重み係数であり、 $fac2$  は、残差信号の符号化ビットに割り当てられた重み係数であり、 $round$  は、は切り捨てを表し、 $F$  は、仮想スピーカ信号のチャンネル数であり、 $R$  は残差信号のチャンネル数を表し、 $numbit$  は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数の合計である。

40

【 0 0 7 2 】

第五の態様によれば、本出願の一実施形態は、コンピュータ可読記憶媒体を提供する。本コンピュータ可読記憶媒体は、命令を格納する。この命令がコンピュータ上で実行されると、そのコンピュータは、第一の態様もしくは第二の態様における方法を実行することが可能になる。

【 0 0 7 3 】

第六の態様によれば、本出願の一実施形態は、命令を含むコンピュータプログラム製品を提供する。本コンピュータプログラム製品がコンピュータ上で実行されると、そのコン

50

コンピュータは、第一の態様もしくは第二の態様における方法を実行することが可能になる。

【0074】

第七の態様によれば、本出願の一実施形態は、第一の態様における方法において生成されるビットストリームを含む、コンピュータ可読記憶媒体を提供する。

【0075】

第八の態様によれば、本出願の一実施形態は、通信装置を提供する。本通信装置は、端末機器もしくはチップなどの、実体を含み得る。本通信装置は、プロセッサおよびメモリを含む。このメモリは、命令を格納するように構成され、プロセッサは、メモリにおける命令を実行するように構成されて、本通信装置が第一の態様もしくは第二の態様の実装の何れか一つにおける方法を実行することを可能にする。

10

【0076】

第九の態様によれば、本出願は、チップシステムを提供する。本チップシステムは、前述の態様における機能を実装する際、例えば、前述の方法におけるデータおよび/もしくは情報の送信または処理を実行する際に、音声エンコーダもしくは音声デコーダのサポートを行うように構成されるプロセッサを含む。可能な設計では、本チップシステムは、メモリをさらに含む。このメモリは、音声エンコーダもしくは音声デコーダに必要となるプログラム命令およびデータを格納するように構成される。本チップシステムは、チップを含んでもよいし、またはチップおよび別の個別コンポーネントを含んでもよい。

【0077】

前述の技術的解決策から、本出願の実施形態には、以下の利点があることが分かる。

20

【0078】

本出願の本実施形態では、最初に、三次元音声信号の現行フレームに対して、線形分解が実行されて、線形分解結果を取得する。次いで、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータが取得される。最後に、音場分類パラメータに基づいて、現行フレームの音場分類結果が決定される。本出願の本実施形態では、三次元音声信号の現行フレームに対して、線形分解が実行されて、現行フレームの線形分解結果を取得する。次いで、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータが取得される。そのため、音場分類パラメータに基づいて、現行フレームの音場分類結果が決定され、音場分類結果に基づいて、現行フレームの音場分類を実装することができる。本出願の本実施形態では、三次元音声信号に対して音場分類が実行されて、三次元音声信号を正確に識別する。

30

【図面の簡単な説明】

【0079】

【図1】本出願の実施形態による、音声処理システムの構成構造を表す模式図である。

【図2a】本出願の実施形態による、音声エンコーダおよび音声デコーダが端末機器に使用される模式図である。

【図2b】本出願の実施形態による、音声エンコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。

【図2c】本出願の実施形態による、音声デコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。

40

【図3a】本出願の実施形態による、マルチチャネルエンコーダおよびマルチチャネルデコーダが端末機器に使用される模式略図である。

【図3b】本出願の実施形態による、マルチチャネルエンコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。

【図3c】本出願の実施形態による、マルチチャネルデコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。

【図4】本出願の実施形態による、三次元音声信号処理方法を表す模式図である。

【図5】本出願の実施形態による、三次元音声信号処理方法を表す模式図である。

【図6】本出願の実施形態による、三次元音声信号処理方法を表す模式図である。

【図7】本出願の実施形態による、三次元音声信号処理方法を表す模式図である。

50

【図 8】本出願の実施形態による、ハイブリッド H O A エンコーダの符号化を表す模式的フローチャートである。

【図 9】本出願の実施形態による、H O A 信号の符号化モードの決定を表す模式的フローチャートである。

【図 10】本出願の実施形態による、ハイブリッド H O A デコーダの復号化を表す模式的フローチャートである。

【図 11】本出願の実施形態による、M P ベースの H O A エンコーダの符号化を表す模式的フローチャートである。

【図 12】本出願の実施形態による、音声符号化装置の構成構造を表す模式図である。

【図 13】本出願の実施形態による、音声復号化装置の構成構造を表す模式図である。

10

【図 14】本出願の実施形態による、別の音声符号化装置の構成構造を表す模式図である。

【図 15】本出願の実施形態による、別の音声復号化装置の構成構造を表す模式図である。

【発明を実施するための形態】

【0080】

以下、図面を参照しつつ、本出願の実施形態について説明する。

【0081】

本出願の明細書、請求項、および添付図面では、「第一 ( f i r s t ) 」および「第二 ( s e c o n d ) 」などの用語は、同様の対象を区別することが意図されているが、必ずしも特定の順序もしくは配列を示していない。このような態様に使用される用語は、適切な状況に応じて交換可能であり、これは、本出願の実施形態において同じ属性を有する対象を説明する際に使用される、単なる識別態様に過ぎないと理解されるべきである。さらに、用語「含む ( i n c l u d e ) 」、「含む ( c o n t a i n ) 」、および他の任意の変形は、非排他的な包含をカバーすることを意味しており、一連のユニットを含む、プロセス、方法、システム、製品、もしくは機器は、必ずしもこれらに限定されないが、明示的に列挙されていない、またはそのようなプロセス、方法、システム、製品、または機器に固有となる他の単位が含まれる。

20

【0082】

音 ( s o u n d ) は、物体の振動によって生成される連続的な波である。振動によって音波を発する物体は、音源と呼ばれる。音波が媒体 (例えば、空気、固体、もしくは液体など) を介して伝播すると、人間もしくは動物の聴覚器官は、その音を感知することができる。

30

【0083】

音波の特徴には、音調、音響強度、および音色が含まれる。音調は、音の高さを表す。音響強度は、音の強さを表す。音響強度は、音圧もしくは音数とも呼ばれる。音響強度の単位は、デシベル ( d B ) である。音色は、音質とも呼ばれる。

【0084】

音波の周波数は、音調のピッチを決める。周波数が高いほど、ピッチが高くなる。物体が 1 秒間に振動する回数は、周波数と呼ばれ、周波数の単位は、ヘルツ ( H z ) である。人間の耳によって認識される音の周波数は、20 H z から 20 , 000 H z まで及ぶ。

【0085】

音波の振幅は音の強さは、音響強度の強さを決める。振幅が大きいほど、音響強度が大きいことを示す。音源からの距離が近いほど、音響強度が大きいことを示す。

40

【0086】

音波の波形は、音色を決める。音波の波形には、方形波、ノコギリ波、サイン波、およびパルス波が含まれる。

【0087】

音は、音波の特徴に基づいて、規則的な音および不規則な音に分けられる。不規則音とは、音源の不規則な振動によって生成される音である。不規則な音とは、例えば、人間の仕事、勉強、および休憩などに影響を与える騒音である。規則的な音とは、音源の規則的な振動によって生成される音である。規則的な音には、会話および音楽が含まれる。音は

50

電氣的に表現されると、規則的な音は、時間 - 周波数領域において連続的に変化する、アナログ信号である。このアナログ信号は、音声信号（音響信号）と呼ばれることもある。音声信号は、会話、音楽、および効果音を搬送する情報担体である。

【 0 0 8 8 】

人間の聴覚は、空間における音源の位置分布を識別することができるため、空間において音を聴取する場合、傾聴者は、音の音調、音響強度、および音色だけでなく、音の位置も感知することができる。

【 0 0 8 9 】

聴覚システム体験に対する注目および品質要件の高まりに伴い、音の縦方向の奥行き、没入感、および空間の感覚を高めるための三次元音声技術が登場している。そのため、傾聴者は、前後左右の音源から発せられる音を聴取し、傾聴者が位置する空間が、その音源によって生成される空間音場（音場と呼ばれる）によって囲まれているように感じ、音が周囲に広がるように感じることができる。三次元音声技術は、映画館もしくはコンサートホールなどの、場所にいるかのように傾聴者に感じさせる、「没入型」のあるステレオ効果を生み出す。

【 0 0 9 0 】

三次元音声技術とは、人間の耳の外にある空間がシステムとして想定され、鼓膜によって受け取られる信号が耳の外にあるシステムによって、音源から発せられる音をフィルタ抽出および出力することによって取得される、三次元音声信号となる技術である。例えば、人間の耳の外にあるシステムは、システム衝撃応答  $h(n)$  として定義され得て、任意の音源は、 $x(n)$  として定義され得て、鼓膜によって受け取られる信号は、 $x(n)$  および  $h(n)$  の畳み込み結果である。本出願の実施形態では、三次元音声信号は、高次アンビソニックス（HOA）信号、もしくは一次アンビソニックス（FOA）信号であり得る。三次元音声は、三次元音響効果、空間音声、三次元音場再構成、仮想 3D 音声、またはバイノーラル音声などと呼ばれることもある。

【 0 0 9 1 】

音波は、波数  $k = w / c$ 、および角周波数  $w = 2 \pi f$  を有する理想的な媒体中を伝播する。 $f$  は、音波の周波数であり、 $c$  は、音速である。音圧は、式（1）を満たし、

【 0 0 9 2 】

【数 8】

$$\nabla^2$$

【 0 0 9 3 】

は、ラプラス演算子である。

【 0 0 9 4 】

【数 9】

$$\nabla^2 p + k^2 p = 0 \quad (1)$$

【 0 0 9 5 】

人間の耳の外側にある空間系は、球体であり、傾聴者は、その球体の中心にいと仮定される。球体の外側からの音は、球体の表面に投影され、球体の外側の音は、フィルタ抽出される。音源は、球面上に分布していると仮定される。球体の表面上において音源によって生成される音場は、元の音源によって生成される音場のフィッティングを行うために使用され、すなわち、三次元音声技術は、音場フィッティング法である。具体的には、球面座標系において式（1）の方程式を解き、受動球面領域において式（1）の方程式を次の式（2）のように解く。すなわち、

【 0 0 9 6 】

【数 10】

$$p(r, \theta, \varphi, k) = s \sum_{m=0}^{\infty} (2m+1) j_m^{kr}(kr) \sum_{0 \leq n \leq m, \sigma=\pm 1} Y_{m,n}^{\sigma}(\theta_s, \varphi_s) Y_{m,n}^{\sigma}(\theta, \varphi) \quad (2)$$

【 0 0 9 7 】

r は球面半径を表し、θ は水平角を表し、φ は仰角を表し、k は波数を表し、s は理想平面波の振幅を表し、m は次数番号（H O A 信号の次数とも呼ばれる）を表す。

【 0 0 9 8 】

【 数 1 1 】

$$j^m j_m^{kr}(kr)$$

10

【 0 0 9 9 】

は、球面ベッセル関数を表し、球面ベッセル関数は、放射基底関数とも呼ばれ、最初の j は、虚数単位を表し、

【 0 1 0 0 】

【 数 1 2 】

$$(2m+1) j^m j_m^{kr}(kr)$$

20

【 0 1 0 1 】

は、角度によって変化しない。

【 0 1 0 2 】

【 数 1 3 】

$$Y_{m,n}^{\sigma}(\theta, \varphi)$$

30

【 0 1 0 3 】

は、θ、φ の方向における球面調和関数を表し、

【 0 1 0 4 】

【 数 1 4 】

$$Y_{m,n}^{\sigma}(\theta_s, \varphi_s)$$

【 0 1 0 5 】

は、音源の方における球面調和関数を表す。三次元音声信号の係数は、式（ 3 ）を満たす。すなわち、

【 0 1 0 6 】

【 数 1 5 】

$$B_{m,n}^{\sigma} = s \cdot Y_{m,n}^{\sigma}(\theta_s, \varphi_s) \quad (3)$$

40

【 0 1 0 7 】

式（ 3 ）を式（ 2 ）に代入し、式（ 2 ）を式（ 4 ）に変形することができる。

【 0 1 0 8 】

【 数 1 6 】

50

$$p(r, \theta, \varphi, k) = \sum_{m=0}^{\infty} j^m j_m^k(kr) \sum_{0 \leq n \leq m, \sigma=\pm 1} B_{m,n}^{\sigma} Y_{m,n}^{\sigma}(\theta, \varphi) \quad (4)$$

【 0 1 0 9 】

【 数 1 7 】

$B_{m,n}^{\sigma}$

10

【 0 1 1 0 】

は、N 次三次元音声信号の係数を表し、音場を近似的に記述するために使用される。音場とは、媒体において音波が存在する領域である。N は、1 以上の整数である。例えば、N の値は、2 から 6 まで範囲の整数である。本出願の実施形態における三次元音声信号の係数は、H O A 係数もしくはアンビソニック係数であり得る。

【 0 1 1 1 】

三次元音声信号は、音場における音源の空間的位置情報を搬送する情報担体であり、空間における傾聴者の音場を記述する。式 ( 4 ) は、音場が球面調和関数として球面上に展開できること、すなわち音場が、複数の平面波の重なりに分解することができることを示している。そのため、三次元音声信号によって記述される音場は、複数の平面波の重ね合わせを使用することによって表現することができ、三次元音声信号の係数に基づいて、音場を再構成することができる。

20

【 0 1 1 2 】

5 . 1 チャネル音声信号もしくは 7 . 1 チャネル音声信号と比較して、N 次 H O A 信号は、 $(N + 1)^2$  個のチャネルを有する。そのため、H O A 信号には、音場の空間情報を記述するために使用される大量のデータが含まれる。収集機器 (例えば、マイクロフォンなど) が三次元音声信号を再生機器 (例えば、スピーカーなど) に送信する場合、大きな帯域幅を消費する必要がある。現在、エンコーダは、空間スクイーズドサラウンド音声コーディング (S 3 A C) 法、指向性音声コーディング (D i r A C) 法、もしくは仮想スピーカー選択に基づく符号化法を使用することによって、三次元音声信号を圧縮および符号化して、ビットストリームを取得し、そのビットストリームを再生機器に送信し得る。仮想スピーカー選択に基づく符号化法は、マッチ投影 (M P) 符号化法と呼ばれることもある。以下では、仮想スピーカー選択に基づく符号化法を説明のための例として使用する。再生装置は、ビットストリームを復号化し、三次元音声信号を再構成し、再構成された三次元音声信号を再生する。これは、三次元音声信号を再生装置に送信するためのデータ量および帯域幅占有を低減する。

30

【 0 1 1 3 】

三次元音声信号については、現状では、三次元音声信号の音場を分類することができない。三次元音声信号の音場を如何に分類するかは、本出願の実施形態において解決されるべき技術的課題である。本出願の実施形態では、三次元音声信号の音場分類を実装するために、三次元音声信号に対して線形分解が実行される。これは、三次元音声信号の音場分類を正確に実装し、現行フレームの音場分類結果を取得することができる。

40

【 0 1 1 4 】

また、現行のエンコーダでは、三次元音声信号を圧縮化および符号化する場合、高い圧縮率を取得することができない。そのため、異なる音場の三次元音声信号に対して圧縮符号化を実行するために、圧縮率を如何に高めるかは、本出願の実施形態において解決されるべき別の課題となる。

【 0 1 1 5 】

本出願の一実施形態は、音声符号化技術を提供し、特に、三次元音声信号を対象とした三次元音声符号化技術を提供する。具体的には、従来の音声符号化システムを改善するた

50

めに、より少ないチャネル数を使用することによって、三次元音声信号を表現する符号化技術を提供する。音声コーディング（または一般にコーディングと呼ばれる）には、音声符号化および音声復号化というの二つの部分が含まれる。音声符号化は、伝送元側で実行され、元の音声进行处理（例えば、圧縮など）することを含んで、音声を表現するために必要とされるデータ量を削減する。これは、保存および／もしくは伝送の効率を改善する。音声復号化は、伝送先側で実行され、エンコーダに対する逆処理を含んで、元の音声を再構成する。符号化部分および復号化部分は、コーディングとも呼ばれる。以下に、添付の図面を参照して、本出願の実施形態の実装を詳細に説明する。

#### 【0116】

本出願の実施形態における技術的解決策は、種々の音声処理システムに適用され得る。図1は、本出願の一実施形態による、音声処理システムの構成構造を示す模式図である。音声処理システム100は、音声符号化装置101および音声復号化装置102を含み得る。音声符号化装置101は、ビットストリームを生成するように構成され得る。その後、音声符号化ビットストリームは、音声伝送チャネルを通じて音声復号化装置102に伝送され得る。音声復号化装置102は、ビットストリームを受信し、次いで、音声復号化装置102の音声復号化機能を実行して、再構成された信号を取得し得る。

#### 【0117】

本出願の本実施形態では、音声符号化装置は、音声通信を必要とする各種端末機器、ならびにトランス符号化を必要とする無線装置およびコアネットワーク装置に利用され得る。例えば、音声符号化装置は、端末機器、無線機器、もしくはコアネットワーク機器の音声エンコーダであり得る。同様に、音声復号化装置は、音声通信を必要とする各種の端末機器、ならびにトランス符号化を必要とする無線機器およびコアネットワーク機器に利用され得る。例えば、音声復号化装置は、端末機器、無線機器、もしくはコアネットワーク機器の音声デコーダであり得る。例えば、音声エンコーダは、無線アクセスネットワーク、コアネットワークにおけるメディアゲートウェイ、トランス符号化機器、メディアリソースサーバ、移動端末、および固定ネットワーク端末などを含み得る。あるいは、音声エンコーダは、仮想現実（VR）ストリーミングメディアサービスに使用される音声エンコーダであり得る。

#### 【0118】

本出願の本実施形態では、仮想現実ストリーミング（VRストリーミング）メディアサービスに適用可能な音声コーディング（音声符号化および音声復号化）モジュールが、例として使用される。エンドツーエンドの音声信号処理手順は、以下を含む。すなわち、音声信号Aが収集モジュールを通過した後、前処理（音声前処理）の動作が実行される。前処理操作は、以下を含む。すなわち、信号の低周波数部分をフィルタ抽出するステップであって、フィルタ抽出は、境界点として20Hzもしくは50Hzを使用することによって実行され得る、ステップ。および、信号の方位情報を抽出するステップ。その後、符号化（音声符号化）およびカプセル化（ファイル／セグメントのカプセル化）が実行され、デコーダ側に信号が受け渡される（デリバリー）。デコーダ側は、最初に、カプセル化解除（ファイル／セグメントのカプセル化解除）を実行し、次いで、復号化（音声復号化）を実行し、復号化された信号に対してバイノーラルレンダリング（音声レンダリング）を実行する。レンダリングを通じて取得される信号は、傾聴者のヘッドセット（ヘッドフォン）へのマッピングが行われ、ヘッドセットは、独立したヘッドセット、もしくはメガネデバイス上のヘッドセットであり得る。

#### 【0119】

図2aは、本出願の一実施形態による、音声エンコーダおよび音声デコーダが端末機器に使用される模式図である。各端末機器は、音声エンコーダ、チャネルエンコーダ、音声デコーダ、およびチャネルデコーダを含み得る。具体的には、チャネルエンコーダは、音声信号に対してチャネル符号化を実行するように構成され、チャネルデコーダは、音声信号に対してチャネル復号化を実行するように構成される。例えば、第一の端末機器20は、第一の音声エンコーダ201、第一のチャネルエンコーダ202、第一の音声デコーダ

203、および第一のチャンネルデコーダ204を含み得る。第二の端末機器21は、第二の音声デコーダ211、第二のチャンネルエンコーダ212、第二の音声デコーダ213、および第二のチャンネルデコーダ214を含み得る。第一の端末機器20は、無線もしくは有線の第一のネットワーク通信機器22に接続され、第一のネットワーク通信機器22は、無線または有線の第二のネットワーク通信機器23に接続され、第二の端末機器21は、無線または有線の第二のネットワーク通信機器23に接続される。無線または有線のネットワーク通信機器は、一般に、信号伝送機器、例えば、通信基地局もしくはデータ交換機器であり得る。

#### 【0120】

音声通信では、送信端として機能する端末機器は、最初に、音声収集を実行し、収集された音声信号に対して音声符号化を実行し、次いで、チャンネル符号化を実行し、無線ネットワークもしくはコアネットワークを介して、デジタルチャンネルにて符号化された信号を送信する。受信端として機能する端末機器は、受信信号に基づいてチャンネル復号化を実行して、ビットストリームを取得し、次いで、音声復号化を介して音声信号を復元する。受信端にある端末機器は、音声再生を実行する。

10

#### 【0121】

図2bは、本出願の一実施形態による、音声エンコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。無線機器もしくはコアネットワーク機器25は、以下を含む。すなわち、チャンネルデコーダ251、別の音声デコーダ252、本出願の本実施形態において提供される音声エンコーダ253、およびチャンネルエンコーダ254。別の音声デコーダ252は、音声デコーダ以外の別の音声デコーダである。無線機器もしくはコアネットワーク機器25では、最初に、チャンネルデコーダ251がその機器に入力する信号に対してチャンネル復号化を実行し、次いで、別の音声デコーダ252が音声復号化を実行する。その後、本出願の本実施形態において提供される音声エンコーダ253が音声符号化を実行し、最後に、チャンネルエンコーダ254が、音声信号に対してチャンネル符号化を実行し、次いで、チャンネル符号化が完了した後、符号化された音声信号を送信する。別の音声デコーダ252は、チャンネルデコーダ251によって復号化されたビットストリームに対して音声復号化を実行する。

20

#### 【0122】

図2cは、本出願の一実施形態による、音声デコーダが無線機器もしくはコアネットワーク機器に使用される模式図である。無線機器もしくはコアネットワーク機器25は、以下を含む。すなわち、チャンネルデコーダ251、本出願の本実施形態において提供される音声デコーダ255、別の音声エンコーダ256、およびチャンネルエンコーダ254。別の音声エンコーダ256は、音声エンコーダ以外の別の音声エンコーダである。無線機器もしくはコアネットワーク機器25では、最初に、チャンネルデコーダ251がその機器に入力する信号に対してチャンネル復号化を実行し、次いで、音声デコーダ255が受信された音声符号化ビットストリームを復号化する。その後、別の音声エンコーダ256が音声符号化を実行し、最後に、チャンネルエンコーダ254が音声信号に対してチャンネル符号化を実行し、次いで、チャンネル符号化が完了した後、符号化された音声信号を送信する。無線装置もしくはコアネットワーク装置では、トランス符号化を実装する必要がある場合、対応する音声符号化処理を実行する必要がある。無線機器は、通信における高周波関連機器であり、コアネットワーク機器は、通信におけるコアネットワーク関連機器である。

30

40

#### 【0123】

本出願の幾つかの実施形態では、音声符号化装置は、音声通信を必要とする各種端末機器、ならびにトランス符号化を必要とする無線装置およびコアネットワーク装置に利用され得る。例えば、音声符号化装置は、端末機器、無線装置、もしくはコアネットワーク装置のマルチチャンネルエンコーダであり得る。同様に、音声復号化装置は、音声通信を必要とする各種端末機器、ならびにトランス符号化を必要とする無線装置およびコアネットワーク装置に利用され得る。例えば、音声復号化装置は、端末機器、無線装置、もしくはコアネットワーク装置のマルチチャンネルデコーダであり得る。

50

## 【 0 1 2 4 】

図 3 a は、本出願の一実施形態による、端末機器へのマルチチャンネルエンコードおよびマルチチャンネルデコードの適用を示す模式図である。各端末機器は、マルチチャンネルエンコード、チャンネルエンコード、マルチチャンネルデコード、およびチャンネルデコードを含み得る。マルチチャンネルエンコードは、本出願の実施形態において提供される音声符号化法を実行し得て、マルチチャンネルデコードは、本出願の実施形態において提供される音声復号方法を実行し得る。具体的には、チャンネルエンコードは、マルチチャンネル信号に対してチャンネル符号化を実行するように構成され、チャンネルデコードは、マルチチャンネル信号に対してチャンネル復号化を実行するように構成される。例えば、第一の端末機器 3 0 は、第一のマルチチャンネルエンコード 3 0 1、第一のチャンネルエンコード 3 0 2、第一のマルチチャンネルデコード 3 0 3、および第一のチャンネルデコード 3 0 4 を含み得る。第二の端末機器 3 1 は、第二のマルチチャンネルエンコード 3 1 1、第二のチャンネルエンコード 3 1 2、第二のマルチチャンネルデコード 3 1 3、および第二のチャンネルデコード 3 1 4 を含み得る。第一の端末機器 3 0 は、無線もしくは有線の第一のネットワーク通信機器 3 2 に接続され、第一のネットワーク通信機器 3 2 は、デジタルチャンネルを介して無線もしくは有線の第二のネットワーク通信機器 3 3 に接続され、第二の端末機器 3 1 は、無線もしくは有線の第二のネットワーク通信機器 3 3 に接続される。無線もしくは有線のネットワーク通信機器は、一般に、信号伝送機器、例えば、通信基地局もしくはデータ交換機器であり得る。音声通信では、送信端として機能する端末機器は、収集されたマルチチャンネル信号に対してマルチチャンネル符号化を実行し、次いで、チャンネル符号化を実行し、無線ネットワークもしくはコアネットワークを介して、デジタルチャンネルにて符号化された信号を送信する。受信端として機能する端末機器は、受信信号に基づいて、チャンネル復号化を実行して、マルチチャンネル信号符号化のビットストリームを取得し、次いで、マルチチャンネル復号化を介してマルチチャンネル信号を復元する。受信端にある端末機器は、再生を実行する。

## 【 0 1 2 5 】

図 3 b は、本出願の一実施形態による、無線機器もしくはコアネットワーク機器へのマルチチャンネルエンコードの適用を示す模式図である。無線機器もしくはコアネットワーク機器 3 5 は、以下を含む。すなわち、チャンネルデコード 3 5 1、別の音声デコード 3 5 2、マルチチャンネルエンコード 3 5 3、およびチャンネルエンコード 3 5 4。図 3 b は、図 2 b と同様であり、詳細については、本明細書では改めて説明しない。

## 【 0 1 2 6 】

図 3 c は、本出願の一実施形態による、無線機器もしくはコアネットワーク機器へのマルチチャンネルデコードの適用を示す模式図である。無線機器もしくはコアネットワーク機器 3 5 は、以下を含む。すなわち、チャンネルデコード 3 5 1、マルチチャンネルデコード 3 5 5、別の音声エンコード 3 5 6、およびチャンネルエンコード 3 5 4。図 3 c は、図 2 c と同様であり、詳細については、本明細書では改めて説明しない。

## 【 0 1 2 7 】

音声符号化は、マルチチャンネルエンコードの一部であり得て、音声復号化は、マルチチャンネルデコードの一部であり得る。例えば、収集されたマルチチャンネル信号に対してマルチチャンネル符号化を実行するステップは、収集されたマルチチャンネル信号を処理して、音声信号を取得するステップであり得る。次いで、取得された音声信号は、本出願の実施形態において提供される方法に従って符号化される。デコード側は、マルチチャンネル信号に基づいて、ビットストリームを符号化し、復号化を実行して、音声信号を取得し、アップミックス処理後にマルチチャンネル信号を復元する。そのため、本出願の実施形態は、端末機器、無線装置、もしくはコアネットワーク装置におけるマルチチャンネルエンコードおよびマルチチャンネルデコードにも適用され得る。無線機器もしくはコアネットワーク機器では、トランス符号化を実装する必要がある場合、対応するマルチチャンネル符号化処理を実行する必要がある。

## 【 0 1 2 8 】

最初に、本出願の実施形態において提供される三次元音声信号処理方法について説明す

る。本方法は、端末機器によって実行され得る。例えば、端末機器は、音声符号化装置（以下、エンコーダ側もしくはエンコーダと呼ばれる）であり得る。代替的に、端末機器が三次元音声信号処理装置であり得ることは、限定されない。図4に示されるように、三次元音声信号処理方法は、主に、以下のステップを含む。

【0129】

401：三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得する。

【0130】

エンコーダ側は、三次元音声信号を取得し得る。例えば、三次元音声信号は、シーン音声信号であり得る。具体的には、三次元音声信号は、時間領域信号であってもよいし、または周波数領域信号であってもよい。また、三次元音声信号は、代替的に、ダウンサンプリングを介して取得された信号であってもよい。

10

【0131】

本出願の幾つかの実施形態では、三次元音声信号は、高次アンビソニックスHOA信号もしくは一次アンビソニックスFOA信号を含む。代替的に、三次元音声信号が別種類の信号であり得ることは、限定されない。これは、本出願の単なる一例に過ぎず、本出願の本実施形態に対する限定が意図されていない。

【0132】

例えば、三次元音声信号は、時間領域のHOA信号であってもよいし、または周波数領域のHOA信号であってもよい。別の例として、三次元音声信号は、HOA信号の全チャンネルを含み得るか、または一部のHOAチャンネル（例えば、FOAチャンネルなど）を含み得る。また、三次元音声信号は、HOA信号の全サンプリング点であり得るか、またはダウンサンプリングを介して取得された解析対象のHOA信号における $1/Q$ ダウンサンプリング点であり得る。Qはダウンサンプリング間隔であり、 $1/Q$ はダウンサンプリングレートである。

20

【0133】

本出願の本実施形態では、三次元音声信号は、複数のフレームを含む。以下では、例として三次元音声信号の1フレームの処理を使用する。例えば、フレームが現行フレームである場合、三次元音声信号の現行フレームより前には前のフレームが存在し、現行フレームの後には次のフレームが存在する。また、本出願の本実施形態では、三次元音声信号における現行フレーム以外の別のフレームの処理方法も、現行フレームの処理のための方法と同様である。以下では、例として現行フレームの処理を使用する。

30

【0134】

本出願の本実施形態では、三次元音声信号の現行フレームを取得した後、最初に、現行フレームに対して線形分解を実行して、現行フレームの線形分解結果を取得する。複数の線形分解方式が存在し、これは、以下で詳細に説明される。

【0135】

本出願の幾つかの実施形態では、ステップ401における線形分解結果を取得するために、三次元音声信号の現行フレームに対して線形分解を実行するステップは、以下を含む。すなわち、

40

A1：現行フレームに対して特異値分解を実行して、現行フレームに対応する特異値を取得するステップであって、線形分解結果は特異値を含む、ステップ。

A2：現行フレームに対して主成分分析を実行して、現行フレームに対応する第一の特徴値を取得するステップであって、線形分解結果は第一の特徴値を含む、ステップ。または、

A3：現行フレームに対して独立成分分析を実行して、現行フレームに対応する第二の特徴値を取得するステップであって、線形分解結果は第二の特徴値を含む、ステップ。

【0136】

線形分解方法は、複数存在する。例えば、線形分解は、次のうちの少なくとも一つを含み得る。すなわち、特異値分解（SVD）、主成分分析（PCA）、および独立成分分析（ICA）。異なる線形分解方法では、取得される線形分解結果は、異なる表現方式を有

50

し、これは、詳細に後述される。

#### 【 0 1 3 7 】

ステップ A 1 では、線形分解は、特異値分解であり得る。例えば、三次元音声信号が H O A 信号であると仮定される。H O A 信号は、行列 A を形成し、行列 A は、 $L * K$  行列であり、 $L$  は H O A 信号のチャンネル数に等しく、 $K$  は現行フレームにおける H O A 信号の各チャンネルの信号点数である。例えば、信号点数は、以下を含み得る。すなわち、周波数の個数、時間領域におけるサンプリング点の個数、またはダウンサンプリング後の周波数の個数、もしくはサンプリング点の個数。行列 A に対して特異値分解が実行され、次の関係が満たされる。すなわち、

$$A = U \Sigma V^T$$

10

#### 【 0 1 3 8 】

$U$  は  $L * L$  行列であり、 $V$  は  $K * K$  行列であり、上付き文字  $T$  は、行列  $V$  の転置であり、 $*$  は乗算を表す。 $\Sigma$  は  $L * K$  対角行列であり、行列の主対角上の各要素は、特異値分解によって取得される行列 A の特異値であり、主対角の外側にある要素は、全て 0 である。対角行列  $\Sigma$  の主対角上にある要素、すなわち行列 A の特異値は、 $v[i]$  として表され、 $i = 0, 1, \dots, \min(L, K) - 1$  とする。

#### 【 0 1 3 9 】

三次元音声信号がダウンサンプリングを介して取得された H O A 信号である場合、 $K$  はダウンサンプリング後の現行フレームにおける H O A 信号の各チャンネルの信号点数であることは、留意されるべきである。例えば、信号点数は、サンプリング点の個数であってもよいし、または周波数の個数であってもよい。

20

#### 【 0 1 4 0 】

ステップ A 2 では、線形分解は、代替的に、主成分分析であり、特徴値を取得し得る。以降の実施形態では、別の特徴値から区別するために、主成分分析を介して取得される特徴値は、第一の特徴値として定義される。主成分分析の具体的な実装については、本明細書では改めて説明しない。

#### 【 0 1 4 1 】

ステップ A 3 では、線形分解は、代替的に、独立成分分析であり、第二の特徴値を取得し得る。独立成分分析の具体的な実装については、本明細書では改めて説明しない。

#### 【 0 1 4 2 】

30

本出願の本実施形態では、現行フレームの線形分解を前述の実装 A 1 ないし A 3 の何れか一つにて実装して、複数種類の線形分解結果を取得することができる。

#### 【 0 1 4 3 】

4 0 2 : 線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得する。

#### 【 0 1 4 4 】

現行フレームの線形解析結果を取得した後、エンコーダ側は、線形分解結果を解析して、現行フレームに対応する音場分類パラメータを取得する。音場分類パラメータは、現行フレームの線形分解結果を分析することによって取得され、音場分類パラメータは、現行フレームの音場分類結果を決定するために使用される。線形分解結果の異なる特定の実装に基づいて、音場分類パラメータは、複数の実装を有し得る。

40

#### 【 0 1 4 5 】

本出願の本実施形態では、一つもしくは複数の線形分解結果が存在し得る。例えば、線形分解結果は、特異値を含み、その特異値は、 $v[i]$  であり、 $i = 0, 1, \dots, \min(L, K) - 1$  とする。現行フレームの特異値が一つのみである場合、 $i$  の値は、一つのみ、すなわち  $v[0]$  のみである。現行フレームに複数の特異値がある場合、複数の  $i$  の値、すなわち  $v[i]$  が存在し、 $i = 1, \dots, \min(L, K) - 1$  とする。

#### 【 0 1 4 6 】

本出願の本実施形態では、線形分解結果が二つある場合、取得される音場分類パラメータは一つである。線形分解結果の数を  $N$  とすると、取得される音場分類パラメータの数は

50

N - 1 個となり、N の値は限定されない。

【 0 1 4 7 】

本出願の幾つかの実施形態では、ステップ 4 0 2 における線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するステップは、以下を含む。すなわち、B 1 : 現行フレームの ( i + 1 ) 番目の線形解析結果に対する、現行フレームの i 番目の線形解析結果の比を取得するステップであって、i は、正の整数である、ステップ。および、

B 2 : この比に基づいて、現行フレームに対応する i 番目の音場分類パラメータを取得するステップ。

【 0 1 4 8 】

エンコーダ側は、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得し得る。例えば、現行フレームの線形分解結果が複数あり、その複数の線形解析結果のうちの連続する二つの線形解析結果は、現行フレームにおける i 番目の線形解析結果および ( i + 1 ) 番目の線形解析結果として表現される。この場合、現行フレームの ( i + 1 ) 番目の線形解析結果に対する、現行フレームの i 番目の線形解析結果の比は計算され得て、具体的な i の値は限定されない。

【 0 1 4 9 】

任意選択として、i 番目の線形解析結果および ( i + 1 ) 番目の線形解析結果は、現行フレームにおける二つの連続する線形解析結果である。

【 0 1 5 0 】

この比が取得された後、現行フレームに対応する i 番目の音場分類パラメータは、現行フレームの ( i + 1 ) 番目の線形解析結果に対する、現行フレームの i 番目の線形解析結果の比に基づいて、取得され得る。i 番目の音場分類パラメータは、( i + 1 ) 番目の線形解析結果に対する、i 番目の線形解析結果の比に基づいて、計算することができると分かる。( i + 1 ) 番目の音場分類パラメータは、( i + 2 ) 番目の線形解析結果に対する、( i + 1 ) 番目の線形解析結果の比に基づいて計算され得て、残りは、類推によって推測することができる。線形解析結果および音場分類パラメータの間には対応関係がある。

【 0 1 5 1 】

一実装では、( i + 1 ) 番目の線形解析結果に対する i 番目の線形解析結果の比が、i 番目の音場分類パラメータとして使用され得る。( i + 1 ) 番目の線形解析結果に対する i 番目の線形解析結果の比が取得された後、その比に対して複数の計算方式がさらに実行され得て、これにより、i 番目の音場分類パラメータが取得され得る。例えば、プリセット調整係数に基づいて、その比に対して乗算演算が実行されて、i 番目の音場分類パラメータを取得する。

【 0 1 5 2 】

例えば、特異値分解が線形分解に使用される場合、音場分類パラメータに基づいて、特異値が特異値分解を介して取得され得て、隣接する二つの特異値間の比パラメータが計算され、音場分類パラメータとして使用される。

【 0 1 5 3 】

例えば、特異値間の比  $temp[i]$  が計算され、音場分類パラメータとして使用される。i = 0 , 1 , ... , min ( L , K ) - 2 の場合、 $temp[i]$  は、以下を満たす。すなわち、

$$temp[i] = v[i] / v[i + 1]$$

【 0 1 5 4 】

P C A もしくは I C A が線形分解に対して使用される場合、音場分類パラメータは、特徴値に基づいて決定され得る。音場分類パラメータを計算するための方法は、特異値間の比  $temp$  を計算するための方法と同様である。あるいは、連続する二つの特徴値の比は、線形分解を介して取得される特徴値に基づいて計算され、その比は、音場分類パラメータとして使用される。

【 0 1 5 5 】

線形分解を介して取得される特徴値もしくは特異値の個数が2を超える場合、音場分類パラメータは、ベクトルとなることは、留意されるべきである。それ以外の場合、音場分類パラメータは、スカラーとなる。例えば、 $v[i]$ に対して、 $i$ の値が2に等しい場合、計算された $temp[i]$ はスカラーとなり、すなわち $temp$ の値は一つのみ存在する。 $v[i]$ に対して、 $i$ の値が2を超える場合、計算された $temp[i]$ はベクトルとなり、 $temp$ は、少なくとも二つの要素を含む。

【0156】

403：音場分類パラメータに基づいて、現行フレームの音場分類結果を決定する。

【0157】

本発明の本実施形態では、現行フレームに対応する音場分類パラメータを取得した後、エンコーダ側は、音場分類パラメータに基づいて、現行フレームに対して音場分類を実行し得る。現行フレームに対応する音場分類パラメータは、現行フレームに対応する音場の分類に必要とされるパラメータを示し得るため、現行フレームの音場分類結果は、音場分類パラメータに基づいて取得され得る。

【0158】

本出願の幾つかの実施形態では、音場分類結果は、音場種別および不均一型音源数のうちの少なくとも一つを含み得る。

【0159】

音場種別は、現行フレームのものである音場種別であり、現行フレームに対して音場分類が実行された後に決定される。音場種別を分類する方法は、複数ある。例えば、音場種別は、第一の音場種別および第二の音場種別に分類され得る。あるいは、音場種別は、第一の音場種別、第二の音場種別、および第三の音場種別などに分類され得る。具体的には、分類することができる音場種別の個数は、用途シナリオに基づいて決定され得る。別の例として、音場種別には、不均一型音場および分散型音場が含まれ得る。不均一型音場とは、音場において異なる位置および/もしくは方向を有する点音源が存在することを意味し、分散型音場とは、不均一型音源を含まない音場である。例えば、異なる位置および/もしくは方向を有する点音源は不均一型音源であり、不均一型音源を含む音場は不均一型音場であり、不均一型音源を含まない音場は分散型音場である。

【0160】

不均一型音源は、異なる位置および/もしくは方向を有する点音源であり、現行フレームに含まれる不均一型音源数は、不均一型音源数と呼ばれる。現行フレームの音場は、不均一型音源数に基づいて分類することもできる。

【0161】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在する。音場分類結果には、音場種別が含まれる。

【0162】

ステップ403では、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下を含む。すなわち、

複数の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合、音場種別が分散型音場であると判定するステップ。または

複数の音場分類パラメータの値のうちの少なくとも一つが予め設定される不均一型音源判定条件を満たす場合、音場種別が不均一型音場であると判定するステップ。

【0163】

音場種別には、不均一型音場および分散型音場が含まれる。本発明の本実施形態では、分散型音源判定条件および不均一型音源判定条件が予め設定される。分散型音源判定条件は、音場種別が分散型音場であるか否かを判定するために使用され、不均一型音源判定条件は、音場種別が不均一型音場であるか否かを判定するために使用される。現行フレームにおける複数の音場分類パラメータが取得された後、複数の音場分類パラメータの値およびプリセット条件に基づいて、判定が実行される。分散型音源判定条件および不均一型音源判定条件の具体的な実装は、本明細書では限定されない。

10

20

30

40

50

## 【 0 1 6 4 】

複数の音場分類パラメータが取得された後、エンコーダ側は、複数の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合、音場種別が分散型音場であると判定する。例えば、現行フレームは、N個の音場分類パラメータに対応する。N個の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合にのみ、現行フレームの音場種別が分散型音場であると判定される。

## 【 0 1 6 5 】

複数の音場分類パラメータが取得された後、エンコーダ側は、複数の音場分類パラメータの値のうちの少なくとも一つが予め設定される不均一型音源判定条件を満たす場合、音場種別が不均一型音場である判定する。例えば、現行フレームは、N個の音場分類パラメータに対応する。N個の音場分類パラメータの値のうちの少なくとも一つが予め設定される不均一型音源判定条件を満たす場合にのみ、音場種別が不均一型音場であると判定される。

10

## 【 0 1 6 6 】

さらに、本出願の幾つかの実施形態では、分散型音源判定条件は、以下を含む。すなわち、音場分類パラメータの値が予め設定される不均一型音源判定閾値未満であること。または、不均一型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることが含まれること。

## 【 0 1 6 7 】

不均一型音源判定閾値は、プリセット閾値であり得て、具体的な値は限定されない。分散型音源判定条件には、音場分類パラメータの値が予め設定される不均一型音源判定閾値未満であることが含まれる。そのため、複数の音場分類パラメータの値が全て予め設定される不均一型音源判定閾値未満である場合、音場種別が分散型音場であると判定される。不均一型音源判定条件には、音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることが含まれる。そのため、複数の音場分類パラメータの値のうちの少なくとも一つが予め設定される不均一型音源判定閾値以上である場合、音場種別が不均一型音場であると判定される。

20

## 【 0 1 6 8 】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在する。

30

## 【 0 1 6 9 】

音場分類結果は、音場種別を含み、または音場分類結果は、不均一型音源数および音場種別を含む。

## 【 0 1 7 0 】

ステップ403では、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下を含む。すなわち、

C1：複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップ。および、

C2：現行フレームに対応する不均一型音源数に基づいて、音場種別を決定するステップ。

## 【 0 1 7 1 】

現行フレームに対応する複数の音場分類パラメータを取得した後、エンコーダ側は、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得し得る。不均一型音源は、異なる位置および/もしくは方向を有する点音源であり、現行フレームに含まれる不均一型音源数は、不均一型音源数と呼ばれる。現行フレームの音場は、不均一型音源数に基づいて分類することができる。現行フレームに対応する不均一型音源数が、音場種別を決定取するために取得された後、現行フレームに対応する音場種別は、現行フレームに対応する不均一型音源数を分析することによって決定され得る。

40

## 【 0 1 7 2 】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在する。

## 【 0 1 7 3 】

50

音場分類結果は、不均一型音源数を含む。

【0174】

ステップ403では、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下を含む。すなわち、

D1：複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップ。

【0175】

現行フレームに対応する複数の音場分類パラメータを取得した後、エンコーダ側は、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得し得る。不均一型音源は、異なる位置および/もしくは方向を有する点音源であり、現行フレームに含まれる不均一型音源数は、不均一型音源数と呼ばれる。

10

【0176】

さらに、本出願の幾つかの実施形態では、複数の音場分類パラメータは、 $temp[i]$ 、 $i = 0, 1, \dots, \min(L, K) - 2$ であり、 $L$ は現行フレームのチャンネル数を表し、 $K$ は現行フレームの各チャンネルに対応する信号点の数であり、 $\min$ は最小値を選択する演算を表す。例えば、信号点の個数は、周波数の個数、時間領域におけるサンプリング点の個数、またはダウンサンプリング後の時間領域における周波数の個数もしくはサンプリング点の個数であり得る。

【0177】

ステップC1もしくはステップD1では、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するステップは、以下を含む。すなわち、 $i = 0$ から次の判定手順を順次実行するステップ。

20

$temp[i]$ が予め設定される不均一型音源判定閾値を超えるか否かを判定するステップ。および、

本判定手順において、 $temp[i]$ が不均一型音源判定閾値未満である場合、 $i$ の値を $i + 1$ に更新して、次の判定手順を継続するステップ。または

$temp[i]$ が本判定手順における不均一型音源判定閾値以上である場合、本判定手順の実行を終了し、1を加えた本判定手順における $i + 1$ が不均一型音源数に等しいと判定するステップ。

【0178】

30

具体的には、エンコーダ側は、音場分類パラメータに基づいて、不均一型音源数を推定し、音場種別を判定し得る。

【0179】

音場種別には、不均一型音場と分散型音場が含まれる。不均一型音場とは、音場内に位置や方向が異なる点音源が存在することをいう。分散型音場とは、不均一型音源を含まない音場である。

【0180】

音場分類パラメータの値が全て分散型音源判定条件を満たす場合、音場種別は、分散型音場である。

【0181】

40

音場分類パラメータの値が不均一型音源判定条件を満たす場合、音場種別は不均一型音場であると判定される。不均一型音源数は、不均一型音源判定条件を満たす、音場分類パラメータの値のうちの値の順序番号に基づいて推定され得る。

【0182】

例えば、特異値間の比 $temp[i]$ は、音場分類パラメータとして使用される場合、音場種別および不均一型音源数は、音場分類パラメータに基づいて推定され、 $temp[i]$ の値は、 $i = 0$ から順次決定される。 $i$ の値を $m$ とする場合、 $m$ 番目の音場分類パラメータの値は、 $temp[m]$ と表現される。 $m$ 番目の音場分類パラメータが $temp[m]_{TH1}$ を満たす場合、音場種別は不均一型音場であり、現行フレームの音場に $(m + 1)$ 個の不均一型音源が存在する。 $temp[m]_{TH1}$ を満たす場合、音場種別は

50

分散型音場である。mの値の範囲は、 $[0, 1, \dots, \min(L, K) - 2]$ であり、 $TH1$ は、予め設定される不均一型音源判定閾値であり、 $TH1$ の値は定数であり、例えば、 $TH1$ の値は、30もしくは100であり得る。本出願の本実施形態では、 $TH1$ の値は限定されない。

#### 【0183】

本出願の幾つかの実施形態では、ステップC2における現行フレームに対応する不均一型音源数に基づいて、音場種別を決定するステップは、以下を含む。すなわち、不均一型音源数が第一のプリセット条件を満たす場合、音場種別が第一の音場種別であると判定するステップ。または、不均一型音源数が第一のプリセット条件を満たさない場合、音場種別が第二の音場種別であると判定するステップ。

10

#### 【0184】

第一の音場種別に対応する不均一型音源数は、第二の音場種別に対応する不均一型音源数とは相違する。

#### 【0185】

具体的には、音場種別は、不均一型音源数の差異に基づいて、第一の音場種別および第二の音場種別という二種類に分類され得る。エンコーダ側は、第一のプリセット条件を取得する。すなわち、不均一型音源数が第一のプリセット条件を満たすか否かを判定すること。および、不均一型音源数が第一のプリセット条件を満たす場合、音場種別が第一の音場種別であると判定すること。または、不均一型音源数が第一のプリセット条件を満たさない場合、音場種別が第二の音場種別であると判定すること。本出願の本実施形態では、現行フレームの音場種別の分割を実装して、現行フレームの音場種別が第一の音場種別もしくは第二の音場種別に属することを正確に識別するために、不均一型音源数が第一のプリセット条件を満たすか否かが判定され得る。

20

#### 【0186】

本出願の幾つかの実施形態では、第一のプリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、かつ、第二の閾値が第一の閾値を超えることを含む。または第一のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、かつ、第二の閾値が第一の閾値を超えることを含む。

30

#### 【0187】

第一の閾値および第二の閾値の具体的な値は、限定されないで、用途シナリオに基づいて、具体的に決定され得る。第二の閾値は、第一の閾値を超える。そのため、第一の閾値および第二の閾値は、プリセット範囲を構成し得て、第一のプリセット条件は、不均一型音源数がプリセット範囲内に収まることであってもよいし、または第一のプリセット条件は、不均一型音源数がプリセット範囲を超えることであってもよい。不均一型音源数は、第一のプリセット条件における第一の閾値および第二の閾値に基づいて決定されて、不均一型音源数が第一のプリセット条件を満たすか否かを判定して、現行フレームの音場種別が第一の音場種別もしくは第二の音場種別に属することを正確に識別し得る。

#### 【0188】

40

例えば、第一の閾値が0であり、第二の閾値が3であり、不均一型音源数がnとして表現される。この場合、第一のプリセット条件は、 $0 < n < 3$ であってもよいし、または第一のプリセット条件は、 $n = 3$ もしくは $n = 0$ であってもよい。

#### 【0189】

本出願の幾つかの実施形態では、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップは、以下をさらに含む。すなわち、音場分類パラメータ、および三次元音声信号の特徴を含む別のパラメータに基づいて、現行フレームの音場分類結果を決定するステップ。

#### 【0190】

三次元音声信号の特徴を示す別のパラメータには、複数の実装がある。例えば、三次元

50

音声信号の特徴を示す別のパラメータは、以下のうちの少なくとも一つを含み得る。すなわち、三次元音声信号のエネルギー比パラメータ、三次元音声信号の高周波解析パラメータ、および三次元音声信号の低周波特徴解析パラメータなど。

【0191】

図5に示されるように、本出願の一実施形態による三次元音声信号処理方法は、主に、以下のステップを含む。

【0192】

501：三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得するステップ。

【0193】

502：線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するステップ。

【0194】

503：音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するステップ。

【0195】

ステップ501ないしステップ503の実装は、前述の実施形態におけるステップ401ないしステップ403の実装と同様であり、ステップ501ないしステップ503については、本明細書では改めて詳細に説明しない。

【0196】

504：音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップ。

【0197】

エンコーダ側は、ステップ501ないしステップ503を実行し得る。現行フレームの音場分類結果を取得した後、エンコーダ側は、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定し得る。符号化モードは、三次元音声信号の現行フレームを符号化する際に使用されるモードである。複数の符号化モードが存在し、異なる符号化モードは、現行フレームの異なる音場分類結果に基づいて使用され得る。本発明の本実施形態では、適切な符号化モードは、現行フレームの異なる音場分類結果に対して選択され、これにより、現行フレームは、その符号化モードを使用することによって符号化される。これは、音声信号の圧縮効率および聴覚品質を改善する。

【0198】

さらに、本出願の幾つかの実施形態では、ステップ503における音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、

E1：音場分類結果が不均一型音源数を含むか、または音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームに対応する符号化モードを決定するステップ。

E2：音場分類結果が音場種別を含むか、または音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップ。または、

E3：音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップ。

【0199】

ステップE1では、エンコーダ側が現行フレームの不均一型音源数を取得した後、その不均一型音源数は、現行フレームに対応する符号化モードを決定するために使用され得る。ステップE2では、エンコーダ側が現行フレームの音場種別を取得した後、その音場種別は、現行フレームに対応する符号化モードを決定するために使用され得る。ステップE3では、エンコーダ側が不均一型音源数および音場種別を取得した後、それらの不均一型音源数および音場種別は、現行フレームに対応する符号化モードを決定するために使用さ

10

20

30

40

50

れ得る。そのため、エンコーダ側は、現行フレームの音場分類結果に基づいて、対応する符号化モードを決定するために、不均一型音源数および／もしくは音場種別に基づいて、現行フレームに対応する符号化モードを決定し得て、これにより、決定された符号化モードは、三次元音声信号の現行フレームに適合させることができる。これは、符号化効率を改善する。

【 0 2 0 0 】

さらに、本出願の幾つかの実施形態では、ステップ E 1 に不均一型音源数に基づいて、現行フレームに対応する符号化モードを判定するステップは、以下を含む。すなわち、不均一型音源数が第二のプリセット条件を満たす場合、符号化モードが第一の符号化モードであると判定されるステップ。または、不均一型音源数が第二のプリセット条件を満たさない場合、符号化モードが第二の符号化モードであると判定されるステップ。

10

【 0 2 0 1 】

第一の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、第二の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、第一の符号化モードおよび第二の符号化モードは、相違する符号化モードである。仮想スピーカー選択に基づく H O A 符号化モードは、マッチ投影 ( M P ) に基づく H O A 符号化モードと呼ばれることもある。

【 0 2 0 2 】

具体的には、符号化モードは、不均一型音源数の差異に基づいて、第一の符号化モードおよび第二の符号化モードという二種類に分類され得る。エンコーダ側は、第二のプリセット条件を取得する。すなわち、不均一型音源数が第二のプリセット条件を満たすか否かを判定すること。不均一型音源数が第二のプリセット条件を満たす場合、符号化モードが第一の符号化モードであると判定すること。不均一型音源数が第二のプリセット条件を満たさない場合、符号化モードが第二の符号化モードであると判定すること。本出願の本実施形態では、現行フレームの符号化モードの分割を実装して、現行フレームの符号化モードが第一の符号化モードもしくは第二の符号化モードに属することを正確に識別するために、不均一型音源数が第二のプリセット条件を満たすか否かが判定され得る。

20

【 0 2 0 3 】

例えば、第一の符号化モードが仮想スピーカー選択に基づく H O A 符号化モードである場合、第二の符号化モードは、指向性音声コーディングに基づく H O A 符号化モードである。あるいは、第一の符号化モードが指向性音声コーディングに基づく H O A 符号化モードである場合、第二の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モードであり、第一の符号化モードおよび第二の符号化モードの具体的な実装は、用途シナリオに基づいて決定され得る。

30

【 0 2 0 4 】

例えば、本出願の本実施形態では、音場分類結果は、エンコーダ側によって選択される符号化モードを決定するために使用される。例えば、音場分類結果は、H O A 信号の符号化モードを決定するために使用され得る。例えば、符号化モードは、音場種別に基づいて決定される。不均一型音源に属する H O A 信号は、符号化モード A に対応するエンコーダを使用することによる符号化に適しており、分散型音場に属する H O A 信号は、符号化モード B に対応するエンコーダを使用することによる符号化に適している。別の例として、符号化モードは、不均一型音源数に基づいて決定される。不均一型音源数が符号化モード X を使用するための判定条件を満たす場合、符号化は、符号化モード X に対応するエンコーダを使用することによって実行される。別の例として、符号化モードは、代替的に、音場種別および不均一型音源数に基づいて、選択的に決定される。音場種別が分散型音場である場合、符号化は、符号化モード C に対応するエンコーダを使用することによって実行される。音場種別が不均一型音場であり、不均一型音源数が符号化モード X を使用する判定条件を満たす場合、符号化は、符号化モード X に対応するエンコーダを使用すること

40

50

よって実行される。符号化モードA、符号化モードB、符号化モードC、および符号化モードXには、複数の異なる符号化モードが含まれ得る。本出願の本実施形態では、異なる音場分類結果は、異なる符号化モードに対応する。これは、本出願の本実施形態では限定されない。例えば、符号化モードXは、不均一型音源数がプリセット閾値未満である場合、符号化モード1であり得るか、または不均一型音源数がプリセット閾値以上である場合、符号化モード2であり得る。

【0205】

本出願の幾つかの実施形態では、第二のプリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または

10

第二のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

【0206】

第一の閾値および第二の閾値の具体的な値は、限定されないで、用途シナリオに基づいて具体的に決定され得る。第二の閾値は、第一の閾値を超える。そのため、第一の閾値および第二の閾値は、プリセット範囲を構成し、第二のプリセット条件は、不均一型音源数がプリセット範囲内に収まることであってもよく、または第二のプリセット条件は、不均一型音源数がプリセット範囲を超えることであってもよい。不均一型音源数が第二のプリセット条件を満たすか否かを判定し、現行フレームの音場種別が第一の音場種別もしくは第二の音場種別に属することを正確に識別するために、不均一型音源数は、第二の閾値、および第一のプリセット条件における第二の閾値に基づいて決定され得る。

20

【0207】

例えば、第一の閾値が0であり、第二の閾値が3であり、不均一型音源数が $n$ として表現される。この場合、第二のプリセット条件は、 $0 < n < 3$ であってもよいし、または第二のプリセット条件は、 $n \geq 3$ もしくは $n = 0$ であってもよい。

【0208】

本出願の本実施形態では、第一のプリセット条件は、異なる音場種別を識別するための条件セットであり、第二のプリセット条件は、異なる符号化モードを識別するための条件セットであることは、留意されるべきである。第一のプリセット条件および第二のプリセット条件は、同じ条件内容を含んでもよいし、または異なる条件内容を含んでもよい。換言すると、第一のプリセット条件および第二のプリセット条件は、異なるプリセット条件であってもよいし、または同一のプリセット条件であってもよい。ただし、実際の使用時には差異が生じる可能性があると考えられる。第一のプリセット条件および第二のプリセット条件は、第一および第二という数詞を使用することによって区別される。

30

【0209】

本出願の幾つかの実施形態では、ステップE2における音場種別に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、音場種別が不均一型音場である場合、仮想スピーカー選択に基づいて、符号化モードがH O A符号化モードであると判定するステップ。または、音場種別が分散型音場である場合、符号化モードが指向性音声コーディングに基づくH O A符号化モードであると判定するステップ。

40

【0210】

音場に不均一型音源がほとんど無い音場、および分散型音場については、指向性音声に基づくH O A符号化モードは、仮想スピーカー選択に基づくH O A符号化モードよりも低い圧縮効率を有する。しかしながら、音場に複数の不均一型音源が存在する音場については、仮想スピーカー選択に基づくH O A符号化モードは、指向性音声に基づくH O A符号化モードよりも低い圧縮効率を有する。本出願の本実施形態では、音場種別が不均一型音源である場合、符号化モードは、仮想スピーカー選択に基づくH O A符号化モードであると判定される。音場種別が分散型音場である場合、符号化モードは、指向性音声符号化に基づくH O A符号化モードであると判定される。本出願の本実施形態では、異なるタイプ

50

の音声信号に対して最大の圧縮効率を取得するという要件を満たすために、対応する符号化モードは、現行フレームの音場分類結果に基づいて選択され得る。

【0211】

本出願の幾つかの実施形態では、ステップ503における音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップは、以下を含む。すなわち、

F1：現行フレームの音場分類結果に基づいて、現行フレームに対応する初期符号化モードを決定するステップ。

F2：現行フレームが位置するハングオーバー時間枠を取得するステップであって、ハングオーバー時間枠は、現行フレームの初期符号化モードと、現行フレームより前のN-1個のフレームの符号化モードとを含み、Nは、ハングオーバーの長さである。および、

F3：現行フレームの初期符号化モードと、N-1個のフレームの符号化モードとに基づいて、現行フレームの符号化モードを決定するステップ。

【0212】

ステップF1では、初期符号化モードは、音場分類結果に基づいて決定される符号化モードであり得る。例えば、現行フレームの符号化モードは、ステップE1ないしステップE3における前述の実装のうちの何れか一つに基づいて決定され得て、その符号化モードは、F1における初期符号化モードとして使用され得る。初期符号化モードが取得された後、ハングオーバー時間枠は、現行フレームとハングオーバー時間枠のウィンドウサイズとに基づいて取得される。ハングオーバー時間枠には、現行フレームの初期符号化モードと、現行フレームより前のN-1個のフレームの符号化モードとが含まれ、Nはハングオーバー時間枠に含まれるフレームの数を表す。最後に、現行フレームの符号化モードは、ハングオーバー時間枠におけるN個のフレームに個別に対応する符号化モードに基づいて決定される。ステップF3において取得される現行フレームの符号化モードは、現行フレームを符号化する際に使用される符号化モードであり得る。本出願の本実施形態では、現行フレームの符号化モードを取得するために、現行フレームの初期符号化モードが、ハングオーバー時間枠に基づいて修正される。これは、連続するフレームの符号化モードが、頻繁に切り替わらなくなり、符号化効率を改善する。

【0213】

例えば、現行フレームの初期符号化モードが取得された後、確実に、連続するフレームの符号化モードが頻繁に切り替わらないようにするために、ハングオーバー時間枠処理が、現行フレームに対して実行され得る。ハングオーバー時間枠の処理方法は、複数存在する。これは、本出願の本実施形態では限定されない。例えば、処理方式は、長さがN個のフレームであるエンコーダ選択識別子を、ハングオーバー時間枠に保存するステップであって、N個のフレームは、現行フレームと、現行フレームより前のN-1個のフレームとのエンコーダ選択識別子を含む、ステップと、エンコーダ選択識別子が指定された閾値まで累積されると、現行フレームの符号化タイプ指示識別子を更新するステップとし得る。任意選択として、ハングオーバー時間枠処理に加えて、他の後処理が、現行フレームに対する修正を実行するために使用され得る。例えば、初期符号化モードは、初期分類として使用され、初期分類が、会話分類結果、および音声信号の信号対雑音比などの、特徴に基づいて変更され、変更された結果が、符号化モードの最終結果として使用される。

【0214】

図6に示されるように、本出願の一実施形態による三次元音声信号処理方法は、主に、以下のステップを含む。

【0215】

601：三次元音声信号の現行フレームに対して線形分解を実行して、線形分解結果を取得する。

【0216】

602：線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得する。

【0217】

10

20

30

40

50

6 0 3 : 音場分類パラメータに基づいて、現行フレームの音場分類結果を決定する。

【 0 2 1 8 】

ステップ 6 0 1 ないしステップ 6 0 3 の実装は、前述の実施形態におけるステップ 4 0 1 ないしステップ 4 0 3 の実装と同様であり、ステップ 6 0 1 ないしステップ 6 0 3 については、本明細書では改めて詳細に説明しない。

【 0 2 1 9 】

6 0 4 : 音場分類結果に基づいて、現行フレームに対応する符号化パラメータを決定する。

【 0 2 2 0 】

エンコーダ側は、ステップ 6 0 1 ないしステップ 6 0 3 を実行し得る。現行フレームの音場分類結果を取得した後、エンコーダ側は、音場分類結果に基づいて、現行フレームに対応する符号化パラメータを決定し得る。符号化パラメータは、三次元音声信号の現行フレームを符号化する際に使用されるパラメータである。複数の符号化パラメータが存在し、現行フレームの異なる音場分類結果に基づいて、異なる符号化パラメータが使用され得る。本出願の本実施形態では、現行フレームの異なる音場分類結果に対して、適切な符号化パラメータが選択され、これにより、その符号化パラメータに基づいて、現行フレームが符号化される。これは、音声信号の圧縮効率および聴覚品質を改善する。

10

【 0 2 2 1 】

さらに、本出願の幾つかの実施形態では、符号化パラメータは、以下のうちの少なくとも一つを含む。すなわち、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の符号化ビット数、残差信号の符号化ビット数、もしくは最適スピーカーを探索するための投票回数。

20

【 0 2 2 2 】

仮想スピーカー信号および残差信号は、三次元音声信号に基づいて生成される信号である。

【 0 2 2 3 】

具体的には、エンコーダ側は、現行フレームの音場分類結果に基づいて、現行フレームの符号化パラメータを決定し得て、その符号化パラメータは、現行フレームを符号化するために使用され得る。符号化パラメータには、複数の実装がある。例えば、符号化パラメータは、以下のうちの少なくとも一つを含む。すなわち、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の符号化ビット数、残差信号の符号化ビット数、もしくは最適スピーカーを探索するための投票回数。チャンネル数は、伝送チャンネル数とも呼ばれる。チャンネル数は、信号の符号化時に割り当てられた伝送チャンネル数であり、符号化ビット数は、信号の符号化時に割り当てられた符号化ビット数である。

30

【 0 2 2 4 】

本出願の本実施形態において提供される仮想スピーカーを選択するための方法では、エンコーダは、仮想スピーカーを探索するための計算負荷を軽減し、エンコーダの計算負荷を低減するために、現行フレームの仮想スピーカー係数に基づいて、候補仮想スピーカーセットにおける各仮想スピーカーに投票し、投票値に基づいて、現行フレームの仮想スピーカーを選択する。最適スピーカーを探索するための投票回数とは、最適スピーカーを探索する際に必要とされる投票回数である。可能な実装では、投票回数は、事前に構成されてもよく、または現行フレームの音場分類結果に基づいて決定されてもよい。例えば、最適スピーカーを探索するための投票回数は、三次元音声信号に基づいて、仮想スピーカー信号を決定するプロセスにおいて、仮想スピーカーを探索するための投票回数である。

40

【 0 2 2 5 】

また、本出願の本実施形態における仮想スピーカー信号および残差信号は、三次元音声信号に基づいて生成される信号である。例えば、第一の目標仮想スピーカーは、第一のシーン音声信号に基づいて、予め設定される仮想スピーカーセットから選択され、仮想スピーカー信号は、第一のシーン音声信号と、第一の目標仮想スピーカーの属性情報とに基づ

50

いて生成される。第一の目標仮想スピーカークの属性情報と第一の仮想スピーカーク信号とに基づいて、第二のシーン音声信号が取得され、第一のシーン音声信号および第二シーン音声信号に基づいて、残差信号が生成される。

【0226】

本出願の幾つかの実施形態では、投票回数は、次の関係を満たす。すなわち、

1  $I \leq d$

【0227】

$I$  は、投票回数であり、 $d$  は、音場分類結果に含まれる不均一型音源数である。

【0228】

エンコーダ側は、現行フレームの不均一型音源数に基づいて、最適スピーカークを探索するための投票回数を決定する。投票回数は、現行フレームの不均一型音源数以下であり、これにより、投票回数は、現行フレームの音場分類の実際の状況に適合することができる。これは、現行フレームが符号化される際に、最適スピーカークを探索するための投票回数を決定する必要があるという課題を解決する。

10

【0229】

例えば、投票回数  $I$  は、次のルールに従う必要がある。すなわち、最小の投票回数は 1 であり、最大の投票回数は、スピーカークの総数を超えることなく、最大の投票回数は、仮想スピーカーク信号のチャンネル数を超えない。例えば、スピーカークの総数は、エンコーダにおける仮想スピーカークセット生成ユニットによって取得される 1024 個のスピーカークであり、仮想スピーカーク信号のチャンネル数は、エンコーダによって送信される仮想スピーカーク信号の数、すなわち、 $N$  個の最適スピーカークによって対応して生成される  $N$  個の伝送チャンネルの数である。通常、仮想スピーカーク信号のチャンネル数は、スピーカークの総数未満である。投票回数を推定するための方法は、以下の通りである。現行フレームの音場において、音場分類結果で取得される不均一型音源数に基づいて、最適スピーカークを探索するための投票回数  $I$  を決定するステップ。投票回数  $I$  は、 $1 \leq I \leq d$  の関係を満たす。 $d$  は音場に含まれる異なる方向における音源数、すなわち、音場分類結果において推定される不均一型音源数である。例えば、 $I = d$  である。あるいは、投票回数  $I = \min(d, \text{スピーカークの総数}, \text{仮想スピーカーク信号のチャンネル数}, \text{プリセット投票回数})$  である。投票回数  $I$  は、 $\min(d, \text{スピーカークの総数}, \text{仮想スピーカーク信号のチャンネル数}, \text{投票ラウンドの事前設定数})$  に基づいて取得され得て、これにより、エンコーダ側は、 $I$  の値に基づいて、最適スピーカークを探索するための投票回数を決定し得る。

20

30

【0230】

本出願の幾つかの実施形態では、音場分類結果には、不均一型音源数および音場種別が含まれる。

【0231】

音場種別が不均一型音源である場合、仮想スピーカーク信号のチャンネル数は、次の関係を満たす。すなわち、

$F = \min(S, PF)$

ここで、 $F$  は仮想スピーカーク信号のチャンネル数であり、 $S$  は不均一型音源数であり、 $PF$  はエンコーダによって予め設定される仮想スピーカーク信号のチャンネル数である。または、音場種別が分散型音場である場合、仮想スピーカーク信号のチャンネル数は、次の関係を満たす。すなわち、

40

$F = 1$

ここで、 $F$  は、仮想スピーカーク信号のチャンネル数である。

【0232】

仮想スピーカーク信号のチャンネル数は、仮想スピーカーク信号を送送するためのチャンネル数であり、仮想スピーカーク信号のチャンネル数は、不均一型音源数および音場種別に基づいて決定され得る。前述の計算方式では、音場種別が分散型音場である場合、現行フレームの符号化効率を改善するために、仮想スピーカーク信号のチャンネル数は 1 であると判定される。音場種別が不均一型音源である場合、 $\min$  は、最小値を選択する演算、すなわち、仮

50

想スピーカー信号のチャンネル数としてSおよびPFから最小値を選択する演算を表し、これにより、仮想スピーカー信号のチャンネルは、現行フレームの音場分類の実際の状況に適合することができる。これは、現行フレームを符号化する際に、仮想スピーカー信号のチャンネル数を決定する必要があるという課題を解決する。

#### 【0233】

本出願の幾つかの実施形態では、音場種別が分散型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = \max(C - 1, PR)$$

ここで、PRはエンコードによって予め設定される残差信号のチャンネル数であり、Cはエンコードによって予め設定される残差信号のチャンネル数と、エンコードによって予め設定される仮想スピーカー信号のチャンネル数との合計である。または、音場種別が不均一型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = C - F$$

ここで、Rは残差信号のチャンネル数、Cはエンコードによって予め設定される残差信号のチャンネル数とエンコードによって予め設定される仮想スピーカー信号のチャンネル数の合計、Fはチャンネル数である。仮想スピーカー信号の。

#### 【0234】

仮想スピーカー信号のチャンネル数が取得された後、残差信号のチャンネル数は、残差信号のプリセットチャンネル数と、残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計とに基づいて計算され得る。PRの値は、エンコード側において予め設定され得て、Rの値は、 $\max(C - 1, PR)$ の計算式に従って取得され得る。残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計は、エンコード側において予め設定される。なお、Cは伝送チャンネルの総数と呼ばれることもある。

#### 【0235】

本出願の幾つかの実施形態では、仮想スピーカー信号のチャンネル数が取得された後、残差信号のチャンネル数は、仮想スピーカー信号のチャンネル数と、残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計とに基づいて計算され得る。残差信号のプリセットチャンネル数と仮想スピーカー信号のプリセットチャンネル数との合計は、エンコード側において予め設定される。なお、Cは伝送チャンネルの総数と呼ばれることもある。

#### 【0236】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含む。

#### 【0237】

仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、Fは仮想スピーカー信号のチャンネル数であり、Sは不均一型音源数であり、PFはエンコードによって予め設定される仮想スピーカー信号のチャンネル数である。

#### 【0238】

仮想スピーカー信号のチャンネル数は、仮想スピーカー信号を伝送するためのチャンネル数であり、仮想スピーカー信号のチャンネル数は、不均一型音源数に基づいて決定され得る。前述の計算方式では、minは、最小値を選択する演算、すなわち、仮想スピーカー信号のチャンネル数としてSおよびPFから最小値を選択する演算を表し、これにより、仮想スピーカー信号のチャンネル数は、現行フレームの音場分類の実際の状況に適合することができる。これは、現行フレームを符号化する際に、仮想スピーカー信号のチャンネル数を決定する必要があるという課題を解決する。

#### 【0239】

本出願の幾つかの実施形態では、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = C - F$$

ここで、Rは残差信号のチャンネル数であり、Cはエンコーダによって予め設定される残差信号のチャンネル数とエンコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、Fは仮想スピーカ信号のチャンネル数である。例えば、CはPFおよびPRの合計である。

【0240】

仮想スピーカ信号のチャンネル数が取得された後、残差信号のチャンネル数は、仮想スピーカ信号のチャンネル数と、残差信号のプリセットチャンネル数と仮想スピーカ信号のプリセットチャンネル数との合計とに基づいて計算され得る。残差信号のプリセットチャンネル数と仮想スピーカのプリセットチャンネル数との合計は、エンコーダ側において予め設定される。なお、Cは伝送チャンネルの総数と呼ばれることもある。

10

【0241】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。

【0242】

仮想スピーカ信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比に基づいて取得される。

【0243】

残差信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比に基づいて取得される。

20

【0244】

伝送チャンネルの符号化ビット数には、仮想スピーカ信号の符号化ビット数、および残差信号の符号化ビット数が含まれ、不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比を増加させることによって取得される。

【0245】

エンコーダ側は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比を予め設定し、不均一型音源数を取得し、不均一型音源数が仮想スピーカ信号のチャンネル数以下であるか否かを判定する。不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比は、増加され得て、増加された初期比は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比として定義される。伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比は、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数を計算するために使用され得る。前述の計算方式では、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数は、現行フレームの音場分類の実際の状況に適合することができる。これは、現行フレームを符号化する際に、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数を決定する必要があるという課題を解決する。

30

【0246】

例えば、エンコーダ側は、音場分類結果に基づいて、仮想スピーカ信号および残差信号のビット割り当て方式を決定し、伝送チャンネル信号を仮想スピーカ信号グループおよび残差信号グループに分割し、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比として、仮想スピーカ信号グループのプリセット割り当ての割合を使用する。不均一型音源数 仮想スピーカ信号のチャンネル数である場合、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比は、プリセット調整値に基づいて増加され、増加した比は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比として使用される。例えば、増加した比は、プリセット調整値および初期比の合計に等しくなる。

40

【0247】

50

本出願の幾つかの実施形態では、伝送チャネルの符号化ビット数に対する残差信号の符号化ビット数の比 =  $1 - 0$  - 伝送チャネルの符号化ビット数に対する仮想スピーカ信号の符号化ビット数の比である。

【0248】

本出願の幾つかの実施形態では、前述のステップを実行することに加えて、エンコーダ側によって実行される方法は、以下をさらに含み得る。すなわち、現行フレームおよび音場分類結果を符号化するステップ、および符号化された現行フレームおよび音場分類結果をビットストリームに書き込むステップ。

【0249】

音場分類結果は、ビットストリームに符号化され得る。エンコーダ側がビットストリームをデコーダ側に送信した後、デコーダ側は、ビットストリームに基づいて、音場分類結果を取得し得る。デコーダ側は、ビットストリームを解析することによって、ビットストリームにて搬送される音場分類結果を取得し、音場分類結果に基づいて、現行フレームの音場分布状態を取得し、これにより、現行フレームは、三次元音声信号を取得するために復号化され得る。

【0250】

本出願の幾つかの実施形態では、現行フレームおよび音場分類結果を符号化するステップは、具体的には、以下を含む。すなわち、現行フレームを直接符号化するか、または現行フレームを最初に処理するステップ。および、仮想スピーカ信号および残差信号を取得した後、仮想スピーカ信号および残差信号を符号化するステップ。例えば、エンコーダ側は、具体的にはコアエンコーダであり得る。コアエンコーダは、仮想スピーカ信号、残差信号、および音場分類結果を符号化して、ビットストリームを取得する。ビットストリームは、音声信号符号化ビットストリームと呼ばれることもある。

【0251】

本出願の本実施形態において提供される三次元音声信号処理方法は、音声符号化法および音声復号化法を含み得る。音声符号化法は、音声符号化装置によって実行され、音声復号化法は、音声復号化装置によって実行され、音声符号化装置は、音声復号化装置と通信し得る。図4ないし図6は、音声符号化装置によって実行される。以下に、本技術の一実施形態による、音声復号化装置（デコーダ側と呼ばれる）によって実行される三次元音声信号処理方法について説明する。図7に示されるように、本方法は、主に、以下のステップを含み得る。

【0252】

701：ビットストリームを受信する。

【0253】

デコーダ側は、エンコーダ側からビットストリームを受信する。ビットストリームは、音場分類結果を含む。

【0254】

702：ビットストリームを復号化して、現行フレームの音場分類結果を取得する。

【0255】

デコーダ側は、ビットストリームを解析し、ビットストリームから現行フレームの音場分類結果を取得する。音場分類結果は、図4ないし図6に示される実施形態に従ってエンコーダ側によって取得される。

【0256】

703：音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得する。

【0257】

音場分類結果を取得した後、デコーダ側は、音場分類結果に基づいて、ビットストリームを解析して、復号化された現行フレームの三次元音声信号を取得する。本出願の実施形態では、現行フレームの復号プロセスは限定されない。本出願の本実施形態では、デコーダ側は、音場分類結果に基づいて、現行フレームを復号化し得る。音場分類の結果は、ビ

10

20

30

40

50

ットストリームにおける現行フレームを復号化するために使用することができる。そのため、デコーダ側は、現行フレームの音場に適応した復号化方式において復号化を実行して、エンコーダ側によって送信される三次元音声信号を取得する。これは、エンコーダ側からデコーダ側への音声信号の伝送を実装する。

【0258】

例えば、デコーダ側は、ビットストリームにて伝送される音場分類結果に基づいて、エンコーダ側の符号化モードおよび/もしくは符号化パラメータと一致する復号化モードおよび/もしくは復号化パラメータを決定することができる。エンコーダ側が符号化モードおよび/もしくは符号化パラメータをデコーダ側に伝送する方式と比較して、符号化ビット数が削減される。

10

【0259】

本出願の幾つかの実施形態では、ステップ703における音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するステップは、以下を含む。すなわち、  
G1：音場分類結果に基づいて、現行フレームの復号化モードを決定するステップ。および、  
G2：復号化モードに基づいて、復号化された現行フレームの三次元音声信号を取得するステップ。

【0260】

復号化モードは、前述の実施形態における符号化モードに対応する。ステップG1の実装は、前述の実施形態におけるステップ504と同様である。詳細については、本明細書では改めて説明しない。復号化モードを取得した後、デコーダ側は、復号化モードに基づいて、ビットストリームを復号化して、復号化された現行フレームの三次元音声信号を取得し得る。

20

【0261】

さらに、本出願の幾つかの実施形態では、ステップG1における音場分類結果に基づいて、現行フレームの復号化モードを決定するステップは、以下を含む。すなわち、音場分類結果が不均一型音源数を含むか、または音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームの復号化モードを決定するステップ。

音場分類結果が音場種別を含むか、または音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームの復号化モードを決定するステップ。または、

30

音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームの復号化モードを決定するステップ。

【0262】

前述のステップの実装は、前述の実施形態におけるステップE1ないしステップE3の実装と同様である。詳細については、本明細書では改めて説明しない。

【0263】

本出願の幾つかの実施形態では、不均一型音源数に基づいて、現行フレームの復号化モードを決定するステップは、以下を含む。すなわち、

40

不均一型音源数がプリセット条件を満たす場合、復号化モードが第一の復号化モードであると判定するステップ。または

不均一型音源数がプリセット条件を満たさない場合、復号化モードが第二の復号化モードであると判定するステップ。

【0264】

第一の復号化モードは、仮想スピーカー選択に基づくHOA復号化モード、もしくは指向性音声コーディングに基づくHOA復号化モードであり、第二の復号化モードは、仮想スピーカー選択に基づくHOA復号化モード、もしくは指向性音声コーディングに基づくHOA復号化モードであり、第一の復号化モードおよび第二の復号化モードは、相違する復号化モードである。

50

## 【 0 2 6 5 】

プリセット条件は、異なる復号化モードを識別するために、デコーダ側によって設定される条件であり、プリセット条件の実装は限定されないことは、留意されるべきである。

## 【 0 2 6 6 】

本出願の幾つかの実施形態では、プリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または

プリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

## 【 0 2 6 7 】

本出願の幾つかの実施形態では、ステップ 7 0 3 における音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するステップは、以下を含む。すなわち、H 1 : 音場分類結果に基づいて、現行フレームの復号化パラメータを決定するステップ。および、

H 2 : 復号化パラメータに基づいて、復号化された現行フレームの三次元音声信号を取得するステップ。

## 【 0 2 6 8 】

復号化パラメータは、前述の実施形態における符号化パラメータに対応する。ステップ H 1 の実装は、前述の実施形態におけるステップ 6 0 4 と同様である。詳細については、本明細書では改めて説明しない。復号化パラメータを取得した後、デコーダ側は、復号化パラメータに基づいて、ビットストリームを復号化して、復号化された現行フレームの三次元音声信号を取得し得る。

## 【 0 2 6 9 】

本出願の幾つかの実施形態では、復号化パラメータは、以下のうちの少なくとも一つを含む。すなわち、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の復号化ビット数、仮想スピーカー信号の符号化ビット数、もしくは残差信号の復号化ビット数。

## 【 0 2 7 0 】

仮想スピーカー信号および残差信号は、ビットストリームを復号化することによって取得される。

## 【 0 2 7 1 】

本出願の幾つかの実施形態では、音場分類結果には、不均一型音源数および音場種別が含まれる。

## 【 0 2 7 2 】

音場種別が不均一型音源である場合、仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min ( S , P F )$$

ここで、F は仮想スピーカー信号のチャンネル数であり、S は不均一型音源数であり、P F は、デコーダによって予め設定される仮想スピーカー信号のチャンネル数である。または、音場種別が分散型音場である場合、仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = 1$$

ここで、F は仮想スピーカー信号のチャンネル数である。

## 【 0 2 7 3 】

本出願の幾つかの実施形態では、音場種別が分散型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = \max ( C - 1 , P R )$$

ここで、P R はデコーダによって予め設定される残差信号のチャンネル数であり、C はデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカー信号のチャンネル数との合計である。または

10

20

30

40

50

音場種別が不均一型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = C - R$$

ここで、 $R$ は残差信号のチャンネル数であり、 $C$ はデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$ は仮想スピーカ信号のチャンネル数である。

【0274】

デコーダによって予め設定される仮想スピーカ信号のチャンネル数は、エンコーダによって予め設定される仮想スピーカ信号のチャンネル数と等しいことは、留意されるべきである。同様に、デコーダによって予め設定される残差信号のチャンネル数は、エンコーダによって予め設定される残差信号のチャンネル数と等しくなる。

10

【0275】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含む。

【0276】

仮想スピーカ信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、 $F$ は仮想スピーカ信号のチャンネル数であり、 $S$ は不均一型音源数であり、 $PF$ は、デコーダによって予め設定される仮想スピーカ信号のチャンネル数である。

【0277】

本出願の幾つかの実施形態では、残差信号のチャンネル数は、次の関係を満たす。すなわち、

20

$$R = C - F$$

ここで、 $R$ は残差信号のチャンネル数であり、 $C$ はデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$ は仮想スピーカ信号のチャンネル数である。

【0278】

復号化パラメータの実装は、前述の実施形態における符号化パラメータの実装と同様であることは、留意されるべきである。詳細については、本明細書では改めて説明しない。

【0279】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。

30

【0280】

仮想スピーカ信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。

【0281】

残差信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。

【0282】

伝送チャンネルの復号化ビット数には、仮想スピーカ信号の復号化ビット数および残差信号の復号化ビット数が含まれ、不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の初期比を増加させることによって取得される。

40

【0283】

本出願の実施形態における前述の解決策をより良く理解して実装するために、対応する用途シナリオを例として使用することによって、具体的な説明を以下に提供する。

【0284】

本出願の本実施形態では、三次元音声信号がHOA信号である例を使用する。本出願の本実施形態におけるHOA信号のための音場分類方法は、ハイブリッドHOAエンコーダに適用される。図8は、基本的な符号化手順を示している。エンコーダ側は、符号化対象

50

のH O A信号に対して分類を実行して、現行フレームの符号化対象のH O A信号が、仮想スピーカー選択に基づくH O A符号化スキームに適しているか、もしくは指向性音声コーディングD i r A Cに基づくH O A符号化スキームに適しているかを判定し、音場分類結果に基づいて、現行フレームのH O A符号化モードを決定する。具体的には、H O Aエンコーダは、エンコーダ選択ユニットを含む。エンコーダ選択ユニットは、符号化対象のH O A信号に対して音場分類を実行し、現行フレームの符号化モードを決定する。そして、符号化モードに基づいて、符号化のためのエンコーダAもしくはエンコーダBを選択して、最終的な符号化されたビットストリームを取得する。エンコーダAおよびエンコーダBは、異なる種類のエンコーダを示し、各種類のエンコーダは、現行フレームの音場種別に適応される。音場種別に適応するエンコーダが符号化のために使用されると、信号の圧縮率を改善することができる。

10

**【0285】**

符号化対象のH O A信号に対して音場分類を実行し、符号化モードを決定する具体的なプロセスは、以下を含む。すなわち、符号化対象のH O A信号に対して音場分類を実行して、音場分類結果を取得するステップ。および、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するステップ。

**【0286】**

現行フレームの符号化モードは、現行フレームのエンコーダの選択方式を示す。エンコーダ選択識別子を決定するための基準は、エンコーダAおよびエンコーダBが適用可能であるH O A信号の音場種別に基づいて決定され得る。例えば、エンコーダAによって処理される信号タイプは、不均一型音場を有し、かつ不均一型音源数が3未満であるH O A信号であり、エンコーダBによって処理される信号タイプは、不均一型音場を有し、かつ不均一型音源数が3以上であるH O A信号である。あるいは、エンコーダBによって処理される信号タイプは、分散型音場を有するH O A信号、もしくは不均一型音源数が3以上であるH O A信号である。

20

**【0287】**

確実に、連続するフレーム間の符号化モードが頻繁に切り替わらないようにするために、ハングオーバー時間枠処理が、音場分類結果に対して実行されることもあることは、留意されるべきである。複数のハングオーバー時間枠の処理方法が存在する。これは、本出願の本実施形態では限定されない。例えば、処理方式は、長さがN個のフレームであるエンコーダ選択識別子をハングオーバー時間枠に保存するステップであって、N個のフレームは、現行フレームと、現行フレームより前のN - 1個のフレームとのエンコーダ選択識別子を含む、ステップと、エンコーダ選択識別子が指定された閾値まで累積されると、現行フレームの符号化タイプ指示識別子を更新するステップとし得る。任意選択として、ハングオーバー時間枠処理に加えて、他の処理が、音場分類結果の修正を実行するために使用され得る。

30

**【0288】**

図9に示されるように、H O A信号の符号化モードを決定する手順は、主に、以下を含む。

40

**【0289】**

S 0 1 : 分岐対象のH O A信号を取得する。

**【0290】**

S 0 2 : H O A信号に対してダウンサンプリングを実行する。

**【0291】**

解析対象のH O A信号に対してダウンサンプリングを実行するステップが任意選択のステップであることは限定されない。

**【0292】**

計算の複雑さを軽減するために、ダウンサンプリングが、解析対象のH O A信号に対して実行される。解析対象のH O A信号は、時間領域のH O A信号であってもよいし、また

50

は周波数領域のH O A信号であってもよい。解析対象のH O A信号には、全チャンネルが含まれてもよいし、または一部のH O Aチャンネル（例えば、F O Aチャンネルなど）が含まれてもよい。例えば、解析対象のH O A信号は、全サンプリング点であってもよいし、1 / Qのダウンサンプリング点であってもよい。例えば、本実施形態では、1 / 1 2 0のダウンサンプリング点を使用される。

【 0 2 9 3 】

例えば、現行フレームにおけるH O A信号の次数は3であり、H O A信号のチャンネル数は1 6であり、現行フレームのフレーム長は2 0ミリ秒（m s）であり、すなわち、現行フレームの信号には、9 6 0個のサンプリング点が含まれる。現行フレームの符号化対象のH O A信号が1 / 1 2 0のダウンサンプリングによって処理された後、信号の各チャンネルは、8個のサンプリング点を含む。換言すると、H O A信号は、1 6チャンネルを有し、各チャンネルは、8個のサンプリング点を有し、音場種別解析の入力信号、すなわち解析対象のH O A信号を構成する。

10

【 0 2 9 4 】

S 0 3：ダウンサンプリングを介して取得される信号に基づいて、音場種別解析を実行する。

【 0 2 9 5 】

H O A信号に対してダウンサンプリングが実行された後、H O A信号の不均一型音源数を分析することによって、音場種別が取得される。

【 0 2 9 6 】

20

例えば、本出願の本実施形態における音場種別解析は、H O A信号に対して線形分解を実行するステップと、線形分解を介して線形分解結果を取得するステップと、次いで、線形分解結果に基づいて、音場分類結果を取得するステップとし得る。

【 0 2 9 7 】

例えば、不均一型音源数は、線形分解結果に基づいて取得することができる。例えば、線形分解結果は、特徴値を含み得る。不均一型音源数が特徴値間の比に基づいて推定されることは、具体的には、以下を含む。すなわち、解析対象のH O A信号に対して特異値分解を実行して、特異値 $v[i]$ を取得するステップ。ここで、 $i = 0, 1, \dots, \min(L, K) - 1$ とする。

【 0 2 9 8 】

30

LはH O A信号のチャンネル数に等しく、Kは現行フレームにおける各チャンネルの信号点の個数である。例えば、信号点の個数は、周波数の個数であり得る。本実施形態では、 $L = 16$ 、 $K = 8$ 、および $\min(L, K) = 8$ である。

【 0 2 9 9 】

特異値 $v$ 間の比 $temp[i]$ が計算され、音場分類パラメータとして使用される。すなわち、

$$temp[i] = v[i] / v[i + 1]$$

ここで、 $i = 0, 1, \dots, \min(L, K) - 2$ とする。

【 0 3 0 0 】

不均一型音源判定閾値は1 0 0であり、不均一型音源数 $n$ は、次の方式において推定され得る。

40

$temp[i]$ が $i = 0$ から1 0 0を超えるか否かを判定するステップ。および、 $temp[i]$ が1 0 0以上であり、かつ、 $temp[i] - 1 0 0$ が満たされる場合、判定するステップを停止するステップ。それ以外の場合、 $i = i + 1$ として、判定するステップの実行を継続するステップ。判定するステップを停止する場合、不均一型音源数 $n$ は、判定するステップを停止する際のシーケンス番号 $i$ に1を加えたものに等しくなる。例えば、 $i = 0$ である際に、 $temp[0] - 1 0 0$ である場合、判定するステップは停止され、不均一型音源数 $n$ は、1に等しくなる。それ以外の場合、 $i$ は1に設定され、 $1 = 1$ である際に、判定するステップは、継続して実行される。 $i = 1$ 、かつ $temp[1] - 1 0 0$ である場合、判定するステップは停止され、不均一型音源数 $n$ は、 $i + 1 = 2$ に等し

50

くなる。

【 0 3 0 1 】

S 0 4 : 音場種別の解析結果に基づいて、予測符号化モードを決定する。

【 0 3 0 2 】

予測符号化モードは、不均一型音源数  $n$  に基づいて決定される。

【 0 3 0 3 】

$0 < n < 3$  である場合、予測符号化モードは、符号化モード 1 になる。

【 0 3 0 4 】

$n = 3$  もしくは  $n = 0$  である場合、予測符号化モードは、符号化モード 2 になる。

【 0 3 0 5 】

例えば、符号化モード 1 は、仮想スピーカー選択に基づく H O A 符号化モードであり得る。符号化モード 2 は、指向性音声コーディング D i r A C に基づく H O A 符号化方式であり得る。

【 0 3 0 6 】

S 0 5 : 予測符号化モードに基づいて、実際の符号化モードを決定する。

【 0 3 0 7 】

現行フレームの予測符号化モードが決定された後、次いで、実際の符号化モードが決定される。例えば、ハングオーバー時間枠は、実際の符号化モードを決定するために使用される。ハングオーバー時間枠では、ハングオーバー時間枠における複数のフレームの予測符号化モード 2 が指定された閾値まで累積されると、現行フレームの実際の符号化モードは、符号化モード 2 となる。それ以外の場合、現行フレームの実際の符号化モードは、符号化モード 1 となる。

【 0 3 0 8 】

例えば、ステップ S 0 3 における現行フレームの符号化モード判定結果と、現行フレームより 9 フレーム前の符号化モード結果とを含む、ハングオーバー時間枠における 1 0 フレーム分の予測符号化モード結果が存在する。1 0 フレームの予測符号化モード結果のうちで、符号化モードが符号化モード 2 であるフレームが、7 フレームまで蓄積される場合、現行フレームの実際の符号化モードは、符号化モード 2 として決定される。

【 0 3 0 9 】

S 0 6 : 最終的な符号化モードを取得する。

【 0 3 1 0 】

エンコーダ側に相当するハイブリッド H O A デコーダの基本的な復号化手順を図 1 0 に示す。デコーダ側は、エンコーダ側からビットストリームを取得し、次いで、そのビットストリームを解析して、現行フレームの H O A 復号化モードを取得する。対応する復号化スキームは、現行フレームの H O A 復号化モードに基づいて、復号化のために選択されて、再構成された H O A 信号を取得する。具体的には、デコーダ側は、デコーダ選択ユニットを含む。デコーダ選択ユニットは、ビットストリームを解析し、復号化モードを決定し、復号化モードに基づいて、復号化のためのデコーダ A もしくはデコーダ B を選択して、再構成された H O A 信号を取得する。デコーダ A およびデコーダ B は、異なる種類のデコーダを示し、各種類のデコーダは、現行フレームの音場種別に適応される。音場種別に適応したデコーダが復号化のために使用されると、H O A 信号は、正しく再構成することができる。

【 0 3 1 1 】

前述の説明から、音場分類が、符号化対象の H O A 信号に対して実行され、符号化モードが、音場分類の結果に基づいて決定され、これにより、異なる符号化モードが、適切な信号タイプに対して使用されて、種々のタイプの信号に対して最大の圧縮効率を取得することが分かる。

【 0 3 1 2 】

以下に、本出願の一実施形態による、仮想スピーカー選択に基づく H O A エンコーダについて説明する。図 1 1 は、基本的な符号化手順を示している。

10

20

30

40

50

**【 0 3 1 3 】**

エンコーダ側は、以下を含み得る。すなわち、仮想スピーカー構成ユニット、符号化解析ユニット、仮想スピーカーセット生成ユニット、仮想スピーカー選択ユニット、仮想スピーカー信号生成ユニット、コアエンコーダ処理ユニット、信号再構成ユニット、残差信号生成ユニット、選択ユニット、および信号補償ユニット。以下に、エンコーダ側に含まれるユニットの機能を個別に説明する。本出願の本実施形態では、図 1 1 に示されるエンコーダ側が、一つの仮想スピーカー信号もしくは複数の仮想スピーカー信号を生成し得る。複数の仮想スピーカー信号を生成する手順は、図 1 1 に示されるエンコーダの構成に基づく生成を複数回の間実行し得る。以下では、一例として、一つの仮想スピーカー信号を生成する手順を使用する。

10

**【 0 3 1 4 】**

仮想スピーカー構成ユニットは、仮想スピーカーセットにおいて仮想スピーカーを構成して、複数の仮想スピーカーを取得するように構成される。

**【 0 3 1 5 】**

仮想スピーカー構成ユニットは、エンコーダ構成情報に基づいて、仮想スピーカー構成パラメータを出力する。エンコーダ構成情報には、H O A 次数、符号化ビットレート、およびユーザ定義情報などが含まれるが、これらに限定されない。仮想スピーカー構成パラメータには、仮想スピーカーの個数、仮想スピーカーの H O A 次数、および仮想スピーカーの位置座標などが含まれるが、これらに限定されない。

**【 0 3 1 6 】**

仮想スピーカー構成ユニットによって出力される仮想スピーカー構成パラメータは、仮想スピーカーセット生成ユニットの入力として使用される。

20

**【 0 3 1 7 】**

符号化解析ユニットは、符号化対象の H O A 信号に対して符号化解析を実行するように構成され、例えば、符号化対象の H O A における、音源数、指向性、および符号化対象信号の分散度などの、特徴を含む音場分布を解析する。この特徴は、目標仮想スピーカーを選択する方法を決定するための決定条件の一つとして使用される。

**【 0 3 1 8 】**

本出願の本実施形態では、エンコーダ側が、代替的に、符号化解析ユニットを含み得ないことは、限定されない。換言すると、エンコーダ側は、入力信号を解析しないが、デフォルト構成を使用して、目標仮想スピーカーを選択する方法を決定し得る。

30

**【 0 3 1 9 】**

エンコーダ側は、符号化対象の H O A 信号を取得する。例えば、エンコーダ側は、実際の収集機器から記録される H O A 信号、もしくは人工音声プロジェクトを使用することによって合成される H O A 信号をエンコーダの入力として使用し得る。また、エンコーダによって入力される符号化対象の H O A 信号は、時間領域の H O A 信号であってもよいし、または周波数領域の H O A 信号であってもよい。

**【 0 3 2 0 】**

仮想スピーカーセット生成ユニットは、仮想スピーカーセットを生成するように構成される。仮想スピーカーセットは、複数の仮想スピーカーを含み得て、仮想スピーカーセットにおける仮想スピーカーは、「候補仮想スピーカー」と呼ばれることもある。

40

**【 0 3 2 1 】**

仮想スピーカーセット生成ユニットは、仮想スピーカー構成パラメータに基づいて、指定された候補仮想スピーカーの H O A 係数を生成する。候補仮想スピーカーの H O A 係数を生成するには、候補仮想スピーカーの座標（すなわち、位置座標もしくは位置情報）と候補仮想スピーカーの H O A 次数とが必要とされる。候補仮想スピーカーの座標を決定する方法には、等距離原理に従って K 個の仮想スピーカーを生成するステップと、聴覚原理に従って不均等に分布される K 個の候補仮想スピーカーを生成するステップとが含まれるが、これらに限定されない。以下に、均等に分布される一定数の仮想スピーカーを生成するステップの例を説明する。

50

## 【 0 3 2 2 】

均等に分布される候補仮想スピーカースの座標は、候補仮想スピーカースの数に基づいて生成され、例えば、ほぼ均等な仮想スピーカースの配置は、数値反復計算法を使用することによって取得される。

## 【 0 3 2 3 】

仮想スピーカースセット生成ユニットによって出力される候補仮想スピーカースの H O A 係数は、仮想スピーカース選択ユニットの入力として使用される。

## 【 0 3 2 4 】

仮想スピーカース選択ユニットは、符号化対象の H O A 信号に基づいて、仮想スピーカースセットにおける複数の候補仮想スピーカースから目標仮想スピーカースを選択するように構成され、目標仮想スピーカースは、「符号化対象の H O A 信号に適合する仮想スピーカース」、もしくは適合仮想スピーカースと呼ばれることがある。

10

## 【 0 3 2 5 】

仮想スピーカース選択ユニットは、符号化対象の H O A 信号を、仮想スピーカースセット生成ユニットによって出力される、候補仮想スピーカースの H O A 係数と適合させ、指定された適合仮想スピーカースを選択する。

## 【 0 3 2 6 】

本発明の本実施形態では、音場分類結果を取得するために、音場分類が、符号化対象の H O A 信号に対して実行され、符号化パラメータが、音場分類結果に基づいて決定される。

## 【 0 3 2 7 】

符号化解析ユニットは、符号化対象の H O A 信号に基づいて、符号化解析を実行するように構成され、この解析は、以下を含む。すなわち、符号化対象の H O A 信号に基づいて、音場分類を実行するステップ。音場の分類方法については、前述の実施形態を参照されたい。詳細については、本明細書では改めて説明しない。

20

## 【 0 3 2 8 】

符号化パラメータは、音場分類結果に基づいて決定される。符号化パラメータは、仮想スピーカース信号のチャンネル数、残差信号のチャンネル数、もしくは仮想スピーカース選択に基づく H O A 符号化スキームにおいて最適スピーカースを探索するための投票回数のうちの少なくとも一つを含み得る。

## 【 0 3 2 9 】

具体的には、仮想スピーカース選択ユニットは、最適スピーカースを探索するために決定される投票回数と、仮想スピーカース信号のチャンネルとに基づいて、符号化対象の H O A 係数を、仮想スピーカースセット生成ユニットによって出力される、候補仮想スピーカースの H O A 係数と適合させ、最適仮想スピーカースを選択し、最適仮想スピーカースの H O A 係数を取得する。最適仮想スピーカースの数は、仮想スピーカース信号のチャンネル数に等しくなる。

30

## 【 0 3 3 0 】

仮想スピーカース選択ユニットは、投票に基づく最適スピーカース探索方法を使用することによって、符号化対象の H O A 係数を、仮想スピーカースセット生成ユニットによって出力される、候補仮想スピーカースの H O A 係数に適合させ、最適仮想スピーカースを選択し、音場分類結果に基づいて、最適スピーカースを探索するための投票回数 I を決定し得る。

40

## 【 0 3 3 1 】

投票回数 I は、次の規則に従う必要がある。すなわち、最小の投票回数は 1 であり、最大の投票回数は、スピーカースの総数（例えば、仮想スピーカースセット生成ユニットによって取得される 1 0 2 4 個のスピーカースなど）と、仮想スピーカース信号のチャンネル数（エンコードによって送信される仮想スピーカース信号の数、すなわち、N 個の最適スピーカースによって対応して生成される N 個の伝送チャンネル）とをを超えない。通常、仮想スピーカース信号のチャンネル数は、スピーカースの総数未満である。

## 【 0 3 3 2 】

投票回数を推定するための方法は、次の通りである。すなわち、

50

音場分類結果で取得される、音場における不均一型音源数に基づいて、スピーカーを選択するための投票回数  $I$  を決定するステップ。

【0333】

投票回数  $I$  は、 $1 \leq I \leq d$  を満たす。  $d$  は、音場に含まれる異なる方向における音源数、すなわち、音場分類結果において推定される不均一型音源数である。例えば、 $I = d$  である。

【0334】

仮想スピーカー信号のチャンネル数および残差信号のチャンネル数は、音場種別に基づいて決定される。

【0335】

次いで、本出願の実施形態は、適応仮想スピーカー信号のチャンネル数  $F$  を選択するための方法を提供する。

【0336】

音場種別が不均一型音場である場合、 $F = \min(S, PF)$  となる。ここで、 $S$  は音場における不均一型音源数であり、 $PF$  はエンコーダによって予め設定される仮想スピーカー信号のチャンネル数である。

【0337】

音場種別が分散型音場である場合、 $F = 1$  となる。

【0338】

次いで、本出願の一実施形態は、適応残差信号のチャンネル数  $R$  を選択するための方法を提供する。

【0339】

音場種別が分散型音源場である場合、 $R = \max(C - 1, PR)$  となる。ここで、 $C$  は予め設定される伝送チャンネルの総数であり、 $PR$  はエンコーダによって予め設定される残差信号数である。例えば、 $C$  は  $PF$  および  $PR$  の合計である。

【0340】

音場種別が不均一型音源である場合、 $R = C - F$  となる。

【0341】

音場分類結果に基づいて、仮想スピーカー信号および残差信号のビット割り当てを決定するための方法は、次の通りである。

【0342】

不均一型音源数 仮想スピーカー信号のチャンネル数である場合、残差信号のエネルギーが低いため、より多くのビットが、仮想スピーカー信号のチャンネルに割り当てられ得る。

【0343】

幾つかの実施形態では、仮想スピーカー信号および残差信号は、二つのグループ、すなわち、仮想スピーカー信号グループおよび残差信号グループに分割される。不均一型音源数 仮想スピーカー信号のチャンネル数である場合、プリセット調整値に基づいて、仮想スピーカー信号グループの予め設定される割り当ての割合が増加され、仮想スピーカー信号グループの増加した割り当ての割合が、仮想スピーカー信号グループの割り当ての割合として使用される。

【0344】

残差信号グループの割り当ての割合 =  $1 - 0$  - 仮想スピーカー信号グループの割り当ての割合である。

【0345】

仮想スピーカー信号生成ユニットは、符号化対象の  $HOA$  係数と、最適仮想スピーカーの  $HOA$  係数とに基づいて、仮想スピーカー信号を算出する。

【0346】

信号再構成ユニットは、仮想スピーカー信号と、最適仮想スピーカーの  $HOA$  係数とに基づいて、 $HOA$  信号を再構成する。

【0347】

10

20

30

40

50

残差信号生成ユニットは、ステップ 1 において決定された残差信号のチャンネル数、符号化対象の H O A 係数、および H O A 信号再構成ユニットによって出力される再構成 H O A 信号に基づいて、残差信号を算出する。

【 0 3 4 8 】

N 次のアンビソニック係数を有する残差信号と比較して、N 次のアンビソニック係数未満であるチャンネル数が、送信対象の残差信号として選択される場合、情報損失が発生するため、信号補償ユニットは、送信されない残差信号に対して情報補償を実行する必要がある。

【 0 3 4 9 】

仮想スピーカ信号は、高い振幅もしくはエネルギーを有し、送信対象の残差信号は、低い振幅もしくはエネルギーを有する。そのため、選択ユニットは、利用可能な全てのビットを仮想スピーカ信号および送信対象の残差信号に事前に割り当てる。取得されたビット事前割り当て情報は、処理のためにコアエンコーダを誘導するために使用される。

【 0 3 5 0 】

コアエンコーダ処理ユニットは、伝送チャンネルに対してコアエンコーダ処理を実行し、伝送ビットストリームを出力する。伝送チャンネルには、仮想スピーカ信号のチャンネルおよび残差信号のチャンネルが含まれる。

【 0 3 5 1 】

符号化パラメータは、音場分類結果に基づいて決定される。符号化パラメータは、仮想スピーカ選択に基づく H O A 符号化スキームにおける、仮想スピーカ信号のビット割り当ておよび残差信号のビット割り当てのうちの少なくとも一つをさらに含み得る。仮想スピーカ信号のビット割り当ておよび残差信号のビット割り当てが、音場分類結果に基づいて決定される場合、音場分類結果に基づいて、仮想スピーカ信号および残差信号のビット割り当てを決定する必要がある。

【 0 3 5 2 】

幾つかの実施形態では、音場分類結果に基づいて仮想スピーカ信号および残差信号のビット割り当てを決定するための方法は、次の通りである。すなわち、仮想スピーカ信号のチャンネル数は F であり、仮想スピーカのチャンネル数は R であり、残差信号のチャンネル数は R であり、仮想スピーカ信号および残差信号を符号化するために使用することができるビットの総数は、`numbit` である。

【 0 3 5 3 】

一つの方式では、最初に、仮想スピーカ信号の符号化ビットの総数と残差信号の符号化ビットの総数が決定され、次いで、各チャンネルの符号化ビット数が決定される。例えば、仮想スピーカ信号の符号化ビットの総数は、

【 0 3 5 4 】

【数 1 8 】

$$\text{core}_{\text{numbit}} = \text{round} \left( \frac{fac1 * F * \text{numbit}}{fac1 * F + fac2 * R} \right)$$

【 0 3 5 5 】

である。

【 0 3 5 6 】

`fac1` は、仮想スピーカ信号の符号化ビットに割り当てられる重み係数であり、`fac2` は、残差信号の符号化ビットに割り当てられる重み係数であり、`round ( )` は、切り捨てを表す。例えば、`fac1 > fac2` である。例えば、`fac1 = 2`、かつ、`fac2 = 1` である。

【 0 3 5 7 】

残差信号の符号化ビットの総数は、

$$\text{res\_numbit} = \text{numbit} - \text{core\_numbit}$$

になる。

【 0 3 5 8 】

次いで、仮想スピーカ信号の各チャネルの符号化ビットは、仮想スピーカ信号のビット割当基準に従って割り当てられ、残差信号の各チャネルの符号化ビットは、残差信号のビット割当基準に従って割り当てられる。

【 0 3 5 9 】

あるいは、残差信号の符号化ビットの総数は、

【 0 3 6 0 】

【 数 1 9 】

$$\text{res\_numbit} = \text{round}\left(\text{fac2} * R * \frac{\text{numbit}}{\text{fac1} * F + \text{fac2} * R}\right)$$

10

【 0 3 6 1 】

になる。

【 0 3 6 2 】

$\text{fac1}$  は、仮想スピーカ信号の符号化ビットに割り当てられる重み係数であり、 $\text{fac2}$  は、残差信号の符号化ビットに割り当てられる重み係数であり、 $\text{round}()$  は、切り捨てを表す。例えば、 $\text{fac1} > \text{fac2}$  である。例えば、 $\text{fac1} = 2$ 、かつ、 $\text{fac2} = 1$  である。

20

【 0 3 6 3 】

次いで、仮想スピーカ信号の符号化ビットの総数は、

$$\text{core\_numbit} = \text{numbit} - \text{res\_numbit}$$

となる。

【 0 3 6 4 】

次いで、仮想スピーカ信号の各チャネルの符号化ビットは、仮想スピーカ信号のビット割当基準に従って割り当てられ、残差信号の各チャネルの符号化ビットは、残差信号のビット割当基準に従って割り当てられる。

【 0 3 6 5 】

また、各チャネルの符号化ビット数は、代替的に、直接決定され得る。例えば、各仮想スピーカ信号の符号化ビット数は、

30

【 0 3 6 6 】

【 数 2 0 】

$$\text{core\_ch\_numbit} = \text{round}\left(\text{fac1} * \frac{\text{numbit}}{\text{fac1} * F + \text{fac2} * R}\right)$$

【 0 3 6 7 】

となる。

【 0 3 6 8 】

各残差信号の符号化ビット数は、

40

【 0 3 6 9 】

【 数 2 1 】

$$\text{res\_ch\_numbit} = \text{round}\left(\text{fac2} * \frac{\text{numbit}}{\text{fac1} * F + \text{fac2} * R}\right)$$

【 0 3 7 0 】

となる。

【 0 3 7 1 】

50

仮想スピーカ信号および残差信号の符号化に最終的に使用されるビット割り当て結果は、前述の方法を使用することによって取得される調整ビット割り当て結果に基づいて決定され得る。仮想スピーカ信号および残差信号を符号化するためのビット割り当て結果を取得した後、コアエンコーダ処理ユニットは、ビット割り当て結果に基づいて、仮想スピーカ信号および残差信号を符号化する。

【 0 3 7 2 】

音場分類が、符号化対象のH O A信号に対して実行され、符号化パラメータが、音場分類結果に基づいて決定され、符号化対象信号が、決定された符号化パラメータに基づいて符号化される。符号化パラメータは、仮想スピーカ信号のチャンネル数、残差信号のチャンネル数、仮想スピーカ信号のビット割り当て、残差信号のビット割り当て、もしくは仮想スピーカ選択に基づくH O A符号化スキームにおける最適スピーカを探索するための投票回数のうちの少なくとも一つを含む。符号化パラメータの説明については、前述の内容を参照されたい。詳細については、本明細書では改めて説明しない。

10

【 0 3 7 3 】

前述の例から、本出願の本実施形態では、音場分類が、符号化対象のH O A信号に対して実行され、これにより、H O A信号を符号化するために、適切な符号化モードおよび/もしくは符号化パラメータが、符号化対象のH O A信号における異なる特徴に基づいて選択されることが分かる。これは、圧縮効率および聴覚品質を改善する。

【 0 3 7 4 】

デコーダ側によって実行される復号化手順については、本出願の実施形態では詳細に説明しない。

20

【 0 3 7 5 】

簡単に説明するために、前述の方法の実施形態は、一連の動作として表現されることは、留意されるべきである。しかしながら、当業者は、本出願によれば、幾つかのステップが他の順序で、もしくは同時に実行され得るため、本出願が記載された動作の順序に限定されないことを理解するはずである。さらに、本明細書において説明される実施形態は、全て実施形態の例に属しており、関与する動作およびモジュールは、本出願によって必ずしも必要とされないことも、さらに当業者には理解されるべきである。

【 0 3 7 6 】

本出願の実施形態の解決策をより適切に実装するために、解決策を実装するための関連装置が、以下にさらに提供される。

30

【 0 3 7 7 】

図12は、本出願の一実施形態による、三次元音声信号処理装置を示している。例えば、三次元音声信号処理装置は、具体的には音声符号化装置1200であり、線形解析モジュール1201、パラメータ生成モジュール1202、および音場分類モジュール1203を含み得る。

【 0 3 7 8 】

線形解析モジュールは、三次元音声信号に対して線形分解を実行して、線形分解結果を取得するように構成される。

【 0 3 7 9 】

40

パラメータ生成モジュールは、線形分解結果に基づいて、現行フレームに対応する音場分類パラメータを取得するように構成される。

【 0 3 8 0 】

音場分類モジュールは、音場分類パラメータに基づいて、現行フレームの音場分類結果を決定するように構成される。

【 0 3 8 1 】

本出願の幾つかの実施形態では、三次元音声信号は、高次アンビソニックスH O A信号、もしくは一次アンビソニックスF O A信号を含む。

【 0 3 8 2 】

本出願の幾つかの実施形態では、線形解析モジュールは、以下を行うように構成される

50

。すなわち、現行フレームに対して特異値分解を実行して、現行フレームに対応する特異値を取得することであって、線形分解結果は、特異値を含む、こと。現行フレームに対して主成分分析を実行して、現行フレームに対応する第一の特徴値を取得することであって、線形分解結果は、第一の特徴値を含む、こと。または、現行フレームに対して独立成分分析を実行して、現行フレームに対応する第二の特徴値を取得することであって、線形分解結果は第二の特徴値を含むこと。

【 0 3 8 3 】

本出願の幾つかの実施形態では、複数の線形分解結果が存在し、複数の音場分類パラメータが存在する。

【 0 3 8 4 】

パラメータ生成モジュールは、以下を行うように構成される。すなわち、現行フレームの  $(i + 1)$  番目の線形解析結果に対する、現行フレームの  $i$  番目の線形解析結果の比を取得することであって、 $i$  は、正の整数である、こと。および、その比に基づいて、現行フレームに対応する  $i$  番目の音場分類パラメータを取得すること。

【 0 3 8 5 】

任意選択として、 $i$  番目の線形解析結果および  $(i + 1)$  番目の線形解析結果は、現行フレームにおける連続する二つの線形解析結果である。

【 0 3 8 6 】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在し、音場分類結果は、音場種別を含む。音場分類モジュールは、以下を行うように構成される。すなわち、複数の音場分類パラメータの値が全て予め設定される分散型音源判定条件を満たす場合、音場種別が分散型音場であると判定すること。または、複数の音場分類パラメータの値のうちの少なくとも一つが予め設定される不均一型音源判定条件を満たす場合、音場種別が不均一型音場であると判定すること。

【 0 3 8 7 】

本出願の幾つかの実施形態では、分散型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値未満であることを含む。または、不均一型音源判定条件は、音場分類パラメータの値が予め設定される不均一型音源判定閾値以上であることを含む。

【 0 3 8 8 】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在する。

【 0 3 8 9 】

音場分類結果は、音場種別を含み、または音場分類結果は、不均一型音源数および音場種別を含む。

【 0 3 9 0 】

音場分類モジュールは、以下を行うように構成される。すなわち、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得すること。および、現行フレームに対応する不均一型音源数に基づいて、音場種別を決定すること。

【 0 3 9 1 】

本出願の幾つかの実施形態では、複数の音場分類パラメータが存在する。

【 0 3 9 2 】

音場分類結果は、不均一型音源数を含む。

【 0 3 9 3 】

音場分類モジュールは、複数の音場分類パラメータの値に基づいて、現行フレームに対応する不均一型音源数を取得するように構成される。

【 0 3 9 4 】

本出願の幾つかの実施形態では、複数の音場分類パラメータは、 $temp[i]$ 、 $i = 0, 1, \dots, min(L, K) - 2$  であり、 $L$  は現行フレームのチャンネル数を表し、 $K$  は現行フレームの各チャンネルに対応する信号点の個数であり、 $min$  は最小値を選択する演算を表す。

10

20

30

40

50

## 【 0 3 9 5 】

音場分類モジュールは、以下の判定処理を  $i = 0$  から順次実行するように構成される。

すなわち、

$temp[i]$  が予め設定される不均一型音源判定閾値を超えるか否かを判定するステップ。および、

$temp[i]$  が本判定手順における不均一型音源判定閾値未満である場合、 $i$  の値を  $i + 1$  に更新し、次の判定手順の実行を継続するステップ。または

$temp[i]$  が本判定手順における不均一型音源判定閾値以上である場合、本判定手順の実行を終了し、本判定手順における  $i$  に 1 を加えたものは、不均一型音源数に等しいと判定するステップ。

10

## 【 0 3 9 6 】

本出願の幾つかの実施形態では、現行フレームに対応する不均一型音源数に基づいて、音場種別を決定するステップは、以下を含む。すなわち、

不均一型音源数が第一のプリセット条件を満たす場合、音場種別が第一の音場種別であると判定するステップ。または

不均一型音源数が第一のプリセット条件を満たさない場合、音場種別が第二の音場種別であると判定するステップ。

## 【 0 3 9 7 】

第一の音場種別に対応する不均一型音源数は、第二の音場種別に対応する不均一型音源数とは相違する。

20

## 【 0 3 9 8 】

本出願の幾つかの実施形態では、第一のプリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または

第一のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

## 【 0 3 9 9 】

本出願の幾つかの実施形態では、音声符号化装置は、符号化モード決定モジュール（図 12 に図示されない）をさらに含む。符号化モード決定モジュールは、音場分類結果に基づいて、現行フレームに対応する符号化モードを決定するように構成される。

30

## 【 0 4 0 0 】

可能な実装では、符号化モード決定モジュールは、以下を行うように構成される。すなわち、音場分類結果が不均一型音源数を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームに対応する符号化モードを決定すること。音場分類結果が音場種別を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームに対応する符号化モードを決定すること。または、音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームに対応する符号化モードを決定すること。

## 【 0 4 0 1 】

本出願の幾つかの実施形態では、符号化モード決定モジュールは、以下を行うように構成される。不均一型音源数が第二のプリセット条件を満たす場合、符号化モードが第一の符号化モードであると判定すること。または、不均一型音源数が第二のプリセット条件を満たさない場合、符号化モードが第二の符号化モードであると判定されること。

40

## 【 0 4 0 2 】

第一の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、第二の符号化モードは、仮想スピーカー選択に基づく H O A 符号化モード、もしくは指向性音声コーディングに基づく H O A 符号化モードであり、第一の符号化モードおよび第二の符号化モードは、相違する符号化モードである。

50

## 【 0 4 0 3 】

本出願の幾つかの実施形態では、第二のプリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または

第二のプリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

## 【 0 4 0 4 】

本出願の幾つかの実施形態では、符号化モード決定モジュールは、以下を行うように構成される。すなわち、音場種別が不均一型音場である場合、符号化モードが仮想スピーカ選択に基づく符号化モード H O A であると判定すること。または、音場種別が分散型音場である場合、符号化モードが指向性音声コーディングに基づく H O A 符号化モードであると判定すること。

10

## 【 0 4 0 5 】

本出願の幾つかの実施形態では、符号化モード決定モジュールは、以下を行うように構成される。すなわち、現行フレームの音場分類結果に基づいて、現行フレームに対応する初期符号化モードを決定すること。現行フレームが位置するハングオーバー時間枠を取得することであって、ハングオーバー時間枠は、現行フレームの初期符号化モードと、現行フレームより前の N - 1 個のフレームの符号化モードを含み、N はハングオーバー時間枠の長さである、こと。および現行フレームの初期符号化モードと、N - 1 個のフレームの符号化モードとに基づいて、現行フレームの符号化モードを決定すること。

20

## 【 0 4 0 6 】

本出願の幾つかの実施形態では、音声符号化装置は、符号化パラメータ決定モジュール（図 1 2 に図示されない）をさらに含む。符号化パラメータ決定モジュールは、音場分類結果に基づいて、現行フレームに対応する符号化パラメータを決定するように構成される。

## 【 0 4 0 7 】

本出願の幾つかの実施形態では、符号化パラメータは、仮想スピーカ信号のチャンネル数、残差信号のチャンネル数、仮想スピーカ信号の符号化ビット数、残差信号の符号化ビット数、もしくは最適スピーカを探索するための投票回数のうちの少なくとも一つを含む。

## 【 0 4 0 8 】

仮想スピーカ信号および残差信号は、三次元音声信号に基づいて生成される信号である。

30

## 【 0 4 0 9 】

本出願の幾つかの実施形態では、投票回数は次の関係を満たす。すなわち、

$$1 \leq d$$

## 【 0 4 1 0 】

I は、投票回数であり、d は、音場分類結果に含まれる不均一型音源数である。

## 【 0 4 1 1 】

本出願の幾つかの実施形態では、音場分類結果には、不均一型音源数および音場種別が含まれる。

40

## 【 0 4 1 2 】

音場種別が不均一型音源である場合、仮想スピーカ信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、F は仮想スピーカ信号のチャンネル数であり、S は不均一型音源数であり、PF はエンコーダによって予め設定される仮想スピーカ信号のチャンネル数である。または音場種別が分散型音場である場合、仮想スピーカ信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = 1$$

ここで、F は仮想スピーカ信号のチャンネル数である。

50

## 【 0 4 1 3 】

本出願の幾つかの実施形態では、音場種別が分散型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = \max(C - 1, PR)$$

ここで、 $PR$ はエンコーダによって予め設定される残差信号のチャンネル数であり、 $C$ はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計である。または

音場種別が不均一型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = C - F$$

ここで、 $R$ は残差信号のチャンネル数であり、 $C$ はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$ は仮想スピーカ信号のチャンネル数である。

## 【 0 4 1 4 】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含む。

## 【 0 4 1 5 】

仮想スピーカ信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、 $F$ は仮想スピーカ信号のチャンネル数であり、 $S$ は不均一型音源数であり、 $PF$ はエンコーダによって予め設定される仮想スピーカ信号のチャンネル数である。

## 【 0 4 1 6 】

本出願の幾つかの実施形態では、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = C - F$$

ここで、 $R$ は残差信号のチャンネル数であり、 $C$ はエンコーダによって予め設定される残差信号のチャンネル数と、エンコーダによって予め設定される仮想スピーカ信号のチャンネル数との合計であり、 $F$ は仮想スピーカ信号のチャンネル数である。

## 【 0 4 1 7 】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。

## 【 0 4 1 8 】

仮想スピーカ信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比に基づいて取得される。

## 【 0 4 1 9 】

残差信号の符号化ビット数は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比に基づいて取得される。

## 【 0 4 2 0 】

伝送チャンネルの符号化ビット数には、仮想スピーカ信号の符号化ビット数および残差信号の符号化ビット数が含まれ、不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合には、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の比は、伝送チャンネルの符号化ビット数に対する、仮想スピーカ信号の符号化ビット数の初期比を増加させることによって取得される。

## 【 0 4 2 1 】

本出願の幾つかの実施形態では、音声符号化装置は、符号化モジュール（図 1 2 に図示されない）をさらに含む。符号化モジュールは、現行フレームおよび音場分類結果を符号化し、符号化された現行フレームおよび音場分類結果をビットストリームに書き込むように構成される。

## 【 0 4 2 2 】

前述の実施形態における例から、最初に、線形分解が、三次元音声信号の現行フレームに対して実行されて、線形分解結果を取得することが分かる。次いで、現行フレームに対

10

20

30

40

50

応する音場分類パラメータが、線形分解結果に基づいて取得される。最後に、現行フレームの音場分類結果が、音場分類パラメータに基づいて決定される。本出願の本実施形態では、線形分解が、三次元音声信号の現行フレームに対して実行されて、現行フレームの線形分解結果を取得する。次いで、現行フレームに対応する音場分類パラメータが、線形分解結果に基づいて取得される。そのため、現行フレームの音場分類結果が、音場分類パラメータに基づいて決定され、音場分類結果に基づいて、現行フレームの音場分類を実装することができる。本出願の本実施形態では、音場分類が、三次元音声信号に対して実行されて、三次元音声信号を正確に識別する。

【 0 4 2 3 】

図 1 3 は、本出願の一実施形態による、三次元音声信号処理装置を示している。例えば、三次元音声信号処理装置は、具体的には音声復号化装置 1 3 0 0 であり、受信モジュール 1 3 0 1、復号化モジュール 1 3 0 2、および信号生成モジュール 1 3 0 3 を含み得る。

10

【 0 4 2 4 】

受信モジュールは、ビットストリームを受信するように構成される。

【 0 4 2 5 】

復号化モジュールは、ビットストリームを復号化して、現行フレームの音場分類結果を取得するように構成される。

【 0 4 2 6 】

信号生成モジュールは、音場分類結果に基づいて、復号化された現行フレームの三次元音声信号を取得するように構成される。

20

【 0 4 2 7 】

本出願の幾つかの実施形態では、信号生成モジュールは、音場分類結果に基づいて、現行フレームの復号化モードを決定し、復号化モードに基づいて、復号化された現行フレームの三次元音声信号を取得するように構成される。

【 0 4 2 8 】

本出願の幾つかの実施形態では、信号生成モジュールは、以下を行うように構成される。すなわち、音場分類結果が不均一型音源数を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数に基づいて、現行フレームの復号化モードを決定すること。音場分類結果が音場種別を含むか、もしくは音場分類結果が不均一型音源数および音場種別を含む場合、音場種別に基づいて、現行フレームの復号化モードを決定すること。または、音場分類結果が不均一型音源数および音場種別を含む場合、不均一型音源数および音場種別に基づいて、現行フレームの復号化モードを決定すること。

30

【 0 4 2 9 】

本出願の幾つかの実施形態では、信号生成モジュールは、以下を行うように構成される。すなわち、不均一型音源数がプリセット条件を満たす場合、復号化モードが第一の復号化モードであると判定すること。または、不均一型音源数がプリセット条件を満たさない場合、復号化モードが第二の復号化モードであると判定すること。

【 0 4 3 0 】

第一の復号化モードは、仮想スピーカー選択に基づく H O A 復号化モード、もしくは指向性音声コーディングに基づく H O A 復号化モードであり、第二の復号化モードは、仮想スピーカー選択に基づく H O A 復号化モード、もしくは指向性音声コーディングに基づく H O A 復号化モードであり、第一の復号化モードおよび第二の復号化モードは、相違する復号化モードである。

40

【 0 4 3 1 】

本出願の幾つかの実施形態では、プリセット条件は、不均一型音源数が第一の閾値を超えるか、もしくは第二の閾値未満であること、および第二の閾値が第一の閾値を超えることを含む。または、プリセット条件は、不均一型音源数が第一の閾値以下であるか、もしくは第二の閾値以上であること、および第二の閾値が第一の閾値を超えることを含む。

【 0 4 3 2 】

50

本出願の幾つかの実施形態では、信号生成モジュールは、音場分類結果に基づいて、現行フレームの復号化パラメータを決定し、復号化パラメータに基づいて、復号化された現行フレームの三次元音声信号を取得するように構成される。

【0433】

本出願の幾つかの実施形態では、復号化パラメータは、以下のうちの少なくとも一つを含む。すなわち、仮想スピーカー信号のチャンネル数、残差信号のチャンネル数、仮想スピーカー信号の復号化ビット数、もしくは残差信号の復号化ビット数のうちの少なくとも一つを含む。

【0434】

仮想スピーカー信号および残差信号は、ビットストリームを復号化することによって取得される。

10

【0435】

本出願の幾つかの実施形態では、音場分類結果には、不均一型音源数および音場種別が含まれる。

【0436】

音場種別が不均一型音源である場合、仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、Fは仮想スピーカー信号のチャンネル数であり、Sは不均一型音源数であり、PFはデコーダによって予め設定される仮想スピーカー信号のチャンネル数である。または、音場種別が分散型音場である場合、仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

20

$$F = 1$$

ここで、Fは仮想スピーカー信号のチャンネル数である。

【0437】

本出願の幾つかの実施形態では、音場種別が分散型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = \max(C - 1, PR)$$

ここで、PRはデコーダによって予め設定される残差信号のチャンネル数であり、Cはデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカー信号のチャンネル数との合計である。または、音場種別が不均一型音場である場合、残差信号のチャンネル数は、次の関係を満たす。すなわち、

30

$$R = C - F$$

ここで、Rは残差信号のチャンネル数であり、Cはデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカー信号のチャンネル数との合計であり、Fは仮想スピーカー信号のチャンネル数である。

【0438】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含む。

【0439】

40

仮想スピーカー信号のチャンネル数は、次の関係を満たす。すなわち、

$$F = \min(S, PF)$$

ここで、Fは仮想スピーカー信号のチャンネル数であり、Sは不均一型音源数であり、PFはデコーダによって予め設定される仮想スピーカー信号のチャンネル数である。

【0440】

本出願の幾つかの実施形態では、残差信号のチャンネル数は、次の関係を満たす。すなわち、

$$R = C - F$$

ここで、Rは残差信号のチャンネル数であり、Cはデコーダによって予め設定される残差信号のチャンネル数と、デコーダによって予め設定される仮想スピーカー信号のチャンネル数と

50

の合計であり、Fは仮想スピーカ信号のチャンネル数である。

【0441】

本出願の幾つかの実施形態では、音場分類結果は、不均一型音源数を含むか、または音場分類結果は、不均一型音源数および音場種別を含む。

【0442】

仮想スピーカ信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。

【0443】

残差信号の復号化ビット数は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比に基づいて取得される。

10

【0444】

伝送チャンネルの復号化ビット数には、仮想スピーカ信号の復号化ビット数および残差信号の復号化ビット数が含まれ、不均一型音源数が仮想スピーカ信号のチャンネル数以下である場合には、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の比は、伝送チャンネルの復号化ビット数に対する、仮想スピーカ信号の復号化ビット数の初期比を増加させることによって取得される。

【0445】

前述の実施形態における例から、音場分類結果をビットストリームにおける現行フレームを復号化するために使用できることが分かる。そのため、デコーダ側は、現行フレームの音場に適応した復号化方式において復号化を実行して、エンコーダ側によって送信される三次元音声信号を取得する。これは、エンコーダ側からデコーダ側への音声信号の伝送を実装する。

20

【0446】

本装置のモジュール/ユニット間の情報交換、およびその実行プロセスなどの内容は、本出願の方法の実施形態と同じ考え方に基づいており、本出願の方法の実施形態と同じ技術的效果を生み出すことは、留意されるべきである。特定の内容については、本出願の方法の実施形態における前述の説明を参照されたい。詳細については、本明細書では改めて説明しない。

【0447】

本出願の実施形態は、コンピュータ記憶媒体をさらに提供する。本コンピュータ記憶媒体は、プログラムを格納し、そのプログラムは、前述の方法の実施形態において説明されるステップの一部または全部を実行する。

30

【0448】

以下に、本出願の実施形態による、別の音声符号化装置について説明する。図14を参照されたい。音声符号化装置1400は、以下を含む。すなわち、受信機1401、送信機1402、プロセッサ1403、およびメモリ1404（音声符号化装置1400において、プロセッサ1403は一つもしくは複数存在し得て、図14では、一つのプロセッサが、一例として使用される）。本出願の幾つかの実施形態では、受信機1401、送信機1402、プロセッサ1403、およびメモリ1404は、バスを介して、もしくは別の方法において接続され得る。図14では、バスを介した接続が、一例として使用される。

40

【0449】

メモリ1404は、読取専用メモリおよびランダムアクセスメモリを含み、プロセッサ1403に対して命令およびデータを提供し得る。メモリ1404の一部は、不揮発性ランダムアクセスメモリ（NVRAM）をさらに含む。メモリ1404は、オペレーティングシステムおよび動作命令、実行可能モジュールもしくはデータ構造、またはそのサブセット、またはその拡張セットを格納する。動作命令には、種々の動作を実現するために使用される、種々の動作命令が含まれ得る。オペレーティングシステムは、種々の基本サービスを実装し、ハードウェアベースのタスクを処理するために、種々のシステムプログラムが含まれ得る。

50

## 【 0 4 5 0 】

プロセッサ 1 4 0 3 は、音声符号化装置の動作を制御し、プロセッサ 1 4 0 3 は、中央処理ユニット（ＣＰＵ）と呼ばれることもある。特定の用途中に、音声符号化装置のコンポーネントは、バスシステムを介して結合される。データバスに加えて、バスシステムは、電力バス、制御バス、およびステータス信号バスなどをさらに含み得る。ただし、説明を明確にするために、図における種々の種類のバスは、バスシステムとして表記されている。

## 【 0 4 5 1 】

本出願の実施形態に開示される方法は、プロセッサ 1 4 0 3 に適用され得るか、またはプロセッサ 1 4 0 3 を使用することによって実装され得る。プロセッサ 1 4 0 3 は、集積回路チップとし得て、信号処理能力を有する。実装プロセスでは、前述の方法におけるステップは、プロセッサ 1 4 0 3 におけるハードウェア集積論理回路を使用することによって、またはソフトウェアの形式における命令を使用することによって実装され得る。プロセッサ 1 4 0 3 は、汎用プロセッサ、デジタルシグナルプロセッサ（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、フィールドプログラマブルゲートアレイ（ＦＰＧＡ）もしくは別のプログラマブルロジックデバイス、ディスクリートゲートまたはトランジスタロジックデバイス、またはディスクリートハードウェアコンポーネントとし得て、本出願の実施形態に開示される方法、ステップ、および論理ブロック図を実装または実行し得る。汎用プロセッサは、マイクロプロセッサであってもよいし、またはプロセッサは、従来の任意のプロセッサなどであってもよい。本出願の実施形態を参照して開示される方法のステップは、ハードウェア復号化プロセッサを使用することによって直接実行および達成されてもよいし、または復号化プロセッサにおけるハードウェアおよびソフトウェアモジュールの組み合わせを使用することによって実行および達成されてもよい。ソフトウェアモジュールは、ランダムアクセスメモリ、フラッシュメモリ、読取専用メモリ、プログラマブル読取専用メモリ、電氣的に消去可能なプログラマブルメモリ、もしくはレジスタなどの、当技術分野において成熟した記憶媒体に配置され得る。その記憶媒体は、メモリ 1 4 0 4 に配置され、プロセッサ 1 4 0 3 は、メモリ 1 4 0 4 における情報を読み取り、プロセッサ 1 4 0 3 におけるハードウェアと組み合わせて、本方法のステップを完了する。

## 【 0 4 5 2 】

受信機 1 4 0 1 は、入力デジタル情報もしくは文字情報を受信し、音声符号化装置の設定および機能制御と関連した信号入力を生成するように構成される。送信機 1 4 0 2 は、ディスプレイ画面のような表示機器を含み得て、外部インターフェースを介してデジタル情報もしくは文字情報を出力するように構成され得る。

## 【 0 4 5 3 】

本出願の本実施形態では、プロセッサ 1 4 0 3 は、図 4 ないし図 6 に示される実施形態における音声符号化装置によって実行される方法を実行するように構成される。

## 【 0 4 5 4 】

以下に、本出願の実施形態による、別の音声復号化装置について説明する。図 1 5 を参照されたい。音声復号化装置 1 5 0 0 は、以下を含む。すなわち、受信機 1 5 0 1、送信機 1 5 0 2、プロセッサ 1 5 0 3、およびメモリ 1 5 0 4（音声復号化装置 1 5 0 0 におけるプロセッサ 1 5 0 3 は、一つもしくは複数存在し得て、図 1 5 では、一つのプロセッサが、一例として使用される）。本出願の幾つかの実施形態では、受信機 1 5 0 1、送信機 1 5 0 2、プロセッサ 1 5 0 3、およびメモリ 1 5 0 4 は、バスを介して、もしくは別の方法において接続され得る。図 1 5 では、バスを介した接続が、一例として使用される。

## 【 0 4 5 5 】

メモリ 1 5 0 4 は、読取専用メモリおよびランダムアクセスメモリを含み、プロセッサ 1 5 0 3 に対して命令およびデータを提供し得る。メモリ 1 5 0 4 の一部は、ＮＶＲＡＭをさらに含み得る。メモリ 1 5 0 4 は、オペレーティングシステムおよび動作命令、実行可能モジュールもしくはデータ構造、またはそのサブセット、またはその拡張セットを格

10

20

30

40

50

納する。その動作命令には、種々の動作を実装するために使用される、種々の動作命令が含まれ得る。オペレーティングシステムには、種々の基本サービスを実装し、ハードウェアベースのタスクを処理するために、種々のシステムプログラムが含まれ得る。

【 0 4 5 6 】

プロセッサ 1 5 0 3 は、音声復号化装置の動作を制御し、プロセッサ 1 5 0 3 は、C P U と呼ばれることもある。特定の用途中に、音声復号化装置のコンポーネントは、バスシステムを介して結合される。データバスに加えて、バスシステムは、電力バス、制御バス、およびステータス信号バスなどをさらに含み得る。ただし、説明を明確にするために、図における種々の種類のバスは、バスシステムとして表記されている。

【 0 4 5 7 】

本出願の実施形態に開示される方法は、プロセッサ 1 5 0 3 に適用され得て、またはプロセッサ 1 5 0 3 を使用することによって実装され得る。プロセッサ 1 5 0 3 は、集積回路チップとし得て、信号処理能力を有する。実装プロセスでは、前述の方法におけるステップは、プロセッサ 1 5 0 3 におけるハードウェア集積論理回路を使用することによって、またはソフトウェアの形式における命令を使用することによって実装され得る。前述のプロセッサ 1 5 0 3 は、本出願の実施形態で開示される方法、ステップ、および論理ブロック図を実装または実行するために、汎用プロセッサ、D S P、A S I C、F P G A もしくは別のプログラマブルロジックコンポーネント、ディスクリートゲートもしくはトランジスタロジックデバイス、またはディスクリートハードウェアコンポーネントとし得る。汎用プロセッサは、マイクロプロセッサであってもよいし、またはプロセッサは、従来の任意のプロセッサなどであってもよい。本出願の実施形態を参照して開示される方法のステップは、ハードウェア復号化プロセッサを使用することによって直接実行および達成されてもよいし、または復号化プロセッサにおけるハードウェアおよびソフトウェアモジュールの組み合わせを使用することによって実行および達成されてもよい。ソフトウェアモジュールは、ランダムアクセスメモリ、フラッシュメモリ、読取専用メモリ、プログラマブル読取専用メモリ、電氣的に消去可能なプログラマブルメモリ、もしくはレジスタなどの、当技術分野において成熟した記憶媒体に配置され得る。その記憶媒体は、メモリ 1 5 0 4 に配置され、プロセッサ 1 5 0 3 は、メモリ 1 5 0 4 における情報を読み取り、プロセッサ 1 5 0 3 におけるハードウェアと組み合わせて、本方法のステップを完了する。

【 0 4 5 8 】

本出願の本実施形態では、プロセッサ 1 5 0 3 は、図 7 に示される実施形態における音声復号化装置によって実行される方法を実行するように構成される。

【 0 4 5 9 】

別の可能な設計では、音声符号化装置もしくは音声復号化装置が、端末におけるチップである場合、チップは、処理ユニットおよび通信ユニットを含む。処理ユニットは、例えば、プロセッサであってもよく、通信ユニットは、例えば、入出力インターフェース、ピン、もしくは回路であってもよい。この処理ユニットは、記憶ユニットに保存されたコンピュータ実行可能命令を実行し得て、これにより、端末におけるチップは、第一の態様の実装の何れか一つにおける音声符号化法、もしくは第二の態様の実装の何れか一つにおける音声復号方法を実行する。任意選択として、記憶ユニットは、チップにおける記憶ユニット、例えば、レジスタもしくはバッファである。あるいは、記憶ユニットは、端末内であるがチップの外部にある記憶ユニット、例えば、読取専用メモリ ( R O M )、静的な情報および命令を保存することができる別種類の静的記憶機器、またはランダムアクセスメモリ ( R A M ) とし得る。

【 0 4 6 0 】

上述されるプロセッサは、汎用中央処理ユニット、マイクロプロセッサ、A S I C、または第一の態様もしくは第二の態様における方法のプログラム実行を制御するように構成される、一つもしくは複数の集積回路であり得る。

【 0 4 6 1 】

さらに、上述される装置の実施形態は、単なる一例に過ぎないことは、留意されるべき

10

20

30

40

50

である。別個の部品として説明されるユニットは、物理的に別個であることもあれば、またはそうでないこともあり、ユニットとして表示される部品は、物理的なユニットであることもあれば、またはそうでないこともあり、一つの位置に配置されていることもあれば、または複数のネットワークユニットに分散されることもある。幾つもしくは全てのモジュールは、実施形態の解決策の目的を達成するために、実際の要件に基づいて選択され得る。さらに、本出願によって提供される装置の実施形態に関する添付図面では、モジュール間の接続関係は、モジュールが相互に通信接続を有することを示し、これらは、具体的には、一つまたは複数の通信バス、または信号ケーブルとして実装され得る。

#### 【 0 4 6 2 】

前述の実装の説明に基づいて、当業者は、本出願が、必要な汎用ハードウェアに加えてソフトウェアによって、または専用集積回路、専用CPU、専用メモリ、および専用コンポーネントなどを含む専用ハードウェアによって実装され得ることを明確に理解し得る。メモリ、専用コンポーネントなど。一般に、コンピュータプログラムによって実行することができる任意の機能は、対応するハードウェアを使用することによって容易に実装することができる。さらに、同一の機能を実現するために使用される、具体的なハードウェア構成は、種々の形態、例えば、アナログ回路、デジタル回路、もしくは専用回路の形態にあり得る。ただし、本出願に関しては、ほとんどの場合、ソフトウェアプログラムの実装がより良い実装である。このような理解に基づいて、本出願の本質的な技術的解決策、もしくは従来技術に寄与する部分は、ソフトウェア製品の形態において実装され得る。コンピュータソフトウェア製品は、コンピュータにおけるフロッピーディスク、USBフラッシュドライブ、リムーバブルハードディスク、ROM、RAM、磁気ディスク、もしくは光ディスクなどの、可読記憶媒体に格納され、コンピュータ機器（パーソナルコンピュータ、サーバ、もしくはネットワーク装置であってもよい）に、本出願の実施形態に説明される方法を実行するように指示するために、幾つかの命令を含む。

#### 【 0 4 6 3 】

前述の実施形態の全てもしくは幾つかは、ソフトウェア、ハードウェア、ファームウェア、もしくはそれらの任意の組み合わせを使用することによって実装され得る。ソフトウェアが実施形態を実装するために使用される場合、実施形態の全部または一部は、コンピュータプログラム製品の形態において実装され得る。

#### 【 0 4 6 4 】

コンピュータプログラム製品には、一つまたは複数のコンピュータ命令が含まれる。コンピュータプログラム命令がコンピュータ上にロードされ実行されると、本出願の実施形態による手順または機能が、全てまたは部分的に生成される。コンピュータは、汎用コンピュータ、専用コンピュータ、コンピュータネットワーク、もしくは他のプログラム可能な装置であり得る。コンピュータ命令は、コンピュータ可読記憶媒体に格納されてもよいし、またはコンピュータ可読記憶媒体から別のコンピュータ可読記憶媒体に送信されてもよい。例えば、コンピュータ命令は、有線（例えば、同軸ケーブル、光ファイバー、もしくはデジタル加入者線（DSL）など）もしくは無線（例えば、赤外線、無線、もしくはマイクロ波など）方式において、ウェブサイト、コンピュータ、サーバ、もしくはデータセンターから、別のウェブサイト、コンピュータ、サーバ、もしくはデータセンターに送信されることがある。コンピュータ可読記憶媒体は、コンピュータによってアクセス可能な任意の使用可能な媒体、または一つもしくは複数の使用可能な媒体を統合するサーバまたはデータセンターなどの、データ記憶装置であり得る。使用可能な媒体は、磁気媒体（例えば、フロッピーディスク、ハードディスク、もしくは磁気テープなど）、光学媒体（例えば、DVDなど）、半導体媒体（例えば、ソリッドステートディスク（SSD）など）であり得る。

10

20

30

40

50

【図面】

【図 1】

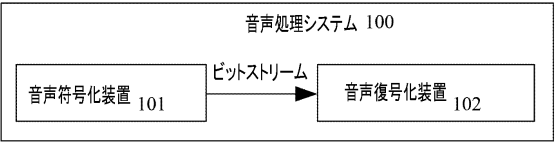


FIG. 1

【図 2 a】

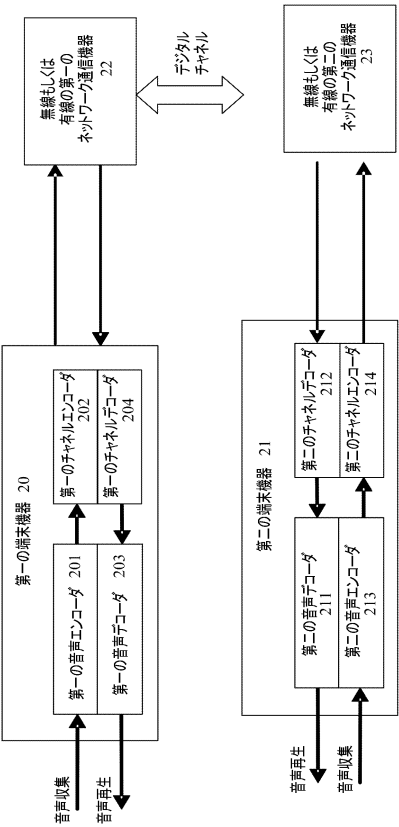


FIG. 2a

10

20

30

40

50

【図 2 b】

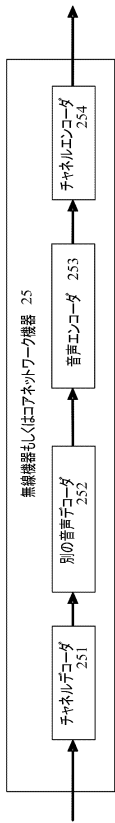


FIG. 2b

【図 2 c】

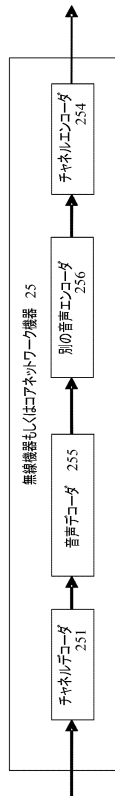


FIG. 2c

【図 3 a】

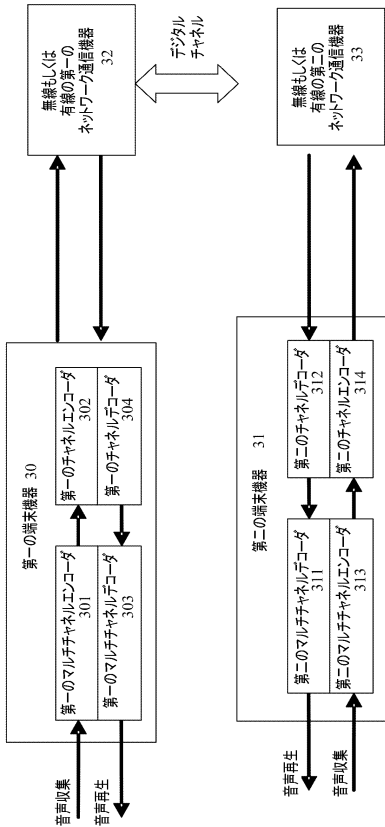


FIG. 3a

【図 3 b】

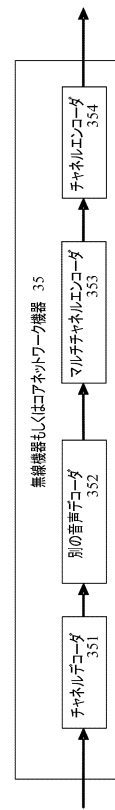


FIG. 3b

10

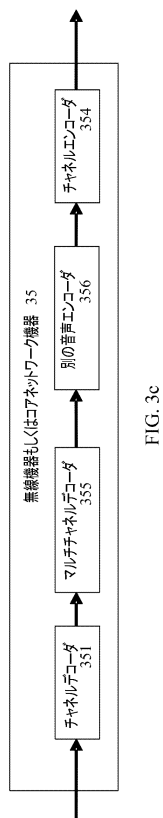
20

30

40

50

【図 3 c】



【図 4】

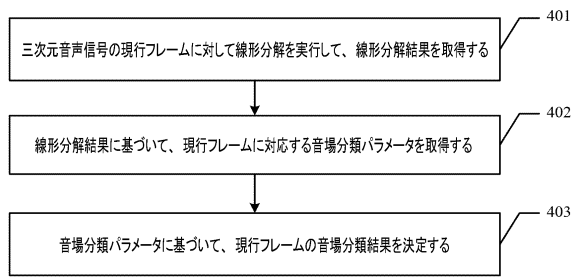


FIG. 4

10

【図 5】

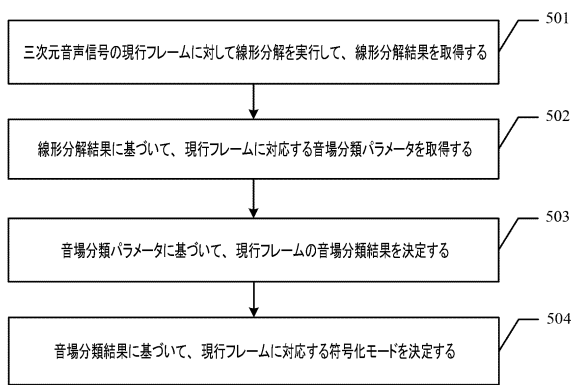


FIG. 5

【図 6】

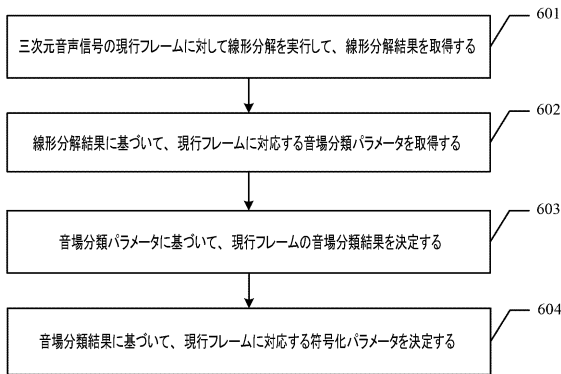


FIG. 6

30

40

50

【図 7】

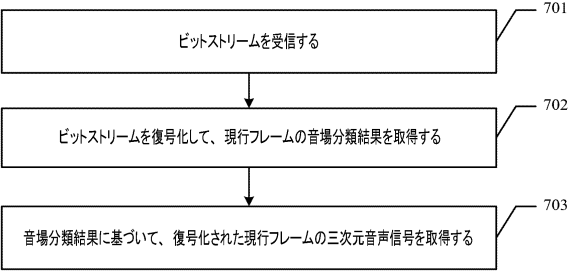


FIG. 7

【図 8】

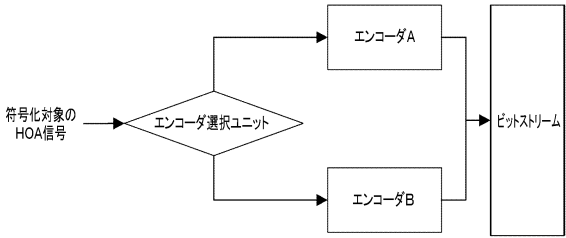


FIG. 8

10

【図 9】

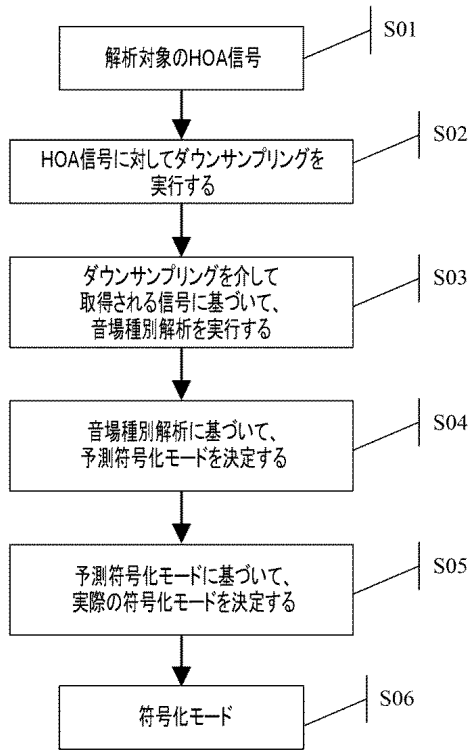


FIG. 9

【図 10】

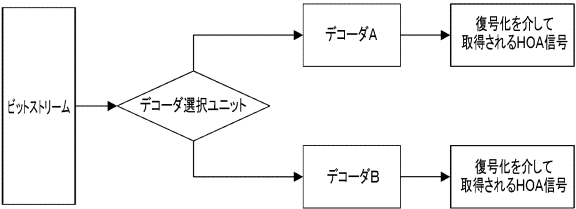


FIG. 10

20

30

40

50

【図 1 1】

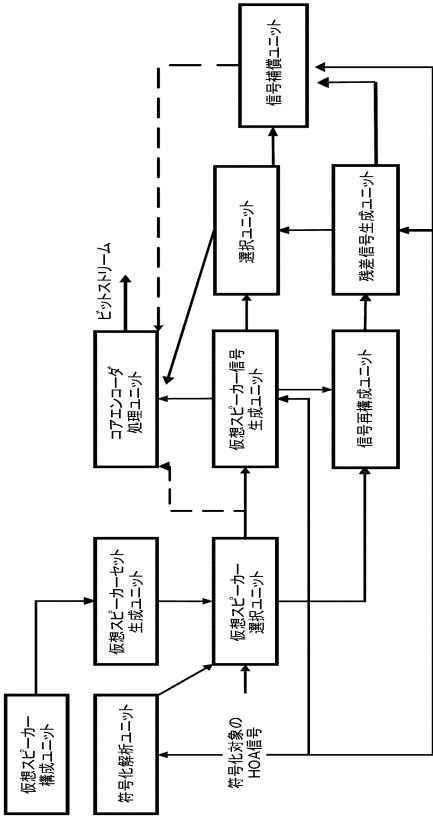


FIG. 11

【図 1 2】

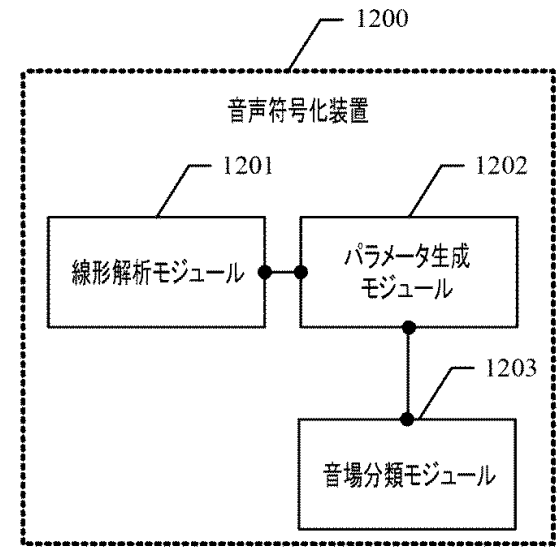


FIG. 12

【図 1 3】

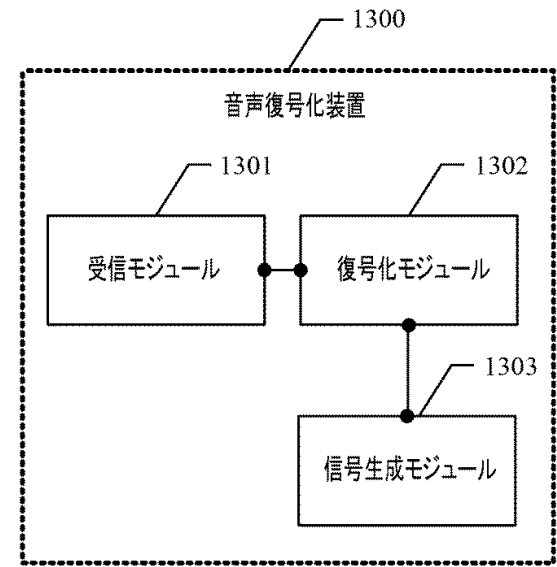


FIG. 13

【図 1 4】

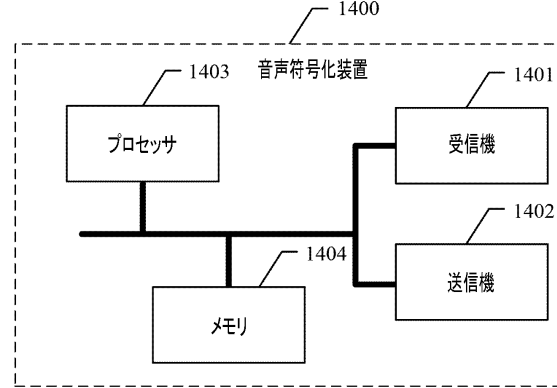


FIG. 14

10

20

30

40

50

【図 15】

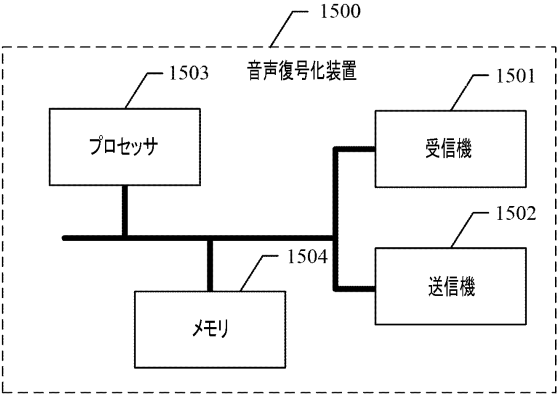


FIG. 15

10

20

30

40

50

## フロントページの続き

- (74)代理人 100115635  
弁理士 窪田 郁大
- (72)発明者 高 原  
中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍崗区坂田 華為総部▲ベン▼公楼
- (72)発明者 ▲劉▼ ▲帥▼  
中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍崗区坂田 華為総部▲ベン▼公楼
- (72)発明者 王 ▲賓▼  
中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍崗区坂田 華為総部▲ベン▼公楼
- (72)発明者 王 ▲ジョー▼  
中華人民共和国 5 1 8 1 2 9 広東省深▲チェン▼市龍崗区坂田 華為総部▲ベン▼公楼
- (72)発明者 曲 天▲書▼  
中華人民共和国 1 0 0 8 7 1 ベイジン ハイディエン ディストリクト イーホーユエン ロード  
ナンバー 5
- (72)発明者 徐 佳浩  
中華人民共和国 1 0 0 8 7 1 ベイジン ハイディエン ディストリクト イーホーユエン ロード  
ナンバー 5
- 審査官 山下 剛史
- (56)参考文献 特表2016-523468(JP,A)  
欧州特許出願公開第3706119(EP,A1)  
特表2016-524727(JP,A)  
特表2017-501440(JP,A)  
特表2017-513383(JP,A)  
特表2017-525318(JP,A)
- (58)調査した分野 (Int.Cl., DB名)  
G10L 19/00-25/93  
H04S 1/00-7/00