

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5115818号
(P5115818)

(45) 発行日 平成25年1月9日(2013.1.9)

(24) 登録日 平成24年10月26日(2012.10.26)

(51) Int.Cl. F I
G 1 O L 21/034 (2013.01) G 1 O L 21/02 3 O 1 B
G 1 O L 21/0364 (2013.01) G 1 O L 21/02 3 O 2 B

請求項の数 3 (全 15 頁)

<p>(21) 出願番号 特願2008-263472 (P2008-263472) (22) 出願日 平成20年10月10日 (2008.10.10) (65) 公開番号 特開2010-91897 (P2010-91897A) (43) 公開日 平成22年4月22日 (2010.4.22) 審査請求日 平成23年10月6日 (2011.10.6)</p>	<p>(73) 特許権者 504145342 国立大学法人九州大学 福岡県福岡市東区箱崎六丁目10番1号 (74) 代理人 100085257 弁理士 小山 有 (72) 発明者 中島 祥好 福岡県福岡市東区箱崎6-10-1 国立 大学法人九州大学内 (72) 発明者 上田 和夫 福岡県福岡市東区箱崎6-10-1 国立 大学法人九州大学内 審査官 井上 健一</p>
--	---

最終頁に続く

(54) 【発明の名称】 音声信号強調装置

(57) 【特許請求の範囲】

【請求項1】

入力された音声信号を複数の周波数帯域に分割する帯域分割部と、前記帯域分割部で分割されたそれぞれの周波数帯域内の信号を複数の時間フレームに分割する時間フレーム分割部と、前記時間フレーム分割部で分割されたそれぞれの時間フレーム内の平均パワーを算出するパワー算出部と、前記パワー算出部で算出されたそれぞれの時間フレーム内の平均パワーを互いに比較する比較部と、前記比較部の比較結果に基づいて前記時間フレーム分割部で分割されたそれぞれの信号の増幅度を決定する増幅度決定部と、前記時間フレーム分割部で分割されたそれぞれの信号を前記増幅度決定部で決定された増幅度で増幅する増幅部と、前記増幅部で増幅されたそれぞれの周波数帯域内の信号を加算する加算部を備えたことを特徴とする音声信号強調装置。

【請求項2】

請求項1に記載の音声信号強調装置において、前記比較部の出力がパワーの増加を示した場合には前記増幅度決定部が増幅度を増すと共に、以降の時間フレーム内の信号に対する増幅度を減ずることを特徴とする音声信号強調装置。

【請求項3】

入力された音声信号を複数の周波数帯域に分割する第1の帯域分割部と、前記第1の帯域分割部で分割されたそれぞれの周波数帯域内の信号のパワーエンベロープ信号を抽出するパワーエンベロープ抽出部と、前記入力された音声信号のゼロクロス波を生成するゼロクロス波生成部と、前記ゼロクロス波生成部で生成されたゼロクロス波を複数の周波数帯

域に分割する第2の帯域分割部と、前記パワーエンベロープ抽出部で抽出されたそれぞれの帯域のパワーエンベロープと、前記第2の帯域分割部で分割されたゼロクロス波のそれぞれの周波数帯域内の信号を乗算する乗算部と、前記乗算部で乗算されたそれぞれの周波数帯域内の信号を加算する加算部を備えたことを特徴とする音声信号強調装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、公共音響設備、メガホン、インターフォン、電話、放送、音声ガイド装置などで、残響や背景騒音があっても明瞭な音声を提供するための音声信号強調装置に関する。

10

【背景技術】

【0002】

病院内でスピーカーから流れる患者の呼び出し音声や、駅構内で流される発車番線や行く先を知らせるアナウンス、イベント会場などでメガホン（ハンドスピーカー）を通して流される様々な情報などの、各種音声伝達装置から伝えられる様々な音声情報を正確に聞き取ることは、現代社会において文化的な生活を営むために必要不可欠なものとなっている。

【0003】

病院、駅、イベント会場などの公共空間は、利用者の利便性を考慮しつつも、省スペース、低コストを意識して設計されていることは言うまでもない。さらに昨今はデザイン性も重視されるようになってきているため、狭い空間内に複雑な構造の壁面が多数存在し、そのような狭小、複雑な空間内に大勢の利用者が存在するような状況が散見される。

20

【0004】

このような空間内で音を出すと、壁や天井にぶつかって音が反射し、原音（出した音）と反射音が重なる。音が飛び回り「ワーン」と長く残ってしまい、さらに原音と反射音（1次、2次、・・・）が重なってしまう、いわゆる“残響”という現象が発生する。

【0005】

大勢の人が集まる公共空間は、人のざわめき声やBGM（バックグラウンドミュージック）など、元々が背景騒音の多い場所である。その中で利用者に的確に情報を提供するために、放送装置やメガホンなどを用いて、大音量で繰り返しアナウンスが流されるわけであるが、その場合に多量の残響が発生し、「音は聞こえるが、うるさいばかりで何を言っているのかさっぱりわからない」といった不快感を感じる人は多い。

30

【0006】

一方、携帯電話などの携帯型音声通信機器においても、話者もしくは聴取者が存在する環境に残響や背景騒音が存在すれば、不快感が伴い、会話が困難になることは言うまでもない。この場合は特に通信を行うので、情報伝送量をできる限り少なくした上で明瞭な音声を提供することが求められる。

【0007】

特に、聴力が低下してきた高齢者では、このような不快感はさらに大きく、場合によっては気分が悪くなってしまうケースがあることも知られており、残響や背景騒音の多い空間においても、音量を上げずとも利用者に正確に音声情報を伝達できる手段が求められていることは言うまでもないことである。

40

【0008】

特許文献1には、入力された音声信号から複数の時間フレームによってそれぞれでフレーム信号を抽出するフレーム分割部と、フレーム信号のそれぞれで平均パワーまたは音圧レベルを算出するパワー算出部と、フレーム信号間で平均パワーまたは音圧レベルを互いに比較する比較部と、比較部の比較結果に基づいて音声信号が子音であるか否かを判定する子音判定部と、子音判定部が子音と判断した場合は音声信号の増幅対象点または増幅対象幅を増幅すると共に、子音または音節の端点でないと判断した場合は増幅しない増幅部とを備えたことを主要な特徴とする子音加工装置、音声情報伝達装置及び子音加工方法に

50

関する記載がある。

【 0 0 0 9 】

特許文献 2 には、信号の第 1 の周波数帯域内の第 1 の残響特性を識別するように動作する信号分析論理回路と、信号分析論理回路にตอบสนองし、第 1 の周波数帯域内の該信号を減衰するように動作可能な減衰論理回路とを備え、入力信号の周波数帯域を分析し、残響が検出された場合、残響を低減し削除するために、残響周波数帯域を減衰させ得ることを特徴とする残響評価及び抑制システムに関する記載がある。

【 0 0 1 0 】

特許文献 3 には、複数のマイクロホンを備えたマイクロホンアレーと、マイクロホンアレーによって得られる複数のマイクロホン信号から、目的の音声信号が強調された信号を生成する適応ビームフォーマと、適応ビームフォーマの出力信号上の雑音を抑圧する雑音低減装置とを備えており、適応ビームフォーマとして、固定ビームフォーマ、適応ブロッキング行列および適応外乱キャンセラを備え、固定ビームフォーマおよび適応外乱キャンセラが入力信号の S N R に応じて適応制御されるロバスト一般化サイドローブ・キャンセラが用いられており、雑音低減装置として、GMM に基づくウィナーフィルタを用いて、雑音を抑圧する単一チャンネル雑音低減装置が用いられていることを特徴とする音声強調装置に関する記載がある。

【特許文献 1】特開 2 0 0 7 - 2 1 9 1 8 8

【特許文献 2】特開 2 0 0 6 - 1 5 7 9 2 0

【特許文献 3】特開 2 0 0 7 - 9 3 6 3 0

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 1 】

病院、駅、イベント会場など、大勢の人が集まる公共空間において、残響や背景騒音の影響を受けず、さらに音量を上げずともアナウンス等の音声情報を提供する技術が求められている。このような技術においては、様々な施設にローコストで手軽に設置できる必要がある。

【 0 0 1 2 】

特許文献 1 には、子音加工装置、音声情報伝達装置及び子音加工方法に関する記載がある。この方法は音声の子音部分のみを抽出、強調する方法であるので、騒音が多い場所などでも音量を上げずに明瞭な音声を提供できる利点があるものの、残響の多い空間では、その効力を十分に発揮できないという問題があった。

【 0 0 1 3 】

特許文献 2 には、残響評価及び抑制システムに関する記載がある。この方法は、対象となる室内の残響特性を事前に評価し、残響が強いと評価された周波数帯域の音成分を減衰させるものであるが、残響の評価を常にし続けねばならず、また評価に誤りがあると対象音声の音質を劣化させてしまうという問題があった。

【 0 0 1 4 】

特許文献 3 には、マイクロホンアレーを用いた音声強調装置に記載がある。この方法は、様々な背景騒音の中から目的音声を高精度に抽出できる利点があるものの、複数のマイクロホンを要する上に複雑な演算が必要となるためにシステムが大型化し、一般的な公共空間に設置するにはコスト面、技術面で困難さが伴うという問題があった。

【課題を解決するための手段】

【 0 0 1 5 】

上記の課題を解決するために、本発明は、公共音響設備、メガホン、インターフォン、電話、放送、音声ガイド装置などで、残響や背景騒音があっても明瞭な音声を提供するための音声信号強調装置に関して、以下の構成とした。

【 0 0 1 6 】

入力された音声信号を複数の周波数帯域に分割する帯域分割部と、前記帯域分割部で分割されたそれぞれの周波数帯域内の信号を複数の時間フレームに分割する時間フレーム分

10

20

30

40

50

割部と、前記時間フレーム分割部で分割されたそれぞれの時間フレーム内の平均パワーを算出するパワー算出部と、前記パワー算出部で算出されたそれぞれの時間フレーム内の平均パワーを互いに比較する比較部と、前記比較部の比較結果に基づいて前記時間フレーム分割部で分割されたそれぞれの信号の増幅度を決定する増幅度決定部と、前記時間フレーム分割部で分割されたそれぞれの信号を前記増幅度決定部で決定された増幅度で増幅する増幅部と、前記増幅部で増幅されたそれぞれの周波数帯域内の信号を加算する加算部を備える構成とした。これにより、残響や背景騒音が存在する公共空間においても、音量を上げることなく、ローコストで技術的困難さを伴うことなく明瞭で自然な音声を提供することが可能となる。

【0017】

また、前記比較部の出力がパワーの増加を示した場合には前記増幅度決定部が増幅度を増すと共に、以降の時間フレーム内の信号に対する増幅度を減ずることを特徴とする構成とした。これにより、残響に対してより頑健となり、ローコストで技術的困難さを伴うことなく明瞭で自然な音声を提供することが可能となる。

【0018】

また、入力された音声信号を複数の周波数帯域に分割する第1の帯域分割部と、前記第1の帯域分割部で分割されたそれぞれの周波数帯域内の信号のパワーエンベロープ信号を抽出するパワーエンベロープ抽出部と、前記入力された音声信号のゼロクロス波を生成するゼロクロス波生成部と、前記ゼロクロス波生成部で生成されたゼロクロス波を複数の周波数帯域に分割する第2の帯域分割部と、前記パワーエンベロープ抽出部で抽出されたそれぞれの帯域のパワーエンベロープと、前記第2の帯域分割部で分割されたゼロクロス波のそれぞれの周波数帯域内の信号を乗算する乗算部と、前記乗算部で乗算されたそれぞれの周波数帯域内の信号を加算する加算部を備えることを特徴とする構成とした。これにより、情報伝送量をより少なくした上で、ローコストで技術的困難さを伴うことなく明瞭で自然な音声を提供することが可能となる。

【発明の効果】

【0019】

本発明の音声信号強調装置を用いれば、残響が存在することによって従来ではかき消されていた音声のスペクトル変化を、残響下でも十分に聞き取れるようになる。

【0020】

人間が音声内容を理解するためには、音声中に含まれる音節の端点（子音および母音の始まりないし終わりの部分）が重要な役割を担っていることが知られている。

【0021】

この音節の端点では、音が物理的には弱い場合が多く、騒音にかき消される可能性が高い。残響の多い場所では、母音定常部の残響が音節の端点をかき消すこともありえるわけであるが、本発明によって端点を強調（さらに端点以外の部分を抑制）することによって、この問題は解決され、明瞭な音声を提供することが可能となる。

【0022】

特に、最近の聴覚心理学分野の研究により、ヒトが音声を聴取する際には、複数の周波数帯域のパワーの時間的な変化を重要な情報源としていることが明らかになってきている。よって、本発明によって入力音声を複数の周波数帯域に分割し、それぞれの帯域内の信号の音節の端点を強調（さらに端点以外の部分を抑制）することによって、明瞭な音声を提供することが可能となるのである。

【0023】

さらに、本発明は、周波数帯域分割と時間フレーム分割以外の演算は、少量の基本的な四則演算のみで構成されており、極めて小規模なシステム構成が実現可能である。携帯電話などの通信機器に搭載する際には、通常よりも情報伝送量を抑えることも可能であり、極めて汎用性の高い技術であると言える。

【発明を実施するための最良の形態】

【0024】

10

20

30

40

50

以下、本発明を実施するための最良の形態を図面に基づいて詳細に説明する。なお、以下の説明において、同一機能を有するものは同一の符号とし、その繰り返しの説明は省略する。

【0025】

図1は、本発明の第1の実施形態におけるシステムのブロック図であり、入力された音声信号を複数の周波数帯域に分割する帯域分割部1と、前記帯域分割部1で分割されたそれぞれの周波数帯域内の信号を複数の時間フレームに分割する時間フレーム分割部2と、前記時間フレーム分割部2で分割されたそれぞれの時間フレーム内の平均パワーを算出するパワー算出部3と、前記パワー算出部3で算出されたそれぞれの時間フレーム内の平均パワーを互いに比較する比較部4と、前記比較部4の比較結果に基づいて前記時間フレーム分割部2で分割されたそれぞれの信号の増幅度を決定する増幅度決定部5と、前記時間フレーム分割部2で分割されたそれぞれの信号を前記増幅度決定部5で決定された増幅度で増幅する増幅部6と、前記増幅部6で増幅されたそれぞれの周波数帯域内の信号を加算する加算部7から構成されている。

10

【0026】

図2を用いて、帯域分割部1および時間フレーム分割部2の動作を、さらに詳細に説明する。ここでは、帯域分割部1は1つの低域通過フィルタと3つの帯域通過フィルタで構成されており、その通過周波数帯域は、(1) 600 Hz 以下、(2) 600-1800 Hz、(3) 1800-3400 Hz、(4) 3400-8000 Hzの4帯域となっている。これは、各国語の音声の分析結果から、音声コミュニケーションの基本に関わると考えられている4帯域である。

20

【0027】

各周波数帯域の信号(時間波形)を、それぞれ、 $x_{600}[t]$ 、 $x_{1800}[t]$ 、 $x_{3400}[t]$ 、 $x_{8000}[t]$ とし、時間フレーム分割部2にて、これらを時間フレームに分割する。

【0028】

図2では、2種類の時間フレーム(30msと120ms)で分割し、フレームの重なり合いはないものとしている。当然のことながら、重なり合いを持たせ、その重なり合いの部分を長くすれば、本発明の音声信号強調の時間分解能が高精度になる。

【0029】

パワー算出部3では、時間フレーム分割部2で分割されたフレーム内の平均パワーを求める。例えば、(1) 600 Hz 以下の帯域の出力を30msの時間フレームで分割した際の平均パワー $I_{n_30_600}[T]$ は、次式から求められる。

30

【0030】

【数1】

$$I_{n_30_600}[T] = \frac{\sum_{t=0}^N \{x_{600}[t]\}^2}{N}$$

(Nは30msのサンプル数で、例えばサンプリング周波数が16kHzであれば480)

同様にして、120msの時間フレームで分割した際の平均パワー $I_{n_120_600}[T]$ も求め、比較部4によって両者を比較する。具体的には、音の強さが120msの範囲で局所的に増している($I_{n_30_600}(T) > I_{n_120_600}(T)$)か、減じている($I_{n_30_600}(T) < I_{n_120_600}(T)$)かを判定し、増幅度決定部5において、前者であればパワーを一層増すことにより、音の強さの時間変化を強調する。さらに加えて、後者であれば音の強さを一層減ずることにより、音の強さの時間変化はさらに強調される。例えば時間波形 $x_{600}[t]$ に対して、増幅度決定部5において(数2)のような数式で増幅度 $v_{600}[t]$ を決定し、増幅部6によって、(数3)のように増幅を行い、増幅波形 $p_{600}[t]$ を得る。

40

【0031】

【数 2】

$$v_{600}[t] = (In_{30_600}(t) / In_{120_600}(t))^{1.5}$$

【0032】

【数 3】

$$p_{600}[t] = x_{600}(t) \times v_{600}[t]$$

$x_{1800}[t]$ 、 $x_{3400}[t]$ 、 $x_{8000}[t]$ に関しても同様の処理を行い、それぞれの出力波形 $p_{1800}[t]$ 、 $p_{3400}[t]$ 、 $p_{8000}[t]$ を得た後に、加算部 7 において、本発明の出力波形 $y[t]$ を得る。 10

【0033】

【数 4】

$$y[t] = p_{600}[t] + p_{1800}[t] + p_{3400}[t] + p_{8000}[t]$$

なお、本発明において、時間フレーム分割部 2 で分割する時間フレームは、時刻 t を中心として対称になっている必要はなく、例えば、30ms の時間フレームを、 $t - 15 \sim t + 15$ ms、120ms の時間フレームを $t - 90 \sim t + 30$ ms と配置しても良い。この場合は、時刻 t における増幅度を決定する際に、必要となる未来方向の音情報の量が制限されるので、実時間信号処理における遅れ時間を最小限に止めることが可能となり、これはまた残響の影響を少なくするのに好都合となる。 20

【0034】

また、本例における時間フレームの分割では、時間波形を矩形の時間窓で切り出していることになっているが、当然のことながら、ガウス形、指数関数形などの形状の時間窓関数を乗じて切り出す（時間フレーム分割する）ことも可能である。

【0035】

図 3 には、本発明を用いて作成した強調音声の一例を示す。図の上段は、音声（発話内容/ASA/）の時間波形、下段はサウンドスペクトログラム（横軸が時間、縦軸が周波数で、エネルギーの強弱を色の濃淡で示している）である。一番左が原音声であり、その隣に 4 帯域に分割した時間波形が並べて示されている。4 帯域に分割された音声の時間的コントラストがそれぞれ強調されて、最終的にはすべて合成（加算）されて出力されている様子が見える。 30

【0036】

図 3 で適用された強調処理の様子を図 4 に示す。600 ~ 1800Hz の帯域では、低 ~ 中周波数帯域に主要な成分を有する/ASA/の母音/A/の部分が強調され、3400Hz ~ 8000Hz の帯域では高周波数帯域に主要な成分を有する無声子音の/S/が特に強調されている。強調処理においては、各音節の端点では増幅度が増し、以降は増幅度が減らされている。

【0037】

図 5 は、本発明の第 2 の実施形態におけるシステムのブロック図であり、入力された音声信号を複数の周波数帯域に分割する第 1 の帯域分割部 8 と、前記第 1 の帯域分割部 8 で分割されたそれぞれの周波数帯域内の信号のパワーエンベロープ信号を抽出するパワーエンベロープ抽出部 9 と、前記入力された音声信号のゼロクロス波を生成するゼロクロス波生成部 10 と、前記ゼロクロス波生成部 10 で生成されたゼロクロス波を複数の周波数帯域に分割する第 2 の帯域分割部 11 と、前記パワーエンベロープ抽出部 9 で抽出されたそれぞれの帯域のパワーエンベロープと、前記第 2 の帯域分割部 11 で分割されたゼロクロス波のそれぞれの周波数帯域内の信号を乗算する乗算部 12 と、前記乗算部 12 で乗算されたそれぞれの周波数帯域内の信号を加算する加算部 13 から構成されている。 40

【0038】

図 6 を用いて、本発明の第 2 の実施形態の動作を、さらに詳細に説明する。ここでは、 50

第1の帯域分割部8は4つの低域通過フィルタおよび帯域通過フィルタで構成されており、その通過周波数帯域は、(1) 600 Hz 以下、(2) 600-1800 Hz、(3) 1800-3400 Hz、(4) 3400-8000 Hzの4帯域となっている。これは、各国語の音声の分析結果から、音声コミュニケーションの基本に関わると考えられている4帯域である。

【0039】

パワーエンベロープ抽出部9は、入力音声のパワーエンベロープを抽出する。ここでは、このパワーエンベロープを1 ms の時間フレーム内(時間窓内)の平均パワーとして、例えば(数1)と同様の演算により算出し、それを時間軸上にプロットしている。

【0040】

一方、ゼロクロス波生成部10は、入力音声のゼロクロス波を抽出する。ここでゼロクロス波とは、時間波形の瞬時振幅値が正なら+1、ゼロなら0、負なら-1の符号に変換した波形である。

【0041】

第2の帯域分割部11は、ゼロクロス波生成部10で生成されたゼロクロス波を複数の周波数帯域に分割する。なお、本実施例では、第1の帯域分割部8と同様のフィルタ群によって分割を行っている。

【0042】

パワーエンベロープ抽出部9と第2の帯域分割部11の出力は、乗算部12で互いに対応する周波数帯域の出力同士が乗算され、加算部13にて全ての帯域の出力が加算され出力される。

【0043】

ここで、第2の帯域分割部11で帯域分割された出力は、第2の帯域分割部11における低域通過フィルタおよび帯域通過フィルタの作用によって、そのパワーに時間的な変化が生ずる場合がある。この場合は、各フィルタの出力の短時間平均パワーを求めた後に、そのパワーが一定値になるように出力波形を増幅もしくは減衰するような処理を加えて、各帯域の出力のパワーを一定値にすれば、より効果的な出力音声を得られる。

【0044】

ゼロクロス波で音声のピッチの有無、および、ピッチがある場合はその変化が、パワーエンベロープで音声の強弱の情報が伝わるので、両者だけで言語の内容は完全に伝わる。ゼロクロス波の情報量は1ビットであり、パワーエンベロープは時間フレーム内の平均パワーであるので、本発明によれば、言語の内容が完全に伝わった上で、情報量は原音声の15分の1程度に圧縮されることとなる。

【0045】

さらに、本実施例の帯域で周波数分割を行えば、1オクターブ以上の周波数帯域が同時に強度変化を示す。ここには信号の冗長性があり、結果として耐雑音性が強くなる(背景騒音の中でも聞き取りやすくなる)ので、劣悪な騒音環境下で音声を伝える必要がある場合に特に有効である。

【0046】

なお、本実施例における実施の形態1と実施の形態2は縦続に接続して使用することが可能である。この場合は、残響に頑健で背景騒音にも強く、さらに情報量が1/15程度の音声生成が可能となる。

【図面の簡単な説明】

【0047】

【図1】本発明の第1の実施形態におけるシステムのブロック図

【図2】本発明の第1の実施形態における帯域分割部1および時間フレーム分割部2の詳細動作図

【図3】本発明の第1の実施形態を用いて作成した強調音声の一例

【図4】図3で適用された強調処理の様子

【図5】本発明の第2の実施形態におけるシステムのブロック図

【図6】本発明の第2の実施形態の詳細動作図

10

20

30

40

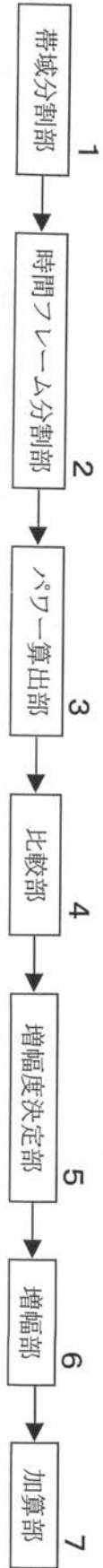
50

【符号の説明】

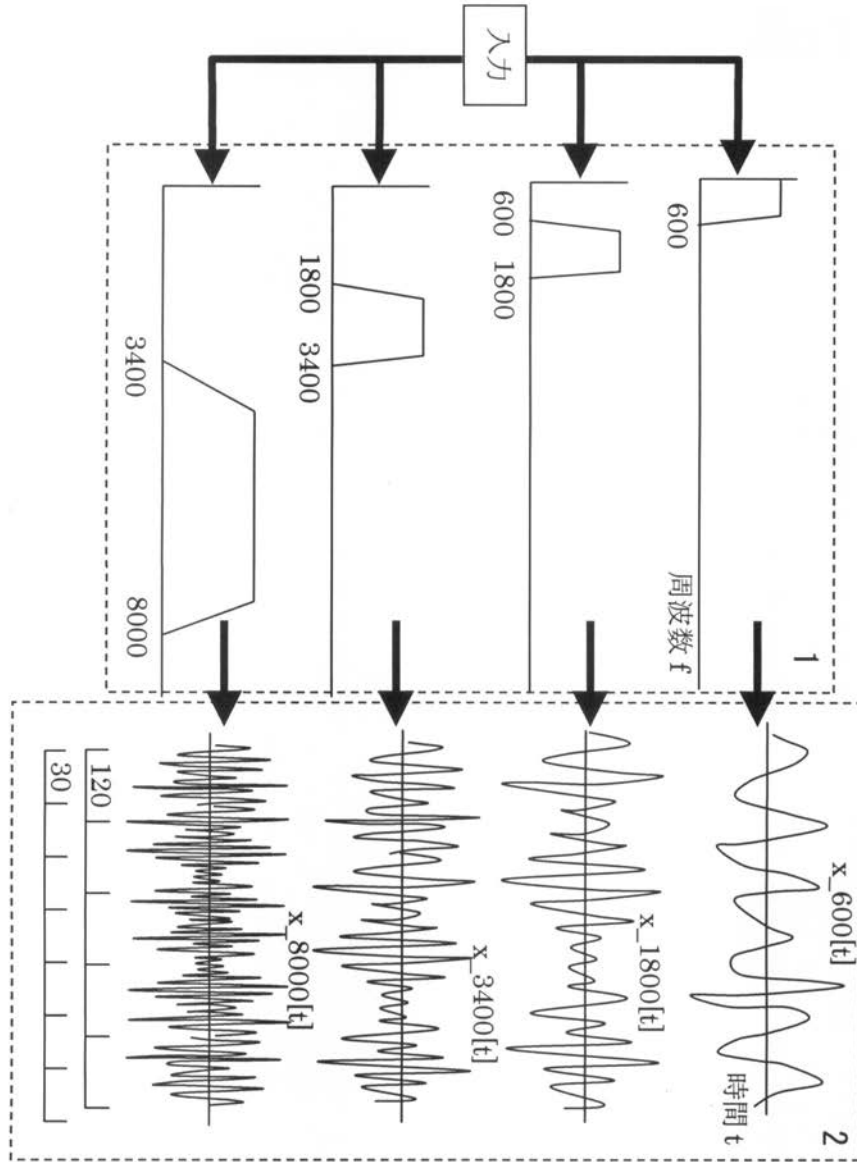
【0048】

1 ... 帯域分割部、 2 ... 時間フレーム分割部、 3 ... パワー算出部、 4 ... 比較部、
5 ... 増幅度決定部、 6 ... 増幅部、 7 ... 加算部、 8 ... 第1の帯域分割部、 9 ... パワーエンベロープ抽出部、 10 ... ゼロクロス波生成部、 11 ... 第2の帯域分割部、 12 ... 乗算部、 13 ... 加算部。

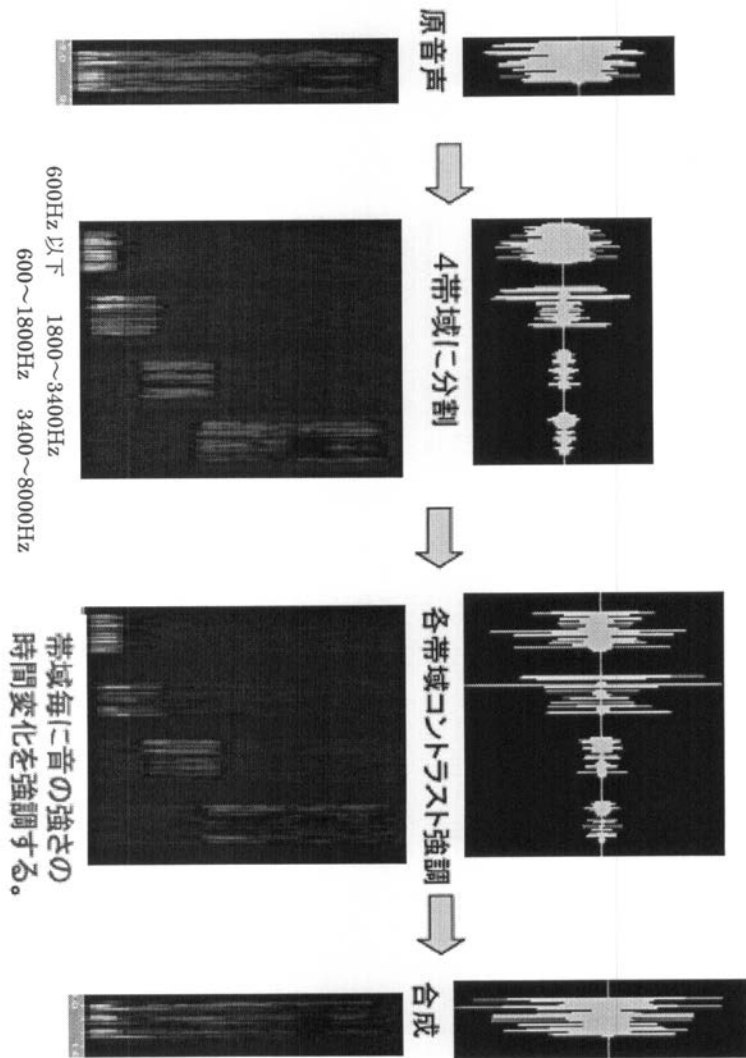
【図1】



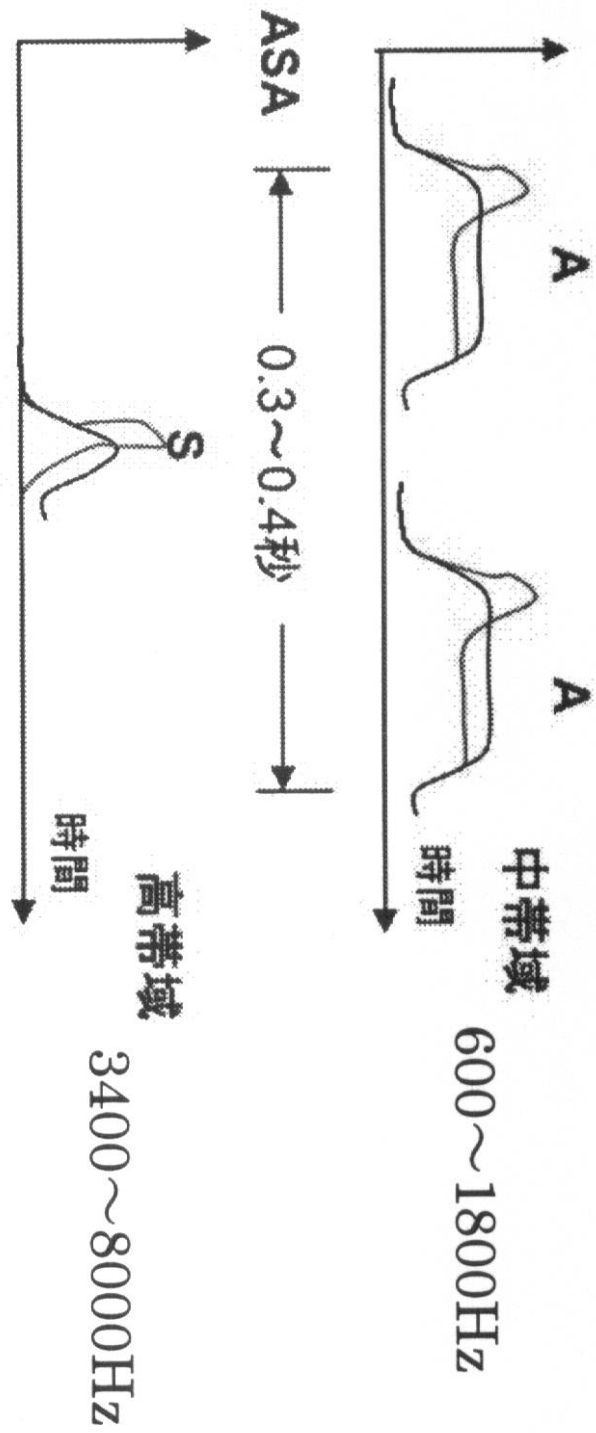
【図2】



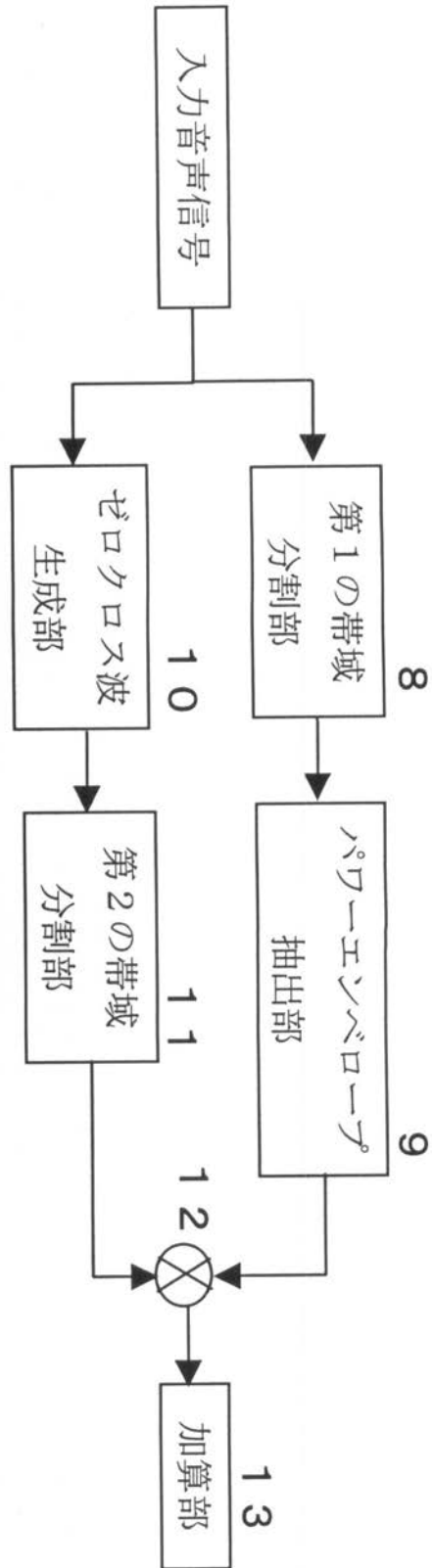
【 図 3 】



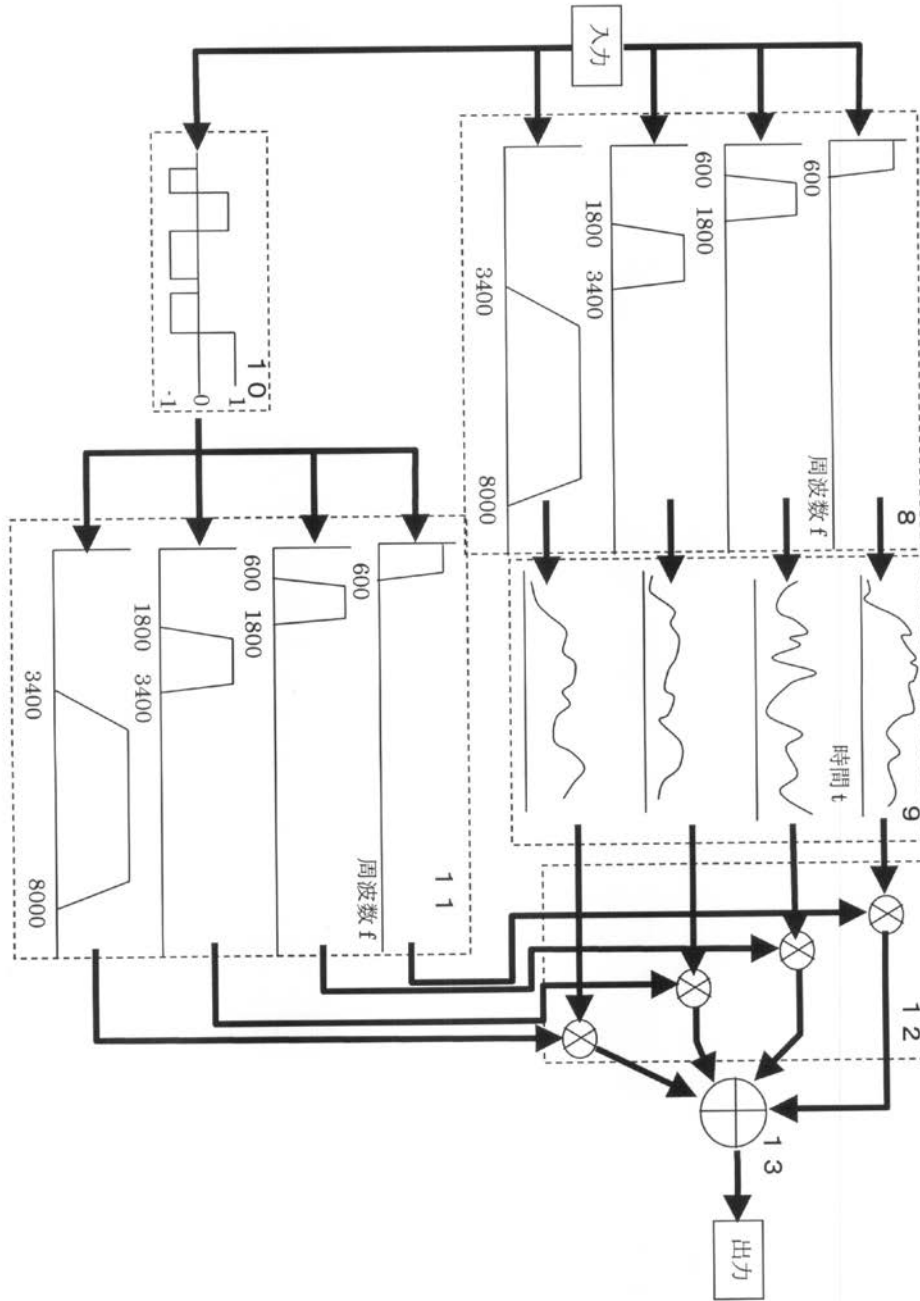
【 図 4 】



【図5】



【図6】



フロントページの続き

- (56)参考文献 特開2005-175674(JP,A)
特開2008-085520(JP,A)
特開2007-219188(JP,A)
特開2001-051687(JP,A)
特開2001-356800(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 21/02

G10L 19/02