

(19) 日本国特許庁 (JP)

(12) 特許公報 (B2)

(11) 特許番号

**特許第3660599号**

(P3660599)

(45) 発行日 平成17年6月15日 (2005.6.15)

(24) 登録日 平成17年3月25日 (2005.3.25)

(51) Int. Cl.<sup>7</sup>

F I

G 1 0 L 11/02  
G 1 0 H 7/02  
G 1 0 L 11/00  
G 1 0 L 15/04

G 1 0 L 3/00 5 1 3 A  
G 1 0 L 11/00 1 0 1 F  
G 1 0 L 11/02  
G 1 0 H 7/00 5 2 1 A

請求項の数 8

(全 10 頁)

(21) 出願番号 特願2001-66712(P2001-66712)  
(22) 出願日 平成13年3月9日(2001.3.9)  
(65) 公開番号 特開2002-268657(P2002-268657A)  
(43) 公開日 平成14年9月20日(2002.9.20)  
審査請求日 平成14年8月13日(2002.8.13)

(73) 特許権者 000004226  
日本電信電話株式会社  
東京都千代田区大手町二丁目3番1号  
(74) 代理人 100066153  
弁理士 草野 卓  
(74) 代理人 100100642  
弁理士 稲垣 稔  
(72) 発明者 千喜良 和明  
東京都千代田区大手町二丁目3番1号 日  
本電信電話株式会社内  
(72) 発明者 守谷 健弘  
東京都千代田区大手町二丁目3番1号 日  
本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 音響信号の立ち上がり・立ち下がり検出方法及び装置並びにプログラム及び記録媒体

(57) 【特許請求の範囲】

【請求項 1】

入力された音響信号を解析フレーム単位に分割し、前記解析フレーム内をさらに複数の解析区間に分割し、

各解析区間内の音響信号のエネルギーを算出し、各解析区間のエネルギーの前記解析フレーム内の変動を近似した直線あるいは曲線を計算して求め、

近似した直線あるいは曲線の傾きと予め設定された閾値と比較することにより前記解析フレーム内に音響信号の立ち上がり・立ち下がりのあることを検出することを特徴とする音響信号の立ち上がり・立ち下がり検出方法。

【請求項 2】

請求項 1 に記載の音響信号の立ち上がり・立ち下がり検出方法において、

複数区間のエネルギーの変動を近似する直線あるいは曲線は最小二乗法により求めることを特徴とする音響信号の立ち上がり・立ち下がり検出方法。

【請求項 3】

入力された音響信号を解析フレーム単位に分割し、前記解析フレーム内をさらに複数の解析区間に分割するフレーム解析区間抽出部と、

各解析区間内の音響信号のエネルギーを算出するエネルギー生成部と、

各解析区間のエネルギーの前記解析フレーム内の変動を近似した直線あるいは曲線を計算して求めるエネルギー近似値計算部と、

近似した直線あるいは曲線の傾きと予め設定された閾値と比較することにより解析フレ

ーム内に音響信号の立ち上がり・立ち下がりのあることを検出する検出部とを備えたことを特徴とする音響信号の立ち上がり・立ち下がり検出装置。

【請求項 4】

請求項 3 に記載の音響信号の立ち上がり・立ち下がり検出装置において、

エネルギー近似値計算部における複数区間のエネルギーの変動を近似する直線あるいは曲線を最小二乗法により求めることを特徴とする音響信号の立ち上がり・立ち下がり検出装置。

【請求項 5】

入力された音響信号を解析フレーム単位に分割し、前記解析フレーム内をさらに複数の解析区間に分割する処理と、

各解析区間の音響信号のエネルギーを算出し、前記解析フレーム内のエネルギーの変動を近似した直線あるいは曲線を計算して求める処理と、

近似した直線あるいは曲線の傾きと予め設定された閾値と比較することにより解析フレーム内に音響信号の立ち上がり・立ち下がりのあることを検出する処理をコンピュータに実行させる音響信号の立ち上がり・立ち下がり検出プログラム。

【請求項 6】

請求項 5 に記載の音響信号の立ち上がり・立ち下がり検出プログラムにおいて、

複数区間のエネルギーの変動を近似する直線あるいは曲線を計算して求める処理は最小二乗法を用いて求める処理を有する音響信号の立ち上がり・立ち下がり検出プログラム。

【請求項 7】

入力された音響信号を解析フレーム単位に分割し、前記解析フレーム内をさらに複数の解析区間に分割する処理と、

各解析区間の音響信号のエネルギーを算出し、前記解析フレーム内のエネルギーの変動を近似する直線あるいは曲線を計算して求める処理と

近似した直線あるいは曲線の傾きと予め設定された閾値と比較することにより解析フレーム内に音響信号の立ち上がり・立ち下がりのあることを検出する処理をコンピュータに実行させる音響信号の立ち上がり・立ち下がり検出プログラムを記録した記録媒体。

【請求項 8】

請求項 7 に記載の音響信号の立ち上がり・立ち下がり検出プログラムを記録した記録媒体において、

複数区間のエネルギーの変動を近似する直線あるいは曲線を計算して求める処理は最小二乗法を用いて求める処理を有する音響信号の立ち上がり・立ち下がり検出プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、音響信号（楽音、音声信号）の立ち上がり・立ち下がりの検出方法に関し、音響信号の符号化、音声認識、音声通信等に用いられる。

【0002】

【従来の技術】

アタック検出技術は、例えば音響信号の符号化技術において変換長を決定する際に利用されている。

楽音符号化技術の代表的な例としてはTwinVQ(Transform domain Weighted INterleave Vector Quantization)、AAC(Advanced Audio Coding)、ATRAC3(Adaptive TRansform Acoustic Coding)等があげられる。これらの符号化技術は主に量子化によって表現ビット数を削減するが、量子化効率を上げるために、音響信号を時間領域から周波数領域へ変換してから量子化を行う。この変換は一般的に直交変換を用いて行うが変換長の選択が重要である。より長い変換フレームを採用し、フレームを一度に変換すれば周波数領域でパワーが集中しやすく後に行う量子化効率が高くなるが、同一フレーム内で信号の特性が著しく変化する場合には、量子化によって特性の異なる信号同士が干渉しあうため、聴感上のひずみ

10

20

30

40

50

が目立つようになる（これをプリエコーといい、高圧縮率時に顕著となる）。逆に変換長を短くすれば音響信号の特性が変化した場合でもその影響（干渉）が及ぶ範囲は時間的に短い。したがって音響信号の特性が激しく変化する（アタックが存在する）場合には短い変換長で時間領域／周波数領域変換を行い、通常はそれよりも長い変換長で変換を行うことが対策として考えられる。

#### 【0003】

前述のTwinVQ方式では変換フレーム内の各区間ごとの信号エネルギーのばらつき度（相加平均と相乗平均の比）を主な尺度としてアタックの検出を行っている。また、AAC方式では聴覚的エントロピーを用いた検出方法が提案されている。聴覚的エントロピーは複数の周波数帯域部分ごとに聴覚モデルを用いて計算した最小可聴ノイズから計算され、このエントロピーの大小によってアタックの存在を判定し、アタックの検出を行っている。

10

#### 【0004】

##### 【発明が解決しようとする課題】

従来技術のようにエネルギーのばらつきを選択基準とすると、一定の割合でエネルギーが増加している場合（エネルギーの変動の例：1,2,3,4,5）でも、急激にエネルギーが変動している場合（エネルギー変動の例：1,5,4,3,2）でも同じ基準値を生成することになる。後者は明らかにアタックであり、前者の場合はアタックとはいえないためエネルギーのばらつきは尺度としての的確ではない場合がある。また、聴覚エントロピーは計算するために高次のFFT(Fast Fourier Transform)を行う必要があるなど演算量の面で不利である。この発明の課題は音響信号の立ち上がり及び立ち下がりをも的確に、少ない演算量、処理量で検出することである。

20

#### 【0005】

##### 【課題を解決するための手段】

音響信号の立ち上がり及び立ち下がりの検出判定として、解析フレーム内のエネルギーの変動を直線あるいは曲線により近似し、（1）この近似した値と実際の値の誤差の和をとり閾値と比較して行う、（2）近似した直線あるいは曲線の傾きを閾値と比較して行う。また、高域信号部分（例えば、4 kHz以上）のエネルギーだけを対象としてエネルギー変動を直線あるいは曲線により近似し、同様に検出判定を行う。

#### 【0006】

##### 【発明の実施の形態】

図1を参照してこの発明の概要を説明する。

30

（1）入力された音響信号を解析フレーム単位に分割し、解析フレーム内をさらに複数の解析区間(1,2,・・・,j)に分割し、各区間内の音響信号のエネルギーを算出し( $E_1, E_2, \dots, E_j$ )、

#### 【0007】

##### 【数1】

各区間ごとのエネルギーの変動を直線で近似し( $\hat{E}_j = a_0 + a_1 j$ )、

各区間ごとの近似値と音響信号のエネルギーの誤差を加算した値

$(\sum (E_j - \hat{E}_j))$  と予め設定された閾値  $E_{th}$  と比較することにより

40

解析フレーム内の音響信号の立ち上がり・立ち下がりを検出する。

#### 【0008】

（2）（1）でエネルギーの変動を直線で近似した傾き  $a_1$  と予め設定された閾値  $A_{th}$  と比較することにより解析フレーム内の音響信号の立ち上がり・立ち下がりを検出する。

#### 【0009】

図2, 3を参照してこの発明を詳細に説明する。

図2に概要（1）で説明した音響信号の立ち上がり・立ち下がり検出装置の構成、図3に

50

入力された音響信号を解析フレーム単位に分割し、解析フレーム内をさらに複数の区間に分割するための説明図を示す。

#### フレーム解析区間抽出部

(1)入力された音響信号列を解析フレーム単位に分割し、解析フレーム内をさらに複数の解析区間に分割する。解析フレームは標本化点を2048点とすると、解析フレーム内を図3に示すように1,2,・・・,8個の解析区間jを重ね合わせて抽出する場合、第j区間は $((j+2) \times 2048) \div 16 + 1$ 点目から $((j+4) \times 2048 \div 16)$ 点目となる。

【0 0 1 0】

図3からわかるように第j区間と第j+1区間は $2048 / 16 = 128$ 点重複する。通常楽音符号化方式における時間領域/周波数領域変換部では重ね合わせ直交変換(LOT:Lapped Orthogonal Transform)が利用されているため、第i変換フレームと第i+1変換フレームとは1024点重複している。したがって、この変換フレームをそのままこの発明の解析フレームとした場合には、第i解析フレーム内の第8区間は第i+1解析フレーム内の第1区間と128点重複している。

10

(2)解析区間内の音響信号(サンプル値)を $[x_k]$ , $(k=0,1,2,\dots,255)$ とするとき、各解析区間内のサンプル値にハミング窓をかけると $w_k = x_k \times (0.54 - 0.46 \times \cos(2 \times k / 256))$ となる。

(3)各区間の $[w_k]$ を周波数領域に変換する。これは256点のFFTを用いて行うことができる。周波数成分は $[f_k]$ , $(k=0,1,2,\dots,127)$ となる。また、高域成分は例えば $[f_k]$ , $(k=64,65,66,\dots,127)$ とすることができる(この高域成分の設定は実験等により行う)。

20

#### 解析区間内エネルギー生成部

(4)FFT前(時間領域)とFFT後(周波数領域)のそれぞれについて解析区間内サンプル値の二乗和(エネルギー)G,Eを計算する。

【0 0 1 1】

【数2】

$$G = \sum_{k=0}^{255} w_k^2$$

$$E = \sum_{k=0}^{127} f_k^2 \quad E = \sum_{k=64}^{127} f_k^2 \quad (\text{高域部分})$$

30

#### 直線によるエネルギー近似値計算部

(4)次に $E_j, G_j$ の変動を直線あるいは曲線により近似する。

【0 0 1 2】

【数3】

$E_j, G_j$ の近似値 $\hat{E}_j, \hat{G}_j$ をとすると、近似式は以下のよう表すことができる。

40

$$\hat{E}_j = a_0 + a_1 j + a_2 j^2 + \dots + a_{N-1} j^{N-1}$$

$$\hat{G}_j = b_0 + b_1 j + b_2 j^2 + \dots + b_{N-1} j^{N-1}$$

このとき

$$\sum_{j=1}^8 (E_j - \hat{E}_j)^2 \quad \text{および} \quad \sum_{j=1}^8 (G_j - \hat{G}_j)^2$$

50

を最小にする $a_n, b_n$ は最小二乗法により求める。

【0013】

【数4】

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{N-1} \end{pmatrix} = \begin{pmatrix} \sum 1 & \sum j & \sum j^2 & \cdots & \sum j^{N-1} \\ \sum j & \sum j^2 & \sum j^3 & \cdots & \sum j^N \\ \sum j^2 & \sum j^3 & \sum j^4 & \cdots & \sum j^{N+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum j^{N-1} & \sum j^N & \sum j^{N+1} & \cdots & \sum j^{2N-2} \end{pmatrix}^{-1} \begin{pmatrix} \sum E_j \\ \sum jE_j \\ \sum j^2 E_j \\ \vdots \\ \sum j^{N-1} E_j \end{pmatrix} \quad 10$$

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_{N-1} \end{pmatrix} = \begin{pmatrix} \sum 1 & \sum j & \sum j^2 & \cdots & \sum j^{N-1} \\ \sum j & \sum j^2 & \sum j^3 & \cdots & \sum j^N \\ \sum j^2 & \sum j^3 & \sum j^4 & \cdots & \sum j^{N+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum j^{N-1} & \sum j^N & \sum j^{N+1} & \cdots & \sum j^{2N-2} \end{pmatrix}^{-1} \begin{pmatrix} \sum G_j \\ \sum jG_j \\ \sum j^2 G_j \\ \vdots \\ \sum j^{N-1} G_j \end{pmatrix} \quad 20$$

ここで、 $\Sigma$ は $\sum_{j=1}^8$ である。

例えば、 $N=2$ で、解析区間数が8の場合、これはエネルギーの変動を直線近似することになり、以下のようなになる。

【0014】

【数5】

$$a_0 = \frac{17 \sum_{j=1}^8 E_j - 3 \sum_{j=1}^8 j E_j}{28}, \quad a_1 = \frac{2 \sum_{j=1}^8 j E_j - 9 \sum_{j=1}^8 E_j}{84}$$

$$b_0 = \frac{17 \sum_{j=1}^8 G_j - 3 \sum_{j=1}^8 j G_j}{28}, \quad b_1 = \frac{2 \sum_{j=1}^8 j G_j - 9 \sum_{j=1}^8 G_j}{84} \quad 30$$

となる。

二乗演算器・加算器

誤差の二乗和は、以下に示すようになる。

【0015】

【数6】

$$e = \sum_{j=1}^8 (E_j - \hat{E}_j)^2 = \sum_{j=1}^8 E_j^2 - a_0 \sum_{j=1}^8 E_j - a_1 \sum_{j=1}^8 j E_j$$

$$g = \sum_{j=1}^8 (G_j - \hat{G}_j)^2 = \sum_{j=1}^8 G_j^2 - b_0 \sum_{j=1}^8 G_j - b_1 \sum_{j=1}^8 j G_j \quad 50$$

## 【0016】

なお、上記の例では誤差の二乗和を求めているが、誤差の絶対値を用いることもできる。

## 立ち上がり・立ち下がり検出部

(5)音響信号の立ち上がり・立ち下がりの有無は $e, g, a_1, b_1$ の大小で判断する。

$e$ あるいは $g$ が大きいときは直線による近似がうまくいかなかったときであるから(すなわち、誤差の和が大きい場合)、解析フレーム内に音響信号の立ち上がり・立ち下がりが存在すると判定する目安となる。また、 $a_1, b_1$ が大きいとき(すなわち、傾きが大きい場合)、解析フレーム内でエネルギーの変動が激しいときであるから音響信号の立ち上がり・立ち下がりが存在すると判定する目安となる。したがって、実験結果等により求めた閾値 $A_{th}, B_{th}, E_{th}, G_{th}$ と比較し、 $a_1 > A_{th}, b_1 > B_{th}, e > E_{th}, g > G_{th}$ のときに音響信号の立ち上がり・立ち下がりが存在すると判定する。また、判定を例えば、 $e > E_{th}$ かつ $a_1 > A_{th}$ かつ $b_1 > B_{th}$ とすれば検出精度を上げることができる。

10

## 【0017】

なお、 $a_1 > A_{th}$ あるいは $b_1 > B_{th}$ の判定は直線によるエネルギー近似値計算部において計算された傾き $a_1$ あるいは $b_1$ を閾値 $A_{th}, B_{th}$ と比較することにより音響信号の立ち上がり・立ち下がりを検出することができる。

この発明は、例えば符号化技術において変換フレーム内の変換長の決定に利用することができる。

変換フレームを解析フレームとしてこの発明による音響信号の立ち上がり・立ち下がり検出を行い、音響信号の立ち上がり・立ち下がりが存在するフレームでは変換長を短く、また、音響信号の立ち上がり・立ち下がりが存在しないならばフレーム全体を変換長とすればよい。

20

## 【0018】

図4に示すように、音響信号の立ち上がり(アタック)が存在する音を、長い変換領域を用いて符号化するとアタックのエネルギーが変換領域全体にわたって拡散されてしまい、復号後にプリエコーと呼ばれる音質の劣化を引き起こす。このような場合に精度の高いアタック検出法を用いて変換長を短くすればプリエコーを抑えることができ、復号時に音質が向上する。

本方式を実際に符号化・復号化装置に組み込み品質評価試験を行った結果を図5に示す。

## 【0019】

コーデックはMPEG-4/Audioのリファレンスソフトウェアを利用している。

試験はリファレンスソフトによる復号音と、本方式による変換長切換えを実装したリファレンスソフトによる復号音を比較するものである。この試験では判定の条件を $E_{th} > 0.2$ かつ $G_{th} > 0.2$ とした。表の数値は、従来方式に比べて本方式の音を採点したものである。採点の範囲は、悪い(-3)から良い(+3)までである。上限と下限は評価の分布を正規分布で近似し、分布を95%で切り捨てたときの上限と下限である。

30

## 【0020】

音声のように急激に音が変化する場合にはアタックによって変換長を切り換える方法が有効であるが、本方式では音声(歌唱:女性、独語:男性、英語等)において音質が平均値0より上回り向上する傾向がある。

40

また、この発明の音響信号の立ち上がり・立ち下がり検出装置をCPUやメモリ等を有するコンピュータと、アクセス主体となるユーザが利用する利用者端末と記録媒体から構成することができる。

記録媒体は、CD-ROM、磁気ディスク、半導体メモリ等の機械読み取り可能な

記録媒体であり、ここに記録された、あるいは通信回線等で配信された音響信号の立ち上がり・立ち下がり検出プログラムはコンピュータに読み取られ、コンピュータの動作を制御し、コンピュータ上に実施の形態における各構成要素、すなわち、フレーム解析区間抽出部、解析区間内エネルギー生成部、直線あるいは曲線によるエネルギー近似値計算部等を実現する。

## 【0021】

50

【発明の効果】

本発明によれば、なめらかなエネルギー変動（エネルギー変動の例：1,2,3,4,5）は直線で近似できるためアタック無しと判定される。また、アタックが存在する例（エネルギー変動の例：1,5,4,3,2）は近似が困難なため、誤差が大きくなりアタックが存在すると判定される。

また、エネルギーの変動が大きい場合でも、高域信号にエネルギーが少ない場合は、短い変換長を用いたときの量子化雑音がマスキングされないため、聴覚上のひずみが目立ってしまう。したがってこの場合には長い変換長を選択することが必要である。本発明では高域信号のエネルギーの変動にも着目しているため高域信号のエネルギー変動が無い場合には長い変換長が選択される。

10

【0022】

以上から本発明に音響信号の立ち上がり・立ち下がり検出法を用いれば少ない演算量で音響信号の特性に適した変換長を選択することが可能となり、符号化技術変換長の選択に応用すればプリエコーが軽減され、復号後の音質が向上する。

【図面の簡単な説明】

【図1】本発明の概要を説明するための図。

【図2】本発明の音響信号の立ち上がり・立ち下がり検出装置の構成を示すブロック図。

【図3】音響信号と解析フレーム、解析区間の関係を示す図。

【図4】アタック音に対する長・短変換長における復号後の出力信号を示す図。

【図5】本発明の音響信号の立ち上がり・立ち下がり検出法を適用した符号化・復号化による品質評価試験の結果を示す図。

20

【図1】

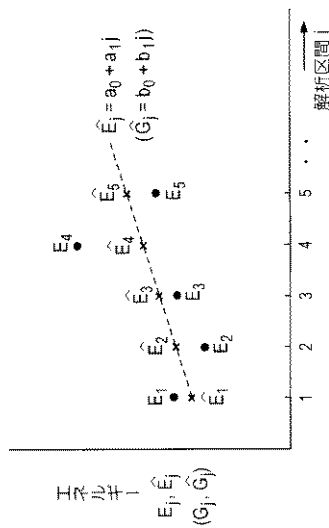


図1

【図2】

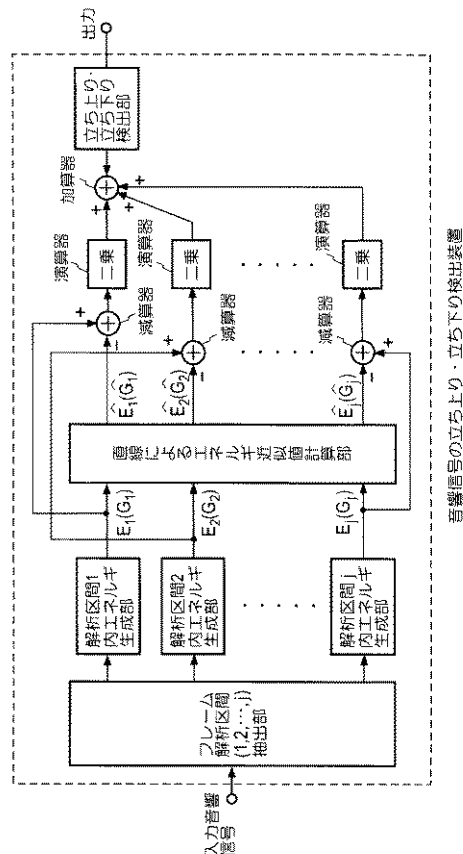


図2

【図3】

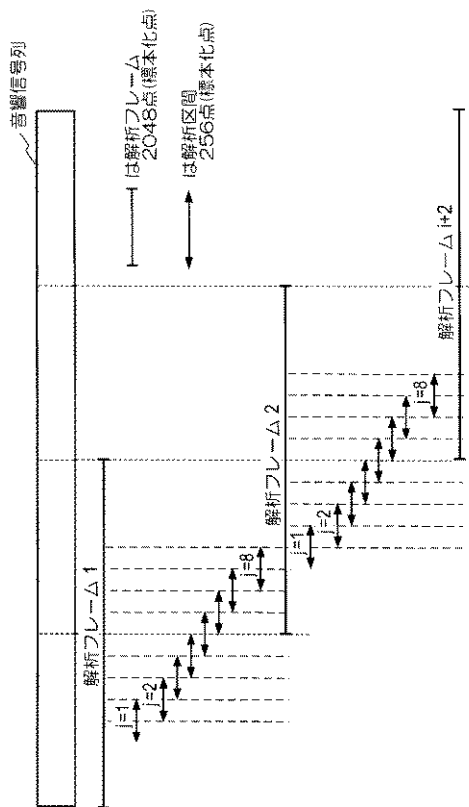


図3

【図4】

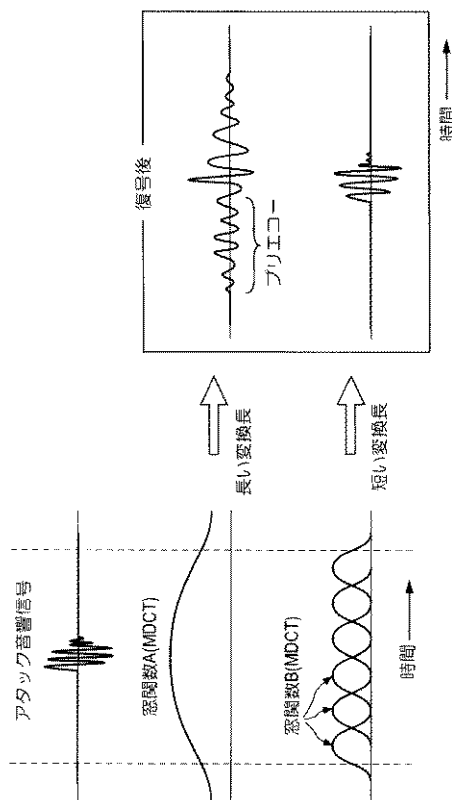


図4

【図5】

	平均	下限	上限
歌謡：女性	0.375	-0.135	0.885
独語：男性	0.438	-0.112	0.987
英語：女性	0.062	-0.469	0.594
英語：男性	0.125	-0.258	0.508
カスターネット	-0.312	-0.688	0.063
ハグハイブ	0.062	-0.392	0.517
グロッケン	0.188	-0.257	0.632
total	0.134	-0.026	0.294

図5



-----  
フロントページの続き

- (72)発明者 神 明夫  
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 森 岳至  
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 山下 剛史

- (56)参考文献 特開平08-272394(JP,A)  
特開2000-315091(JP,A)  
特開2000-172283(JP,A)  
特開2000-259197(JP,A)

- (58)調査した分野(Int.Cl.<sup>7</sup>, DB名)  
G10L 11/00-21/06  
G10H 7/02  
JICSTPlusファイル(JOIS)