

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3869823号
(P3869823)

(45) 発行日 平成19年1月17日(2007.1.17)

(24) 登録日 平成18年10月20日(2006.10.20)

(51) Int. Cl.	F I
G 1 0 L 21/04 (2006.01)	G 1 0 L 21/04 1 2 0 C
G 1 0 L 11/00 (2006.01)	G 1 0 L 21/04 1 2 0 E
	G 1 0 L 11/00 1 0 1 H
	G 1 0 L 11/00 1 0 1 Z

請求項の数 2 (全 9 頁)

<p>(21) 出願番号 特願2003-288293 (P2003-288293)</p> <p>(22) 出願日 平成15年8月6日(2003.8.6)</p> <p>(65) 公開番号 特開2005-55778 (P2005-55778A)</p> <p>(43) 公開日 平成17年3月3日(2005.3.3)</p> <p>審査請求日 平成16年6月24日(2004.6.24)</p> <p>(出願人による申告) 国等の委託研究の成果に係る特許出願(平成15年度通信・放送機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」)に関する委託研究、産業活力再生特別措置法第30条の適用を受けるもの)</p>	<p>(73) 特許権者 393031586 株式会社国際電気通信基礎技術研究所 京都府相楽郡精華町光台二丁目2番地2</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 ニ ジンフ 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 津崎 実 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p>(72) 発明者 河井 恒 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内</p> <p style="text-align: right;">最終頁に続く</p>
---	---

(54) 【発明の名称】 音声の周波数特性の等化装置

(57) 【特許請求の範囲】

【請求項1】

処理対象となる音声と基準となる音声との間のパワースペクトル密度(PSD)の差分を算出するための手段と、

前記差分をケプストラムにより表される周波数特性の特徴空間に変換するための手段と、

前記変換するための手段により出力される、前記ケプストラムにより表された差分をメルケプストラムに変換するための手段と、

前記メルケプストラムに変換するための手段により出力される、メルケプストラムにより表された差分を、周波数帯域ごとの利得を定めるフィルタパラメータとして受け、前記処理対象となる音声を入力として受ける様に接続されるMLSA(mel-logarithmic spectral approximation)フィルタとを含む、音声の周波数特性の等化装置。

【請求項2】

処理対象となる音声と基準となる音声との間のPSDの差分を算出するための手段と、

前記PSDの差分を、第1の次数を有する差分のケプストラムに変換するための手段と、

前記差分のケプストラムに対し、前記第1の次数のケプストラムから前記第1の次数よりも小さな第2の次数を有するメルケプストラムへの周波数軸ワーピングを行なうための第1のメルワーピング手段と、

10

20

前記第 1 のメルワーピング手段の出力に対し、前記第 1 のメルワーピング手段による周波数軸ワーピングの逆変換を行ない、逆変換された差分のケプストラムを出力するための第 2 のメルワーピング変換手段と、

第 2 のメルワーピング変換手段の出力する前記逆変換された差分のケプストラムをスペクトルに変換することにより、平滑化された PSD の差分を出力するための手段と、

前記平滑化された PSD の差分を、周波数帯域ごとの利得を定めるフィルタパラメータとして受け、前記処理対象となる音声を入力として受ける様に接続される FIR (finite impulse response) フィルタを含む、音声の周波数特性の等化装置。

【発明の詳細な説明】

10

【技術分野】

【0001】

この発明は、音声の補正技術に関し、特に、波形素片接続型音声合成システム等において、入力される音声の音質を、ターゲット音声に近い音質に補正するための技術に関する。

【背景技術】

【0002】

コンピュータ技術及びデータコミュニケーション技術の発達に伴い、人間と機械との間のインターフェイスが重要となっている。人間にとっては、人と話をするのと同様に機械とのコミュニケーションを行なえることが望ましく、そのための技術開発が進められている。

20

【0003】

人間から機械への情報の伝達としては、音声認識、画像認識等の認知技術が主として用いられる。また機械から人間への情報の伝達方法は種々あるが、中でも音声合成技術が用いられる機会が増加している。音声応答システム、音声翻訳システム、コンピュータゲーム等が代表的な応用例である。さらに、近年のロボット等の開発の進展に伴い、音声認識及び画像認識と音声合成とを組み合わせることで、人間とロボットとのコミュニケーションを人間同士のコミュニケーションと同様に実現することが期待される。

【0004】

音声合成では、如何にして自然な音声を合成するかが重要である。最近では、数十時間規模の大規模な音声コーパスを使用して音声素片を作成しておき、入力されるテキストデータに応じて適切な音声素片を選択し接続する、いわゆる音声素片接続型音声合成が主流となっている。この技術では、如何に自然に音声波形素片を接続するかが重要となる。

30

【0005】

上述した様に現在の波形素片音声合成システムでは、音質向上のために大規模な音声コーパスを使用している。多くの場合、単一の話者の音声を長期間かけて収録する。場合によってはその収録に数ヶ月から数年の期間を必要とする。

【0006】

こうした場合、録音時期が異なると、録音系の特性が経年変化し、そのために録音された音声を再生した場合、その音質が変化してしまうことがある。波形接続を行なう場合、その様に互いに異なる音質の音声を接続すると、合成された音声が不自然なものとなる問題がある。

40

【0007】

こうした問題を解決するための音声の補正技術に関し、一つの提案が非特許文献 1 においてなされている。図 4 に、非特許文献 1 に記載されたチャネル等化装置のブロック図を示す。図 4 を参照して、この装置 200 は、ソース音声 30 を受け、ソース音声 30 の発話内容でターゲット音声 32 とほぼ同じ周波数特性の音声を発生するためのものである。なお、本明細書では、ターゲット音声 32 は予め録音されていた、基準となる音声を指す。ソース音声 30 は、ターゲット音声 32 とは別の時期に録音された音声であり、録音系の特性の経年変化により、その周波数特性がターゲット音声 32 とは異なっている可能性

50

があるものとする。

【0008】

この装置200は、ソース音声30のパワースペクトル密度(PSD)を生成するためのソースPSD生成部34と、ターゲット音声32のPSDを生成するためのターゲットPSD生成部36と、ソース音声30のPSDとターゲット音声32のPSDとの差分(ターゲットPSD/ソースPSD)を計算するための除算部38と、除算部38の出力に基づくLPC(線形予測係数)分析の結果を用いたIIR(Infinite Impulse Response)フィルタでソース音声30をフィルタリングし等化処理済みの音声212を出力するためのLPCフィルタ部210を含む。

【0009】

ソースPSD生成部34及びターゲットPSD生成部36は同様の構成を有する。ソースPSD生成部34は、ソース音声30のデータに含まれる各音声フレームを検出するための音声フレーム検出部50と、音声フレーム検出部により検出された各音声フレームに所定の窓掛け処理を行なうための窓掛け処理部52と、窓掛け処理部52により窓掛け処理された音声フレームデータから高速フーリエ変換(FFT)により当該音声フレームのPSDを算出するためのパワースペクトル算出部54と、パワースペクトル算出部54により算出された、所定期間のソース音声30のフレームのPSDの平均を算出するためのフレーム平均部56を含む。

【0010】

ターゲットPSD生成部36も同様に、音声フレーム検出部60と、窓掛け処理部62

10

20

【0011】

LPCフィルタ部210は、除算部38の出力に逆FFT(IFFT)処理を行なうためのIFFT部220と、IFFT部220の出力に対しLPC変換を行なうためのLPC変換部222と、LPC変換部222の出力するLPC係数により決定されるフィルタパラメータを持ち、ソース音声30に対するフィルタリング処理を行なってソース音声30の周波数特性をターゲット音声32の周波数特性に等化させるためのIIRフィルタ224を含む。

【0012】

ソース音声30とターゲット音声32との周波数特性の差分を除算部38で算出し、その差分に対するLPC変換を行なってIIRのフィルタパラメータを設定する。このチャンネル等化装置200により、ソース音声30の周波数特性をターゲット音声32のそれとほぼ等しいものに等化できる。

30

【0013】

【非特許文献1】ユー シ、エリック チャン、フ ペン、ミン チュウ、「接続型TTSシステムのための、大規模音声データベースについてのパワースペクトル密度に基づくチャンネル等化」、ICSLP2002予稿集、pp.2369-2372、米国、2002(Yu Shi, Eric Chang, Hu Peng, and Min Chu, "Power Spectral Density Based Channel Equalization Of Large Speech Database For Concatenative TTS System, Proc of ICSLP2002, pp.2369-2372, USA, 2002)

40

【発明の開示】

【発明が解決しようとする課題】

【0014】

図4に示す従来の等化装置については、その有効性が示されている。

【0015】

しかし従来法では、LPC変換における次数をどの様を選択すべきかについて、困難な問題がある。すなわち、LPC変換の次数を小さくすると、補正の効果がほとんどなくなる一方、次数を大きくすると音質の劣化が甚だしくなるという問題がある。そのためLP

50

C変換の次数を適切な値に決めるのが困難である。

【課題を解決するための手段】

【0016】

本発明の第1の局面に係る音声の周波数特性の等化装置は、処理対象となる音声と基準となる音声との間のパワースペクトル密度(PSD)の差分を算出するための手段と、差分をケプストラムにより表される周波数特性の特徴空間に変換するための手段と、変換するための手段により出力される、ケプストラムにより表された差分をメルケプストラムに変換するための手段と、メルケプストラムに変換するための手段により出力される、メルケプストラムにより表された差分を、周波数帯域ごとの利得を定めるフィルタパラメータとし、処理対象となる音声を入力として受けるように接続されるMLSAフィルタとを含む。

10

【0019】

本発明の第2の局面に係る音声の周波数特性の等化装置は、処理対象となる音声と基準となる音声との間のPSDの差分を算出するための手段と、PSDの差分を、第1の次数を有する差分のケプストラムに変換するための手段と、差分のケプストラムに対し第1の次数のケプストラムから第1の次数よりも小さな第2の次数を有するメルケプストラムへの周波数軸ワーピングを行なうための第1のメルワーピング手段と、第1のメルワーピング手段の出力に対し、第1のメルワーピング手段による周波数軸ワーピングの逆変換を行ない、逆変換された差分のケプストラムを出力するための第2のメルワーピング変換手段と、第2のメルワーピング変換手段の出力する逆変換された差分のケプストラムをスペクトルに変換することにより、平滑化されたPSDの差分を出力するための手段と、平滑化されたPSDの差分を、周波数帯域ごとの利得を定めるフィルタパラメータとして受け、処理対象となる音声を入力として受けるように接続されるFIRフィルタとを含む。

20

【発明の効果】

【0020】

処理対象となる音声と基準となる音声とのパワースペクトル密度の差分がケプストラムにより表される周波数特性の特徴空間に変換される。それをさらにメルスケールに変換してMLSAフィルタを設定する。又は、ケプストラムに変換した後、メルワーピング及びその逆変換を行なって逆変換されたケプストラムを得て、それをさらにスペクトルに戻すことでPSDの差分を平滑化し、そのPSDの差分でフィルタを設定する。こうして設定されたフィルタは人間の聴覚特性に近い特性を持つ。またこうして設定されるフィルタの特性は、LPC変換によるフィルタと異なり、パラメータ次数に敏感でない。フィルタの精度を高める様にパラメータの算出を行なう場合にも、音質の劣化が生じることがない。また従来チャンネル等価装置と同程度の音質で、処理対象となる音声の周波数特性を基準となる音声の周波数特性に等化させることができる。

30

【発明を実施するための最良の形態】

【0021】

[第1の実施の形態]

図1に、本発明の第1の実施の形態に係るチャンネル等価装置20のブロック図を示す。図1において、図4と同じ部品には同じ参照番号を付してある。それらの機能も同一である。従ってそれらについての詳細な説明は繰返さない。

40

【0022】

図1に示すチャンネル等価装置20が図4のチャンネル等価装置200と異なるのは、図4のLPCフィルタ部210に代えて、PSDの差分を平滑化したフィルタパラメータで設定されたFIR(Finite Impulse Response)フィルタを用いて等化を行なうケプストラムフィルタ部40を含む点である。

【0023】

ケプストラムフィルタ部40は、除算部38の出力するソース音声30とターゲット音声32とのPSDの平均の差分について平方根をとりさらにその対数を算出するための対

50

数算部 70 と、対数算出部 70 の出力に対し m 次の IFFT 処理を実行することにより除算部 38 の出力に対するケプストラムを算出するための IFFT 処理部 72 と、IFFT 処理部 72 の出力するケプストラムについて、その横軸（周波数軸）をメルスケールに変換する（メルワーピングする）ための第 1 のメルワーピング部 74 とを含む。

【0024】

第 1 のメルワーピング部 74 での変換をメルワーピング (m 、 n 、 a) と表しているが、 m は変換前のケプストラムの次数、 n は変換後の次数である。ここでは $n < m$ となる様に m と n とが選ばれている。 a は周波数軸伸縮のパラメータである定数であり、サンプリング周波数に応じて定められる。

【0025】

ケプストラムフィルタ部 40 はさらに、第 1 のメルワーピング部 74 の出力に対してメルワーピング (n 、 m 、 $-a$) を実行するための第 2 のメルワーピング部 76 を含む。第 2 のメルワーピング部 76 でのメルワーピングと第 1 のメルワーピング部 74 でのメルワーピングとは、互いに逆変換の関係になる。すなわちこれら二つの処理を直列に実行することにより周波数軸は元の線形軸に戻る。ただし m 、 n の値が $n < m$ となる様に選ばれているため、これら二つの処理を直列に実行した場合、メル変換後の周波数軸上の値の高い部分のケプストラムが除去される。

【0026】

ケプストラムフィルタ部 40 はさらに、第 2 のメルワーピング部 76 の出力に対して FFT 処理を行なうための FFT 処理部 78 と、FFT 処理部 78 の出力に指数変換を行なうための指数変換部 80 と、指数変換部 80 の出力に対し IFFT 処理を行なってフィルタパラメータを出力するための IFFT 処理部 82 とを含む。IFFT 処理部 82 の出力は、第 1 のメルワーピング部 74 及び第 2 のメルワーピング部 76 によるメルワーピングとその逆変換とにより、除算部 38 の出力の PSD が平滑化されたものとなる。

【0027】

ケプストラムフィルタ部 40 はさらに、IFFT 処理部 82 の出力するフィルタパラメータにより設定され、ソース音声 30 に対しフィルタ処理を行なうことにより、ソース音声 30 の周波数特性を補正し、ターゲット音声 32 の周波数特性とほぼ同じ周波数特性の音声 42 として出力するための FIR 84 を含む。

【0028】

チャンネル等価装置 20 は以下の様に動作する。図 2 を参照して、ソース音声 30 の波形 90 に対して図 1 に示す音声フレーム検出部 50、窓掛け処理部 52、パワースペクトル算出部 54、及びフレーム平均部 56 によってソース PSD 100 が得られる。同様にターゲット音声 32 の波形 92 に対して音声フレーム検出部 60、窓掛け処理部 62、パワースペクトル算出部 64、及びフレーム平均部 66 によってターゲット PSD 102 が得られる。除算部 38 が後者を前者で除算することにより、PSD の差分 110 が得られる。

【0029】

この PSD の差分 110 に対し、対数算出部 70、IFFT 処理部 72、及び第 1 のメルワーピング部 74 での処理を行なうことにより、メルケプストラム 120 が得られる。このメルケプストラム 120 に対し、第 2 のメルワーピング部 76、FFT 処理部 78、及び指数変換部 80 での処理を実行することにより、PSD の差分 110 の平滑化された PSD 130 が得られる。この平滑化された PSD 130 に対し IFFT 処理部 82 の処理を行なうことにより、FIR 84 のフィルタパラメータを設定する。この様に設定された FIR 84 を用いてソース音声 30 をフィルタリングすることにより得られる音声 42 の周波数特性は、ターゲット音声 32 の周波数特性とほぼ等しいものとなる。

【0030】

メルワーピングによってケプストラムを一旦メルスケールに変換した後、その逆変換によってその高周波数成分を除去することで、PSD の差分を平滑化している。従ってこうして得られたフィルタパラメータにより設定された FIR 84 は人間の聴覚特性に近い特

10

20

30

40

50

性を持つ。さらに、この様に設定されるFIR84の特性は、LPC変換によるフィルタと異なり、パラメータ次数に敏感でない。フィルタの精度を高める様にフィルタパラメータの算出を行なった場合にも音質の劣化が生じることがなく、従来のチャンネル等価装置200と同程度の音質でソース音声30の周波数特性の等化を行なうことができる。

【0031】

[第2の実施の形態]

上記した第1の実施の形態の装置では、フィルタリングはFIRで行なっている。しかし本発明はその様にFIRを用いるものには限定されない。図1に示す第1のメルワーピング部74の出力するメルケプストラム係数で直接設定できるフィルタを使用する場合、構成はより簡単となる。図3にそうしたフィルタとしてMLSA (mel-logarithmic spectral approximation) フィルタを用いた、本発明の第2の実施の形態に係るチャンネル等価装置140のブロック図を示す。

10

【0032】

図3において、図1及び図4と同一部品には同一の参照符号を付してある。それらの機能も同一である。従ってここではそれらについての詳細な説明は繰返さない。

【0033】

図3を参照して、この第3の実施の形態に係るチャンネル等価装置140が図1に示すチャンネル等価装置20と異なるのは、図1に示すケプストラムフィルタ部40に代えて、MLSAフィルタを含むMLSAフィルタ部150を含む点である。そしてMLSAフィルタ部150が図1のケプストラムフィルタ部40と異なるのは、図1の第2のメルワーピング部76、FFT処理部78、指数変換部80、及びIFFT処理部82に代えて、第2のメルワーピング部76から出力されるメルケプストラムによって直接にフィルタパラメータが設定されるMLSAフィルタ160を含む点である。

20

【0034】

この第2の実施の形態に係るチャンネル等価装置140の動作は、第1のメルワーピング部74の出力によってMLSAフィルタ160が設定される点を除き、第1の実施の形態のチャンネル等価装置20と同じである。

【0035】

チャンネル等価装置140によっても、第1の実施の形態のチャンネル等価装置20と同様の効果を得ることができる。それに加えて、MLSAフィルタ160は、メルケプストラムをパラメータとするフィルタであり、第1のメルワーピング部74の出力によって直接設定できる。従って、第1の実施の形態のチャンネル等価装置20と比較して、メルケプストラムからFIRのフィルタパラメータを作成するための種々の部品が不要となり、回路構成が簡単となる。

30

【0036】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内のすべての変更を含む。

【図面の簡単な説明】

40

【0037】

【図1】第1の実施の形態に係るチャンネル等価装置のブロック図である。

【図2】第1の実施の形態に係るチャンネル等価装置の動作を説明するための模式図である。

【図3】第2の実施の形態に係るチャンネル等価装置のブロック図である。

【図4】従来技術に係るチャンネル等価装置のブロック図である。

【符号の説明】

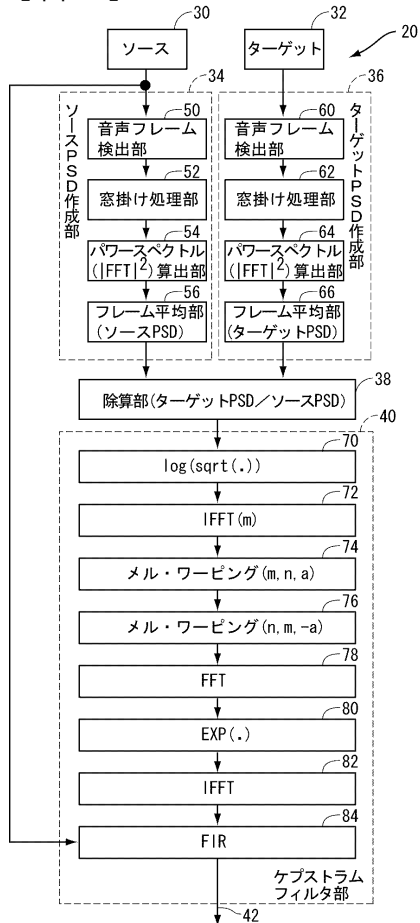
【0038】

20, 140, 200 チャンネル等価装置、30 ソース音声、32 ターゲット音声、34 ソースPSD生成部、36 ターゲットPSD生成部、38 除算部、40 ケ

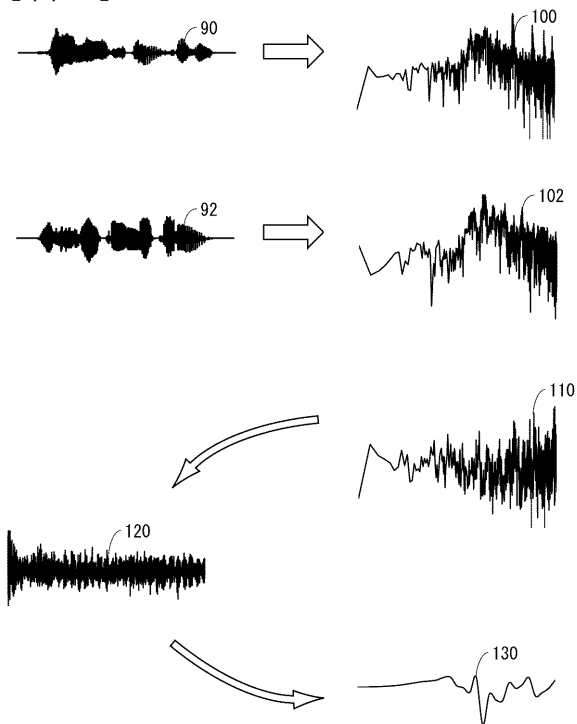
50

プストラムフィルタ部、50, 60 音声フレーム検出部、52, 62 窓掛け処理部、54, 64 パワースペクトル算出部、56, 66 フレーム平均部、74 第1のメルワーピング部、76 第2のメルワーピング部、84 FIR、150 MLSAフィルタ部、160 MLSAフィルタ、210 LPCフィルタ部

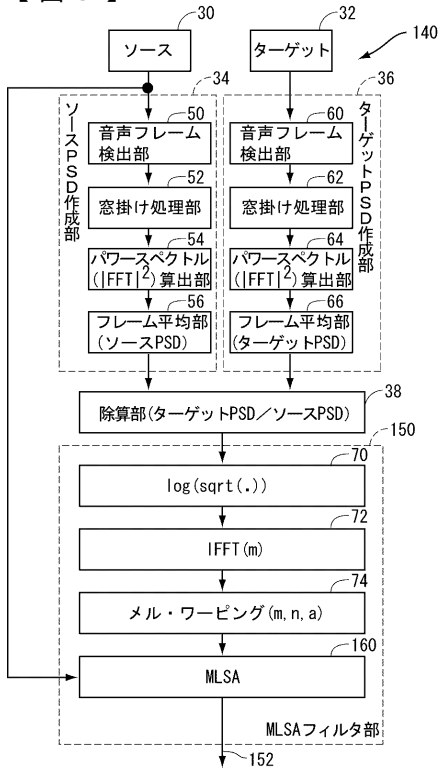
【図1】



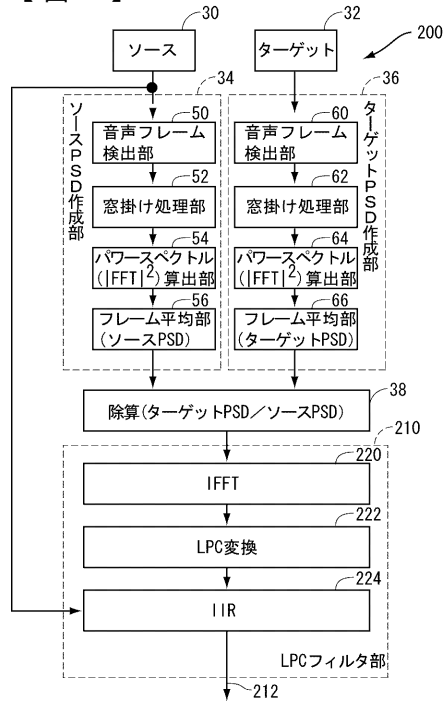
【図2】



【図3】



【図4】



フロントページの続き

審査官 荏原 雄一

(56)参考文献 特開昭62-138900(JP,A)

(58)調査した分野(Int.Cl.,DB名)

G10L 21/04

G10L 11/00