

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4150795号
(P4150795)

(45) 発行日 平成20年9月17日(2008.9.17)

(24) 登録日 平成20年7月11日(2008.7.11)

(51) Int.Cl. F I
HO4R 25/00 (2006.01) HO4R 25/00 L
 HO4R 25/00 M

請求項の数 10 (全 27 頁)

(21) 出願番号 特願2005-14568 (P2005-14568)
 (22) 出願日 平成17年1月21日(2005.1.21)
 (65) 公開番号 特開2006-203683 (P2006-203683A)
 (43) 公開日 平成18年8月3日(2006.8.3)
 審査請求日 平成19年12月14日(2007.12.14)

(73) 特許権者 304020292
 国立大学法人徳島大学
 徳島県徳島市新蔵町2丁目2番地
 (74) 代理人 100104949
 弁理士 豊栖 康司
 (74) 代理人 100074354
 弁理士 豊栖 康弘
 (72) 発明者 赤松 則男
 徳島県徳島市南常三島町2丁目1番地 国立
 大学法人徳島大学工学部知能情報学科内
 審査官 志摩 兆一郎

最終頁に続く

(54) 【発明の名称】 聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器

(57) 【特許請求の範囲】

【請求項1】

音声信号を入力するための音声入力部と、

前記音声入力部で入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換するための凹凸波形変換部と、

前記凹凸波形変換部で得られた音声信号の凹凸波形から、予め登録された人の声に関する登録パターンに従い人の声に対応する声成分を抽出するための声成分抽出部と、

前記声成分抽出部で抽出された声成分の凸部を高く、凹部を低くすることで強調し、強調声波形を生成するための強調処理部と、

強調声波形を出力するための音声出力部と、

を備えることを特徴とする聴覚補助装置。

【請求項2】

請求項1に記載の聴覚補助装置であって、

前記凹凸波形変換部が、凹凸波形に変換するサンプリング個数を2のべき乗とし、前記声成分抽出部が、べき指数を調整することで声成分を抽出することを特徴とする聴覚補助装置。

【請求項3】

請求項1又は2に記載の聴覚補助装置であって、

10

20

前記声成分抽出部が、周波数成分を人の声の高域に対応する成分、低域に対応する成分に応じたべき指数をそれぞれ設定して声成分を抽出することを特徴とする聴覚補助装置。

【請求項 4】

請求項 1 から 3 のいずれかに記載の聴覚補助装置であって、

前記凹凸波形変換部が部分平均値を演算する際に、所定の範囲の加算すべきデータの個数を 2 のべき乗として、前記所定の範囲の加算したデータの個数で除算をビット・シフト演算で行うことを特徴とする聴覚補助装置。

【請求項 5】

請求項 4 に記載の聴覚補助装置であって、

前記凹凸波形変換部が部分平均値を演算する際に、一のデータにつき平均値を求めるために所定の範囲のデータの振幅値を加算した加算値を保持しておき、次のデータの加算値を求める際に、保持された加算値から、不要な振幅値を減算すると共に、必要な振幅値を加算することで、加算値を演算することを特徴とする聴覚補助装置。

10

【請求項 6】

請求項 5 に記載の聴覚補助装置であって、 k 点を中心とする前後 n の区間 $N (= 2n)$ における平均値 α_k を、

【数 1】

$$\alpha_k = \frac{\left(\sum_{i=k+1}^{n+k} x_i \right) + \left(\sum_{i=k-n}^{k-1} x_i \right)}{N}$$

20

として表現する際、平均値の演算において、平均値 α_k を、その前段の位置である ($k-1$) 点における平均値 α_{k-1} を用いて

【数 2】

$$\alpha_k = \alpha_{k-1} + (x_{k-1} + x_{n+k} - x_k - x_{k-1-n}) / N$$

で演算することを特徴とする聴覚補助装置。

30

【請求項 7】

請求項 1 から 6 のいずれかに記載の聴覚補助装置であって、

前記声成分抽出部で抽出された声成分につき、子音が認識されると前記音声出力部から出力される音量を大きくし、子音の後に母音が認識されると、母音から所定時間で音量増幅を解除することを特徴とする聴覚補助装置。

【請求項 8】

入力された音声信号に基づいて人の声を補正して出力する音声信号処理方法であって、音声信号を入力する工程と、

入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換すると共に、得られた音声信号の凹凸波形から人の声に対応する声成分を抽出する工程と、

40

抽出された声成分を強調して強調声波形を生成する工程と、

強調声波形を再生する工程と、

を有することを特徴とする音声処理方法。

【請求項 9】

入力された音声信号に基づいて人の声を補正して出力する音声信号処理プログラムであって、

音声信号を入力する機能と、

50

入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換すると共に、得られた音声信号の凹凸波形から人の声に対応する声成分を抽出する機能と、

抽出された声成分を強調して強調声波形を生成する機能と、

強調声波形を再生する機能と、

をコンピュータに実現させることを特徴とする音声処理プログラム。

【請求項 10】

請求項 9 に記載されるプログラムを格納したコンピュータで読み取り可能な記録媒体又は記録した機器。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば聴力の衰えた高齢者や、難聴者などの聴覚障害者の聴覚を補助する補聴器等の聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器に関する。

【背景技術】

【0002】

聴覚障害者に対し、高い生活環境を提供するために、その聴覚（聴力）を補助する装置として、補聴器がある。補聴器には、例えば小型マイク、増幅器、およびイヤホンからなるものがあるが、このような補聴器は、マイク（小型マイク）に入力された音を、単純に増幅して出力するだけであるため、その出力にはノイズが多く含まれ、さらには会話相手の声や注意すべき物音（重要な環境音）などが、そのノイズに埋もれてしまうこともあり、視覚障害者の聴覚を補助するのに充分とは言えなかった。

【0003】

そこで、人間の音声は、特定の周波数帯域（中音域）に局在していることを利用して、マイクに入力された音声を、中音域を抜き出すバンドパスフィルタを通してから増幅する補聴器が開発されている。しかしながら、このような補聴器でも、会話相手の音声や注意すべき物音などが、快適かつ明瞭に聞こえるとは言い難かった。

【0004】

一方、最近のデジタル信号処理デバイスの発達により、デジタル回路やプロセッサを超小型化することが可能になり、このような技術が、補聴器の分野にも応用されている。デジタル信号処理を応用した補聴器では、アナログ信号の音声信号を A/D 変換し、デジタル信号としてから、このデジタル信号に対し、デジタルフィルタによるフィルタリング、雑音除去、周波数空間処理などのデジタル信号処理を施すことにより、可聴性を高めるようになされている。

【0005】

ここで、図 16 は、従来の聴覚補助装置としての補聴器の一例の構成を示している。この補聴器においては、まずマイク 301 で、周囲の音声やその他の物音を拾い、これを電気信号に変換し、原音声信号 A11 として出力する。この原音声信号 A11 はアナログフィルタ 108 に供給され、ここでは、人間の音声の周波数分布が集中する中音域だけが通過され、他はカットされる。これにより、アナログフィルタ 108 からは、中音域音声信号 A12 が出力される。中音域音声信号 A12 は、A/D 変換器 109 に供給され、そこで A/D 変換され、これによりデジタル信号としての音声信号 A13 にされる。

【0006】

音声信号 A13 はメモリ 302 に供給され、一時記憶される。メモリ 302 は、信号バスを介してデジタルシグナルプロセッサ（DSP）303 に接続されており、この DSP 303 は、メモリ 302 に格納された音声信号に対して、例えばデジタルフィルタリングや、雑音除去、FFT（高速フーリエ変換）等の周波数成分分解処理や周波数空間処理等

10

20

30

40

50

を施す。このような信号処理が施された音声信号は、処理音声信号 A 1 5 として、メモリ 3 0 2 から D / A 変換器 1 1 7 に供給される。D / A 変換器 1 1 7 では、デジタル信号である処理音声信号 A 1 5 が D / A 変換され、アナログ音声信号 A 1 6 にされる。アナログ音声信号 A 1 6 は、増幅器 1 1 8 に供給されて増幅される。そして、増幅器 1 1 8 からは、増幅された音声信号 A 1 7 がイヤホン 3 0 4 に供給され、そこから出力される。以上のようにして、マイク 3 0 1 に入力された音が、使用者（視覚障害者）の耳に届く。

【 0 0 0 7 】

しかしながら、上述したような補聴器では、単一のマイク 3 0 1 に入力された音から人間の音声に相当すると考えられる周波数成分を取り出して、可聴性を高めるようになされているため、会話相手の音声と、そうでない他人の音声とがともに増幅され、使用者が聞こうとしている会話相手の音声聞き取り難くなる課題があった。さらに、人間の音声と、それと同じような周波数成分を有する外部の物音とも区別されずに増幅されるため、やはり使用者が聞こうとしている音が聞き取りにくい課題があった。また、例えば自動車のクラクションや、警報音、電話のベルなどは、生活上重要な環境音（重要音）であり、常時聞こえる状態にあることが望ましいが、上述した補聴器を使用した場合には、このような重要音を聞き逃すおそれもある。

【 0 0 0 8 】

このような課題に対して、特許文献 1 に示す聴覚補助装置が開発されている。この聴覚補助装置は、図 1 7 に示すように、無指向性マイク 1 0 2 L および 1 0 2 R に入力された環境音と、指向性マイク 1 0 7 に入力された会話相手の音声とが、プロセッサユニット 1 5 0 でそれぞれ独立に処理され、いずれか一方が増幅されて、イヤースピーカ 1 1 9 R および 1 1 9 L から出力されるよう構成され、会話相手の音声および注意すべき物音（重要音）が、快適かつ明瞭に聞こえるようにしている。

【 0 0 0 9 】

しかしながら、この補聴器では D S P 等の音声処理回路がフーリエ変換を行っているため、浮動小数点演算などの複雑な演算処理が必要となり、高速で高性能の演算回路が要求され、装置が大型で高価になるという課題があった。補聴器においてはリアルタイム処理が求められるため、高速な処理が可能なハードウェア仕様が要求される一方で、携帯可能な、可能な限り小型軽量化が望まれている。

【特許文献 1】特開平 8 - 7 9 8 9 7 号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 0 】

本発明は、このような問題点を解決するためになされたものである。本発明の主な目的は、音声処理において演算を簡素化し、高速かつ低負荷の処理とすることで実装や組み込みを容易にした聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器を提供することにある。

【課題を解決するための手段】

【 0 0 1 1 】

上記の目的を達成するために、本発明の第 1 の聴覚補助装置は、音声信号を入力するための音声入力部と、前記音声入力部で入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換するための凹凸波形変換部と、前記凹凸波形変換部で得られた音声信号の凹凸波形から、予め登録された人の声に関する登録パターンに従い人の声に対応する声成分を抽出するための声成分抽出部と、前記声成分抽出部で抽出された声成分の凸部を高く、凹部を低くすることで強調し、強調声波形を生成するための強調処理部と、強調声波形を出力するための音声出力部とを備える。これにより、フーリエ変換で音声信号を周波数スペクトルに変換することなく、音声信号から声成分を抽出でき、かつ聞き取りやすく補正できるので、ノイズの音量

を増やさない快適な聴覚補助装置が実現できる。

【0012】

さらにまた、本発明の第2の聴覚補助装置は、前記凹凸波形変換部が、凹凸波形に変換するサンプリング個数を2のべき乗とし、前記声成分抽出部が、べき指数を調整することで声成分を抽出する。これにより、声成分を抽出する処理をビットシフト演算により行うことができるので、処理を低負荷で行うことが可能となる。

【0013】

さらにまた、本発明の第3の聴覚補助装置は、前記声成分抽出部が、周波数成分を人の声の高域に対応する成分、低域に対応する成分に応じたべき指数をそれぞれ設定して声成分を抽出する。これにより、人の声が含まれない高域、低域をカットできるので、効果的に人の声に対応する声成分のみを抽出できる。

【0014】

さらにまた、本発明の第4の聴覚補助装置は、前記凹凸波形変換部が部分平均値を演算する際に、所定の範囲の加算すべきデータの個数を2のべき乗として、前記所定の範囲の加算したデータの個数で除算をビット・シフト演算で行う。これにより、除算の際にビット・シフト演算が可能となり、さらに演算処理を簡素化でき、高速化に寄与し得る。

【0015】

さらにまた、本発明の第5の聴覚補助装置は、前記凹凸波形変換部が部分平均値を演算する際に、一のデータにつき平均値を求めるために所定の範囲のデータの振幅値を加算した加算値を保持しておき、次のデータの加算値を求める際に、保持された加算値から、不要な振幅値を減算すると共に、必要な振幅値を加算することで、加算値を演算する。これにより、各平均値演算において加算値を求める際に、前回のデータについて演算した加算値を利用して、必要なデータの入れ替えによって所望の加算値とすることができ、加算演算を大幅に簡素化でき、演算処理をさらに高速化することが可能となる。

【0016】

さらにまた、本発明の第6の聴覚補助装置は、k点を中心とする前後nの区間N(=2n)における平均値 α_k を、

【0017】

【数3】

$$\alpha_k = \frac{\left(\sum_{i=k+1}^{n+k} x_i \right) + \left(\sum_{i=k-n}^{k-1} x_i \right)}{N}$$

として表現する際、平均値の演算において、平均値 α_k を、その前段の位置である(k-1)点における平均値を用いて

【0018】

【数4】

$$\alpha_k = \alpha_{k-1} + (x_{k-1} + x_{n+k} - x_k - x_{k-1-n}) / N$$

で演算する。これにより、各平均値を前段の平均値を利用して逐次的に求めることができ、演算処理量を大幅に低減して高速かつ低負荷な音声信号特徴量抽出処理を実現する。

【0019】

さらにまた、本発明の第7の聴覚補助装置は、前記声成分抽出部で抽出された声成分につき、子音が認識されると前記音声出力部から出力される音量を大きくし、子音の後に母音が認識されると、母音から所定時間で音量増幅を解除することを特徴とする聴覚補助装置。これにより、子音が聞き取りやすくでき、音量全体を上げずとも明瞭に音声を取り取りできる聴覚補助装置が実現できる。

10

20

30

40

50

【0020】

さらにまた、本発明の音声処理方法は、入力された音声信号に基づいて人の声を補正して出力する音声信号処理方法であって、音声信号を入力する工程と、入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換すると共に、得られた音声信号の凹凸波形から人の声に対応する声成分を抽出する工程と、抽出された声成分を強調して強調声波形を生成する工程と、強調声波形を再生する工程とを有する。これにより、フーリエ変換で音声信号を周波数スペクトルに変換することなく、音声信号から声成分を抽出でき、かつ聞き取りやすく補正できるので、ノイズの音量を増やさずに快適な音声再生が実現できる。

10

【0021】

さらにまた、本発明の音声信号特徴量抽出プログラムは、入力された音声信号に基づいて人の声を補正して出力する音声信号処理プログラムであって、音声信号を入力する機能と、入力された音声信号の振幅を量子化し、量子化された各点のデータにつき、隣接する所定の範囲のデータの振幅値を加算し、これを加算したデータの個数で除算して該データを中心とする部分平均値を求め、各点のデータの振幅値と各々の部分平均値とを比較し、比較結果の真偽に基づいて凹凸波形に変換すると共に、得られた音声信号の凹凸波形から人の声に対応する声成分を抽出する機能と、抽出された声成分を強調して強調声波形を生成する機能と、強調声波形を再生する機能とをコンピュータに実現させる。これにより、フーリエ変換で音声信号を周波数スペクトルに変換することなく、音声信号から声成分を抽出でき、かつ聞き取りやすく補正できるので、ノイズの音量を増やさずに快適な音声再生が実現できる。

20

【0022】

また本発明のコンピュータで読み取り可能な記録媒体又は記録した機器は、上記プログラムを格納するものである。記録媒体には、CD-ROM、CD-R、CD-RWやフレキシブルディスク、磁気テープ、MO、DVD-ROM、DVD-RAM、DVD-R、DVD+R、DVD-RW、DVD+RW、Blue-rayディスク、HDDVD(AOD)等の磁気ディスク、光ディスク、光磁気ディスク、半導体メモリその他のプログラムを格納可能な媒体が含まれる。またプログラムには、上記記録媒体に格納されて配布されるものの他、インターネット等のネットワーク回線を通じてダウンロードによって配布される形態のものも含まれる。さらに記録した機器には、上記プログラムがソフトウェアやファームウェア等の形態で実行可能な状態に実装された汎用もしくは専用機器を含む。さらにまたプログラムに含まれる各処理や機能は、コンピュータで実行可能なプログラムソフトウェアにより実行してもよいし、各部の処理を所定のゲートアレイ(FPGA、ASIC)等のハードウェア、又はプログラム・ソフトウェアとハードウェアの一部の要素を実現する部分的ハードウェア・モジュールとが混在する形式で実現してもよい。

30

【発明の効果】

【0023】

本発明の聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器によれば、人の声を聞き取りやすく強調して出力可能な聴覚補助装置が実現される。それは、音声信号から抽出された人の声成分に対して、強調処理を行っているからである。これにより、ノイズなどを大きくすることなく声の成分の音量を大小調整して、快適な聴覚補助装置が実現できる。

40

【発明を実施するための最良の形態】

【0024】

以下、本発明の実施の形態を図面に基づいて説明する。ただし、以下に示す実施の形態は、本発明の技術思想を具体化するための聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器を例示するものであって、本発明は聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュ

50

ータで読み取り可能な記録媒体並びに記録した機器を以下のものに特定しない。また、本明細書は特許請求の範囲に示される部材を、実施の形態の部材に特定するものでは決してない。特に実施の形態に記載されている構成部品の寸法、材質、形状、その相対的配置等は特に特定の記載がない限りは、本発明の範囲をそのみに限定する趣旨ではなく、単なる説明例にすぎない。なお、各図面が示す部材の大きさや位置関係等は、説明を明確にするため誇張していることがある。さらに以下の説明において、同一の名称、符号については同一もしくは同質の部材を示しており、詳細説明を適宜省略する。さらに、本発明を構成する各要素は、複数の要素を同一の部材で構成して一の部材で複数の要素を兼用する態様としてもよいし、逆に一の部材の機能を複数の部材で分担して実現することもできる。

10

【0025】

本明細書において聴覚補助装置や音声信号処理システムに接続される操作、制御、入出力、表示、その他の処理等のためのコンピュータ、プリンタ、外部記憶装置その他の周辺機器との接続は、例えばIEEE1394、RS-232x、RS-422、RS-423、RS-485、USB等のシリアル接続、パラレル接続、あるいは10BASE-T、100BASE-TX、1000BASE-T等のネットワークを介して電氣的に接続して通信を行う。接続は有線を使った物理的な接続に限られず、IEEE802.1x、OFDM方式等の無線LANやBluetooth等の電波、赤外線、光通信等を利用した無線接続等でもよい。さらに認識対象の音声データや認識後の音声データの保存や設定の保存等を行うための記録媒体には、メモ리카ードや磁気ディスク、光ディスク、光磁気

20

【0026】

図1に、本発明の聴覚補助装置の一例として、補聴器のブロック図を示す。図1(a)に示す聴覚補助装置100は、音声入力部10、A/D(アナログ/デジタル)変換器12、凹凸波形変換部14、声成分抽出部16、強調処理部18、音声出力部20を備える。音声入力部10としては、マイクロホンや音声入力端子等などが利用できる。特に補聴器として利用する場合は、マイクロホンで入力した音声信号をA/D変換器12でデジタル音声信号に変換して声成分抽出部16に送出する。ただ、直接デジタルの音声データを音声入力部10から入力する形態や、ネットワーク接続された外部機器から音声データを入力する方式も適宜採用できる。

30

【0027】

マイクロホン10から取り込まれた音声信号は、アナログフィルタなどの雑音除去装置に入力され、ここで10ms程度の周期でフレーム分析され、周囲環境の雑音や、マイクや伝送路が有する伝送特性雑音を除去される。その後、A/D変換器12でデジタル音声信号に変換して凹凸波形変換部14により量子化され、凹凸波形に変換される。さらにこの凹凸波形から声成分抽出部16で人の声に対応する声成分を抽出し、これを強調処理部18で強調声波形に補正し、音声出力部20から出力される。音声出力部20はスピーカや音声出力端子などである。凹凸波形変換部14、声成分抽出部16、強調処理部18等の各種演算処理部材は、マイクロプロセッサ(MPU)やCPU、LSI、FPGAやASIC等のゲートアレイといった論理回路や中央演算処理装置等のハードウェアやソフト

40

【0028】

特徴量としては、一般にはケプストラム係数が利用され、対数的な変換処理により対数スペクトルを求め、逆フーリエ変換や逆コサイン変換をすることにより算出して抽出することが行われていた。ただ、この方法では周波数スペクトル等への演算が必要となり処理量の負担が大きいので、本実施の形態では各母音(5母音)の特徴を振幅波形から抽出する方式とする。振幅波形から特徴を抽出することで周波数スペクトル等への変換のための演算を省くことが可能であり、演算回数も比較的少ない計算量で済む。得られた特徴量を

50

用いて離散ボロノイ図により領域を分割し、異なるカテゴリの境界座標を算出して最小2乗法による識別境界線を決定する。

【0029】

標準モデルとしては、複数の認識対象語彙毎の特徴量の時系列を確率的な遷移として表現する隠れマルコフモデル(HMM)と呼ばれる方法がある。HMMとは、あらかじめ個人差による音韻や単語の特徴量の時系列をHMMモデルに学習させておき、入力音声モデルに確率値としてどのくらい近いかを捉えて認識する方法である。また、標準モデルとしては、複数の認識対象語彙毎の特徴量の時系列の中の代表的な特徴量の時系列をモデルとしても良いし、さらに特徴量の時系列を時間的あるいは周波数的に正規化(伸縮)することで得られる特徴量の正規化時系列を用いてもよい。例えば、時間軸上で任意の長さ

10

【0030】

本実施形態では、このようにいずれの場合の標準モデルを使用することができる。ただし、いずれの標準モデルを作成する場合でも、標準モデルを作成するための複数の音声データをあらかじめ用意しておき、入力音声の振幅に対して同様の処理を行い凹凸波形に変換して登録しておく必要がある。

【0031】

凹凸波形変換部14は、図1(b)に示すように、振幅算出部22と、平均値演算部24と、比較部26と、変換部28とを備える。振幅算出部22は、入力波形の振幅波形に基づいて量子化を行う。量子化されたデータは平均値演算部24で、各点の平均値を求め

20

【0032】

【数5】

$$\alpha_k = \frac{\left(\sum_{i=k+1}^{n+k} x_i \right) + \left(\sum_{i=k-n}^{k-1} x_i \right)}{N}$$

30

【0033】

ここで、添え字のkは現在参照しているサンプル点である。 x_k はk点における振幅値を表している。またNは凹凸波形を算出するために用いる閾値を計算するための幅を表している。ここでNを2のべき乗とすることで、除算の際に2進数であればビット・シフト演算が利用でき、平均値演算部24での演算を簡単に行うことができるので好ましい。

【0034】

このようにして平均値演算部24で各点につき得られた平均値を、比較部26で振幅値と比較する。具体的には、各点の振幅値 x_k とその平均値 α_k とを比較し、以下の数6のように比較結果を出力する。

40

【0035】

【数6】

$$B_k = \begin{cases} a (\alpha_k < x_k) \\ b (\alpha_k > x_k) \end{cases}$$

50

【 0 0 3 6 】

このように、比較部 2 6 は振幅波形の各点につき平均値を演算し、k 点のサンプル値（振幅値）が各平均値より大か小により得られる信号を出力する。比較結果として x_k が平均値以上であれば a、平均値以下であれば b を比較部 2 6 は出力する。この結果を、変換部 2 8 で凹凸状の波形として出力する。例えば $a = 1$ 、 $b = 0$ とすると、振幅波形は谷か山（0 か 1）の凹凸波形で表現できる。この凹凸波形は、平均値を演算した N 個の領域において、 x_k が平均値以上すなわち波形が凸形になっているのか、あるいは x_k が平均値以下すなわち凹形になっているかを表現している。よって、N の値を変化させれば元の振幅波形の凹凸を粗く、あるいは細かく表現することができ、N をパラメータとして変化させることにより複数個の特徴量を抽出できる。このように振幅波形の特徴を凹凸のみで簡素化して表現することにより、必要な特徴量を抽出でき声成分抽出や音声認識モデルに適用することができる。特に、凹凸のみで表現される特徴量は、あまり高度でない情報処理機能においても音声波形から子音部と母音部の切り出し（Segmentation）が実現できる。また抽出された声成分に基づいて音声出力に処理を加える場合も、実用的な時間間隔で音声出力の大きさを変化させることが可能となる。

10

【 0 0 3 7 】

さらに、平均値の演算において、振幅波形の加算値を求める際に、近接する位置での算出値を利用して演算を簡素化することができる。すなわち、あるデータにつき平均値を求めるために $N (= 2n)$ 個のデータの振幅値を加算した加算値を保持しておく。このとき、平均値 α_k は以下の数 7 のように演算できる。

20

【 0 0 3 8 】

【数 7】

$$\alpha_k = \frac{\sum_{i=k-n}^{k-1} x_i + \sum_{i=k+1}^{n+k} x_i}{N}$$

$$= \frac{\sum_{i=k-n}^{k-2} x_i + x_{k-1} + \sum_{i=k+1}^{n+k-1} x_i + x_{n+k}}{N}$$

30

【 0 0 3 9 】

ここで、平均値 α_k は以下の数 8 のように変形できる。

【 0 0 4 0 】

【数 8】

$$\alpha_k = \frac{\sum_{i=k-n}^{k-2} x_i + \sum_{i=k+1}^{n+k-1} x_i}{N} + \frac{x_{k-1} + x_{n+k}}{N}$$

40

【 0 0 4 1 】

一方、平均値 α_k は上記数 5 で表記されるので、以下の数 9 のようにも表記できる。

【 0 0 4 2 】

【数 9】

$$\alpha_k = \frac{\sum_{i=k-n}^{k-1} x_i + \sum_{i=k+1}^{n+k} x_i}{N}$$

【0043】

したがって、 k を $k-1$ に置き換えると、 $k-1$ すなわち k より 1 つ前のデータに関する平均値 α_{k-1} は以下の数 10 のように表記できる。

10

【0044】

【数 10】

$$\alpha_{k-1} = \frac{x_{k-1-n} + \sum_{i=k-n}^{k-2} x_i + \sum_{i=k}^{n+k-1} x_i}{N}$$

【0045】

さらに、 α_{k-1} を変形すると、次式数 11 のようになる。

20

【0046】

【数 11】

$$\alpha_{k-1} - \frac{x_{k-1-n}}{N} - \frac{x_k}{N} = \frac{\sum_{i=k-n}^{k-2} x_i + \sum_{i=k+1}^{n+k-1} x_i}{N}$$

【0047】

上記数 11 で α_{k-1} に代わって平均値 α_k を求めると、以下の数 12 のようになる。

30

【0048】

【数 12】

$$\alpha_k = \frac{\sum_{i=k-n}^{k-2} x_i + \sum_{i=k+1}^{n+k-1} x_i}{N} + \frac{x_{k-1} + x_{n+k}}{N}$$

【0049】

さらに α_{k-1} を再記すると、数 13 のようになる。

40

【0050】

【数 13】

$$\alpha_{k-1} - \frac{x_{k-1-n}}{N} - \frac{x_k}{N} = \frac{\sum_{i=k-n}^{k-2} x_i + \sum_{i=k+1}^{n+k-1} x_i}{N}$$

【0051】

50

上記数 1 2、数 1 3 を整理すると、以下の数 1 4 のようになる。

【 0 0 5 2 】

【数 1 4】

$$\begin{aligned}\alpha_k &= \alpha_{k-1} + \frac{x_{k-1} + x_{n+k}}{N} - \frac{x_k + x_{k-1-n}}{N} \\ &= \alpha_{k-1} + \frac{x_{k-1} + x_{n+k} - x_k - x_{k-1-n}}{N}\end{aligned}$$

10

【 0 0 5 3 】

数 1 4 から、平均値 α_k はその前段で演算した平均値 α_{k-1} を用いて容易に演算することが可能となる。すなわち、 α_{k-1} が求まると、数 1 4 の 2 項目である以下の数 1 5 に α_{k-1} に加算することで、平均値 α_k を求めることができる。

【 0 0 5 4 】

【数 1 5】

$$\alpha_k = \alpha_{k-1} + (x_{k-1} + x_{n+k} - x_k - x_{k-1-n}) / N$$

20

【 0 0 5 5 】

これにより、逐次的に α_{k+1} 、 α_{k+2} 、 \dots 、 α_{k+n} を求めることができる。上記の演算を図表で表すと、図 2 (b) のようになる。この図に示すように、 α_k と α_{k-1} を算出する際のデータには共通領域があることが分かる。したがって、これらの共通領域のデータに対しては、一度の計算で演算は終了し、この演算結果をメモリなどの記憶手段に格納しておけば、次の演算に利用することができるので、全体の計算時間が短縮される。以上のように、演算を簡素化し、さらに演算量も減らすことで演算処理を極めて低負荷にすることができ、非常に簡単な演算によってすべての平均値 α_k を求めるアルゴリズムが得られる。これによって、音声波形から声成分抽出のための特徴抽出を高速かつ簡単に求めることが可能となり、実用性が極めて高い。

30

【実施例】

【 0 0 5 6 】

次に、実施例としてコンピュータ・シミュレーションにより抽出した特徴量を用いた母音の認識実験の結果を図 3 ~ 図 1 0 に示す。

(音声データ)

【 0 0 5 7 】

本実施例では、音声データから特徴量を抽出する特徴量抽出部 1 4 として、音声の特徴を生かしアナログ処理とデジタル処理部を組み合わせで作られた専用の集積回路 (I C) を使用した。音声の特徴とは、音声を波形で表したとき一般に正領域と負領域において非対称であること、声帯から送り出される圧力はパルス信号の発生、減衰に基づいていることである。これらの点を考慮してマイクロホン 1 0 から得られる音声信号を電圧値で測定すると、同時に正負の電力値の最大を一定時間保持させながら、次の正負の電圧値を検出するまでの時間を検出し、ピッチを検出している。この I C を用いることで、音声波形とピッチを検出することが可能である。図 3 に、I C により取得した音声波形とピッチ情報を示す。

40

【 0 0 5 8 】

また分類は 5 母音から得られる 2 母音の組み合わせを用いて、投票形式で行なわれる。そして音声の振幅波形から認識に有効な特徴の抽出と解析を行なう。本実施例では、成人男性 1 名から 6 7 音素の 1 7 セットの音声データを取得する。サンプリング周波数は 8 1 . 9 2 k H z である。さらに、音声は自然発話と意識発話のデータを取得し、自然発話は

50

一日の時間を問わず発生した音声であり、意識発話は夜の静かな時間帯に音素をはっきりと発声した音声である。音声の長さとして意識発話のピッチ数は自然発話の約 1.5 倍の長さで取得されている。

(母音波形抽出と前処理)

【 0 0 5 9 】

母音識別を行なうために IC により得られているピッチを参照して音素データから母音の定常とする区間を抽出する。そこで、全ピッチ数の 3 分の 2 の位置にあたるピッチを中心に前後の 1 周期分を抽出し、全部で 3 周期分の信号を定常な母音波形として特徴を抽出するために用いる。そして、抽出した 3 周期の母音波形から凹凸波形を生成する。ここで、元の振幅波形である定常とした 3 周期の母音波形を図 4 に、 $N = 256$ として図 4 から抽出した 1 周期分の凹凸波形を図 5 に、 $N = 64$ として図 4 から抽出した 1 周期分の凹凸波形を図 6 に、それぞれ示す。これらの図においては、上記数 6 の $a = 0.8$ 、 $b = 0.2$ としている。図 5 と図 6 を比較すると、 N の小さい図 6 の方が細かな凹凸波形となっており、振幅波形の山谷を細かく抽出していること、および N の大きい図 5 は、振幅波形の山谷を粗く抽出していることが判る。本実施例では、このように N を 256 と 64 の 2 つに設定して得られた凹凸波形を用いて認識実験を行う。なお、 N の値として凹凸波形を生成する幅である 256 と 64 は経験的に決定しているが、これ以外の値とすることもできることはいうまでもない。

【 0 0 6 0 】

このようにして得られた振幅波形と凹凸波形を用いて時間軸上で母音の特徴量を抽出する。特徴量は主に 1 周期分の波形から抽出している。その 1 周期分の波形は 3 周期の波形の始点とした点から類似度 (ユークリッド距離) を計算し、距離の近い 2 周期を選択する。そして時間的に早く存在している波形を選択する。これは中心より後半での母音波形の抽出を行っているために、後半に位置する母音波形よりも前半に位置する母音波形のほうが、より母音の特徴を保った波形であると考えられるためである。このように選択された 1 周期分の波形部分と 3 周期分の凹凸波形から特徴を抽出する。次に、母音を識別するための特徴量について述べる。抽出する特徴量の数は全部で 5 個である。母音を識別するための提案システムは、5 母音の内の 2 母音の各組み合わせから選ばれた母音の投票数により識別したい母音であると決定している。2 母音の組み合わせから選択すべき母音を識別するために、各 2 母音を識別する特徴は組み合わせごとに異なる。2 母音を識別しやすい特徴量を各組み合わせごとに選択することで比較的高い識別率が得られると考えられる。抽出する 5 つの特徴量は、以下の通りである。

【 0 0 6 1 】

(1) 256 凹凸波形を参照し、母音波形の 1 周期の始まりから探索して得られる最初の凸部分の幅

(2) 特徴量 1 で検出された凸幅に存在する振幅波形の面積

(3) 特徴量 1 で検出された凸幅に存在する振幅波形の分散値

(4) 特徴量 1 で検出された凸幅に存在する振幅波形を 0 ~ 1 に正規化を行い、凸幅で生成した正弦波との類似性

(5) 3 周期分の 64 凹凸波形に存在する凸の数

【 0 0 6 2 】

ここで特徴量 4 について詳述する。まず特徴量 4 を抽出するために正弦波を生成する。その正弦波は、

(1) 凸区間の振幅値を 0 ~ 1 に正規化を行う。その振幅値の最大値の位置の検出を行う。

(2) 始点から最大値位置までに 0 から $\pi/2$ までの正弦波を生成する。最大値位置から終点までに $\pi/2$ から π までの正弦波を生成する。

という処理を行なう。正弦波により得られた波形と振幅波形の例を図 7 と図 8 に示す。図 7 は、凸部分の / e / 音素波形と生成した正弦波であり、図 8 は、凸部分の / o / 音素波

10

20

30

40

50

形と生成した正弦波である。特徴量 4 はこの二つの波形から逐次 4 点で角度が計算され、その差の合計が特徴量とされる。これらの特徴量の大半は 1 周期の始めに存在する凸部分の特徴量で構成されている。この部分は母音波形（定常波形）で最も変化があり、また各母音の特徴が現われていると考えられる。この始めに位置する凸部分の検出は 1 周期分の 2 5 6 凹凸波形から検出され、凹幅が最も大きい箇所の次に現われる凸部分である。

（母音認識実験）

【0063】

次に、抽出した特徴量を用いて離散ポロノイ図と最小 2 乗法を適用し母音の識別を行なう。離散ポロノイ図とは空間中に配置された多数の母点の勢力範囲を表す図である。離散ポロノイ図は母点の存在する空間を離散化し、各画素がどの母点に近いかによって空間を分割したものである。本実施例で用いた離散ポロノイ図は逐次添加法を適用している。これは離散化された空間に母点を一つずつ追加していき、新しいポロノイ領域だけを効率よく作成することで、高速に離散ポロノイ図を作成する方法である。本実施例で用いた離散ポロノイ領域は 5 1 6 0 × 5 1 6 0 の領域であり、各母点との対応は得られている母点の最大値を用いて 0 から最大値までを均等に分割している。

（母音認識アルゴリズム）

【0064】

識別アルゴリズムは 2 母音の組み合わせ（ ${}_5C_2$ ）によりなされる。そして、各組み合わせに適した特徴量を用いて母音を選択し、全組み合わせから得られた投票数により識別する母音を決定する。その各組み合わせに適した特徴量（識別のために使用した）を表 1 に示す。

【0065】

【表 1】

表 1: 識別のための特徴量の組み合わせ

/a/-/i/	/a/-/u/	/a/-/e/	/a/-/o/	/i/-/u/
1-2	1-2	1-2	1-2	1-5
/i/-/e/	/i/-/o/	/u/-/e/	/u/-/o/	/e/-/o/
1-2	1-2	1-4	1-2	3-4

【0066】

ここで、識別のための特徴量の組み合わせを示す表 1 において、上段の / a / - / i / 等は各母音の組み合わせを表しており、下段の 1 - 2 等は使用した特徴量の番号（1）および（2）を示している。これらの 2 母音の識別を行なう特徴の選択は全データから抽出した特徴量の分布を調査し、目視による特徴量の分布を確認した後に経験的に決定している。さらに母音の識別を行なうための特徴量の数が 2 個であるのは、単一の特徴量を用いて識別を行なえば簡単な閾値処理になるため処理時間は少なく済むが、同じ母音であっても分布の広がりのために境界付近のデータに関して単一の特徴で識別を行なうと誤認識を起しやすと考えられるからである。また 2 個の特徴量を用いて目視によるデータ分布の確認を行ないやすいことや、直線的な閾値処理ではなく非線形的な識別が可能であり演算回数をできるだけ少なくすることが目的である。

【0067】

本手法で母音識別は離散ポロノイ図による 2 クラスの母音の組み合わせの境界線を求めるものである。得られている特徴量をマッピングしポロノイ領域に分ける。そして、同カテゴリによる領域の統合を図り、2 クラス間の境界部分の座標を得る。そして、その座標により構成される境界線を最小 2 乗法により境界線の関数を算出する。その境界線の関数を用いて母音の識別は 2 クラス間の投票形式により行なわれる。選択された母音の投票数

が単独で最も多く存在した場合のみ、その母音に識別されたとする。一番多い投票数が等しく複数の母音に見られる場合は破棄する。ここで、/i/と/u/以外の組み合わせに対する識別は抽出した特徴量をそのまま用いている。しかし、/i/と/u/の識別に対しては特徴量1の常用対数をとることで数値を変換している。これは特徴量5に対して特徴量1の値のスケールが大きいためであり、最小2乗法を適用して境界線の関数を求めやすくするためである。また比較実験としてマハラノビス距離を用いた認識実験を行なう。これは各組み合わせから母音の選択はマハラノビス距離に基づいて行なうものである。マハラノビス距離は各グループの中心から分散を考慮した距離を示す。本実施例で抽出された母音の特徴は分布に偏りのある特徴量である。そのためマハラノビス距離による識別が有効であると考えられる。意識発話による各母音データ多数は188個であり、自然発話による各母音データは178個である。そして各識別に適した特徴量を用いてボロノイ図とマハラノビスにより得られた自然発話による識別結果と意識発話による識別結果を表2と表3に示す。表2は、離散ボロノイ図を用いた自然発話と意識発話の単独第一候補による認識率を、表3はマハラノビス距離による自然発話と意識発話の単独第一候補による認識率を、それぞれ示している。

【0068】

【表2】

表2: 離散ボロノイ図を用いた自然発話と意識発話の単独第一候補による認識率

自然発話						意識発話					
/a/	/i/	/u/	/e/	/o/	平均	/a/	/i/	/u/	/e/	/o/	平均
99.4	91.5	94.9	91.5	90.4	93.5	98.9	97.3	97.8	98.9	98.4	98.2

【0069】

【表3】

表3: マハラノビス距離による自然発話と意識発話の単独第一候補による認識率

自然発話						意識発話					
/a/	/i/	/u/	/e/	/o/	平均	/a/	/i/	/u/	/e/	/o/	平均
95.2	91.4	82.4	98.4	88.2	91.1	98.4	97.3	90.4	100.0	90.9	95.4

【0070】

以上のように、離散ボロノイ図を適用し境界線を求めることで、マハラノビス距離を用いて行なった認識実験より認識率が向上している。これは離散ボロノイ図により明確な境界線の導出が可能であることで認識率が向上したと考えられ良好な結果を得ることができた。また表3に示すように、マハラノビス距離を用いた認識結果から自然発話と意識発話の両方とも/u/と/e/の識別結果が他の母音に比べて低い。これは/i/と/u/、/e/と/o/の抽出した特長が境界を超えて存在するためである。この境界付近に存在する特徴を持つ母音は、マハラノビス距離により分散を考慮され境界を得ているが、正しい識別が行なえていないことが考えられる。この理由として、現在用いているデータから正確な各母音に関する分散が計算できないこと、あるいは音圧データを用いることで類似している母音の特徴の分布状態に偏りが存在すること等が考えられる。

【0071】

ここで、意識発話に用いた特徴の分布を図9と図10に示す。図9は/i/と/u/に関する特徴分布を示し、図10は/e/と/o/に関する特徴分布を示している。この結果から、自然発話より意識発話の結果が良好であることがわかる。これは意識して声を発声しているために、母音部分の1周期波形(ピッチ)が多く現われ、安定した母音波形を抽出できていると考えられる。このように意識発話を行なうことで比較的良好な識別結果

10

20

30

40

50

が得られる。さらに簡易な演算のみで特徴量の抽出やマハラノビス距離による識別を行なっているため小規模なハードウェア構成で実現可能であると考えられる。しかし、認識精度において離散ポロノイ図を用いた認識精度には及ばない。そのために離散ポロノイ空間の設定を適切に行なうことができれば、離散ポロノイ図を用いた小規模なハードウェアが構築できると考えられる。このように、本実施の形態を適用することで比較的良好な認識精度が得られると考えられる。また、境界線を求めるのではなく勢力図を保持することで特徴量をその勢力図に照らし合わせ、母音の決定も可能である。

【0072】

以上のように、本実施例によってモバイル機器のための音声認識システムが実現可能となる。特に小型のハードウェアのために比較的簡単な演算による母音の振幅波形からの特徴量抽出と5母音の識別を行ない、その有効性を検証した。

10

【0073】

さらに、平均値等による特徴量のスケールの正規化を行なうことで、より小スペースで離散ポロノイ図が適用可能となり、演算時間を少なくすることができる。また、より明確な識別が可能である特徴の抽出にも適用可能である。例えば、識別境界線を2分類を行なうのに適した手法であるSVMを用いることができる。また、母音波形の1周期ごとにも違いがあるために定常であるとする母音波形の特定や母音に対して音声の始まりから終わりまでのピッチ幅の変動等を調査し、標準モデルとして登録しておくことで識別に有効な特徴の抽出や演算時間がさらに改善される。

(補聴機能)

20

【0074】

さらに、本発明は音声信号処理の前処理として広い応用範囲を備えており、入力された音声信号の声成分抽出や音声認識のみならず、音声の出力に際しての処理、例えば音声をより聞き取りやすい形に加工することもできる。この機能を応用すれば、音声をより聞き取りやすく加工する補聴機能として、補聴器等に利用できる。

【0075】

従来の補聴器では単純に検知した音声信号をすべて増幅させて音量を大きくするものであった。しかしながら、補聴器から出力される音量を増加すると、目的とする音以外の雑音も大きく聞こえるため、耳にガンガンと音が入り頭痛障害等を引き起こしたり不快感を伴うことがあった。また、小さなスピーカでは音が割れて音質が劣化する。さらに音量を大きくすると消費電力が大きくなり、長時間の使用が困難になる。さらにまたイヤースピーカの口径が大きくなり、重量も増加し、補聴器全体の形状も大きくなるといった欠点があった。

30

【0076】

これに対して、本発明の音声信号の特徴量抽出機能を応用し、声の聞き取り難い成分を聞き取りやすく加工することができる。すなわち、発音を峻別する部位のみを強調する処理によって、ノイズ成分を大きくすることなく聞き分けの容易な音声に加工できる。この方法では、音声認識のような高度な識別までは不要で、声成分の抽出ができれば十分であるため、より精度を向上させることができ、しかも処理をさらに簡素化できる。

(音声信号の取得手順)

40

【0077】

この音声信号処理方法の手順を、図11のフローチャートに示す。図11に示すように、音声入力部10で音声信号波形を取得した後、声成分抽出部16で人の声成分を抽出し、さらに強調処理部18で凹凸波形を強調して強調声波形を生成して、この強調声波形に基づいて音声出力部20で補正された声を出力する。以下、図12～図13を参照して凹凸波形変換部14及び声成分抽出部16で音声信号波形から声成分を抽出する手順を説明する。図12～図13は、取得した音声信号波形をそれぞれ示している。図12は、音声の最も低い周波数に対応する波形を示しており、その周期は T_1 である。一方、図13は、音声の最も高い周波数に対応する波形を示しており、その周期は T_2 である。人の声に含まれる周波数成分は、男性でも最低の周波数 $f_1 (= 1 / T_1)$ は通常100Hz止ま

50

りであり、100 Hz以下の周波数成分は殆ど無い。このとき、周期 T_1 は $1/100\text{ Hz} = 0.01\text{ 秒} = 10\text{ ms}$ である。そしてこの一周期の正又は負の半サイクルは、その $1/2$ 、すなわち $10\text{ ms}/2 = 5\text{ ms}$ となる。したがって、音声信号波形のサンプリングは、 5 ms (0.005 秒)程度の幅(窓)でサンプリング周期を設定すれば、音声の最も低い周波数 T_1 にも対応してサンプリングが可能となる。本実施の形態では、周波数が100 Hz近傍にある場合の余裕を考慮して、 6 ms に設定する。言い換えると、半波長が 6 ms よりも長い成分は人の声でない成分(すなわちノイズ)であると峻別でき、これによって音声信号から人の声(の低音領域)に相当する成分を抽出することができる。

【0078】

例えば、音楽用CDと同等の音質でサンプリングする場合、 $f = 44.1\text{ kHz}$ すなわち1秒間に44100サンプルを取得する必要がある。この周期でサンプリングする場合、 0.6 ms の区間(窓)で取得されるサンプル数は、 $44100\text{ サンプル} \times 0.006\text{ ms} = 264.6\text{ サンプル}$ となる。したがって、上記の手法でサンプリングする場合は、 $N = 264.6$ に設定すればよいことになる。ここで、上述したように N を2のべき乗に設定すれば、ビットシフトでの演算が可能となり、浮動小数点演算を使用しない整数型演算で高速且つ低負荷に処理することができる。ここで $N = 2^n$ とすると、 $N = 2^8 = 256$ とすれば、上記とほぼ等しい設定が実現できる。この場合は、 $n = 8$ を採用する。以上から、低音の場合は $n = 8$ を採用することで、低い周波数でのサンプリングすなわち音声信号取得に対応できることが判る。

【0079】

同様に、高い周波数 $f_2 (= 1/T_2)$ に対しては、 $n = 1, 2, 3$ のいずれか一を採用できる。これによって、人の声の高い周波数成分よりもさらに高い周波数成分は、人の声でないノイズであると峻別できる。この結果、人の声が含まれない低域と高域をカットして人の声のみを抽出できる。この例では、音声信号の高周波域信号、中周波域信号、低周波域信号に分けて、それぞれ低周波域に対しては $n = 8$ 、中周波域に対しては $n = 5$ 又は 6 、高周波域に対しては $n = 1, 2, 3$ のいずれかに設定している。高域、中域、低域のそれぞれにおいて、 n の値をいずれに設定するかは、使用される環境やユーザの聴力特性などに応じて設定される。また、高域、中域、低域の3つで分ける他、高域と低域のみの2つを利用する方法や4以上に区別する方法も用途やユーザなどに応じて適宜採用できる。

【0080】

以上のように、取得した音声信号に対して、 n の値を高域、中域、低域の3つでそれぞれ変化させて人の声を含む音声信号として高周波域信号、中周波域信号、低周波域信号をそれぞれ抽出することができる。いいかえると、サンプリングの際に n の値を変化させることで、人の声にあたる成分のみを抽出できる。

言い換えると、以上の工程では、音声信号波形から低音と高音をカットして人の声の成分を抽出し、これをデジタル処理して凹凸波形としている。このようにして得られた凹凸パターンは人の声成分を含んでいるため、これを明瞭にするための補正を行う。

(強調声データの生成)

【0081】

次に、このようにして取得された各帯域の音声信号に対して、強調処理部18で音声を聞き取りやすくするための補正処理を行う。具体的な処理を図14～図15に基づいて説明する。従来、補聴器などの聴覚補助装置においては、取得された音声の音量を一律に引き上げることで聞き取りやすくしていた。しかしながら、単にボリュームを上げるだけでは、取得された音声に含まれるノイズも増幅されるため、耳元で喧しく再生されるだけで快適に聴取できるとは言い難い状況であった。そこで本実施の形態においては、デジタル信号で取得された音声信号に対して、人の声のみを抽出し、さらに人の声を聞き取りやすくする処理を加えることによって、音量を変えずに聴取しやすくできる。具体的には、図14に示すように、波形の凸部を持ち上げ、一方凹部を引き下げることによって、実線で示す凹凸を強調した波線で示す強調声波形に変換する。これによって、音のメリハリが強調さ

れ、聴取しやすい音声に補正することが可能となる。また、波形の補正量は、低域で大きく、高域で小さくなるように設定することで、より聴取しやすい音声に補正できる。

【0082】

具体的な演算としては、上記の手法で検出された凹凸波形に対して、凹凸のそれぞれに所定の係数を乗算する。この際、小数点を含む浮動小数点演算では演算処理が複雑化し、特にリアルタイム処理が求められる補聴器などの頂角補助装置においては要求される仕様が高くなる。そこで、整数の加算、乗算のみで演算可能なように、ビットシフトの手法を適用する。

【0083】

まず、凸部の強調においては、 $(1 + 1/2^U)$ を乗算する。例えば、 $U = 3$ の場合は、 $1 + 1/2^3 = 1 + 1/8 = 1 + 0.125 = 1.125$ となり、整数値に3回ビットシフト演算を行い、元の値に加算することで1.125倍に強調できる。また同様に $U = 0$ の場合は、 $1 + 1/2^0 = 1 + 1 = 2$ 、 $U = 1$ の場合は、 $1 + 1/2^1 = 1 + 0.5 = 1.5$ 、 $U = 2$ の場合は、 $1 + 1/2^2 = 1 + 0.25 = 1.25$ 、 $U = 4$ の場合は、 $1 + 1/2^4 = 1 + 0.0625 = 1.0625$ 、 $U = 5$ の場合は、 $1 + 1/2^5 = 1 + 0.03125 = 1.03125$ 等、 U の値を変化させることで強調の程度を調整できる。このように、整数に加算とビットシフト演算を繰り返すことで、高速で低負荷な波形補正が実現できる。

(強調声波形の補正量)

【0084】

本実施の形態では、低域では補正量を大きくし、高域では小さく、中域ではその中間となるように設定している。具体的には、低域では $U = 0$ 又は1、中域では $U = 2$ 、高域では $U = 3$ 又は4を採用している。これにより、凹凸波形の凹凸を強調した聴取しやすい音声波形に変換することができる。図15に示すように、高域、中域、低域毎に所定の周期でサンプリングを行い $t_1 \sim t_5$ に示すようなスペクトルが得られる。さらに実線で示す波形の凸部及び凹部を各々検出し、凹凸を強調して波線で示すような強調声波形を生成する。この強調声波形に従って音声信号をスピーカなどの出力部から出力すると、ノイズなどの成分は強調されることなく、音声の、聞き取りやすいポイント部分のみが音量が大小に変化される結果強調されて再生されるので、極めて聴取しやすい音声となつて、またノイズが強調されない結果S/N比の優れた補聴器として理想的な特性を得ることができる。特に、従来のように単純に音量を上げる方法では背景音なども強調される結果、耳でガンガンと音が鳴る不快な状態となる。これに対して本実施の形態では、音量を殊更上げることなく、僅かに調整するのみで聞き取りやすい形に補正したのみであるため、非常に快適に使用できる。また必要に応じて、補正後の音量を調整する機能を付加することも可能であることはいうまでもない。

【0085】

以上のように、本実施の形態によれば、音声信号波形のフーリエ変換を行うことなく、フーリエ級数と同等の周波数成分が取得でき、しかもそのための演算を整数型の加算、乗算のみで処理でき、極めて高速かつ低負荷な処理とすることができ、安価なハードウェア、ソフトウェア構成においてもリアルタイム処理が可能であり、システムの小型化が可能となる。また、ビットシフトはレジスタで構成でき、またワイヤードロジックで実現できるため、安価且つ簡単な構成とできる。このため、補聴器のような小型化、軽量化が求められる装置への実装には理想的である。また、携帯電話などへの実装も容易に行える利点を得られる。

(子音の強調処理)

【0086】

以上の処理は、音声信号に対して母音と子音を特に区別することなく補正を行っている。ただ、母音よりも子音を強調する処理とすることで、より聞き取りやすくすることもできる。一般には、母音でなく子音の成分を聞き取りやすくすることで、音声の聞き取り難い成分を聞き取りやすくして音の判別が容易になる。すなわち、音声は子音と母音に分か

10

20

30

40

50

れるが、母音は比較的音量が大きく聞き取りやすいのに対し、子音は時間的に短く音量も小さくなりがちで聞き取り難い部分となる。そのため、子音の部分強調すれば、聞き取りやすくすることができる。検出された音声信号から、子音と母音とを区別するには、上述した音声認識技術を適用してもよい。この場合は、具体的な発生音の識別までは不要で、母音と子音の区別ができれば十分であるため、より精度を向上させることができ、しかも処理をさらに簡素化できる。これによって子音と母音とを区別し、さらに再生前に音声信号を加工する。また母音と子音と区別は、フォルマント等に基づいて行うこともできる。例えば、母音についてはサンプリングの周期を大きくして凹凸波形を取得し、凹凸数を計数することで判別できる。

【0087】

10

ここでは、再生中に音量をほぼリアルタイムに調整するアルゴリズムとして、聞き取り難く時間的にも短い子音部がきたときに音量を大きくし、それ以外の部分は通常の音量とする。子音部の直後に表れる母音部の音量は一般に大きいので、子音とその直後の母音により音声は認識される。この際に母音部は1周期から数周期の短時間波形で音声認識は完了する。聴覚と脳機能による短時間の音声認識が完了すると、それ以後は母音の継続であり、その期間において音声の出力信号を非常に小さくする。そして再び子音が入力されると、同様に音声出力を大きくする。この操作により、音声出力を聴感上は大きくしたことになるが、常に大音量の音声を聞いている訳ではないので、不快感は極めて少なくなる。また必要に応じて、母音の音量を絞るように調整してもよい。

【0088】

20

上記の音声信号処理のアルゴリズムにおいて重要な点は、音声の出力を通常とする、あるいは低下させた状態から、子音が入力される時点で音声出力を増加させる処理である。従来のシステムでは音声認識自体の処理量が多く、さらにこのような音量変化の処理を加えると、音声の子音の検出に要する時間と処理量が増大し、実用化は容易でない。これに対して、本発明のアルゴリズムでは整数演算のみで子音部と母音部を検出することが可能であるので、非常に短時間の簡単な信号処理により子音の検出が可能となる。すなわち、高速化が可能な演算処理量に抑えることができるので、補聴器や携帯電話のような携帯型電気機器等の小型のシステムへの組み込みや実装が容易となる。特に専用の処理を行うICを用意し、上記のデジタル信号処理を行わせることにより、携帯機器が保有するあまり高度でない情報処理機能においても音声波形から子音部と母音部の切り出し (Segmentation) が実現され、実用的な処理速度で音声出力の大きさを変化させることが可能となる。このように、本実施の形態によれば長い処理時間と大規模集積回路を必要とする不動小数点演算を用いることなく、整数型の演算のみで高速かつシンプルに実行できるので、実用的な価値は高い。また、音声信号の処理には通常DSP (Digital Signal Processor) 等を用いるが、浮動小数点演算を行うと長時間を要し、しかも集積回路の規模が増加するので、携帯電話などに搭載することは困難である。これに対し以上のアルゴリズムは整数型演算とビット・シフト演算のみで高速に処理することができ、これを回路として組み込むには小規模の集積回路で実現可能である。携帯機器や補聴器などの小型装置に搭載することができることは、実用化に際して大きな特長となる。

30

【0089】

40

さらに、補聴器の使用者に応じた周波数特性に調整することもできる。補聴器のスピーカから出力される音声の周波数特性を、使用者の耳の聴覚特性に合致させるよう調整することにより、聞こえ難い周波数に対して補正をかけることが可能となり、使用者に応じて聞き取りやすい適切な音声で聴くことができるようになる。

【0090】

このように、本発明によれば音声の聞き取りを容易にできる高性能な携帯型補聴器が実現できる。特に小型の補聴器は使用できる電池の大きさも制限され、小消費電力とする必要があるため、本発明のように演算処理量が少なく必要な消費電力も少なく済む聴覚補助装置は、理想的である。

【0091】

50

また、本発明は補聴器以外の携帯機器への実装も容易に行える。すなわち、携帯電話や PHS 等、音声を再生するスピーカを備える機器において、上記の処理を適用することで通話内容をより聞き取りやすくすることができる。特に携帯電話は小型、軽量化や連続駆動時間の長時間化の要求が強く、それでいて高性能化が求められているため、本発明のように低消費電力で音声をクリアに再生できるという優れた機能は非常に実用性がある。また、上記と同様に携帯電話の使用者に応じた周波数特性に再生設定を調整しておくことで、使用者に応じて最適な状態で音声を聞くことのできる携帯電話が実現可能である。これによって、補聴器を利用しない者であっても携帯電話の音声を聞き取りやすくして便利に使用できる。特に、通常の携帯電話ではスピーカから出力される音声等は出力も小さく、健全な聴覚特性の使用者向けに作成されているので、聴覚特性が低下した高齢者には使用し難いことがあったが、本発明を利用して音声を聞き取りやすくした高齢者やシニア向けの携帯電話を実現することが可能となる。

(補聴機能付携帯電話)

【0092】

また、補聴器機能を携帯電話等の携帯電気機器に組み込むことも可能である。このような補聴機能付携帯電話を、例えば、携帯電話の動作モードを切り替えて補聴器として使用可能とできる。これによって、携帯電話を使用するのと同じ姿勢で、すなわち耳元に電話機をおく状態で補聴器として使用することができるので、使用時の不自然さが少なく、周囲の人間も補聴器を使用しているとは気が付きにくく、使用時の抵抗感を和らげることができる。特に近年、携帯電話が普及し、年齢・男女を問わず携帯電話を使用することが通常の行動パターンとして定着した結果、携帯電話を耳に当てることに対して、周囲の人々は特別な反応を示さなくなり、当たり前前の行為として認識されている。他方、特別な形状をした補聴器を取り出して耳に当てることは通常では不自然さがつきまとい、しゃべり方にも多少の変化が表れ、補聴器の使用者も周囲の人々も特別な心理状態に陥りやすく、自然な会話が阻害される可能性がある。このような状況に鑑み、携帯電話に補聴器を組み込むことで、携帯電話を耳に当てて、補聴器として利用していることは外面的には判別できず、ごく普通の日常風景に埋没させることができ、補聴器の使用者の心理的なプレッシャは軽減される。

【0093】

また補聴器を携帯電話と一体とすることで、多くのメリットが生まれる。例えば、携帯電話のスピーカを補聴器のスピーカとして共用することができるので、補聴器専用のイヤースピーカを持ち歩く必要がない。また携帯電話を持ち運ぶ感覚で補聴器を常時携帯できるので、特別に補聴器を持ち歩いているという感覚が少なくなり、補聴器の使用における抵抗感を軽減できる。さらに補聴器を使用しないときには、身体に着けておく必要がないので、使用者の身体的負担が少ない。さらにまた、携帯電話のデジタル処理の集積回路に補聴器機能をもつ集積回路を組み込むことが可能であり、補聴器専用の集積回路を別個に製作する必要がなく、補聴器と携帯電話を個別に持つ場合に比較してコストを安価に抑えることができる。さらに近年は携帯電話用に長期使用可能な充電式電池が開発されており、これら最新のものを使用することができるので、補聴器の電源供給不足の心配が少なくなる。さらにまた、電池に限らず本体も最新の携帯電話をベースに製作できるので、機能的、デザイン的にも最新の携帯電話モデルを採用でき、利便性が高く、コスト的にも安く製作することが可能である。特に、補聴器専用の金型を作製すれば高価になるが、携帯電話と兼用することにより金型代の節約にもなる。

【0094】

この際、携帯電話にはマイク等の集音機を、使用者が通話するために使用するものと、周囲の音を集音するためのものとで2個以上を設けることが好ましい。

【0095】

さらに、補聴機能等のように音声信号の再生側で聞き取りやすく処理する他、音声信号の入力側で同様の処理を行うこともできる。例えば携帯電話において、スピーカのみならずマイク側に本発明を適用し、通話の相手側に向けて送信される音声信号に、上記のよう

10

20

30

40

50

な子音強調処理を行えば、携帯電話の使用者のみならず通話先の相手においても、会話内容を聞きとりやすくできる。さらにまた、電話のみならず、TVやラジオ、TV電話、TV会議システム等、スピーカやイヤホンで音声を再生する機器にも同様に適用できる。

【産業上の利用可能性】

【0096】

本発明の聴覚補助装置、音声信号処理方法、音声処理プログラム及びコンピュータで読み取り可能な記録媒体並びに記録した機器は、補聴器などで音声を聞き取りやすく補正する前処理や後処理に好適に適用できる。

【図面の簡単な説明】

【0097】

【図1】本発明の一実施の形態に係る聴覚補助装置の電子線撮像部の構成を示すブロック図である。

【図2】振幅波形のN個のデータについて平均値 \bar{k} を求める様子を示す説明図である。

【図3】入力される音声波形とそのピッチ情報を示す説明図である。

【図4】定常とした3周期の母音波形を示すグラフである。

【図5】 $N = 256$ として図4から抽出した1周期分の凹凸波形を示すグラフである。

【図6】 $N = 64$ として図4から抽出した1周期分の凹凸波形を示すグラフである。

【図7】凸部分の / e / 音素波形と生成した正弦波の波形を示すグラフである。

【図8】凸部分の / o / 音素波形と生成した正弦波の波形を示すグラフである。

【図9】意識発話に用いた / i / と / u / に関する特徴分布を示すグラフである。

【図10】意識発話に用いた / e / と / o / に関する特徴分布を示すグラフである。

【図11】音声信号処理方法の手順を示すフローチャートである。

【図12】音声の最も低い周波数に対応する波形を示すグラフである。

【図13】音声の最も高い周波数に対応する波形を示すグラフである。

【図14】凹凸波形の凹凸を強調して強調声波形に変換する様子を示すグラフである。

【図15】凹凸波形に基づき補正された強調声波形の一例を示すグラフである。

【図16】従来の補聴器の一例の構成を示すブロック図である。

【図17】従来の補聴器の音声処理回路の構成例を示すブロック図である。

【符号の説明】

【0098】

100 ... 聴覚補助装置

10 ... 音声入力部

12 ... A / D変換器

14 ... 凹凸波形変換部

16 ... 声成分抽出部

18 ... 強調処理部

20 ... 音声出力部

22 ... 振幅算出部

24 ... 平均値演算部

26 ... 比較部

28 ... 変換部

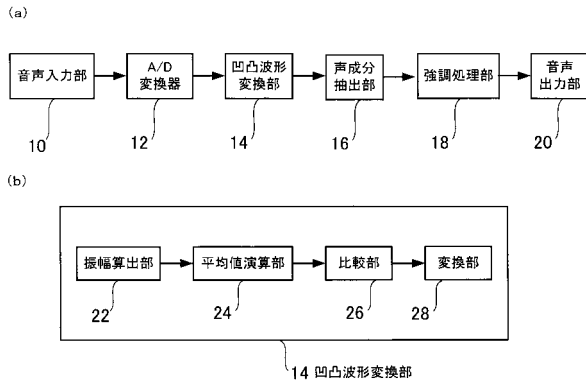
10

20

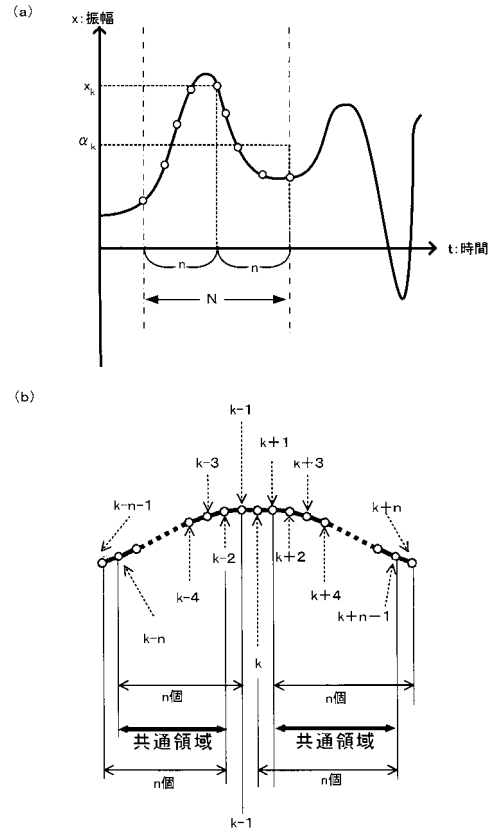
30

40

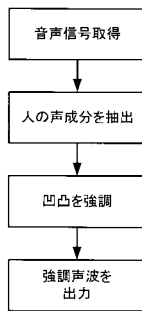
【図1】



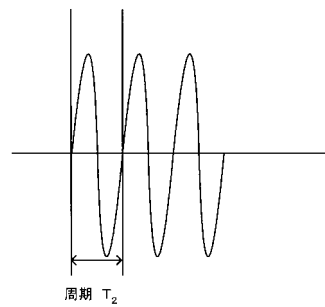
【図2】



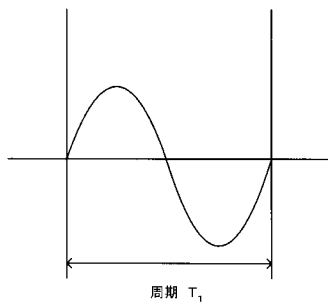
【図1 1】



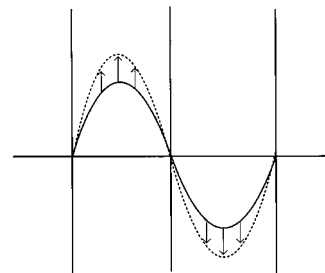
【図1 3】



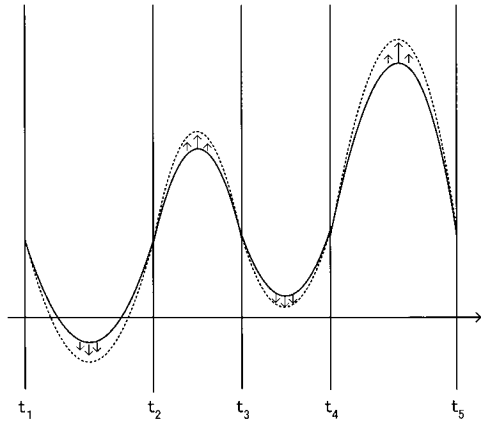
【図1 2】



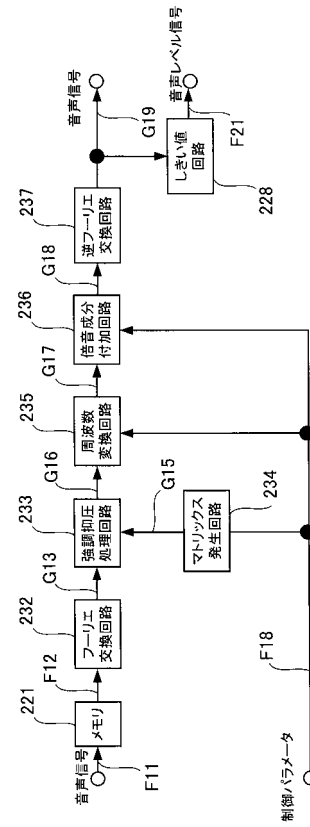
【図1 4】



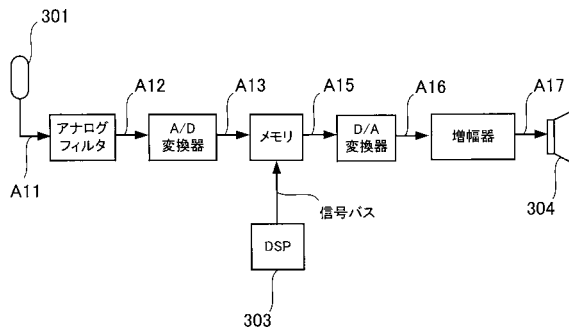
【図15】



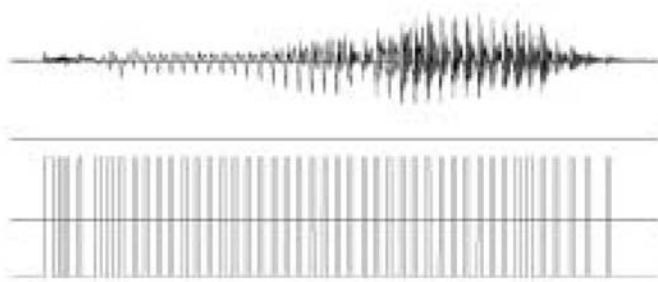
【図17】



【図16】

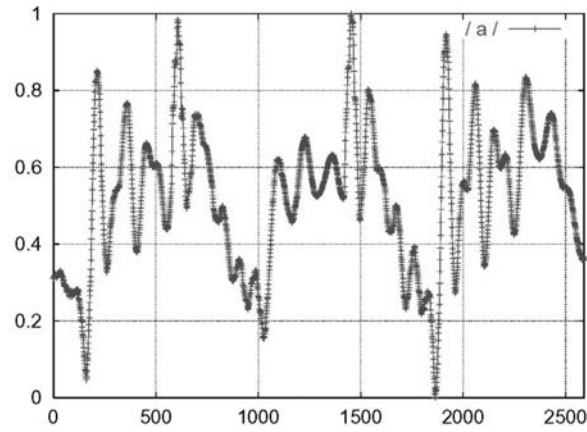


【図3】



取得した音声波形とピッチ

【 図 4 】



定常とした3周期の母音波形(/a/)

【 図 5 】

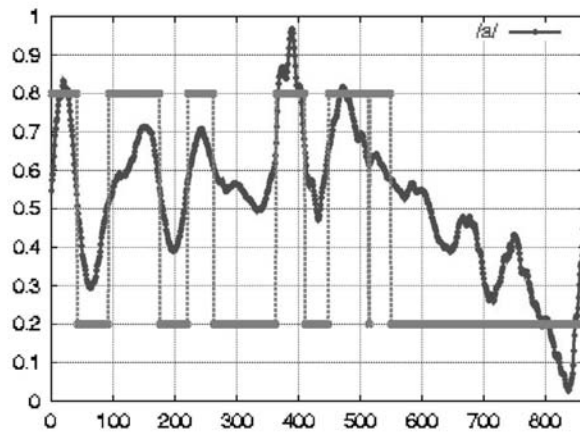


図4から抽出した1周期分の256凹凸波形

【 図 6 】

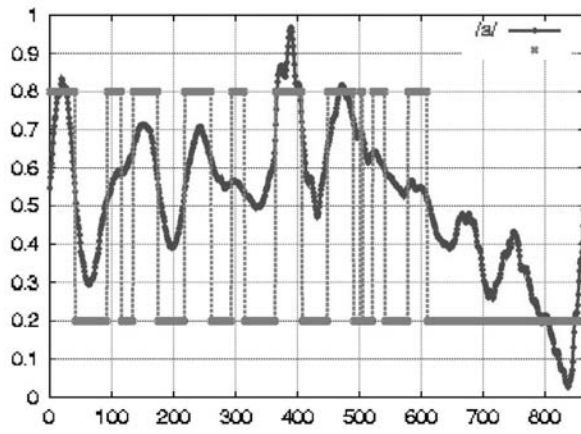
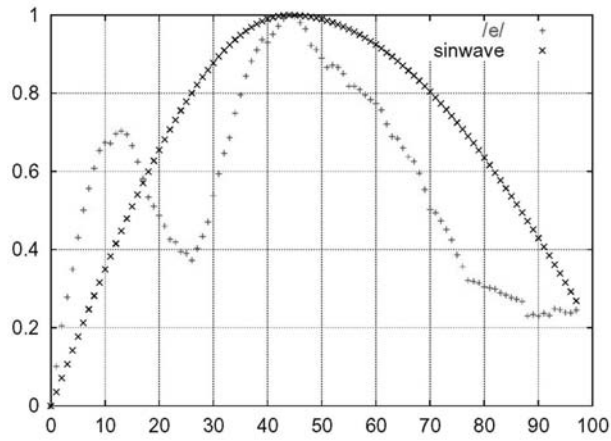


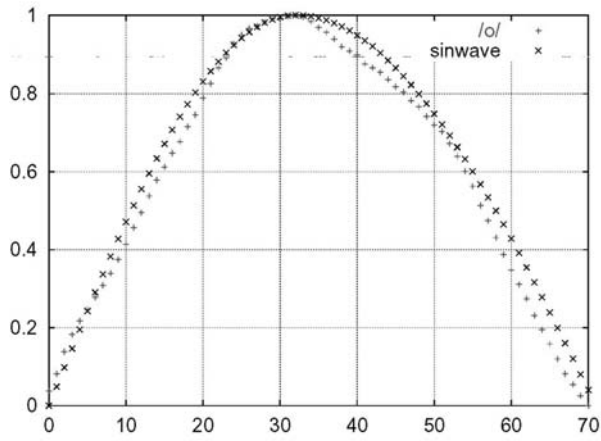
図4から抽出した1周期分の64凹凸波形

【 図 7 】



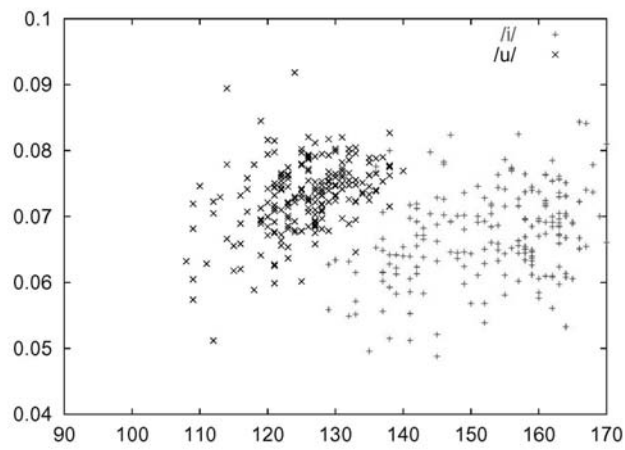
凸部分の/e/音素波形と生成した正弦波

【 図 8 】



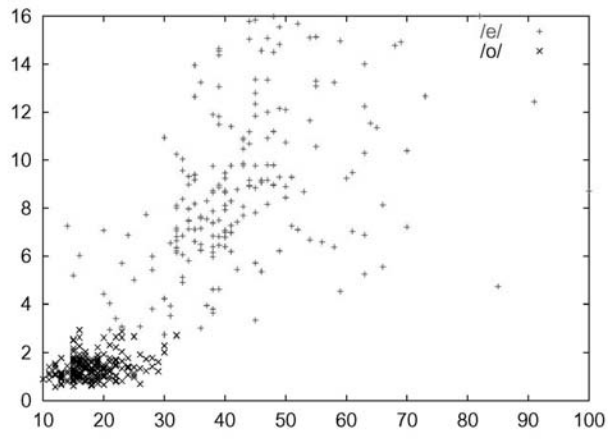
凸部分の/o/音素波形と生成した正弦波

【 図 9 】



/i/と/u/に関する特徴分布

【図10】



/e/と/o/に関する特徴分布

フロントページの続き

- (56)参考文献 特開平08-079897(JP,A)
特開2002-278579(JP,A)
特開平11-041698(JP,A)