

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-39059

(P2010-39059A)

(43) 公開日 平成22年2月18日(2010.2.18)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 0 L 15/04 (2006.01)	G 1 0 L 15/04 3 0 0 C	5 D 0 1 5
G 1 0 L 15/02 (2006.01)	G 1 0 L 15/02 3 0 0 A	
G 1 0 L 15/00 (2006.01)	G 1 0 L 15/04 3 0 0 B	
	G 1 0 L 15/00 2 0 0 J	

審査請求 未請求 請求項の数 13 O L (全 19 頁)

(21) 出願番号	特願2008-199713 (P2008-199713)	(71) 出願人	000010076
(22) 出願日	平成20年8月1日(2008.8.1)		ヤマハ発動機株式会社
			静岡県磐田市新貝2500番地
		(74) 代理人	100104444
			弁理士 上羽 秀敏
		(74) 代理人	100123906
			弁理士 竹添 忠
		(72) 発明者	赤坂 貴志
			静岡県磐田市新貝2500番地 ヤマハ発
			動機株式会社内
		Fターム(参考)	5D015 DD03

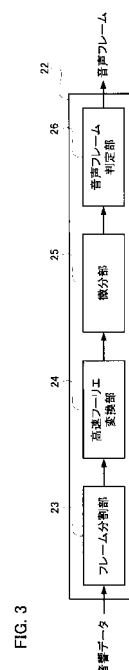
(54) 【発明の名称】 発話区間検出装置

(57) 【要約】

【課題】複数のマイクが不要で、単チャンネルの音声に対しても適用可能な発話区間検出装置を提供する。

【解決手段】発話区間検出装置22は、音響データを所定時間ごとにフレームに分割するフレーム分割部23と、フレーム分割部23により分割された音響データをフレームごとに高速フーリエ変換する高速フーリエ変換部24と、高速フーリエ変換部24によりフーリエ変換された音響データを微分して微分係数を算出する微分部25と、微分部25により算出された微分係数の度数分布に基づいて音声を含む音声フレームを判定する音声フレーム判定部26とを備える。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

音響データを所定時間ごとにフレームに分割するフレーム分割手段と、
前記フレーム分割手段により分割された音響データをフレームごとに周波数変換する周波数変換手段と、
前記周波数変換手段により周波数変換された音響データを微分して微分係数を算出する微分手段と、
前記微分手段により算出された微分係数の度数分布に基づいて音声を含む音声フレームを判定する音声フレーム判定手段とを備える発話区間検出装置。

【請求項 2】

請求項 1 に記載の発話区間検出装置であって、
前記音声フレーム判定手段は、
前記微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、
前記絶対値算出手段により算出された絶対値の分散を算出する分散算出手段と、
前記分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 3】

請求項 1 に記載の発話区間検出装置であって、
前記音声フレーム判定手段は、
前記微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、
前記絶対値算出手段により算出された絶対値の平均を算出する平均算出手段と、
前記平均算出手段により算出された平均を所定のしきい値と比較し、該平均が所定のしきい値よりも大きいフレームを音声フレームと判定する平均比較手段とを含む、発話区間検出装置。

【請求項 4】

請求項 1 に記載の発話区間検出装置であって、
前記音声フレーム判定手段は、
前記微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、
前記絶対値算出手段により算出された絶対値のヒストグラムを作成するヒストグラム作成手段と、
前記ヒストグラム作成手段により作成されたヒストグラムに基づいて所定の階級における度数を算出する度数算出手段と、
前記度数算出手段により算出された度数を所定のしきい値と比較し、該度数が所定のしきい値よりも多いフレームを音声フレームと判定する度数比較手段とを含む、発話区間検出装置。

【請求項 5】

請求項 1 に記載の発話区間検出装置であって、
前記音声フレーム判定手段は、
前記微分手段により算出された微分係数を微分して 2 階微分係数を算出する 2 階微分手段と、
前記 2 階微分手段により算出された 2 階微分係数の分散を算出する分散算出手段と、
前記分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 6】

請求項 1 に記載の発話区間検出装置であって、
前記音声フレーム判定手段は、
前記微分手段により算出された微分係数を微分して 2 階微分係数を算出する 2 階微分手段と、

10

20

30

40

50

前記 2 階微分手段により算出された 2 階微分係数のヒストグラムを作成するヒストグラム作成手段と、

前記ヒストグラム作成手段により作成されたヒストグラムに基づいて所定の階級における度数を算出する度数算出手段と、

前記度数算出手段により算出された度数を所定のしきい値と比較し、該度数が所定のしきい値よりも多いフレームを音声フレームと判定する度数比較手段とを含む、発話区間検出装置。

【請求項 7】

請求項 1 に記載の発話区間検出装置であって、

前記音声フレーム判定手段は、

10

前記微分手段により算出された微分係数を 2 混合正規分布にフィッティングするフィッティング手段と、

前記フィッティング手段により得られた 2 混合正規分布の分散を算出する分散算出手段と、

前記分散算出手段により算出された 2 つの分散の各々を所定のしきい値と比較し、少なくとも 1 つの分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 8】

請求項 1 に記載の発話区間検出装置であって、

前記音声フレーム判定手段は、

20

前記微分手段により算出された微分係数を 2 混合正規分布にフィッティングするフィッティング手段と、

前記フィッティング手段により得られた 2 混合正規分布の平均を算出する平均算出手段と、

前記平均算出手段により算出された 2 つの平均の各々を所定のしきい値と比較し、少なくとも 1 つの平均が所定のしきい値よりも大きいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 9】

請求項 1 に記載の発話区間検出装置であって、

前記音声フレーム判定手段は、

30

前記微分手段により算出された微分係数のうち正又は負の微分係数を抽出する微分係数抽出手段と、

前記微分係数抽出手段により抽出された正又は負の微分係数の分散を算出する分散算出手段と、

前記分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 10】

請求項 1 に記載の発話区間検出装置であって、

前記音声フレーム判定手段は、

40

前記微分手段により算出された微分係数のうち正又は負の微分係数を抽出する微分係数抽出手段と、

前記微分係数抽出手段により抽出された正又は負の微分係数の平均を算出する平均算出手段と、

前記平均算出手段により算出された平均を所定のしきい値と比較し、該平均が所定のしきい値よりも大きいフレームを音声フレームと判定する分散比較手段とを含む、発話区間検出装置。

【請求項 11】

請求項 1 ～ 10 のいずれか 1 項に記載の発話区間検出装置であってさらに、

前記周波数変換手段により周波数変換された音響データのうち前記微分手段による微分

50

の対象となる周波数帯域を所定範囲に制限する帯域制限手段を備える発話区間検出装置。

【請求項 1 2】

請求項 1 ~ 1 1 のいずれか 1 項に記載の発話区間検出装置を備えた音声認識装置。

【請求項 1 3】

請求項 1 2 に記載の音声認識装置を備えた輸送機器。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、発話区間検出装置に関し、さらに詳しくは、音響データの中から音声を含む発話区間を検出する発話区間検出装置に関する。

10

【背景技術】

【0002】

自動二輪車、自動車、船舶、飛行機、ヘリコプタなどの輸送機器においては、ナビゲーションシステム、携帯電話機、オーディオ機器などの電子情報機器が搭載されており、最近では、この機器を音声で操作できるように音声認識装置も搭載されている。しかしながら、輸送機器では風切り音などの雑音が多いため、音声を含む発話区間を検出することが重要な課題である。

【0003】

特許第 3 9 2 5 7 3 4 号公報（特許文献 1）には、ユーザに装着された第 1 及び第 2 マイクからの音声信号を処理する音声信号処理装置が記載されている（段落 0 0 2 6 ~ 0 0 4 6 参照）。この音声信号処理装置は、第 1 及び第 2 フレーム化部、第 1 及び第 2 周波数分析部、クロススペクトル計算部、位相抽出部、位相unwrap処理部、周波数帯域分割部、第 1 ~ 第 N 傾き計算部、ヒストグラム等計算部、及び音声 / 非音声判定部を備える。第 1 及び第 2 フレーム化部は、それぞれ、第 1 及び第 2 マイクからの音声信号をフレーム化する。第 1 及び第 2 周波数分析部は、それぞれ、フレーム化された音声信号をフーリエ変換して周波数関数を算出する。クロススペクトル計算部は、算出された周波数関数に基づいてクロススペクトルを算出する。位相抽出部は、算出されたクロススペクトルから位相を抽出する。位相unwrap処理部は、抽出された位相に基づいてクロススペクトルをunwrap処理する。周波数帯域分割部は、帯域分割（セグメント分割）した位相を第 1 ~ 第 N 傾き計算部にそれぞれ出力する。第 1 ~ 第 N 傾き計算部は、最小 2 乗法を適用することでセグメントごとに傾きを算出する。ヒストグラム等計算部は、算出された傾きのヒストグラムを作成し、各頻度を音声 / 非音声判定部に出力する。音声 / 非音声判定部は、頻度の平均値周辺の所定範囲内に含まれる傾きの頻度が所定のしきい値以上の場合、音声区間と判定し、頻度が所定のしきい値未満の場合、非音声区間と判定する。

20

30

【0004】

しかしながら、この音声信号処理装置はマイクを複数用いるため、数周波数分析部等も複数必要で、処理負荷が大きいという問題がある。また、マイクの設置が煩わしいという問題もある。また、マイクではなく、既に録音されている音声を用いる場合、その音声は複数チャンネルで録音されていなければならない。すなわち、音声単チャンネルで録音されている場合、この音声信号処理装置は使用できないという問題がある。

40

【0005】

また、特開 2 0 0 0 - 1 9 4 3 8 5 号公報（特許文献 2）には、音声区間検出部を備えた音声認識装置が記載されている。この音声区間検出部は、フレームごとの音響分析で求められた特徴パラメータに基づいて、話者が実際に音声を発している音声区間を検出する。しかしながら、この検出方法は本発明と原理的に全く異なる。

【特許文献 1】特許第 3 9 2 5 7 3 4 号公報

【特許文献 2】特開 2 0 0 0 - 1 9 4 3 8 5 号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

50

本発明の目的は、複数のマイクが不要で、単チャネルの音声に対しても適用可能な発話区間検出装置を提供することである。

【課題を解決するための手段及び発明の効果】

【0007】

本発明による発話区間検出装置は、音響データを所定時間ごとにフレームに分割するフレーム分割手段と、フレーム分割手段により分割された音響データをフレームごとに周波数変換する周波数変換手段と、周波数変換手段により周波数変換された音響データを微分して微分係数を算出する微分手段と、微分手段により算出された微分係数の度数分布に基づいて音声を含む音声フレームを判定する音声フレーム判定手段とを備える。ここでいう周波数変換は、時間領域の関数を周波数領域の関数に変換するもので、典型的にはフーリエ変換である。

10

【0008】

本発明によれば、フレームごとに周波数変換された音響データを微分して微分係数を算出し、微分係数の度数分布に基づいて音声フレームを判定しているため、複数のマイクが不要で、単チャネルの音声に対しても適用可能である。

【0009】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、絶対値算出手段により算出された絶対値の分散を算出する分散算出手段と、分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む。

20

【0010】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、絶対値算出手段により算出された絶対値の平均を算出する平均算出手段と、平均算出手段により算出された平均を所定のしきい値と比較し、該平均が所定のしきい値よりも大きいフレームを音声フレームと判定する平均比較手段とを含む。

【0011】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数の絶対値を算出する絶対値算出手段と、絶対値算出手段により算出された絶対値のヒストグラムを作成するヒストグラム作成手段と、ヒストグラム作成手段により作成されたヒストグラムに基づいて所定の階級における度数を算出する度数算出手段と、度数算出手段により算出された度数を所定のしきい値と比較し、該度数が所定のしきい値よりも多いフレームを音声フレームと判定する度数比較手段とを含む。

30

【0012】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数を微分して2階微分係数を算出する2階微分手段と、2階微分手段により算出された2階微分係数の分散を算出する分散算出手段と、分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む。

【0013】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数を微分して2階微分係数を算出する2階微分手段と、2階微分手段により算出された2階微分係数のヒストグラムを作成するヒストグラム作成手段と、ヒストグラム作成手段により作成されたヒストグラムに基づいて所定の階級における度数を算出する度数算出手段と、度数算出手段により算出された度数を所定のしきい値と比較し、該度数が所定のしきい値よりも多いフレームを音声フレームと判定する度数比較手段とを含む。

40

【0014】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数を2混合正規分布にフィッティングするフィッティング手段と、フィッティング手段により得られた2混合正規分布の分散を算出する分散算出手段と、分散算出手段により算出された2つの分散の各々を所定のしきい値と比較し、少なくとも1つの分散が所定のしきい値よりも小

50

さいフレームを音声フレームと判定する分散比較手段とを含む。

【 0 0 1 5 】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数を 2 混合正規分布にフィッティングするフィッティング手段と、フィッティング手段により得られた 2 混合正規分布の平均を算出する平均算出手段と、平均算出手段により算出された 2 つの平均の各々を所定のしきい値と比較し、少なくとも 1 つの平均が所定のしきい値よりも大きいフレームを音声フレームと判定する分散比較手段とを含む。

【 0 0 1 6 】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数のうち正又は負の微分係数を抽出する微分係数抽出手段と、微分係数抽出手段により抽出された正又は負の微分係数の分散を算出する分散算出手段と、分散算出手段により算出された分散を所定のしきい値と比較し、該分散が所定のしきい値よりも小さいフレームを音声フレームと判定する分散比較手段とを含む。

【 0 0 1 7 】

好ましくは、音声フレーム判定手段は、微分手段により算出された微分係数のうち正又は負の微分係数を抽出する微分係数抽出手段と、微分係数抽出手段により抽出された正又は負の微分係数の平均を算出する平均算出手段と、平均算出手段により算出された平均を所定のしきい値と比較し、該平均が所定のしきい値よりも大きいフレームを音声フレームと判定する分散比較手段とを含む。

【 0 0 1 8 】

好ましくは、発話区間検出装置はさらに、周波数変換手段により周波数変換された音響データのうち微分手段による微分の対象となる周波数帯域を所定範囲に制限する帯域制限手段を備える。

【 0 0 1 9 】

この場合、音声を含まない周波数帯域は微分の対象外となるので、音声フレームの判定精度が高くなる。

【 発明を実施するための最良の形態 】

【 0 0 2 0 】

以下、図面を参照し、本発明の実施の形態を詳しく説明する。図中同一又は相当部分には同一符号を付してその説明は繰り返さない。

【 0 0 2 1 】

図 1 は、自動二輪車の走行中に聞こえる風切り音（雑音）の周波数スペクトルである。一方、図 2 は、図 1 の風切り音に音声が含まれる場合の周波数スペクトルである。図 1 と図 2 を比較すると明らかなように、図 2 に示した周波数スペクトルには、0 ~ 2 K H z の範囲に荒い波形が現れる。本発明の実施の形態による発話区間検出装置は、この相違点に着目し、与えられた音響データの中から発話区間を検出する。

【 0 0 2 2 】

[第 1 の実施の形態]

図 3 を参照して、本発明の実施の形態による発話区間検出装置 2 2 は、フレーム分割部 2 3 と、高速フーリエ変換部（ F F T ; Fast Fourier Transform ） 1 4 と、微分部 2 5 と、音声フレーム判定部 2 6 とを備える。フレーム分割部 2 3 は、音響データを所定時間（たとえば 1 0 m s ）ごとにフレームに分割する。高速フーリエ変換部 2 4 は、フレーム分割部 2 3 により分割された音響データをフレームごとに高速フーリエ変換する。微分部 2 5 は、高速フーリエ変換部 2 4 によりフーリエ変換された音響データを微分して微分係数を算出する。音声フレーム判定部 2 6 は、微分部 2 5 により算出された微分係数の度数分布に基づいて音声を含む音声フレームを判定する。

【 0 0 2 3 】

次に、図 4 に示したフロー図を参照し、発話区間検出装置 2 2 の動作を説明する。

【 0 0 2 4 】

まず、フレーム分割部 2 3 は、与えられた音響データを所定時間ごとにフレームに分割

10

20

30

40

50

する (S 1 0)。音響データは、マイクで検知した信号でも、録音された信号でもよい。また、所定時間が重ならないように音響データをフレームに分割してもよいし、所定時間の一部が重なるように音響データをフレームに分割してもよい。

【 0 0 2 5 】

続いて、高速フーリエ変換部 2 4 は、フレーム分割部 2 3 により分割された音響データ $x(t)$ をフレームごとに高速フーリエ変換する (S 1 1)。

【 0 0 2 6 】

続いて、微分部 2 5 は、高速フーリエ変換部 2 4 によりフーリエ変換された音響データ (以下、「パワースペクトル」という。) $X(f)$ を微分して微分係数 (傾き dX/df) を算出する (S 1 2 ~ S 1 4)。

10

【 0 0 2 7 】

より具体的には、周波数変換された音響データのうち微分部 2 5 による微分の対象となる周波数帯域を所定範囲に制限する (S 1 2)。所定範囲としては音声信号が含まれる 0 ~ 2 K H z が好ましい。ただし、これに限定されるものではない。また、帯域制限自体をしなくてもよい。

【 0 0 2 8 】

続いて、周波数 f を微小区間 Δf (たとえば 1 0 0 H z) に分割する (S 1 3)。パワースペクトル $X(f)$ を Δf の幅で離散化された関数 $X_i = X(i \cdot \Delta f)$ とみなし、微分係数をパワースペクトル X_i の差分 D_i として次の式 (1) により算出する (S 1 4)。

20

$$D_i = (X_{i+1} - X_i) / (i+1 - i) = X_{i+1} - X_i \quad \dots (1)$$

【 0 0 2 9 】

なお、厳密には、微分係数 dX/df は次の式 (2) で表される。

$$dX/df = (X(f + \Delta f) - X(f)) / \Delta f \quad \dots (2)$$

【 0 0 3 0 】

微分係数の算出後、音声フレーム判定部 2 6 は、微分部 2 5 により算出された微分係数の度数分布に基づいて音声を含む音声フレームを判定する (S 1 5 ~ S 1 9)。

【 0 0 3 1 】

図 5 は、微分係数の度数分布を示す。音声を含まない雑音のみのフレームの場合、微分係数 0 付近で度数が最大になり、音声を含むフレームの場合、正及び負のある微分係数で度数が最大になる。図 6 は、この微分係数の絶対値の度数分布を示す。音声を含まない雑音のみのフレームの場合、分散は比較的大きくなり、音声を含むフレームの場合、分散は比較的小さくなる。

30

【 0 0 3 2 】

したがって、まず微分係数 D_i の絶対値 $|D_i|$ を算出し (S 1 5)、算出した絶対値 $|D_i|$ の分散 σ を次の式 (3) により算出する (S 1 6)。

【 数 1 】

$$\sigma = \frac{\sum_{i=1}^N (|Di| - \mu)^2}{N} \quad (3)$$

40

【 0 0 3 3 】

ここで、 N は微分係数 D_i の全数である。また、 μ は絶対値 $|D_i|$ の平均であり、次の式 (4) で表される。

【数 2】

$$\mu = \frac{\sum_{i=1}^N |D_i|}{N} \quad (4)$$

【0034】

10

そして、算出した分散 σ を所定のしきい値 σ_t と比較する (S 17)。比較の結果、分散 σ が所定のしきい値 σ_t よりも小さい場合 (S 17でYES)、当該フレームを音声フレームと判定する (S 18)。一方、分散 σ が所定のしきい値 σ_t よりも大きい場合 (S 17でNO)、当該フレームを非音声フレーム (雑音フレーム) と判定する (S 19)。なお、 $\sigma = \sigma_t$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

【0035】

以上のように第 1 の実施の形態によれば、フレームごとにフーリエ変換された音響データを微分して微分係数を算出し、微分係数の度数分布に基づいて音声フレームを判定しているため、複数のマイクが不要で、単チャンネルの音声に対しても適用可能である。

20

【0036】

[第 2 の実施の形態]

上記第 1 の実施の形態は分散 σ を用いている。ただし、分散 σ の代わりに、上記式 (4) で算出される平均 μ を用いてもよい。具体的には図 7 に示すように、微分係数の絶対値 $|D_i|$ の平均 μ を算出し (S 20)、算出した平均 μ を所定のしきい値 μ_t (図 6 参照) と比較する (S 21)。比較の結果、平均 μ が所定のしきい値 μ_t よりも大きい場合 (S 21でYES)、当該フレームを音声フレームと判定する (S 18)。一方、平均 μ が所定のしきい値 μ_t よりも小さい場合 (S 21でNO)、当該フレームを非音声フレームと判定する (S 19)。なお、 $\mu = \mu_t$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

30

【0037】

[第 3 の実施の形態]

また、図 8 に示すように、微分係数の絶対値 $|D_i|$ のヒストグラムを作成し、所定の階級における度数 N を分散 σ や平均 μ の代わりに用いてもよい。図 8 の (a) は音声フレームのヒストグラムを示し、図 8 の (b) は非音声フレームのヒストグラムを示す。ここでは、所定の微分係数 f_m を中心に $\pm \theta$ 内に入る階級の総度数を用いる。所定の微分係数 f_m としては、微分係数の絶対値の平均 μ 、最高度数の微分係数などを用いてもよい。

【0038】

具体的には図 9 に示すように、微分係数の絶対値 $|D_i|$ のヒストグラムを作成し (S 30)、作成したヒストグラムに基づいて所定の階級における度数 N を算出し (S 31)、算出した度数 N を所定のしきい値 N_t と比較する (S 32)。比較の結果、度数 N が所定のしきい値 N_t よりも多い場合 (S 32でYES)、当該フレームを音声フレームと判定する (S 18)。一方、度数 N が所定のしきい値 N_t よりも少ない場合 (S 32でNO)、当該フレームを非音声フレームと判定する (S 19)。なお、 $N = N_t$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

40

【0039】

[第 4 の実施の形態]

上記第 1 の実施の形態は各フレームを 1 階微分し、1 階微分係数の絶対値の分散を算出している。ただし、各フレームを 2 階微分し、2 階微分係数の分散を算出してもよい。具体的には図 10 に示すように、上記ステップ S 13 及び S 14 を繰り返すことにより、微

50

分係数 D_i をさらに微分して2階微分係数を算出する (S 4 0, S 4 1)。図 1 1 は、算出された2階微分係数を示す。非音声フレームの場合、分散は比較的大きくなり、音声フレームの場合、分散は比較的小さくなる。

【0040】

続いて、算出した2階微分係数の分散 σ' を算出し (S 4 2)、算出した分散 σ' を所定のしきい値 $\sigma't$ と比較する (S 4 3)。比較の結果、分散 σ' が所定のしきい値 $\sigma't$ よりも小さい場合 (S 4 3 で YES)、当該フレームを音声フレームと判定する (S 1 8)。一方、分散 σ' が所定のしきい値 $\sigma't$ よりも大きい場合 (S 4 3 で NO)、当該フレームを非音声フレームと判定する (S 1 9)。なお、 $\sigma' = \sigma't$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

10

【0041】

[第5の実施の形態]

また、上記第3及び第4の実施の形態を組み合わせてもよい。すなわち図 1 2 に示すように、微分係数 D_i をさらに微分して2階微分係数を算出した後 (S 4 0, S 4 1)、微分係数の絶対値 $|D_i|$ のヒストグラムを作成し (S 3 0)、作成したヒストグラムに基づいて所定の階級における度数 N を算出し (S 3 1)、算出した度数 N を所定のしきい値 Nt と比較する (S 3 2)。

【0042】

[第6の実施の形態]

図 5 に示したように、音声フレームに関する微分係数の度数分布は、2つの正規分布 (ガウス分布) を合成した2混合正規分布に類似している。そこで、図 1 3 に示すように、微分係数 D_i を2混合正規分布にフィッティングし (S 5 0)、特定された2混合分布の分散 σ_1, σ_2 を算出し (S 5 1)、算出した分散 σ_1, σ_2 の各々を所定のしきい値 σ_t と比較する (S 5 2)。比較の結果、少なくとも1つの分散 σ_1 又は σ_2 が所定のしきい値 σ_t よりも小さい場合 (S 4 3 で YES)、当該フレームを音声フレームと判定する (S 1 8)。一方、少なくとも1つの分散 σ_1 又は σ_2 が所定のしきい値 σ_t よりも大きい場合 (S 4 3 で NO)、当該フレームを非音声フレームと判定する (S 1 9)。なお、 $\sigma_1 = \sigma_t$ 又は $\sigma_2 = \sigma_t$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

20

【0043】

[第7の実施の形態]

上記第6の実施の形態は2混合正規分布の分散 σ_1, σ_2 を用いている。ただし、分散 σ_1, σ_2 の代わりに、2混合正規分布の平均 μ_1, μ_2 を用いてもよい。具体的には図 1 4 に示すように、微分係数 D_i を2混合正規分布にフィッティングした後 (S 5 0)、特定された2混合分布の平均 μ_1, μ_2 (図 5 参照) を算出し (S 6 0)、算出した平均 μ_1, μ_2 を所定のしきい値 μ_{t1}, μ_{t2} (図 5 参照) とそれぞれ比較する (S 6 1)。比較の結果、平均 μ が所定のしきい値 μ_t よりも大きい場合 (S 2 1 で YES)、当該フレームを音声フレームと判定する (S 1 8)。一方、平均 μ が所定のしきい値 μ_t よりも小さい場合 (S 2 1 で NO)、当該フレームを非音声フレームと判定する (S 1 9)。なお、 $\mu = \mu_t$ の場合、当該フレームを音声フレームと判定しても、非音声フレームと判定してもよい。

30

40

【0044】

[第8の実施の形態]

上記第1の実施の形態は微分係数の絶対値 $|D_i|$ を算出している。ただし、図 1 5 に示すように、算出した微分係数 D_i のうち正又は負の微分係数のみを抽出してもよい (S 7 0)。

【0045】

[第9の実施の形態]

上記第8の実施の形態は分散 σ を算出している。ただし、図 1 6 に示すように、抽出した正又は負の微分係数 D_i の平均 μ (図 5 中の μ_1 又は μ_2 に相当) を算出してもよい (

50

S 2 0)。

【 0 0 4 6 】

[第 1 0 の実施の形態]

上記第 1 ～ 第 9 の実施の形態は、周波数 f を微小区間 Δf に分割し、その差分 D_i を算出することにより、微分係数（傾き）を算出している。ただし、 (i, X_i) の前後数点を考慮し、最小二乗法により微分係数を算出してもよい。たとえば、 i を含む K 個の点の傾き D_i は最小二乗法によると一般に次の式（5）で与えられる。

【 数 3 】

$$Di = \frac{\sum_k (k - Im)(X_k - X_m)}{\sum_k (k - Im)^2} \quad (5)$$

10

ここで、 k の範囲は i を含み、全部で K 個ある。また、 Im は次の式（6）で表される i の平均であり、 X_m は次の式（7）で表される X_i の平均である。

【 数 4 】

$$Im = \frac{\sum_k k}{K} \quad (6)$$

20

$$X_m = \frac{\sum_k X_k}{K} \quad (7)$$

【 0 0 4 7 】

さらにこのとき、 $[i - n, i + n]$ の $(2n + 1)$ 個の点で上記式（5）を適用すると、 $Im = i$ などを考慮して、次の式（8）で与えることができる

30

【数 5】

$$\begin{aligned}
& (D_i \text{の分子}) \\
& = (i-n-i)(X_{i-n} - X_m) + (i-(n-1)-i)(X_{i-(n-1)} - X_m) + \cdots \\
& \quad + ((i-1)-i)(X_{i-1} - X_m) + (i-i)(X_i - X_m) + (i+1-i)(X_{i+1} - X_m) + \cdots \\
& \quad + (i+(n-1)-i)(X_{i+n-1} - X_m) + (i+n-i)(X_{i+n} - X_m) \\
& = -nX_{i-n} + nX_m - (n-1)X_{i-(n-1)} + (n-1)X_m + \cdots \\
& \quad - X_{i-1} + X_m + X_{i+1} - X_m + \cdots \\
& \quad + (n-1)X_{i+n-1} - (n-1)X_m + nX_{i+n} - nX_m \\
& = \sum_k k \cdot X_{i+k} \quad (K \text{の範囲: } -n \leq k \leq n)
\end{aligned}$$

10

$$\begin{aligned}
& (D_i \text{の分母}) \\
& = (i-n-i)^2 + (i-(n-1)-i)^2 + \cdots + (i-i)^2 + \cdots \\
& \quad + (i+(n-1)-i)^2 + (i+n-i)^2 \\
& = \sum_k k^2 \quad (k \text{の範囲: } -n \leq k \leq n) \\
& \therefore D_i = \frac{\sum_k k \cdot X_{i+k}}{\sum_k k^2} \quad \cdots (8)
\end{aligned}$$

20

【0048】

[その他の実施の形態]

図1及び図2に示した周波数スペクトルは、横軸の周波数はリニアスケールで、縦軸のパワー（音圧）は対数スケールで表されている。ただし、これらはリニアスケールでも対数スケールでもどちらでもよい。計算を容易にするためには、周波数はリニアスケールであるのが好ましい。高速フーリエ変換に由来する微小区間 Δf をそのまま使用できるからである。たとえば、サンプリング周波数16KHzの音声信号を1024ポイントで高速フーリエ変換した場合、 Δf は次の式(9)で表される。

$$\Delta f = 1 / ((1 / 1600) \times 1024) = 15.625 \text{ Hz} \quad \cdots (9)$$

【0049】

また、上記実施の形態はフレーム単位で音声を含むか否かを判定しているため、誤判定（音声フレームを非音声フレームと判定、又は非音声フレームを音声フレームと判定）が断続的に発生する場合がある。したがって、音声フレームと判定されたフレームをそのまま発話区間として抽出するのではなく、その前後のフレームの判定結果を考慮した上で発話区間を抽出するのが好ましい。

40

【0050】

たとえば図17に示すように、 $f(i)$ が音声フレームと判定されたとき、さらに後続するNフレームについても音声フレームか否かを判定する。その結果、nフレーム（ $0 < n \leq N$ ）が音声フレームと判定されたとき、 $f(i)$ からkフレーム（ $k \geq 0$ ）遡ったフレーム $f(i-k)$ を発話区間の開始フレームとする。発話区間の開始後、j番目のフレーム $f(j)$ （ $j > i$ ）から7フレーム $f(j)$ を含むMフレーム（ $M > 1$ ）が連続して非音声フレームと判定されたとき、フレーム $f(j+M-1)$ を発話区間の最終フレームとする。N, n, k, Mの値は、アプリケーションの特徴、リアルタイム性などを考慮して適宜定められる（特開2000-194385号公報の段落0049参照）。

50

【 0 0 5 1 】

〔 用途 〕

上記実施の形態は典型的には音声認識装置で用いられる。たとえば図 1 8 に示すように、音声認識装置 3 0 は、発話区間検出部 3 2 と、音声分析部 3 4 と、音響ライブラリ 3 5 と、照合部 3 7 とを備える。発話区間検出部 3 2 は、上記実施の形態による発話区間検出装置で構成され、マイクから入力される音声信号の中から発話区間を検出して抽出する。音声分析部 3 4 は、発話区間内の音声信号を分析してその音響的特徴を表す特徴パラメータ系列を抽出する。音響ライブラリ 3 5 は、音響モデル 3 8 と、単語辞書 3 9 とを備える。音響モデル 3 8 は、単語ごとに標準音声パターンの音響的特徴をモデル化したものであり、入力音声パターンとの音響的な類似性の評価を行うための参照情報である。また、単語辞書 3 9 は、音響モデルの接続に関する制約を与えるための情報、たとえばある音素に引き続いて別の音素が出現する確率である。照合部 3 7 は、音声分析部 3 4 により抽出された特徴パラメータ系列と音響ライブラリ 3 5 の認識候補とを照合して、HMM（隠れマルコフモデル）アルゴリズムにより認識候補の尤もらしさを表す尤度を生成する。詳細は、特開 2 0 0 7 - 2 0 6 2 3 9 号公報の記載をここに援用する。

10

【 0 0 5 2 】

また、音声認識装置 3 0 は典型的には自動二輪車に搭載される。ただし、音声認識装置 3 0 は、自動車、船舶、飛行機、ヘリコプタ、その他の輸送機器に搭載されてもよい。以下、自動二輪車の例を説明する。たとえば図 1 9 に示すように、自動二輪車 1 には、車載通信機 2 0 と、車載情報機器 5 0 と、運転者が着用するヘルメット 1 5 A に装備されるヘルメット側無線通信機 3 0 A と、同乗者が着用するヘルメット 1 5 B に装備されるヘルメット側無線通信機 3 0 B とが搭載される。上記音声認識装置 3 0 は車載情報機器 5 0 に内蔵される。

20

【 0 0 5 3 】

自動二輪車 1 は、車体フレーム 2 と、この車体フレーム 2 に対して上下に揺動可能に取り付けられた動力ユニット 3 と、この動力ユニット 3 からの駆動力を得て回転する後輪 4 と、車体フレーム 2 の前部にフロントフォーク 5 を介して取り付けられた操向車輪としての前輪 6 と、フロントフォーク 5 と一体的に回動するハンドル 7 とを備えている。ハンドル 7 には、メイン電源スイッチ 2 8 が備えられている。

【 0 0 5 4 】

動力ユニット 3 は、車体フレーム 2 の中央付近の下部に揺動自在に連結されているとともに、車体フレーム 2 の後部に対しては、リアサスペンションユニット 8 を介して弾性的に結合されている。車体フレーム 2 の中央付近の上部には、運転者用のシート 9 が配置され、さらにその後方には同乗者用のシート 1 0 が配置されている。車体フレーム 2 において、シート 9 とハンドル 7 との間の位置には、運転者が足を置く運転者用ステップ 1 1 が設けられている。また、運転者用のシート 9 の下方には、車体フレーム 2 の両側に、同乗者が足を置くためのステップ 1 2 が設けられている。運転者及び同乗者の乗車状態を検出するために、シート 9 , 1 0 には、それぞれ、運転席着座センサ 1 3 及び同乗者席着座センサ 1 4 が設けられている。

30

【 0 0 5 5 】

車載通信機 2 0 は、同乗者用のシート 1 0 の下方位置において、車体フレーム 2 に固定されている。この車載通信機 2 0 は、同乗者用のシート 1 0 の後方において車体フレーム 2 に固定されたアンテナ 2 1 と接続されており、ヘルメット側無線通信機 3 0 A , 3 0 B との間で無線通信を行う。車載情報機器 5 0 は、ハンドル 7 に固定されており、さらに、車載通信機 2 0 と配線接続されている。車載情報機器 5 0 の例としては、走行経路の音声案内を行うナビゲーションシステム、音楽プレイヤー、ラジオ、携帯電話機の通話音声の中継する電話音声中継装置などを挙げることができる。車載通信機 2 0 及び車載情報機器 5 0 は、車載バッテリー 2 9 からの給電を受けて動作するようになっている。

40

【 0 0 5 6 】

ヘルメット 1 5 A , 1 5 B の内面において、乗員の左右の耳元に対向する位置には、一

50

対のスピーカ 31 が固定されており、乗員の口元に対向する位置にはマイクロフォン 33 が固定されている。一方、帽体の背面には、ヘルメット側無線通信機 30A, 30B が固定されている。このヘルメット側無線通信機 30A, 30B は、アンテナ 36 を備え、スピーカ 31 及びマイクロフォン 33 と接続される。

【0057】

以上、本発明の実施の形態を説明したが、上述した実施の形態は本発明を実施するための例示に過ぎない。よって、本発明は上述した実施の形態に限定されることがなく、その趣旨を逸脱しない範囲内で上述した実施の形態を適宜変形して実施することが可能である。

【図面の簡単な説明】

【0058】

10

【図 1】自動二輪車の走行中に聞こえる風切り音の周波数スペクトルである。

【図 2】図 1 の風切り音に音声が含まれる場合の周波数スペクトルである。

【図 3】本発明の実施の形態による発話区間検出装置の構成を示す機能ブロック図である。

【図 4】図 1 に示した発話区間検出装置の動作を示すフロー図である。

【図 5】音声フレーム及び非音声フレームをフーリエ変換しかつ微分して得られた微分係数の度数分布を示すグラフである。

【図 6】図 5 に示した微分係数の絶対値の度数分布を示すグラフである。

【図 7】本発明の第 2 の実施の形態による発話区間検出装置の動作を示すフロー図である。

20

【図 8】本発明の第 3 の実施の形態で作成されるヒストグラムであり、(a) は音声フレームのヒストグラムであり、(b) は非音声フレームのヒストグラムである。

【図 9】図 8 に示した第 3 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 10】本発明の第 4 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 11】図 10 に示した第 4 の実施の形態で算出される 2 階微分係数の度数分布を示すグラフである。

【図 12】本発明の第 5 の実施の形態による発話区間検出装置の動作を示すフロー図である。

30

【図 13】本発明の第 6 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 14】本発明の第 7 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 15】本発明の第 8 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 16】本発明の第 9 の実施の形態による発話区間検出装置の動作を示すフロー図である。

【図 17】本発明の実施の形態による判定結果に基づいて発話区間を抽出する方法を説明するための図である。

40

【図 18】本発明の実施の形態による発話区間検出装置を備えた音声認識装置の構成を示す機能ブロック図である。

【図 19】図 18 に示した音声認識装置を搭載した自動二輪車の外観構成を示す側面図である。

【符号の説明】

【0059】

22 発話区間検出装置

23 フレーム分割部

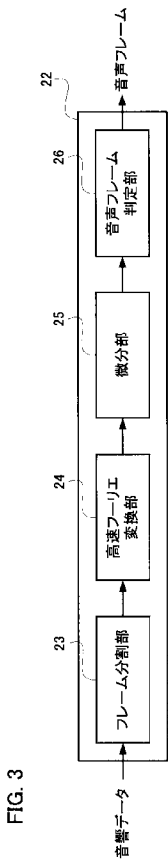
24 高速フーリエ変換部

25 微分部

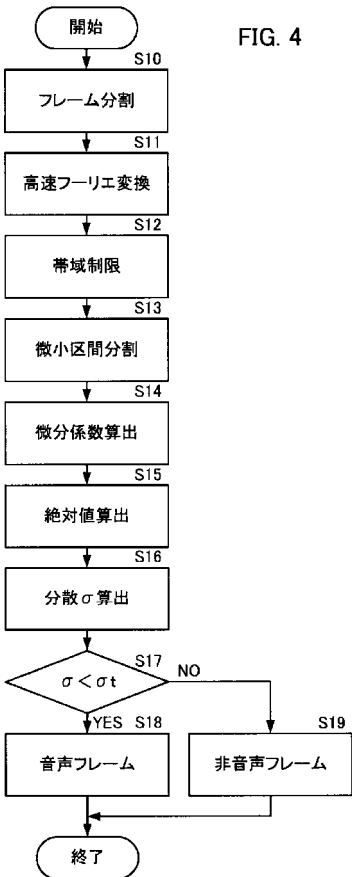
50

- 2 6 音声フレーム判定部
- 3 0 音声認識装置
- 3 2 発話区間検出部

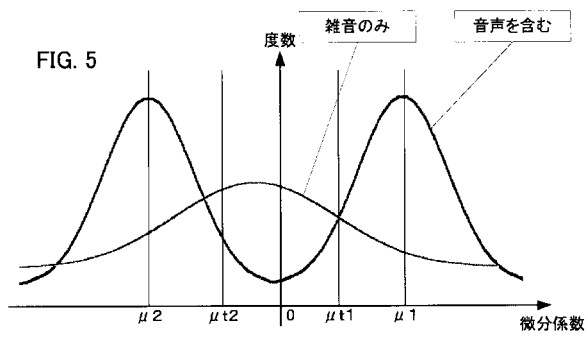
【 図 3 】



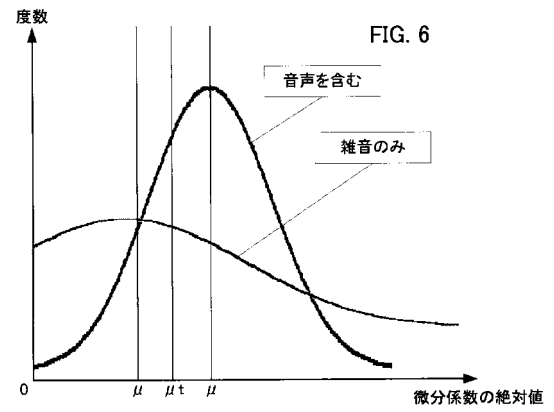
【 図 4 】



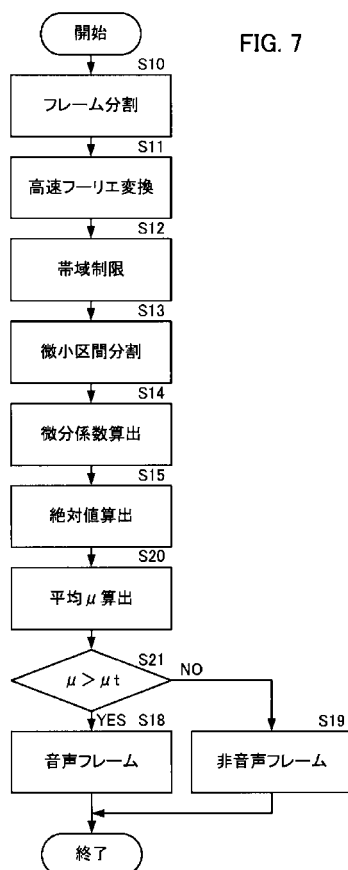
【図 5】



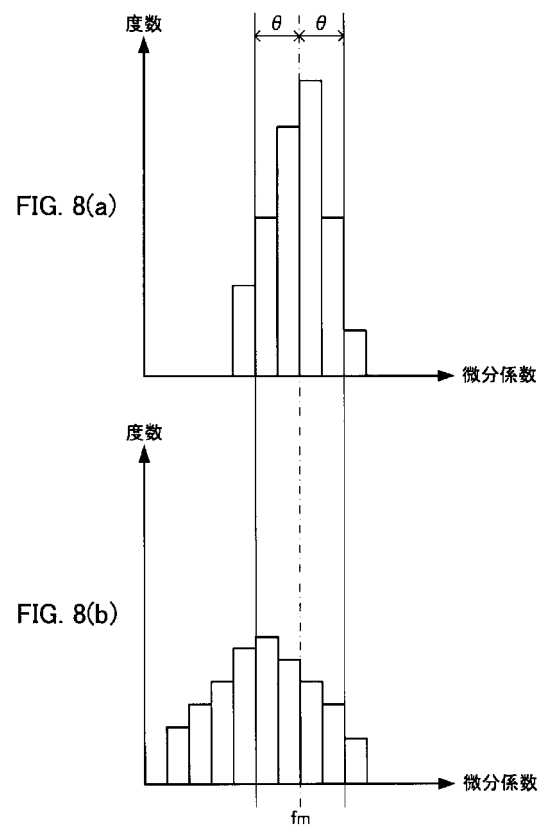
【図 6】



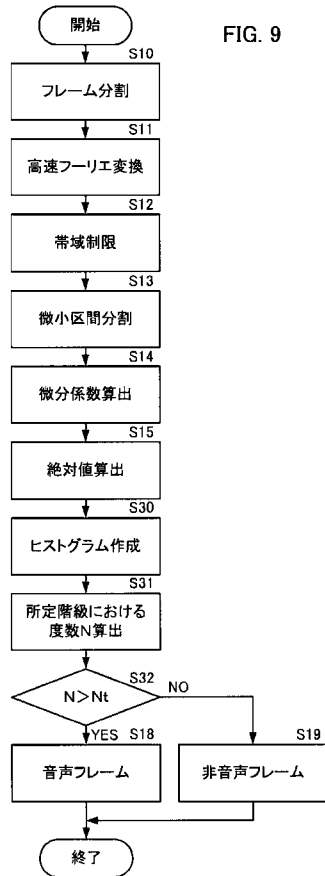
【図 7】



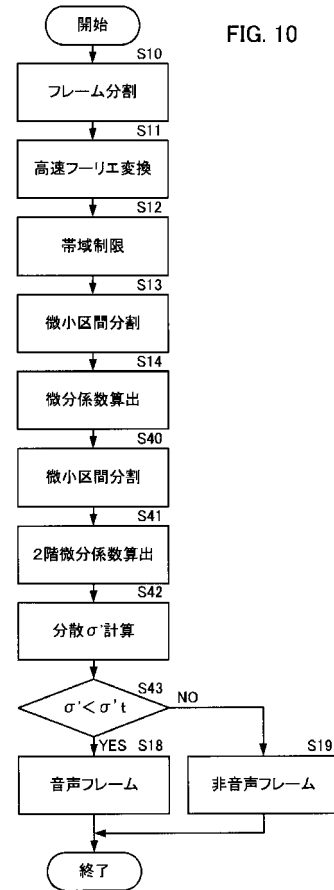
【図 8】



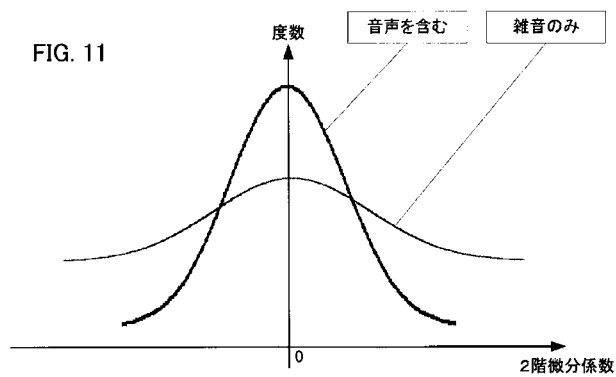
【図 9】



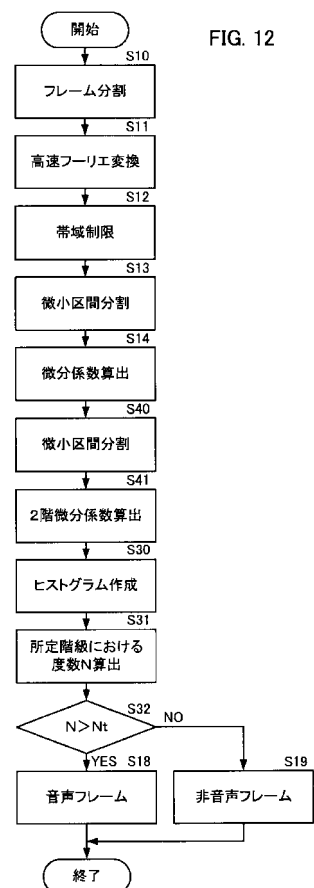
【図 10】



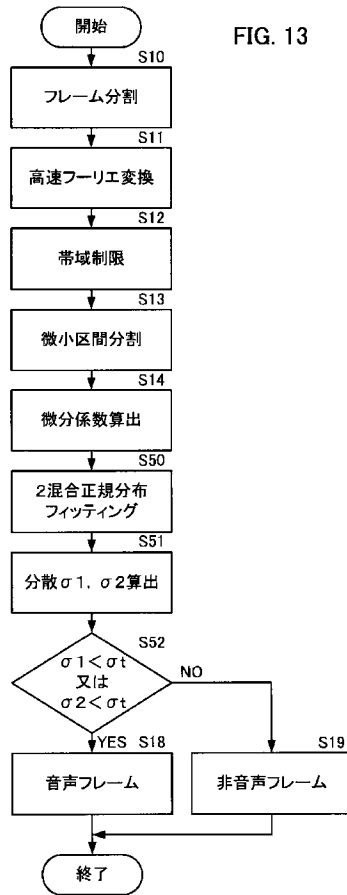
【図 11】



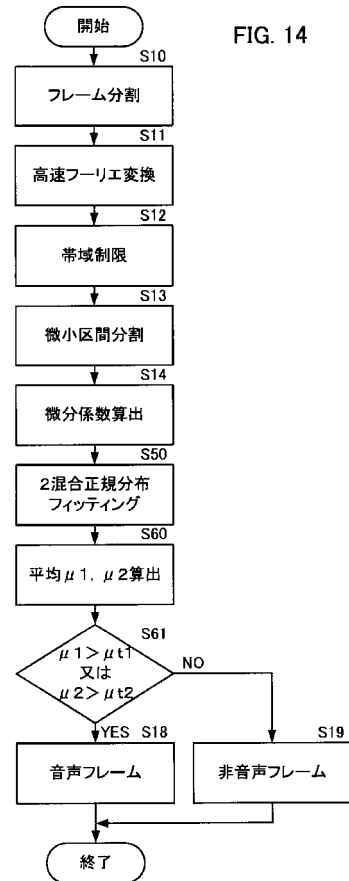
【図 12】



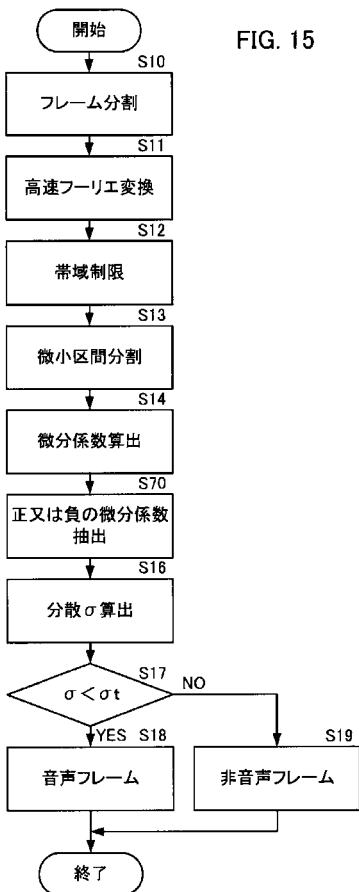
【図 13】



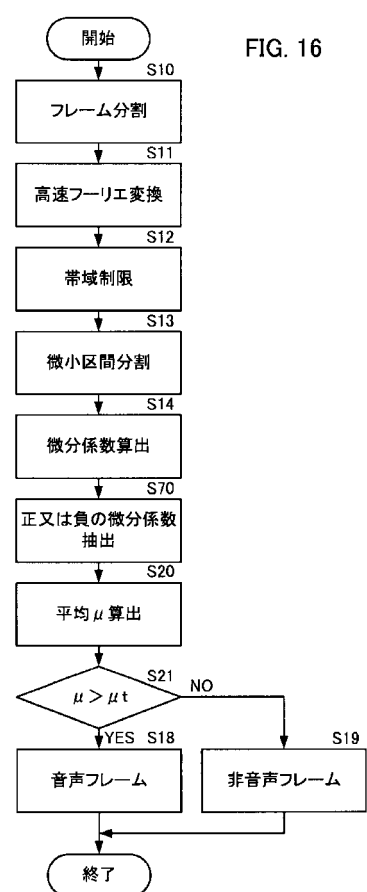
【図 14】



【図 15】



【図 16】



【図 17】

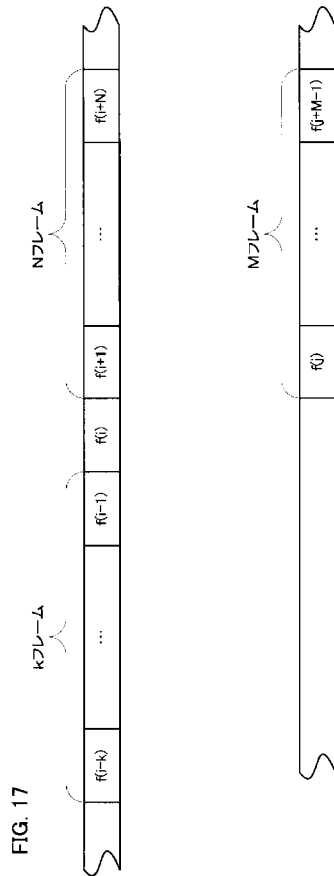
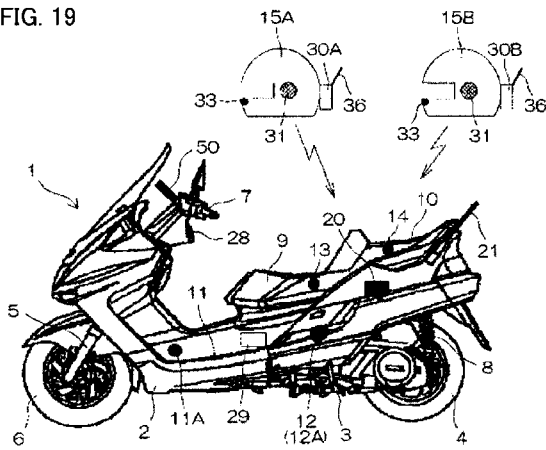


FIG. 17

【図 19】

FIG. 19



【図 18】

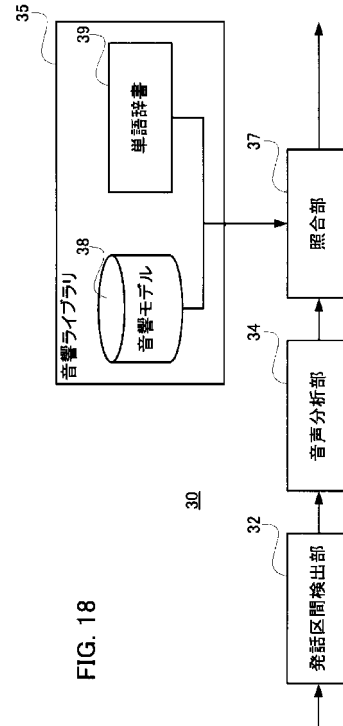
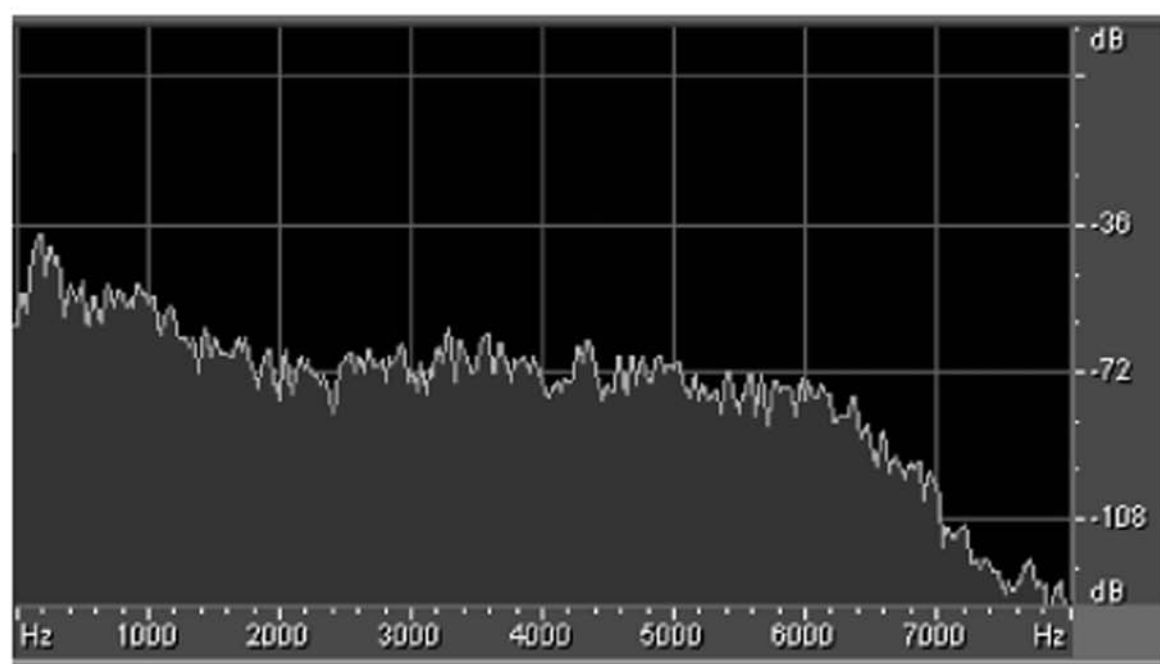


FIG. 18

【図 1】

FIG. 1



【図 2】

FIG. 2

