

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7129331号
(P7129331)

(45)発行日 令和4年9月1日(2022.9.1)

(24)登録日 令和4年8月24日(2022.8.24)

(51)国際特許分類

G 1 0 L	21/0224(2013.01)	F I	G 1 0 L	21/0224
A 6 3 F	13/54 (2014.01)		A 6 3 F	13/54
A 6 3 F	13/60 (2014.01)		A 6 3 F	13/60
G 1 0 L	21/034(2013.01)		G 1 0 L	21/034
G 1 0 L	21/0364(2013.01)		G 1 0 L	21/0364

請求項の数 8 (全12頁)

(21)出願番号 特願2018-241640(P2018-241640)
 (22)出願日 平成30年12月25日(2018.12.25)
 (65)公開番号 特開2020-101767(P2020-101767)
 A)
 (43)公開日 令和2年7月2日(2020.7.2)
 審査請求日 令和3年10月15日(2021.10.15)

(73)特許権者 595000427
 株式会社コーエーテクモゲームス
 神奈川県横浜市西区みなとみらい四丁目
 3番6号
 (74)代理人 100107766
 弁理士 伊東 忠重
 (74)代理人 100070150
 弁理士 伊東 忠彦
 (72)発明者 小池 雅人
 神奈川県横浜市港北区箕輪町一丁目18
 番12号 株式会社コーエーテクモゲー
 ムス内
 審査官 大石 剛

最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

録音された音声データから第1閾値以上の帯域の音が抽出された第1データに基づいて、リップノイズが録音されている第1区間を検出する検出部と、

前記音声データから第2閾値以下の帯域の音が抽出された前記第1区間の第2データに基づいて、前記音声データにおける前記第1区間の音のデータを修正する修正部と、
 を有する情報処理装置。

【請求項2】

第2閾値は、前記第1閾値よりも小さい、
 請求項1に記載の情報処理装置。

【請求項3】

前記検出部は、
 前記第1データにおいて、前記第1区間よりも前の所定時間長の第2区間の音量の最大値よりも、前記第1区間の音量の最大値が第3閾値以上増加しており、
 前記第1区間よりも後の所定数の所定時間長の各区間の音量の最大値の少なくとも一つが前記第2区間の音量の最大値以下の場合、前記第1区間を検出する、
 請求項1または2に記載の情報処理装置。

【請求項4】

前記検出部は、
 前記第1データにおいて、前記第1区間よりも前の所定時間長の第2区間の音量の最大

値よりも、前記第1区間の音量の最大値が第3閾値以上増加しており、

前記第1区間よりも後の所定数の所定時間長の各区間のうち音量の最大値が最も大きい第3区間の音量の最大値よりも、当該各区間のうち前記第3区間よりも後の第4区間の音量の最大値が第4閾値以上低下している場合、前記第1区間を検出する、
請求項1から3のいずれか一項に記載の情報処理装置。

【請求項5】

前記修正部は、

前記第2データに対して、前記第1区間に含まれる最初の時間帯の音量を、前記音声データにおける前記第1区間の前の時間帯の音量に基づいて修正する処理、及び前記第1区間に含まれる最後の時間帯の音量を、前記音声データにおける前記第1区間の後の時間帯の音量に基づいて修正する処理の少なくとも一方を実行する、
請求項1から4のいずれか一項に記載の情報処理装置。

10

【請求項6】

前記修正部は、前記音声データに対して、前記音声データにおける前記第1区間の前の時間帯の音量を、前記第2データの前記第1区間に含まれる最初の時間帯の音量に基づいて修正する処理、及び前記音声データにおける前記第1区間の後の時間帯の音量を、前記第1区間に含まれる最後の時間帯の音量に基づいて修正する処理の少なくとも一方を実行する、

請求項1から5のいずれか一項に記載の情報処理装置。

20

【請求項7】

情報処理装置が、

録音された音声データから第1閾値以上の帯域の音が抽出された第1データに基づいて、リップノイズが録音されている第1区間を検出する処理と、

前記音声データから第2閾値以下の帯域の音が抽出された前記第1区間の第2データに基づいて、前記音声データにおける前記第1区間の音のデータを修正する処理と、
を実行する情報処理方法。

【請求項8】

情報処理装置に、

録音された音声データから第1閾値以上の帯域の音が抽出された第1データに基づいて、リップノイズが録音されている第1区間を検出する処理と、
30

前記音声データから第2閾値以下の帯域の音が抽出された前記第1区間の第2データに基づいて、前記音声データにおける前記第1区間の音のデータを修正する処理と、
を実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置、情報処理方法、及びプログラムに関する。

【背景技術】

【0002】

従来、コンピュータゲーム等において、ゲームの状況に応じて、ゲームのキャラクタのセリフを、予め録音されている音声（ボイス）により出力する技術が知られている（例えば、特許文献1を参照）。

40

【0003】

声優や歌手等の発声者が発声した音声を録音する場合、発声者が口を開閉した際に生じる雑音（リップノイズ）が、発声者の口元に設置されたマイクにより集音される場合がある。この場合、録音されている音声を出力すると、ユーザにとって耳障りな雑音となる場合がある。従来、音声を修正する者が、録音されている音声をスピーカに出力させてリップノイズを耳で聞き取り、リップノイズが発生した時間の音声の波形を画面に表示させ、波形を手入力により修正することが知られている。

【先行技術文献】

50

【特許文献】

【0004】

【文献】特開2017-184842号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、従来技術では、職人の経験と勘に基づいて手作業により周波数成分や音量等を修正するため、作業に手間がかかると共に、修正の品質にばらつきがあるという問題がある。

【0006】

そこで、一側面では、より適切に音声を修正することができる技術を提供することを目的とする。

【課題を解決するための手段】

【0007】

一つの案では、録音された音声データから第1閾値以上の帯域の音が抽出された第1データに基づいて、リップノイズが録音されている第1区間を検出する検出部と、前記音声データから第2閾値以下の帯域の音が抽出された前記第1区間の第2データに基づいて、前記音声データにおける前記第1区間の音のデータを修正する修正部と、を有する。

【発明の効果】

【0008】

一側面によれば、より適切に音声を修正することができる。

【図面の簡単な説明】

【0009】

【図1】実施形態に係る情報処理装置のハードウェア構成例を示す図である。

【図2】実施形態に係る情報処理装置の機能ブロック図である。

【図3】実施形態に係る情報処理装置10の処理の一例を示すフローチャートである。

【図4】実施形態に係る高域部分に基づいてリップノイズ区間を特定する処理の一例を示すフローチャートである。

【図5】実施形態に係る高域部分に基づいてリップノイズ区間を特定する処理について説明する図である。

【図6】実施形態に係る特定したノイズ区間の音声を修正する処理の一例を示すフローチャートである。

【図7A】ノイズ区間を含む所定区間における、録音された音声データの一例を示す図である。

【図7B】実施形態に係るリップノイズ区間の中低域の音を抽出する処理について説明する図である。

【図7C】実施形態に係るリップノイズ区間の音を修正する処理について説明する図である。

【発明を実施するための形態】

【0010】

以下、図面に基づいて本発明の実施形態を説明する。

【0011】

<ハードウェア構成>

図1は、実施形態に係る情報処理装置10のハードウェア構成例を示す図である。図1に示す情報処理装置10は、それぞれバスBで相互に接続されているドライブ装置100、補助記憶装置102、メモリ装置103、CPU104、インターフェース装置105、表示装置106、及び入力装置107等を有する。

【0012】

情報処理装置10での処理を実現するゲームプログラムは、記録媒体101によって提供される。ゲームプログラムを記録した記録媒体101がドライブ装置100にセットさ

10

20

30

40

50

れると、ゲームプログラムが記録媒体 101 からドライブ装置 100 を介して補助記憶装置 102 にインストールされる。但し、ゲームプログラムのインストールは必ずしも記録媒体 101 より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置 102 は、インストールされたゲームプログラムを格納すると共に、必要なファイルやデータ等を格納する。

【0013】

メモリ装置 103 は、例えば、DRAM (Dynamic Random Access Memory)、または SRAM (Static Random Access Memory) 等のメモリであり、プログラムの起動指示があった場合に、補助記憶装置 102 からプログラムを読み出して格納する。CPU 104 は、メモリ装置 103 に格納されたプログラムに従って情報処理装置 10 に係る機能を実現する。インターフェース装置 105 は、ネットワークに接続するためのインターフェースとして用いられる。表示装置 106 はプログラムによる GUI (Graphical User Interface) 等を表示する。入力装置 107 は、コントローラ等、キーボード及びマウス等、またはタッチパネル及びボタン等で構成され、様々な操作指示を入力させるために用いられる。

10

【0014】

なお、記録媒体 101 の一例としては、CD - ROM、DVD ディスク、ブルーレイディスク、又は USB メモリ等の可搬型の記録媒体が挙げられる。また、補助記憶装置 102 の一例としては、HDD (Hard Disk Drive)、SSD (Solid State Drive)、又はフラッシュメモリ等が挙げられる。記録媒体 101 及び補助記憶装置 102 のいずれについても、コンピュータ読み取り可能な記録媒体に相当する。

20

【0015】

<機能構成>

次に、図 2 を参照し、情報処理装置 10 の機能構成について説明する。図 2 は、実施形態に係る情報処理装置 10 の機能プロック図である。

【0016】

情報処理装置 10 は、記憶部 11 を有する。記憶部 11 は、例えば、補助記憶装置 102 等を用いて実現される。記憶部 11 は、録音されたセリフの音声データ等を記憶する。

【0017】

また、情報処理装置 10 は、取得部 12、検出部 13、及び修正部 14 を有する。これら各部は、情報処理装置 10 にインストールされた 1 以上のプログラムが、情報処理装置 10 の CPU 104 に実行させる処理により実現される。

30

【0018】

取得部 12 は、録音されたセリフ等の音声データを記憶部 11 から取得する。検出部 13 は、取得部 12 により取得された音声データから第 1 閾値以上の帯域の音を抽出することにより高域の音データ (「第 1 データ」の一例。) を生成する。そして、検出部 13 は、生成した高域の音データから、リップノイズが録音されている各区間を検出する。

【0019】

修正部 14 は、取得部 12 により取得された音声データから、検出部 13 により検出された各区間の第 2 閾値以下の帯域の各音を抽出することにより、中低域の各音データ (「第 2 データ」の一例。) を生成する。そして、修正部 14 は、生成した中低域の各音データに基づいて、取得部 12 により取得された音声データにおける検出部 13 により検出された各区間の音のデータを修正する。これにより、例えば、声優等が発声したセリフの音声データに含まれるリップノイズを低減することができる。

40

【0020】

<処理>

次に、図 3 を参照して、情報処理装置 10 の処理について説明する。図 3 は、実施形態に係る情報処理装置 10 の処理の一例を示すフローチャートである。

【0021】

ステップ S1 において、検出部 13 は、取得部 12 により取得された音声データから高

50

域の音声データを抽出する。ここで、検出部13は、例えば、ハイパスフィルター（アンチエイリアスフィルタ）を用いて、録音された音声データから第1閾値（例えば、9000Hz）未満の周波数成分を除去した音声データを生成してもよい。

【0022】

続いて、検出部13は、抽出した高域の音声データに基づいて、リップノイズが発生している区間（以下で、「リップノイズ区間」と称する。）をそれぞれ特定する（ステップS2）。ここで、検出部13は、例えば、抽出した高域の音声データの音量が、短時間（例えば、0.1秒）で所定比（例えば、12dB）以上変化する各区間（時間帯、時間的な位置）を検出する。なお、この処理については後述する。

【0023】

続いて、検出部13は、各リップノイズ区間の音声に周期性があるか否かを判定する（ステップS3）。ここで、検出部13は、例えば、リップノイズ区間より前の所定期間内の波形から、リップノイズ区間の波形と類似度が高い区間である第1類似区間を特定する。そして、検出部13は、リップノイズ区間と第1類似区間との間の第1距離（時間差）を算出する。また、検出部13は、リップノイズ区間より後の所定期間内の波形から、リップノイズ区間の波形と類似度が高い区間である第2類似区間を特定する。そして、リップノイズ区間と第2類似区間との間の第2距離（時間差）を算出する。

【0024】

そして、検出部13は、第1距離と第2距離との差が所定の閾値未満である場合、リップノイズ区間の音声に周期性があると判定する。また、検出部13は、第1距離と第2距離との一方を整数倍した値と、他方との差が所定の閾値未満である場合も、リップノイズ区間の音声に周期性があると判定する。

【0025】

周期性がある場合（ステップS3でYES）、処理を終了する。なお、修正部14は、周期性がある場合、リップノイズを除去する量を低減してもよい。この場合、修正部14は、周期性があると判定した区間に 대해서は、後述する第2閾値を比較的大きな値（例えば、4000Hz）に決定してリップノイズ区間の音声の修正処理を行うようにしてもよい。

【0026】

周期性がない場合（ステップS3でNO）、修正部14は、取得部12により取得された音声データにおいて、各リップノイズ区間の音声を修正し（ステップS4）、処理を終了する。なお、この処理については後述する。

【0027】

リップノイズ区間の特定処理

次に、図4、及び図5を参照し、図3のステップS2の、抽出した高域の音声データに基づいてリップノイズ区間を特定する処理について説明する。図4は、実施形態に係る高域部分に基づいてリップノイズ区間を特定する処理の一例を示すフローチャートである。図5は、実施形態に係る高域部分に基づいてリップノイズ区間を特定する処理について説明する図である。なお、以下の説明において、音量の最大値は、音量の絶対値の最大値としてもよい。

【0028】

ステップS101において、検出部13は、音量が所定の閾値以上（例えば、12dB）変化し始める各時点を検出する。図5の例では、ステップS1の処理で録音された音声データから抽出された、高域部分のみの音声データの波形500が、横軸を時間、縦軸を音量（dB）として示されている。ここで、検出部13は、図5のように、例えば、ある時点501から所定時間（例えば、0.1秒）の後の時点502までの区間521の最大音量511と、時点502から当該一定時間の後の時点503までの区間522の最大音量512との差が所定の閾値以上変化する場合、時点502を、音量が所定の閾値以上変化し始める時点として検出する。

【0029】

10

20

30

40

50

なお、検出部 13 は、以下の処理を、検出した各時点に対してそれぞれ実行する。そのため、以下では、検出した各時点に含まれる一の時点（以下で、「処理対象の時点」と称する。）に対しての処理について説明する。図 5 の時点 502 は、処理対象の時点の一例である。

【0030】

続いて、検出部 13 は、処理対象の時点から当該一定時間の後の時点から、所定期間（例えば、0.3秒）後までの時間帯（以下で、「リリース区間」とも称する。）に含まれる各区間の音量の最大値の少なくとも一つが、音量が所定の閾値以上変化する前の区間の音量の最大値以下であるか否かを判定する（ステップ S102）。ここで、検出部 13 は、処理対象の時点 502 から当該一定時間の後の時点である時点 503 から当該一定時間後の時点 504 までの区間 523 の最大音量 513、時点 504 から当該一定時間後の時点 505 までの区間 524 の最大音量 514、時点 505 から当該一定時間後の時点 506 までの区間 525 の最大音量 515 を特定する。そして、検出部 13 は、最大音量 513、最大音量 514、及び最大音量 515 の少なくとも一つが、最大音量 511 よりも小さいか否かを判定する。

【0031】

リリース区間に含まれる各区間の音量の最大値の少なくとも一つが、音量が所定の閾値以上変化する前の区間の音量の最大値以下でない場合（ステップ S102 で NO）、ステップ S104 の処理に進む。

【0032】

一方、最大値以下である場合（ステップ S102 で YES）、検出部 13 は、処理対象の時点から、変化する前の音量まで下がった区間の終了時点までを、リップノイズ区間であると特定する（ステップ S103）。この場合、検出部 13 は、図 5 の例では、最大音量 511 以下となるのが区間 525 の最大音量 515 であるため、時点 502 から区間 525 の終了時点である時点 506 までの区間を、リップノイズ区間とする。これは、音量が所定の閾値以上に変化し始める時点 502 から所定期間内において、一定時間内の最大音量が、時点 502 よりも前の一定時間内の最大音量 511 以下に下がる場合、リップノイズによる音量の変化であると考えられるためである。

【0033】

続いて、検出部 13 は、リリース区間において、音量が所定の閾値以上減衰しているか否かを判定する（ステップ S104）。ここで、検出部 13 は、例えば、最大音量 512、最大音量 513、及び最大音量 514 のうち最も大きいものを特定する。そして、区間 523 から 525 のうちの第 1 区間であって、特定した最大音量の区間よりも後の第 1 区間の最大音量が、特定した最大音量よりも所定の閾値（例えば、12dB）以上減衰しているか否かを判定してもよい。これは、リップノイズとセリフ等の発声が同時に発生したような場合の音量の変化であると考えられるためである。

【0034】

音量が所定の閾値以上減衰していない場合（ステップ S104 で NO）、処理対象の時点に対する処理を終了する。一方、音量が所定の閾値以上減衰している場合（ステップ S104 で YES）、検出部 13 は、処理対象の時点から、音量が所定の閾値以上減衰した区間の終了時点までを、リップノイズ区間であると特定する（ステップ S105）。図 5 において、最大音量 512、最大音量 513、及び最大音量 514 のうち最も大きい値は最大音量 513 であり、最大音量 514 は最大音量 513 よりも 12dB 以上低いものとする。この場合、検出部 13 は、時点 502 から、最大音量 514 の区間 524 の終了時点である時点 505 までの区間を、リップノイズ区間と判定する。

【0035】

リップノイズ区間の音声の修正処理

次に、図 6 から図 7C を参照し、図 3 のステップ S4 の、特定したリップノイズ区間の音声を修正する処理について説明する。図 6 は、実施形態に係る特定したリップノイズ区間の音声を修正する処理の一例を示すフローチャートである。図 7A は、リップノイズ区

10

20

30

40

50

間を含む所定区間における、録音された音声データの一例を示す図である。図 7 B は、実施形態に係るリップノイズ区間の中低域の音を抽出する処理について説明する図である。図 7 C は、実施形態に係るリップノイズ区間の音を修正する処理について説明する図である。

【0036】

ステップ S 201において、修正部 14 は、取得部 12 により取得された音声データから、検出部 13 により検出された各リップノイズ区間の低中域の音声データを抽出する。ここで、修正部 14 は、例えば、ローパスフィルター（アンチエイリアスフィルタ）を用いて、録音された音声データから第 1 閾値よりも低い第 2 閾値（例えば、2000 Hz）以上の周波数成分を除去した音声データを生成してもよい。

10

【0037】

ここで、図 7 A の例では、録音された音声データの波形 700 が、横軸を時間、縦軸を音量 (dB) として示されている。図 7 A において、時点 701 から時点 702 の区間 711 が、リップノイズ区間として特定されているものとする。図 7 B の例では、図 7 A の波形 700 のうち、区間 711 の部分の波形 700A、区間 711 より前の部分の波形 700B、区間 711 より後の部分の波形 700C、及びステップ S 201 の処理により抽出された区間 711 の波形 721 が示されている。図 7 B の例では、ステップ S 201 の処理により第 2 閾値以上の周波数成分が除去された波形 721 の各時点における音量の絶対値は、波形 700A の音量の絶対値よりも小さくなっている。

【0038】

続いて、修正部 14 は、抽出した低中域の音声データの音量を、録音された音声データのリップノイズ区間前後の音声データと整合するように修正（調整）する（ステップ S 202）。

20

【0039】

図 7 C の例で、波形 721 のうち、時点 701 から所定時間（例えば、0.03 秒）後の時点 741 までの区間 751 の部分の波形を波形 721A とする。また、波形 721 のうち、時点 702 から所定時間前の時点 742 から、時点 702 までの区間 752 の部分の波形を波形 721B とする。

【0040】

（リップノイズ区間の中低域データの修正）

30

ステップ S 202 において、修正部 14 は、時点 701 における波形 721A の音量 732A と、時点 701 における波形 700A の音量 732B との中点 732 を通り、時点 741 における波形 721A の音量 733 を通るように波形 721A を修正した波形 731A のデータを生成する。

【0041】

そして、修正部 14 は、時点 742 における波形 721B の音量 734 を通り、時点 702 における波形 721B の音量 735A と、時点 702 における波形 700A の音量 735B との中点 735 を通るように波形 721B を修正した波形 731B のデータを生成する。

【0042】

続いて、修正部 14 は、録音された音声データのリップノイズ区間前後の音声データの音量を、抽出した低中域の音声データと整合するように修正する（ステップ S 203）。

40

【0043】

（リップノイズ区間前後の音声データの修正）

また、ステップ S 202 において、修正部 14 は、区間 711 より前の部分の波形 700B を、時点 701 から所定時間（例えば、0.03 秒）前の時点 743 から時点 701 までの区間において、時点 743 の音量 761 と中点 732 を通る波形 771A に修正する。また、区間 711 より後の部分の波形 700C を、時点 702 から、所定時間（例えば、0.03 秒）後の時点 744 までの区間において、中点 735 と時点 744 の音量 762 とを通る波形 771B に修正する。

50

【 0 0 4 4 】

続いて、修正部 1 4 は、ステップ S 2 0 3 の処理で修正した音声データのリップノイズ区間の音声データを、ステップ S 2 0 2 の処理で修正した低中域の音声データに置換し(ステップ S 2 0 4)、処理を終了する。

【 0 0 4 5 】

<変形例>

情報処理装置 1 0 の各機能部は、例えば 1 以上のコンピュータにより構成されるクラウドコンピューティングにより実現されていてもよい。

【 0 0 4 6 】

以上、本発明の実施例について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、特許請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

10

【 符号の説明 】**【 0 0 4 7 】**

1 0 情報処理装置

1 1 記憶部

1 2 取得部

1 3 検出部

1 4 修正部

20

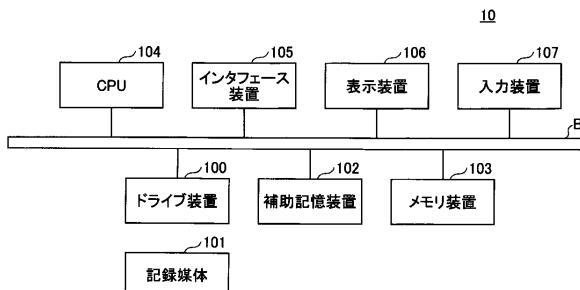
30

40

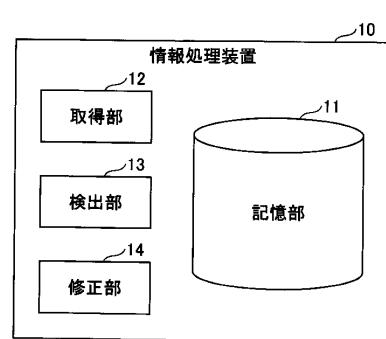
50

【図面】

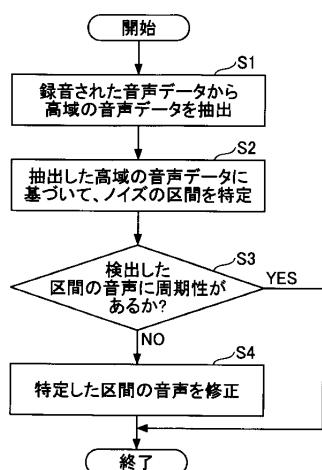
【図 1】



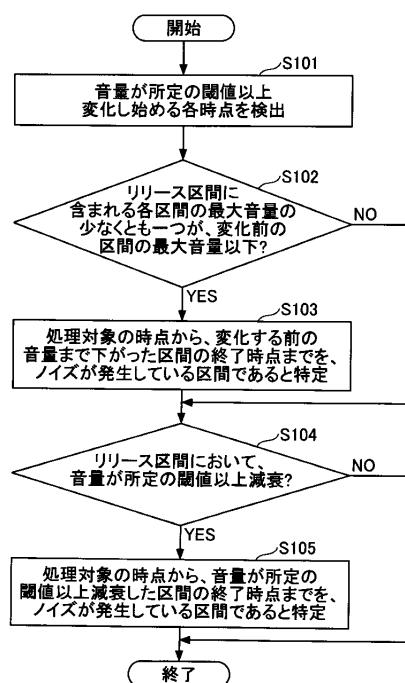
【図 2】



【図 3】



【図 4】



10

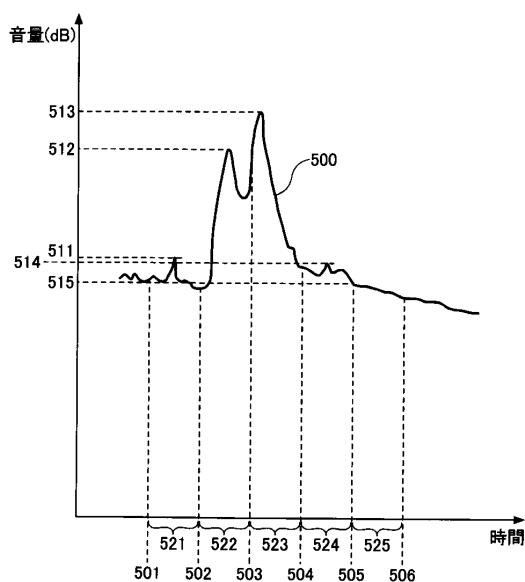
20

30

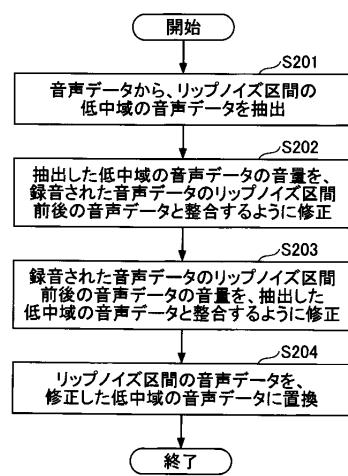
40

50

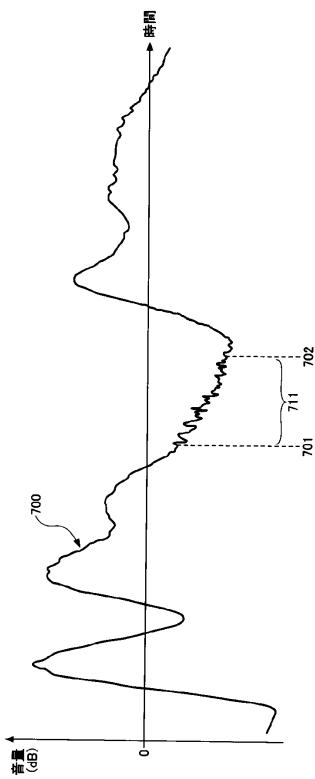
【図 5】



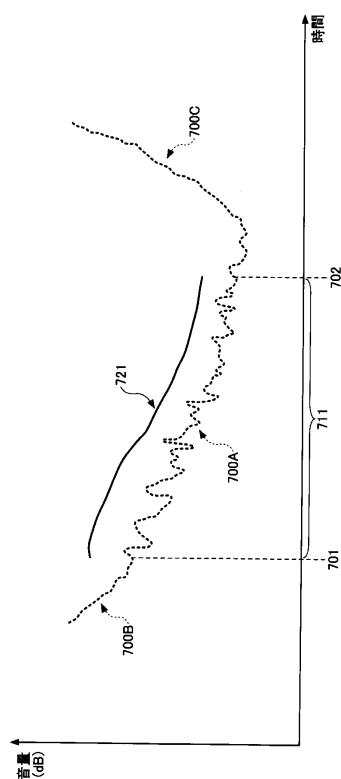
【図 6】



【図 7 A】



【図 7 B】



10

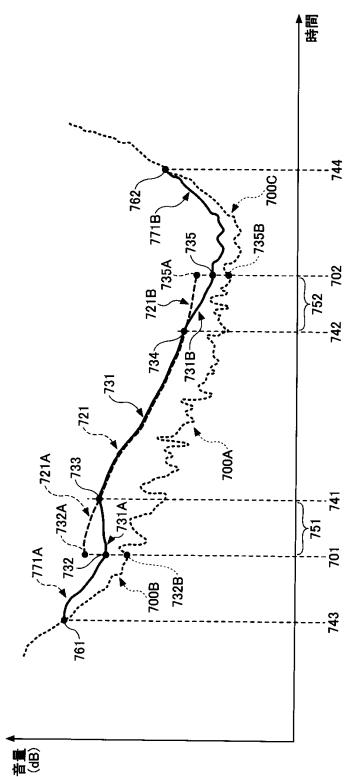
20

30

40

50

【図7C】



10

20

30

40

50

フロントページの続き

(56)参考文献 特開2003-069435 (JP, A)
 特開2013-025291 (JP, A)
 特開2008-158316 (JP, A)
 特開平07-261779 (JP, A)

(58)調査した分野 (Int.Cl., DB名)
 G 1 0 L 2 1 / 0 2 2 4
 A 6 3 F 1 3 / 5 4
 A 6 3 F 1 3 / 6 0
 G 1 0 L 2 1 / 0 3 4
 G 1 0 L 2 1 / 0 3 6 4