

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4876245号
(P4876245)

(45) 発行日 平成24年2月15日(2012.2.15)

(24) 登録日 平成23年12月9日(2011.12.9)

(51) Int.Cl. F I
G 1 0 L 15/10 (2006.01) G 1 0 L 15/10 4 0 0 R
G 1 0 L 15/02 (2006.01) G 1 0 L 15/02 2 0 0 D
G 1 0 L 21/02 (2006.01) G 1 0 L 21/02 3 0 1 A
 G 1 0 L 21/02 3 0 2 B

請求項の数 10 (全 27 頁)

<p>(21) 出願番号 特願2006-40187 (P2006-40187) (22) 出願日 平成18年2月17日(2006.2.17) (65) 公開番号 特開2007-219188 (P2007-219188A) (43) 公開日 平成19年8月30日(2007.8.30) 審査請求日 平成21年1月14日(2009.1.14)</p> <p>特許法第30条第1項適用 2005年12月8日 社団法人電子情報通信学会発行の「電子情報通信学会技術研究報告 信学技報 Vol. 105 No. 479」に発表</p> <p>特許法第30条第1項適用 平成18年2月8日 国立大学法人九州大学主催の「修士論文発表会」において文書をもって発表</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</p> <p>(74) 代理人 100103621 弁理士 林 靖</p> <p>(72) 発明者 中島 祥好 福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内</p> <p>(72) 発明者 安武 達朗 福岡県福岡市東区箱崎六丁目10番1号 国立大学法人九州大学内</p> <p>審査官 前田 祐希</p>
--	--

最終頁に続く

(54) 【発明の名称】 子音加工装置、音声情報伝達装置及び子音加工方法

(57) 【特許請求の範囲】

【請求項1】

1 音節の長さが 1 0 0 m s ~ 4 0 0 m s、子音の長さが数十 m s として表される何れかの長さをもつ音声を対象にして子音強調する子音加工装置であって、
 入力された音声信号に対して前記子音の長さを含む数十 m s の幅を抽出できる時間フレームによって第 2 のフレーム信号を抽出しかつ増幅度算出の対象にされる時点の信号を前記時間フレームの時間軸における範囲の中央位置の信号に設定する第 2 の時間フレームと、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第 2 の時間フレームよりも狭い時間軸の幅を有し前記第 2 のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する前記子音の長さを含む数十 m s の幅の 1 / 3 の幅またはこの 1 / 3 と同等に雑音判定が行える前記子音の長さを含む数十 m s の幅の 1 / 3 程度の幅を備えた時間フレームによって第 1 のフレーム信号を抽出する第 1 の時間フレームと、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第 2 の時間フレームよりも広い時間軸の幅を有し前記第 2 のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する時間フレームによって前記音声信号から前記音節の 1 ~ 3 倍の長さの幅の第 3 のフレーム信号を抽出する第 3 の時間フレームと、を有するフレーム分割部と、前記第 1、第 2 及び第 3 のフレーム信号のそれぞれで平均パワーを算出するパワー算出部と、前記第 1 と第 2 のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否か

10

20

を比較すると共に、前記第2と第3のフレーム信号の平均パワーの比較を行う比較部と、前記比較部の比較結果に基づいて前記差が前記閾値を越えたときには雑音と判定すると共に、前記差が前記閾値を越えずかつ前記第2のフレーム信号の平均パワーが前記第3のフレーム信号の平均パワーより小さい場合に前記音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定する子音判定部と、

前記子音判定部が子音または音節の端点と判断した場合は前記音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅すると共に、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、

前記第1と第2のフレーム信号の平均パワーが、前記音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いられると共に、前記第2と第3のフレーム信号の平均パワーが、前記増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置。

【請求項2】

1音節の長さが100ms～400ms、子音の長さが数十msとして表される何れかの長さをもつ音声を対象にして子音強調する子音加工装置であって、

入力された音声信号に対して前記子音の長さを含む数十msの幅を抽出できる時間フレームによって第2のフレーム信号を抽出しかつ増幅度算出の対象にされる時点の信号を前記時間フレームの時間軸における範囲の中央位置の信号に設定する第2の時間フレームと、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第2

の時間フレームよりも狭い時間軸の幅を有し前記第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する前記子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える前記子音の長さを含む数十msの幅の1/3程度の幅を備えた時間フレームによって第1のフレーム信号を抽出する第1の時間フレームと、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第2の時間フレームよりも広い時間軸の幅を有し前記第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する時間フレームによって前記音声信号から前記音節の1～3倍の長さの幅の第3のフレーム信号を抽出する第3の時間フレームと、を有するフレーム分割部と、

前記第1、第2及び第3のフレーム信号のそれぞれで平均パワーを算出するパワー算出部と、

前記第1と第2のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、前記第2と第3のフレーム信号の平均パワーの比較を行う比較部と、前記比較部の比較結果に基づいて前記差が前記閾値を越えたときには雑音と判定すると共に、前記差が前記閾値を越えずかつ前記第2のフレーム信号の平均パワーが前記第3のフレーム信号の平均パワーより小さい場合に前記音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定する子音判定部と、

前記子音判定部が子音または音節の端点と判断した場合は前記音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅すると共に、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、

前記子音判定部が子音または音節の端点と判断した場合は前記音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号の増幅度を増幅方向に決定すると共に、子音または音節の端点でないと判断した場合は前記音声信号を増幅しない旨決定する増幅度決定部と、

前記増幅度決定部が決定した増幅度に応じて前記音声信号を増幅する増幅部とを備え、前記第1と第2のフレーム信号の平均パワーが、前記音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いられると共に、前記第2と第3のフレーム信号の平均パワーが、前記増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置。

10

20

30

40

50

【請求項 3】

前記比較部が、前記第 1 と第 2 のフレーム信号のデシベル表示した平均パワーの差を算出し、前記閾値をこの差と比較することを特徴とする請求項 1 または 2 に記載された子音加工装置。

【請求項 4】

前記比較部が、各フレーム信号の平均パワーの差に代えて、平均パワーの比率を算出することにより比較することを特徴とする請求項 1 または 2 に記載された子音加工装置。

【請求項 5】

請求項 1 ~ 4 の何れか 1 項に記載の子音加工装置において、前記第 3 の時間フレームは前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第 2 の時間フレームよりも広い時間軸の幅を有し前記第 2 のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する時間フレームによって前記第 3 のフレーム信号を抽出し、前記音声信号の増幅度算出の対象にされる時点の信号が第 2 の時間フレームの時間軸における範囲の中央位置の信号に設定され、前記第 3 のフレーム信号の平均パワーが、前記増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられるのに代えて、前記第 2 の時間フレームに連続して後置される第 4 の時間フレームによって第 4 のフレーム信号を抽出し、前記音声信号の増幅度算出の対象にされる時点の信号が前記第 2 の時間フレームの時間軸における範囲の後端となる前記第 4 の時間フレームとの境界の信号に設定され、前記第 4 のフレーム信号の平均パワーが、前記増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置。

【請求項 6】

前記フレーム分割部に音声信号を入力する前に、所定の周波数成分を通過させる通過特性を有するフィルタ部が設けられたことを特徴とする請求項 1 ~ 5 のいずれかに記載の子音加工装置。

【請求項 7】

請求項 1 ~ 6 のいずれかに記載された子音加工装置と、該子音加工装置からの子音加工された音声信号に基づいて子音強調された音声を入力するスピーカを備えたことを特徴とする音声情報伝達装置。

【請求項 8】

1 音節の長さが $100\text{ms} \sim 400\text{ms}$ 、子音の長さが数十 ms として表される何れかの長さをもつ音声を対象にして子音強調するとき、入力された音声信号に対して増幅度算出の対象にする時点の信号を、前記子音の長さを含む数十 ms の幅を抽出できる第 2 の時間フレームの時間軸における範囲の中央位置の信号に設定し、この第 2 の時間フレームにより前記音声信号から第 2 のフレーム信号を抽出し、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第 2 の時間フレームよりも狭い時間軸の幅を有し前記第 2 のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する前記子音の長さを含む数十 ms の幅の $1/3$ の幅またはこの $1/3$ と同等に雑音判定が行える前記子音の長さを含む数十 ms の幅の $1/3$ 程度の幅を備えた第 1 の時間フレームによって前記音声信号から時間軸上で第 1 のフレーム信号を抽出すると共に、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第 2 のフレームよりも広い時間軸の幅を有し前記第 2 のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する第 3 の時間フレームによって前記音声信号から前記音節の $1 \sim 3$ 倍の長さの幅の第 3 のフレーム信号を抽出し、前記第 1、第 2 及び第 3 のフレーム信号のそれぞれで平均パワーを算出し、前記第 1 と第 2 のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、前記第 2 と第 3 のフレーム信号の平均パワーの比較を行い、この比較結果に基づいて前記差が前記閾値を越えたときには雑音と判定すると共に、前記差が前記閾値を越えずかつ前記第 2 のフレーム信号の平均パワーが前記第 3 のフレーム信号の平均パ

10

20

30

40

50

ワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定し、子音または音節の端点と判断される場合は前記音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅し、子音または音節の端点でないと判断した場合は増幅せず、前記第1と第2のフレーム信号の平均パワーを前記音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いると共に、前記第2と第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法。

【請求項9】

1音節の長さが100ms～400ms、子音の長さが数十msとして表される何れかの長さをもつ音声を対象にして子音強調するとき、入力された音声信号に対して増幅度算出の対象にする時点の信号を、前記子音の長さを含む数十msの幅を抽出できる第2の時間フレームの時間軸における範囲の中央位置の信号に設定し、この第2の時間フレームにより前記音声信号から第2のフレーム信号を抽出し、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第2の時間フレームよりも狭い時間軸の幅を有し前記第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する前記子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える前記子音の長さを含む数十msの幅の1/3程度の幅を備えた第1の時間フレームによって前記音声信号から時間軸上で第1のフレーム信号を抽出すると共に、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第2のフレームよりも広い時間軸の幅を有し前記第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する第3の時間フレームによって前記音声信号から前記音節の1～3倍の長さの幅の第3のフレーム信号を抽出し、前記第1と第2のフレーム信号の平均パワーの比率を計算してこの比率が閾値を越えるか否かを比較すると共に、前記第2と第3のフレーム信号の平均パワーの比較を行い、この比較結果に基づいて前記比率が前記閾値を越えたときには雑音と判定すると共に、前記比率が前記閾値を越えずかつ前記第2のフレーム信号の平均パワーが前記第3のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定し、子音または音節の端点と判断される場合は前記音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅し、子音または音節の端点でないと判断した場合は増幅せず、前記第1と第2のフレーム信号の平均パワーを前記音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いると共に、前記第2と第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法。

【請求項10】

請求項8または9に記載の子音加工方法において、前記増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され前記第2のフレームよりも広い時間軸の幅を有し前記第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を前記音声信号から抽出する第3の時間フレームによって前記第3のフレーム信号を抽出し、かつ増幅度算出の対象にされる時点の信号を第2の時間フレームの時間軸の範囲において中央位置の信号に設定し、前記第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いるのに代えて、前記第2の時間フレームに連続して後置される第4の時間フレームによって前記第4のフレーム信号を抽出し、かつ前記音声信号の増幅度算出の対象にされる時点の信号を前記第2の時間フレームの時間軸の範囲の後端となる前記第4の時間フレームとの境界に設定し、前記第4のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法。

【発明の詳細な説明】

【技術分野】

【0001】

10

20

30

40

50

本発明は、リアルタイムの音声情報伝達が行え、信号処理が簡単で、騒音の多い環境でも、高齢者や聴覚障害者でも子音または音節の端点が聞き取り易く、安価に製造できる子音加工装置と、これを搭載した音声情報伝達装置及び子音加工方法に関する。

【背景技術】

【0002】

聴力が低下した高齢者や聴覚障害者は、一般の人と比べてどうしても言葉を聞き取る力が低下する。このような高齢者や聴覚障害者にとって、飛行場等の構内放送や電車、バス内の案内放送、自動販売機やATMのガイド音声など、周囲に騒音の多い公共空間における各種音声の内容を正確に聞き取るのは難しい。中でも緊急時における避難誘導の音声は、内容が理解できないと大きな事故に繋がりがねない。

10

【0003】

このような言葉の聞き取り力が衰えた高齢者や聴覚障害者のために、受信機を携帯してもらって電波や赤外線などの通信媒体を介して、音声を送信するワイヤレス放送システムが開発されている。このようなシステムでは、発声者は騒音の少ない静かな場所で、口元近くに配置したマイクに向けて発声し、そのクリアな音声通信媒体を介して直接使用者に送られる。従って、使用者はどんな騒音環境下においても、常にクリアな音声を聴取できる。

【0004】

また、補聴器などの聴覚補助機器では、周囲の騒音を抑制する騒音抑制機能 (noise reduction) や、後述するような音声の中の子音のみを強調する子音強調機能 (consonant enhancement) などを備えた機種が提案されている。とくに子音強調は、音声の子音の振幅が母音に比べて小さいために、高齢者や聴覚障害者が、子音部を聞き取ることが困難であるという事実に着目して開発された技術である。

20

【0005】

しかし、このような現象は聴力の低下に由来するものには限られない。聴力が低下していない人間でも、構内放送や案内放送などのアナウンス放送装置、携帯電話、その他の音声情報伝達装置は、騒音の多い環境下では騒音で音声聞き取れなくなる。これに対して、出力の大きなスピーカやイヤホンを搭載して音量を大きくすればよいが、聴覚の許容限界を越えてしまう可能性がある上に、装置の大きさに限界があり、また、音の歪み等が大きくなる。

30

【0006】

このため、音声を聞き取り易くする目的で音声強調の幾つかの手法が提案されるに至った。この音声強調というのは、音声の周波数スペクトルの所定帯域におけるスペクトル振幅を増幅し、聞く側からみて音声の明瞭度を向上させることであり、これは、所定の周波数帯域パワーを増幅すると共に、増幅帯域と異なる帯域のスペクトル振幅を減衰させることである。(特許文献1参照)。

【0007】

ここで、音声の発生するメカニズムについて説明すると、声帯が振動すると、声帯から唇までの声道で音声波が生成され、この声道で生成された音声波を唇及び舌等を介して音声として放射する。すなわち、声帯が一定周期(ピッチ周期)の振動をすると、声道は、肺からの空気を喉等の形状(例えば太さ)で共振させて母音を生成する。このとき、喉等の形状を変化させることで「あ」、「い」、「う」等の母音に調音し、音声波として放射する。そして、唇及び舌等では、口内で破裂音や摩擦音、また、鼻音、その他の音を発生して子音とし、通常は子音と母音とが合わさって空間に放射されるものである。

40

【0008】

ところで、日本語はC V (Consonant-Vowel) 型の言語であり、例えば日本語で「か」という発音は、「くっ」という子音(C)と「あ」という母音(V)とから構成され、子音と母音とが合わさって放射される。また、例えば英語はC V C (Consonant-Vowel-Consonant) 型の言語でもあり、子音、母音、子音の順に配列されることが多いことが知られている。

50

【 0 0 0 9 】

従って、日本語の場合、「ん」「っ」を除いて他の行の発音は、このような子音（C）と母音（V）の組合せなどで発音され、言葉を発するときは、多くの場合各音において、まず唇及び舌等を使って声帯からの音が妨げられて子音が調音され、次いで、声帯からの音が妨げられることなく母音が強い音で発声されることになる。

【 0 0 1 0 】

このため、子音は母音に比べて音声の振幅が小さく、音声情報伝達装置の音声は周囲の環境等では雑音に紛れて聞こえなくなる。このため、例えば、音声を明瞭に聞き分けることが可能な補聴器等の音声情報伝達装置が提案された（特許文献2参照）。

【 0 0 1 1 】

特許文献2の音声情報伝達装置は、外部から音声が入力されるマイクロフォンと、入力された音声信号に基づいて子音明瞭化信号を生成する音声信号処理部と、搬送波信号を生成する搬送波信号発生部と、搬送波信号を子音明瞭化信号に基づいて振幅変調する振幅変調部と、振幅変調された出力信号に基づく機械的振動を伝達する振動子とから構成され、この音声信号処理部が、子音抽出部で音声信号に含まれる子音部を抽出すると共に、反復処理部で抽出された子音部子音部が複数回反復されて音声信号に付加して子音明瞭化信号を生成するものである。

【 0 0 1 2 】

音声信号における子音部は、母音部との間にVOT（Voice Onset Time）が数十ms程度存在する。このVOTは、子音の破裂から声帯が振動するまでの時間であり、無音に近い状態である。したがって、子音部の立ち上がりや母音に比べて振幅は小さく、適当な基準値を設定し、振幅がこの基準値以下となる状態が所定時間（例えば、10ms程度）以上続く領域をVOTとして判別することにより、子音部の残余部分や母音部と区別して、子音部の終期を特定することができるというものである。

【 0 0 1 3 】

同様に、母音部の後、次に続く子音部との間にも、通常は数十ms以上の無音区間が存在する。従って、VOTの検出と同様にしてこの無音区間を判別することにより、次の子音部の始期を特定するものである。

【 0 0 1 4 】

また、特許文献2と同様、上述した特許文献1においても、携帯電話等の受話音声の明瞭度を改善し、入力音声に雑音が含まれる場合に音声品質の劣化及び雑音の増加を抑圧する音声強調装置が提案されている。

【 0 0 1 5 】

特許文献1の音声強調装置は、入力音声信号の音声品質を推定し音声品質推定値（推定S/N比）を出力する音声品質推定部と、音声品質推定部にて出力された音声品質推定値に基づいて、入力音声信号の声道特性の調整（ホルマンントの増幅、アンチホルマンントの減衰）と、入力音声信号の残差信号の強調（ピッチの強調）を行う音声強調処理部とを備えたものである。なお、この残差信号とは、音声波から線形予測可能な部分の除去により分離されたもの音源信号で、これの自己相関を算出することにより、音源のピッチ周期が得られるものである。

【 0 0 1 6 】

【特許文献1】特開2005-331783号公報

【特許文献2】特開2005-287600号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 1 7 】

以上説明したように、子音は母音に比べて音が弱く、音声情報伝達装置の音声は周囲の環境等次第で雑音に紛れて聞こえなくなるという問題があった。

【 0 0 1 8 】

上述したワイヤレス放送システムは、そのようなシステムが備えられている限定された

10

20

30

40

50

公共空間でなければ利用できない上に、使用者が受信機を携帯せねばならない。しかも、この放送システムは大規模で高価であるため、あらゆる公共空間に設けることは難しく、また、使用者全員が受信機を携帯することも困難なため、普及は難しいという問題があった。

【 0 0 1 9 】

また、補聴器等にも問題がある。まず、使用者が装用していなければ役に立たないし、補聴器への入力には音声と周囲の騒音が混在した音となる。従って、騒音抑制機構や子音強調機構によって、この双方の音が混在した中から騒音のみを抑制したり、子音のみを強調したりしなければならない。しかし、これらは騒音の種類など、場合によってうまく動作しないこともあるし、子音強調は静寂下においても難しい。このため従来は振幅エンベロープ、無声閉鎖子音の破裂に伴う無音区間、その他の周波数領域の情報など、複数の手がかりを並列的に利用して子音を検出することで、検出の正確さを上げている。しかし、このための処理は複雑になり、リアルタイム（実時間）若しくはこれに近い時間（準リアルタイム）内に子音強調して音声情報伝達を行うことの障害となっていた。

【 0 0 2 0 】

特許文献 1 の音声強調装置は、入力音声信号の S / N 比を推定し、この S / N 比に基づいて、ホルマント周波数を中心とするホルマントの電力（パワー）に正のゲインを与える処理を行うと共に、アンチホルマント周波数を中心とするアンチホルマントの電力（パワー）に負のゲインを与える処理を行い、また、ピッチを強調して聞き取り易くするものである。しかし、処理が複雑で時間がかかり、高コストで、リアルタイムに近い時間内に音声情報伝達処理を行うには課題が多いものであった。そして、この特許文献 1 の音声強調装置は、日本語の音声は子音と母音の組合せからなり、子音は母音に比べて音が弱いという特性を活かしていない。

【 0 0 2 1 】

これに対して、特許文献 2 の音声情報伝達装置は、音声信号に含まれる子音部を抽出すると共に、抽出された子音部が複数回反復されて音声信号に付加して子音明瞭化信号を生成する。このため、音声の聞き分けの改善にはなったが、各音で子音部分が繰返され、各音の子音部分で遅れが生じ、これが積み重なって、リアルタイムに近い時間内に音声情報伝達が行えるものではなかった。V O T や無音区間の判別のためには、この V O T や無音区間の情報が必要で、このため音声信号を一旦記憶してから処理することが必要であり、処理が複雑で時間がかかる点は特許文献 1 と変わらない。

【 0 0 2 2 】

このように従来技術は、複数の手がかりを並列的に利用して子音の検出を行い、このためその子音強調処理は非常に複雑であり、リアルタイム若しくはこれに近い時間内に音声情報伝達が行えるものではなかった。また、これらの技術は予め強調処理を施した音を記憶しておく必要があり、柔軟性が要求される音声情報伝達装置においては利用が難しい技術であった。

【 0 0 2 3 】

そこで本発明は、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音の多い環境でも、高齢者や聴覚障害者でも子音や音節の端点が聞き取り易く、安価に製造できる子音加工装置と音声情報伝達装置を提供することを目的とする。

【 0 0 2 4 】

また、本発明は、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音の多い環境でも、高齢者や聴覚障害者でも子音や音節の端点が聞き取り易い子音加工方法を提供することを目的とする。

【課題を解決するための手段】

【 0 0 2 5 】

本発明の子音加工装置は、1音節の長さが100ms～400ms、子音の長さが数十msとして表される何れかの長さをもつ音声を対象にして子音強調する子音加工装置であって、入力された音声信号に対して子音の長さを含む数十msの幅を抽出できる時間フレ

10

20

30

40

50

ームによって第2のフレーム信号を抽出しかつ増幅度算出の対象にされる時点の信号を時間フレームの時間軸における範囲の中央位置の信号に設定する第2の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2の時間フレームよりも狭い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える子音の長さを含む数十msの幅の1/3程度の幅を備えた時間フレームによって第1のフレーム信号を抽出する第1の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2の時間フレームよりも広い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する時間フレームによって音声信号から音節の1~3倍の長さの幅の第3のフレーム信号を抽出する第3の時間フレームと、を有するフレーム分割部と、第1、第2及び第3のフレーム信号のそれぞれで平均パワーを算出するパワー算出部と、第1と第2のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、第2と第3のフレーム信号の平均パワーの比較を行う比較部と、比較部の比較結果に基づいて差が閾値を越えたときには雑音と判定すると共に、差が閾値を越えずかつ第2のフレーム信号の平均パワーが第3のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定する子音判定部と、子音判定部が子音または音節の端点と判断した場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅すると共に、子音または音節の端点でないとは判断した場合は増幅しない増幅部とを備え、第1と第2のフレーム信号の平均パワーが、音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いられると共に、第2と第3のフレーム信号の平均パワーが、増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを主要な特徴とする。

10

20

ここで、本発明における「子音の長さを含む数十msの幅を抽出できる時間フレーム」とは、子音を抽出可能な数十msの長さであって、言語等ごとに想定される子音の長さまたはこれより少し長くて、雑音の判定演算、子音または音節の端点の判定演算において、子音または音節の端点を聞き取り易くできるような子音程度の長さの幅の第2の時間フレームのことである。また、「第2の時間フレームよりも狭い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える子音の長さを含む数十msの幅の1/3程度の幅を備えた時間フレーム」とは、第2の時間フレームより長さが短い幅であるが、増幅度算出の対象にされる時点の信号をその時間フレームの時間軸における範囲の中央位置とするとともに、子音を抽出する抽出幅の1/3の幅を中心とする雑音の判定演算において子音または音節の端点を聞き取り易くできる子音を抽出する抽出幅の1/3の幅もしくはこれと同等に雑音判定に用いることが可能な信号を抽出できる子音を抽出する抽出幅の1/3近傍の幅まで含む第1の時間フレームの幅のことである。

30

【発明の効果】

40

【0026】

本発明の子音加工装置、音声情報伝達装置及び子音加工方法によれば、複数の時間フレームによって複数のフレーム信号を抽出し、このフレーム信号の平均パワーを計算して比較するだけで子音強調が行えるから、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声以外の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音または音節の端点が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができる。また、安価に製造できる子音加工装置、音声情報伝達装置を提供することができる。

【発明を実施するための最良の形態】

50

【 0 0 2 7 】

本発明の第 1 の形態は、1 音節の長さが 1 0 0 m s ~ 4 0 0 m s、子音の長さが数十 m s として表される何れかの長さをもつ音声を対象にして子音強調する子音加工装置であって、入力された音声信号に対して子音の長さを含む数十 m s の幅を抽出できる時間フレームによって第 2 のフレーム信号を抽出しかつ増幅度算出の対象にされる時点の信号を時間フレームの時間軸における範囲の中央位置の信号に設定する第 2 の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第 2 の時間フレームよりも狭い時間軸の幅を有し第 2 のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十 m s の幅の 1 / 3 の幅またはこの 1 / 3 と同等に雑音判定が行える子音の長さを含む数十 m s の幅の 1 / 3 程度の幅を備えた時間フレームによって第 1 のフレーム信号を抽出する第 1 の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第 2 の時間フレームよりも広い時間軸の幅を有し第 2 のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する時間フレームによって音声信号から音節の 1 ~ 3 倍の長さの幅の第 3 のフレーム信号を抽出する第 3 の時間フレームと、を有するフレーム分割部と、第 1、第 2 及び第 3 のフレーム信号のそれぞれで平均パワーを算出するパワー算出部と、第 1 と第 2 のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、第 2 と第 3 のフレーム信号の平均パワーの比較を行う比較部と、比較部の比較結果に基づいて差が閾値を越えたときには雑音と判定すると共に、差が閾値を越えずかつ第 2 のフレーム信号の平均パワーが第 3 のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定する子音判定部と、子音判定部が子音または音節の端点と判断した場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅すると共に、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、第 1 と第 2 のフレーム信号の平均パワーが、音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いられると共に、第 2 と第 3 のフレーム信号の平均パワーが、増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置である。この構成により、複数の時間フレームによって複数のフレーム信号を抽出し、このフレーム信号の平均パワーを計算して比較するだけで子音強調が行えるから、並列的に様々の処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音または音節の端点が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができ、安価に製造できる子音加工装置を提供することができる。この子音加工装置は C V 型の日本語でも C V 型以外の外国語でも子音強調が行える汎用性のあるものである。

【 0 0 2 8 】

本発明の第 2 の形態は、1 音節の長さが 1 0 0 m s ~ 4 0 0 m s、子音の長さが数十 m s として表される何れかの長さをもつ音声を対象にして子音強調する子音加工装置であって、入力された音声信号に対して子音の長さを含む数十 m s の幅を抽出できる時間フレームによって第 2 のフレーム信号を抽出しかつ増幅度算出の対象にされる時点の信号を時間フレームの時間軸における範囲の中央位置の信号に設定する第 2 の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第 2 の時間フレームよりも狭い時間軸の幅を有し第 2 のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十 m s の幅の 1 / 3 の幅またはこの 1 / 3 と同等に雑音判定が行える子音の長さを含む数十 m s の幅の 1 / 3 程度の幅を備えた時間フレームによって第 1 のフレーム信号を抽出する第 1 の時間フレームと、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第 2 の時間フレームよりも広い時間軸の幅を有し第 2 のフレーム信号との平均パワーの比較において子音または音節

10

20

30

40

50

の端点の判定をするための信号を音声信号から抽出する時間フレームによって音声信号から音節の1～3倍の長さの幅の第3のフレーム信号を抽出する第3の時間フレームと、を有するフレーム分割部と、第1、第2及び第3のフレーム信号のそれぞれで平均パワーを算出するパワー算出部と、第1と第2のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、第2と第3のフレーム信号の平均パワーの比較を行う比較部と、比較部の比較結果に基づいて差が閾値を越えたときには雑音と判定すると共に、差が閾値を越えずかつ第2のフレーム信号の平均パワーが第3のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定する子音判定部と、子音判定部が子音または音節の端点と判断した場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅すると共に、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、子音判定部が子音または音節の端点と判断した場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号の増幅度を増幅方向に決定すると共に、子音または音節の端点でないと判断した場合は音声信号を増幅しない旨決定する増幅度決定部と、増幅度決定部が決定した増幅度に応じて音声信号を増幅する増幅部とを備え、第1と第2のフレーム信号の平均パワーが、音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いられると共に、第2と第3のフレーム信号の平均パワーが、増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置である。この構成により、第1の形態の作用効果に加えて、増幅度決定部によって増幅度を調整でき、さらに聞き取り容易な子音加工装置を提供することができる。この子音加工装置はC V型の日本語でもC V型以外の外国語でも子音強調が行える汎用性のあるものである。

【0029】

本発明の第3の形態は、第1または第2の形態に従属する形態であって、比較部が、第1と第2のフレーム信号のデシベル表示した平均パワーの差を算出し、閾値をこの差と比較することを特徴とする子音加工装置であり、差を演算するだけであるから容易に信号処理でき、リアルタイムに近い時間内に音声情報伝達が行える。

【0030】

本発明の第4の形態は、第1または第2の形態に従属する形態であって、比較部が、各フレーム信号の平均パワーの差に代えて、平均パワーの比率を算出することにより比較することを特徴とする請求項1または2に記載された子音加工装置であり、比率を演算するだけであるから容易に信号処理でき、リアルタイムに近い時間内に音声情報伝達が行える。

【0032】

本発明の第5の形態は、第1から第4の何れかの形態の子音加工装置において、第3の時間フレームは増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2の時間フレームよりも広い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する時間フレームによって第3のフレーム信号を抽出し、音声信号の増幅度算出の対象にされる時点の信号が第2の時間フレームの時間軸における範囲の中央位置の信号に設定され、第3のフレーム信号の平均パワーが、増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられるのに代えて、第2の時間フレームに連続して後置される第4の時間フレームによって第4のフレーム信号を抽出し、音声信号の増幅度算出の対象にされる時点の信号が第2の時間フレームの時間軸における範囲の後端となる第4の時間フレームとの境界の信号に設定され、第4のフレーム信号の平均パワーが、増幅度算出の対象にされる時点の音声信号に対する子音または音節の端点か否かの判定のために用いられることを特徴とする子音加工装置であり、前記形態の作用効果に加えて、この子音加工装置は、さらに信号処理が簡単になり、日本語等のC V型の言語の子音強調を効果的に行えるものである。

【0037】

本発明の第6の形態は、第1～5の形態の子音加工装置において、平均パワー間のフレーム分割部に音声信号を入力する前に、所定の周波数成分を通過させる通過特性を有するフィルタ部が設けられたことを特徴とする子音加工装置であり、子音強調の明瞭性を増すことができる。

【0039】

本発明の第7の形態は、第1～第6のいずれかの形態に従属する子音加工装置と、該子音加工装置からの子音加工された音声信号に基づいて子音強調された音声を出力するスピーカを備えたことを特徴とする音声情報伝達装置である。この構成により、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音または音節の端点が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができ、安価な音声情報伝達装置を提供することができる。

【0040】

本発明の第8の形態は、1音節の長さが100ms～400ms、子音の長さが数十msとして表される何れかの長さをもつ音声を対象にして子音強調するとき、入力された音声信号に対して増幅度算出の対象にする時点の信号を、子音の長さを含む数十msの幅を抽出できる第2の時間フレームの時間軸における範囲の中央位置の信号に設定し、この第2の時間フレームにより音声信号から第2のフレーム信号を抽出し、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2の時間フレームよりも狭い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える子音の長さを含む数十msの幅の1/3程度の幅を備えた第1の時間フレームによって音声信号から時間軸上で第1のフレーム信号を抽出すると共に、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2のフレームよりも広い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する第3の時間フレームによって音声信号から音節の1～3倍の長さの幅の第3のフレーム信号を抽出し、第1、第2及び第3のフレーム信号のそれぞれで平均パワーを算出し、第1と第2のフレーム信号の平均パワーの差を計算してこの差が閾値を越えるか否かを比較すると共に、第2と第3のフレーム信号の平均パワーの比較を行い、この比較結果に基づいて差が閾値を越えたときには雑音と判定すると共に、差が閾値を越えずかつ第2のフレーム信号の平均パワーが第3のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定し、子音または音節の端点と判断される場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅し、子音または音節の端点でないと判断した場合は増幅せず、第1と第2のフレーム信号の平均パワーを音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いると共に、第2と第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法である。この構成により、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音または音節の端点が聞き取り易くなり、これにより音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができる。この子音加工方法はC V型の日本語でもC V型以外の外国語でも子音強調が行える汎用性のあるものである。

【0041】

本発明の第9の形態は、1音節の長さが100ms～400ms、子音の長さが数十msとして表される何れかの長さをもつ音声を対象にして子音強調するとき、入力された音声信号に対して増幅度算出の対象にする時点の信号を、子音の長さを含む数十msの幅を

10

20

30

40

50

抽出できる第2の時間フレームの時間軸における範囲の中央位置の信号に設定し、この第2の時間フレームにより音声信号から第2のフレーム信号を抽出し、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2の時間フレームよりも狭い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える子音の長さを含む数十msの幅の1/3程度の幅を備えた第1の時間フレームによって音声信号から時間軸上で第1のフレーム信号を抽出すると共に、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2のフレームよりも広い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する第3の時間フレームによって音声信号から音節の1～3倍の長さの幅の第3のフレーム信号を抽出し、第1と第2のフレーム信号の平均パワーの比率を計算してこの比率が閾値を越えるか否かを比較すると共に、第2と第3のフレーム信号の平均パワーの比較を行い、この比較結果に基づいて比率が閾値を越えたときには雑音と判定すると共に、比率が閾値を越えずかつ第2のフレーム信号の平均パワーが第3のフレーム信号の平均パワーより小さい場合に音声信号の増幅度算出の対象にされる時点の信号が子音または音節の端点であると判定し、子音または音節の端点と判断される場合は音声信号の増幅度算出の対象にされる時点の信号またはその時点における所定の時間幅の信号を算出で得られた増幅度で増幅し、子音または音節の端点でないと判断した場合は増幅せず、第1と第2のフレーム信号の平均パワーを音声信号の増幅度算出の対象にされる時点の信号に対する雑音か否かの判定に用いると共に、第2と第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法である。この構成により、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声以外の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音または音節の端点が聞き取り易くなり、これにより音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができる。この子音加工方法はC V型の日本語でもC V型以外の外国語でも子音強調が行える汎用性のあるものである。

【0042】

本発明の第10の形態は、第8または第9の形態の子音加工方法において、増幅度算出の対象にされる時点の信号が時間軸の範囲の中央位置に配置され第2のフレームよりも広い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において子音または音節の端点の判定をするための信号を音声信号から抽出する第3の時間フレームによって第3のフレーム信号を抽出し、かつ増幅度算出の対象にされる時点の信号を第2の時間フレームの時間軸の範囲において中央位置の信号に設定し、第3のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いるのに代えて、第2の時間フレームに連続して後置される第4の時間フレームによって第4のフレーム信号を抽出し、かつ音声信号の増幅度算出の対象にされる時点の信号を第2の時間フレームの時間軸の範囲の後端となる第4の時間フレームとの境界に設定し、第4のフレーム信号の平均パワーを子音または音節の端点か否かの判定のために用いることを特徴とする子音加工方法であり、前記形態の作用効果に加えて、この子音加工方法は、さらに信号処理が簡単になり、日本語等のC V型の言語の子音強調を効果的に行えるものである。

【実施例】

【0043】

(実施例1)

以下、本発明の実施例1における子音加工装置と音声情報伝達装置、子音加工方法について説明する。

【0044】

実施例1の音強調処理装置が行う子音強調は、日本語のように音節がC V型をもつ言語の構造の場合に、すなわち子音(C)の直後に母音(V)が続く頻度が高い場合の子音強調に好適なものである。なお、本明細書において子音強調というが、これは子音だけでなく

10

20

30

40

50

音節の端点も含めて強調するものである。

【0045】

図1は本発明の実施例1における子音加工装置とこれを搭載した音声情報伝達装置の構成図、図2は本発明の実施例1における子音加工装置の処理の説明図、図7(a)は本発明の実施例1における増幅時の増幅度の説明図である。

【0046】

図1において、1は音声信号が入力されると図2に示すように複数の時間フレームでそれぞれフレーム信号を抽出するフレーム分割部である。そして、1aはフレーム分割部1を構成し、子音の長さの1/3程度の幅のフレーム信号を抽出するための第1時間フレーム、また、1bは第1時間フレーム1aを包含し、子音を抽出可能な抽出幅の第2時間フレーム、さらに1cは第2時間フレーム1bを包含し音節の長さの1~3倍程度を抽出可能な第3時間フレームである。ここで、本発明における「子音の長さを含む数十msの幅を抽出できる時間フレーム」とは、子音を抽出可能な数十msの長さであって、言語等ごとに想定される子音の長さまたはこれより少し長く、雑音の判定演算、子音または音節の端点の判定演算において、子音または音節の端点を聞き取り易くできるような子音程度の長さの幅の第2の時間フレームのことである。また、「第2の時間フレームよりも狭い時間軸の幅を有し第2のフレーム信号との平均パワーの比較において雑音判定するための信号を抽出する子音の長さを含む数十msの幅の1/3の幅またはこの1/3と同等に雑音判定が行える子音の長さを含む数十msの幅の1/3程度の幅を備えた時間フレーム」とは、第2の時間フレームより長さが短い幅であるが、増幅度算出の対象にされる時点の信号をその時間フレームの時間軸における範囲の中央位置とするとともに、子音を抽出する抽出幅の1/3の幅を中心とする雑音の判定演算において子音または音節の端点を聞き取り易くできる子音を抽出する抽出幅の1/3の幅もしくはこれと同等に雑音判定に用いることが可能な信号を抽出できる子音を抽出する抽出幅の1/3近傍の幅まで含む第1の時間フレームの幅のことである。

【0047】

第1時間フレーム1a、第2時間フレーム1b、第3時間フレーム1cは方形窓、ハミング窓等の窓関数を乗じる機能を備えたもので、実施例1においては方形窓が採用されている。

【0048】

すなわち、 $t = T$ の時点の音声信号に対して、第1時間フレーム1aは窓関数 $w_1(t) = 1$ （ここで $T - \tau_1 \leq t \leq T + \tau_1$ ）、 $w_1(t) = 0$ （それ以外するとき）で構成され、第2時間フレーム1bは窓関数 $w_2(t) = 1$ （ここで $T - \tau_2 \leq t \leq T + \tau_2$ ）、 $w_2(t) = 0$ （それ以外するとき）、第3時間フレーム1cも窓関数 $w_3(t) = 1$ （ここで $T - \tau_3 \leq t \leq T + \tau_3$ ）、 $w_3(t) = 0$ （それ以外するとき）で構成される。いずれも単位はmsである。

【0049】

なお、第1時間フレーム1a、第2時間フレーム1b、第3時間フレーム1cの中央位置($t = T$)は全て一致する必要はないが、図2に示す音声信号の波形においては中央位置が一致しており、この中央位置がこれらの時間フレームにおける音声信号の増幅対象点である。従って、全ての時点Tについてその時点Tでの増幅度を計算するときこの点が増幅対象点になる。すなわち増幅対象点は音声信号の増幅度算出の対象にされる時点である。そして、これは点でなく、点に代わる所定の幅における信号であってもよい。この場合、本明細書ではこれを増幅対象幅という。点に代わってその時点における所定の時間幅の信号に対する増幅になる。この増幅位置または増幅対象幅は、少なくとも第2時間フレーム1bの中央位置に設定されるのが好適である。なお、第2時間フレーム1bは第3の時間フレーム1cの後端部よりに設けられる方が処理の速さを高める可能性があり、第2時間フレーム1bの中央位置をこの後端部よりに配置することもできる。

【0050】

窓関数の τ_1, τ_2, τ_3 は、子音強調できるように経験的に定められるパラメータで

あり、実施例 1 においては $\tau_1 = 7.5 \text{ ms}$ 程度、 $\tau_2 = 25 \text{ ms}$ 程度（子音を抽出可能な長さ）、 $\tau_3 = 200 \text{ ms}$ 程度（音節の長さの 1 ~ 3 倍程度を抽出可能な長さ）に設定される。日本語の場合、一般的に子音の長さは数十 ms 程度、1 音節の長さは 100 ~ 400 ms の程度である。

【0051】

従って、音声信号 $p(t)$ に対して、第 1 時間フレーム 1 a からは $y_1(t) = w_1(t) \cdot p(t)$ が出力され、第 2 時間フレーム 1 b からは $y_2(t) = w_2(t) \cdot p(t)$ 、第 3 時間フレーム 1 c からは $y_3(t) = w_3(t) \cdot p(t)$ の演算によりフレーム信号が抽出される。デジタル信号の場合は、例えば $y_3(t)$ を説明すると、 $T - \tau_3 < t < T$ 、 $T < t < T + \tau_3$ の間がそれぞれ N 個の時系列値と $t = T$ の時系列値とからなり、全体で $(2N + 1)$ 個の時系列値で演算される。 $y_1(t)$ 、 $y_2(t)$ の時系列値も同様で、 $y_3(t)$ の入力時系列値と重複した値を用いる。

10

【0052】

このように本実施例 1 の子音加工装置は、フレーム分割部 1 などで A/D 変換を行い、後述する増幅度の決定などまで含めてデジタル回路もしくはプロセッサでデジタル処理しているが、アナログ回路を使ってアナログ処理することもできる。なお、プロセッサでデジタル処理するときは、図示はしないが、メモリを設けてプログラムや設定値を格納し、これを読み出して演算する。

【0053】

次に、図 1 において、2 は第 1 時間フレーム 1 a、第 2 時間フレーム 1 b、第 3 時間フレーム 1 c で抽出されたフレーム信号 $y_1(t)$ 、 $y_2(t)$ 、 $y_3(t)$ の平均パワーを計算するパワー算出部である。2 a は第 1 パワー算出部であり、第 1 時間フレーム 1 a から出力された $y_1(t)$ の振幅の 2 乗である平均パワー P_1 をデシベル表示した平均パワー L_1 を演算する。同様に、2 b は第 2 パワー算出部であり、第 2 時間フレーム 1 b から出力された $y_2(t)$ の平均パワー P_2 をデシベル表示した平均パワー L_2 を演算する。さらに、2 c は第 3 パワー算出部であって、第 3 時間フレーム 1 c から出力された $y_3(t)$ の平均パワー P_3 をデシベル表示した平均パワー L_3 を演算する。なお、平均パワー P_i ($i = 1, 2, 3$) は (数 1) で表され、 $2N + 1$ は時系列値のそれぞれの総数である。平均パワー L_1, L_2, L_3 の単位は dB である。

20

【0054】

【数 1】

$$P_i = \frac{\sum_{n=1}^{2N+1} \{y_{in}(t)\}^2}{2N+1}$$

30

【0055】

なお、以下、デシベル表示した平均パワー L_i ($i = 1, 2, 3$) を使って差で説明するが、平均パワー P_i の比率 $K_{ij} = P_i / P_j$ ($i, j = 1, 2, 3; i < j$) を使用して演算することもできる。この比率 K_{ij} を使用した場合の説明は後述の実施例 4 で行う。さらに、平均パワー P_i の対数をとってデシベル表示の L_i ($i = 1, 2, 3$) とするのではなく、平均パワー P_i 自体の差 $P_i - P_j$ を演算しても同様の作用効果が得られる。しかし詳細な説明はデシベル表示の説明に譲って省略する。

40

【0056】

続いて、3 は各フレーム信号の平均パワー L_1, L_2, L_3 の差を計算して比較する比較部であり、4 は比較部 3 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である。また、5 は増幅までの処理に必要な時間だけ音声信号を遅延しあるい

50

はデータをバッファする遅延部、6は子音判定部4が子音と判断したとき第2時間フレーム1bの音声信号の増幅対象点または増幅対象幅の信号の増幅度を変更し、子音でないと判断した場合はその増幅度を変更しない増幅部である。

【0057】

そして、10は音声信号を入力されたとき子音を強調して出力する実施例1の子音加工装置であり、定用途向け集積回路などとして構成される。また、11は音声を入力するためのマイク、12は音声を出力するためのスピーカ、20は子音加工装置10を搭載した音声情報伝達装置である。

【0058】

音声情報伝達装置20は、子音加工装置10によって子音強調された音声信号をスピーカ12から出力し、上述のワイヤレス放送システム、構内放送や案内放送などのアナウンス放送装置、携帯端末等の携帯型情報機器、その他の音声情報伝達装置、補聴器などに利用できる。なお、マイク11を備えていない音声情報伝達装置20の場合、例えば、自動販売機やATMのガイド音声などの場合は、予め録音された音信号について子音加工装置10による音の加工を行えばよい。

【0059】

さて、実施例1の子音加工装置10は、日本語のように子音、母音が続くCV型の構造をもつ言語に有効な装置であり、比較部3はこのような構造を利用して以下のような基準で各フレーム信号の平均パワーの比較を行い、子音判定部4が子音または音節の端点か否かを判定する。

【0060】

すなわち、先ず第1に、デシベル表示の平均パワー L_1 がデシベル表示の平均パワー L_2 より所定の閾値（実施例1では5dB）以上高い場合（すなわち $L_1 > L_2 + 5$ ）は、15ms（子音の長さの約1/3程度）程度のごく狭い幅で振幅が増加しているだけであるから、この増加は雑音の増加とみなす。比較部3は $L_1 - L_2$ を計算し、閾値より大きいか、以下かを算出する。閾値より大きい場合、子音判定部4は音声信号を雑音と判断する。閾値以下の場合、次の基準で判定される。

【0061】

第2に、 $L_1 - L_2$ が閾値（5dB）以下であって、 $L_2 < L_3$ であれば、第2時間フレーム1bにおける50ms（子音の長さより少し長い）の間の平均パワー L_2 が、第3時間フレーム1cの400ms（数音節の長さ）の平均パワー L_3 より低いことになり、ここには子音または音節の端点があると考えられる。

【0062】

すなわち、子音、母音と続くとき、子音または音節の端点は母音に比べて平均パワーが小さいため、 L_2 と L_3 のレベルを比較して L_2 が小さければ第2時間フレーム1bに子音または音節の端点があると推定するものである。この状態を図2に示す。

【0063】

図2において、窓関数 $w_2(t)$ で抽出したフレーム信号の平均パワー L_2 は小さく、窓関数 $w_3(t)$ で抽出したフレーム信号の平均パワー L_3 は平均パワー L_2 に比べて大きい。それ故、第2時間フレーム1bで抽出したフレーム信号は子音または音節の端点であって、前後、あるいは前の部分または後の部分に母音が存在すると推定できる。このとき比較部3は $L_1 - L_2$ を計算し、 $L_2 < L_3$ であれば、子音判定部4は第2時間フレーム1bのフレーム信号が子音または音節の端点と判断し、増幅を行う。

【0064】

これにより実施例1の増幅部6は、以上説明した子音判定部4の判定に基づいて、 $L_1 - L_2 > 5$ の場合には増幅を行わず、 $L_1 - L_2 < 5$ であって、 $L_2 - L_3$ が0～-20dBの範囲内であれば一定の増幅度 α_0 、例えば10dB増幅する。但し、 $L_1 - L_2 < 5$ であっても、 $L_2 - L_3 < -20$ の条件を充たす場合、雑音との判別が難しくなるため増幅部6は増幅の程度を弱める。なお、このとき子音判定部4がこの判断を行うのでもよい。このような増幅特性を図示すると、図7(a)のようになる。きわめて簡単な構成で

10

20

30

40

50

容易に子音強調が可能になる。なお、図7(a)は一例としてあげたもので、急に増幅をやめると、増幅度が不連続に変化して違和感のある音声となるので、一点鎖線のような低減の仕方、若しくは、さらにより滑らかな低減の仕方をするのが好ましい。

【0065】

このように実施例1の子音加工装置、音声情報伝達装置及び子音加工方法は、複数の時間フレームによって複数のフレーム信号を抽出し、このフレーム信号の平均パワーを計算して比較するだけで子音強調が行えるから、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、論理判断が少なく信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、境騒音が増加するのを防ぐことができる。また、構成が簡単で安価に製造できる子音強調処理装置等の子音加工装置、音声情報伝達装置を提供することができる。

10

【0066】

(実施例2)

以下、本発明の実施例2における子音加工装置と音声情報伝達装置、子音加工方法について説明する。実施例2の子音加工装置と音声情報伝達装置は、子音と判断されたときに、比較部の比較結果に応じて増幅度を調整するものである。そして、実施例2の子音強調もとくにCV型の言語の子音強調に好適なものである。

【0067】

図3は本発明の実施例2における子音加工装置とこれを搭載した音声情報伝達装置の構成図、図4(a)は補充現象の説明図、図4(b)は静寂な環境での音と雑音中での音の比較図、図7(b)は本発明の実施例2における増幅時の増幅度の説明図の説明図である。なお、実施例2と実施例1とで同一符号は同様の構成であり、音声信号の処理も比較結果に応じて増幅度を決定する点を除いては実施例1と同様であるから、これらの説明は省略する。

20

【0068】

図3において、1はフレーム分割部、1aは第1時間フレーム、1bは第2時間フレーム、1cは第3時間フレームである。また、2はパワー算出部、2aは第1パワー算出部、2bは第2パワー算出部、2cは第3パワー算出部、3は比較部であり、4は子音判定部、5は遅延部、6は増幅部である。そして、10は実施例2の子音加工装置、11はマイク、12はスピーカ、20は実施例2の音声情報伝達装置である。これらは実施例1と同様の構成である。

30

【0069】

実施例2の特徴的な点は、実施例1が一定の増幅度 α で増幅したのに対して、比較部3の比較結果に応じて増幅度 β を調整する点である。図3において、7は増幅度 β の値を決定する増幅度決定部である。

【0070】

増幅度決定部7は、音比較部3が子音または音節の端点と判断した場合に音声信号の増幅対象点または増幅対象幅の増幅度を決定し、子音または音節の端点でないと判断した場合は音声信号を増幅しない旨決定する。増幅度 β は、デシベル表示で $L_2 < L_3$ であれば、音声信号が雑音ではないと判断されたとき、 L_2 と L_3 のレベル差が大きければ大きいほど大きな値にされる。

40

【0071】

ここで、実施例2で行う増幅度 β について説明する。実施例2の増幅度 β は感音性難聴者に対する補充現象の補正特性と同様の特性を採用する。この補充現象とは、物理的な刺激音の音圧の増加に対して聴覚心理における感覚量である音の大きさの増減が、所定範囲の強さの音に対して健聴者より感音性難聴者の方が大きい現象のことである。

【0072】

このため、補充現象においては、図4(a)に示すように、所定範囲の強さの音に対し

50

て、小さい音ほど大きく補正し、大きな音ほど小さく補正する。これにより、人間の聴覚系が対応し、弱くて聞き取りにくい音を聞き取り易い強さの音に補正することができる。図4(b)は、静寂な環境と雑音中と主観的に認識できる音にどれだけの差が生じるのか、実験したものである。図中Pは静寂な環境で聞いたときの音の強さであり、Nは同一の音を白色雑音中で聞いたときの音の強さである。この結果をみると、Nは図4(a)の補充現象と同様な特性を示している。

【0073】

従って、実施例2で行う増幅は、増幅対象点の音声信号の増幅度を以下のように調整する。先ず第1に、デシベル表示の平均パワー L_1 がデシベル表示の平均パワー L_2 より所定の閾値(実施例2では5dB)以上高い場合(すなわち $L_1 > L_2 + 5$)は、15ms(子音の長さの約1/3程度)程度のごく狭い範囲で振幅が増加しているだけであるから、この増加は雑音の増加とみなし、 $L_1 - L_2$ を計算して、閾値より大きいか、以下かを算出する。閾値より大きい場合、子音判定部4は音声信号を雑音と判断する。閾値以下の場合、次の基準で判定する。

【0074】

第2に、 $L_1 - L_2$ が閾値(5dB)以下であって、 $L_3 - 20 < L_2 < L_3$ であれば、子音判定部4は子音または音節の端点と判断して増幅度を $= c \cdot (L_3 - L_2)$ として決定する。ここで $c = 0.9$ である。なお、デシベル表示した平均パワーの差($L_3 - L_2$)ではなく、平均パワー P_2 、 P_3 の比率 $K_{23} = P_2 / P_3$ でも表現できる。このときは $= (K_{23}^{1/2})^d$ となる。この係数 c の意味については図8を用いて実施例4で説明する。このとき、 L_3 と L_2 のレベル差若しくは K_{23} が大きいときほど大きく増幅することになり、20dBの差を2dBにまで圧縮することができる。なお、 $L_2 - L_3 = -20$ dBのときには増幅度が18dBで最大となる。

【0075】

さらに、 $L_1 - L_2$ が閾値(5dB)以下で、 $L_2 < L_3 - 20$ の場合、 L_2 が L_3 より20dB以上低い場合は、前後の音声信号と比べてパワーが小さく、無理に増幅しても雑音との判別が難しくなるため、徐々に増幅度を低下させる。例えば、図7(b)のように、増幅度を $L_2 - L_3$ が10dB下がるごとに4.5dB下げ、 $L_2 - L_3$ が-60dBのときに増幅度を0とする。しかし、図7(b)は一例としてあげたもので、ステップ状に低下させると、増幅度が不連続に変化するところで違和感のある音声となるので、より滑らかな一点鎖線のような低減の仕方をするのが好ましい。できれば、急激な変化をしないさらに滑らかな低減の仕方をするのが好ましい。

【0076】

なお、以上説明した子音加工装置、音声情報伝達装置及び子音加工方法は、子音強調処理装置、それを搭載した音声情報伝達装置、子音強調方法として有効であり、上述したように $L_3 - 20 < L_2 < L_3$ の場合に増幅度を正の値にして音声信号を増幅方向に増幅したが、逆に増幅度を負にすることにより音声信号を抑制方向にすることもできる。例えば、聴力障害者などに対する聴力検査、外国語学習者などに対する聞き取り訓練等を行う場合に、騒音を長時間にわたって聞かせ続けると聴力の低下、不快感を招くが、この手段、方法によれば、このような聴力検査装置や聞き取り訓練装置に有効な子音抑制処理装置、子音抑制方法となり、音声の加工が可能になる。

【0077】

このように実施例2は、日本語のように重要な情報が音節の始まりに存在することが多いCV型の言語のほかに、英語等のような他の多様な言語に対しても、きわめて簡単な構成で容易に子音強調が可能になる。なお、音環境、使用目的に応じて、時間フレーム1の抽出幅、偏りや、最大ゲインなどのパラメータを変えることができる。

【0078】

実施例2の子音加工装置、音声情報伝達装置及び子音加工方法は、複数の時間フレームによってそれぞれフレーム信号を抽出し、このフレーム信号の平均パワーを計算して比較するだけで子音強調が行えるから、並列的に様々な処理を行う必要がなく、リアルタイム

10

20

30

40

50

に近い時間内に音声情報伝達が行え、論理判断が少なく信号処理が簡単で、騒音下、あるいは音声が必要な音響信号と競合する状況であっても、また、難聴者、高齢者でも子音が聞き取り易くなり、これにより音声全体の強さを減らすことができ、環境が騒音化するのを防ぐことができる。また、多くの言語の子音強調に汎用的に利用でき、増幅度を簡単に調整できるため構成が簡単で安価な子音強調処理装置等の子音加工装置、音声情報伝達装置を提供することができる。

【0079】

また、子音抑制処理装置等として利用して増幅度を負にした場合、音声を子音または音節の端点が聞き取り難い音声に加工することができ、聴力検査、聞き取り訓練等に利用できる。

【0080】

(実施例3)

本発明の実施例3における子音加工装置と音声情報伝達装置、子音加工方法について説明する。実施例3の子音加工装置と音声情報伝達装置も、C V型の言語の子音強調に好適なものである。

【0081】

図5は本発明の実施例3における子音加工装置とこれを搭載した音声情報伝達装置の構成図、図6は本発明の実施例3における子音加工装置の処理の説明図、図7(c)は本発明の実施例3における増幅時の増幅度の説明図である。なお、実施例3と実施例2とで同一符号は同様の構成であり、時間フレームの構成が異なるだけで、その余の点は実施例2と同様であるから、子音加工装置と音声情報伝達装置の基本的な構成の説明は実施例2に譲って省略する。

【0082】

図5は、1はフレーム分割部、1aは第1時間フレーム、1bは第2時間フレームである。また、2はパワー算出部、2aは第1パワー算出部、2bは第2パワー算出部、3は比較部であり、4は子音判定部、5は遅延部、6は増幅部、7は増幅度決定部である。そして、10は実施例2の子音加工装置、11はマイク、12はスピーカ、20は実施例2の音声情報伝達装置である。これらは実施例2と同様の構成である。

【0083】

実施例3の特徴的な点は、実施例2が第3時間フレーム1cによって1~3音節分を抽出して前後、あるいは前の部分または後の部分の音節から増幅の判断を行ったのに対して、第2時間フレーム1bに連続して第4時間フレーム1d(本発明の第4の時間フレーム)を設け、後続の音節と比較することによって増幅度を調整する点である。

【0084】

図5において、1dは第4時間フレームであり、子音を抽出可能な抽出幅を有して第2時間フレーム1bの直後に設けられる。また、2dは第4パワー算出部であり、第4時間フレーム1dから出力されたフレーム信号 $y_4(t)$ (本発明の第4のフレーム信号)の平均パワー P_4 をデシベル表示した平均パワー L_4 を演算する。なお、平均パワー P_4 は(数1)において、 $i=4$ とする。

【0085】

実施例3の第1時間フレーム1aは窓関数 $w_1(t) = 1$ (ここで $T - \tau_1 \leq t \leq T + \tau_1$)、 $w_1(t) = 0$ (それ以外するとき)で構成され、第2時間フレーム1bは窓関数 $w_2(t) = 1$ (ここで $T - \tau_2 \leq t \leq T + \tau_2$)、 $w_2(t) = 0$ (それ以外するとき)、第4時間フレーム1dは窓関数 $w_4(t) = 1$ (ここで $T + \tau_2 \leq t \leq T + \tau_2 + 2\tau_4$)、 $w_4(t) = 0$ (それ以外するとき)で構成される。単位はmsである。 $\tau_2 = \tau_4$ であるが、 τ_2 と τ_4 を異なったパラメータとすることもできる。

【0086】

この τ_1 、 τ_2 、 τ_4 は経験的に定められるもので、実施例3においては $\tau_1 = 7.5$ ms程度、 τ_2 、 $\tau_4 = 25$ ms程度に設定される。従って、 $w_1(t) = 1$ (ここで $T - 7.5 \leq t \leq T + 7.5$)、 $w_1(t) = 0$ (それ以外するとき)で構成され、第2時間

10

20

30

40

50

フレーム 1 b は窓関数 $w_2(t) = 1$ (ここで $T - 25 \leq t \leq T + 25$)、 $w_2(t) = 0$ (それ以外するとき)、第 4 時間フレーム 1 d は窓関数 $w_4(t) = 1$ (ここで $T + 25 \leq t \leq T + 75$)、 $w_4(t) = 0$ (それ以外するとき)となる。

【0087】

次に、実施例 3 で行う増幅について説明する。まず第 1 に、平均パワー L_1 が平均パワー L_2 より所定の閾値 (実施例 2 では 5 dB) 以上高い場合 (すなわち $L_1 > L_2 + 5$) は、15 ms 程度のごく狭い範囲で振幅が増加しているだけであるから、この増加は雑音の増加とみなし、 $L_1 - L_2$ を計算して、閾値より大きいか、以下かを算出する。閾値より大きい場合、子音判定部 4 は音声信号を雑音と判断する。閾値以下の場合、次の基準で判定する。

10

【0088】

第 2 に、 $L_1 - L_2$ が閾値 (5 dB) 以下であって、 $L_4 - 20 < L_2 < L_4$ であれば、子音判定部 4 は子音または音節の端点と判断して増幅度 $G = c \cdot (L_4 - L_2)$ として決定する。ここで $c = 0.72$ である。なお、デシベル表示した平均パワーの差 ($L_4 - L_2$) ではなく、平均パワー P_2 、 P_4 の比率 $K_{24} = P_2 / P_4$ でも表現できる。このとき $G = (K_{24}^{1/2})^d$ となる。d も係数である。係数 c の意味については図 8 を用いて実施例 4 で説明する。これらは、子音と母音が交互に続く配列のとき、子音または音節の端点は母音に比べて平均パワーが小さいため、 L_2 のレベルと L_4 のレベルを比較して L_2 が小さければ、第 2 時間フレーム 1 b に子音あるいは音節の始点があると考え、増幅対象点または増幅対象幅を増幅することを意味する。

20

【0089】

なお、図 6 に示す実施例 3 の増幅対象点は、第 2 時間フレーム 1 b、第 4 時間フレーム 1 d の境界の点である。実施例 1, 2 と同様に、第 1 時間フレーム 1 a、第 2 時間フレーム 1 b の中央位置の音声信号を増幅するのでもよいが、第 2 時間フレーム 1 b、第 4 時間フレーム 1 d を設けた場合、この境界を増幅する方が効果を期待でき、実施例 3 においてはこの境界を増幅対象点としている。また、第 2 時間フレーム 1 b、第 4 時間フレーム 1 d の双方に跨って第 1 時間フレーム 1 a を配置し、境界または付近の増幅対象点または増幅対象幅を増幅するか否かを決定することもできる。このとき、第 1 時間フレーム 1 a を包含する第 5 時間フレーム (図示しない) を設けて、音声信号が雑音であるか否かを判断し、雑音でないと判断された場合にのみ増幅対象点または増幅対象幅を増幅するのが好適である。

30

【0090】

デシベル表示した平均パワー L_4 と L_2 のレベル差が大きいときほど大きく増幅し、20 dB の差を 5.6 dB にまで圧縮することができる。 $L_2 - L_4 = -20$ dB のときには増幅度が 14.4 dB で最大となる。

【0091】

なお、以上説明した子音加工装置、音声情報伝達装置及び子音加工方法は、子音強調処理装置、それを搭載した音声情報伝達装置、子音強調方法として有効であり、 $L_3 - 20 < L_2 < L_3$ の場合に増幅度 G を上げたが、逆に増幅を抑制して増幅度 G を負にすることもできる。例えば、聴力障害者などに対する聴力検査、聞き取り訓練等を行う場合に、騒音を長時間にわたって聞かせ続けると聴力の低下、不快感を招くが、この手段、方法によれば、このような聴力検査装置や聞き取り訓練装置に有効な子音抑制処理装置、子音抑制方法となり、音声の加工が可能になる。

40

【0092】

さらに、 $L_1 - L_2$ が閾値 (5 dB) 以下で、 $L_2 < L_4 - 20$ の場合、 L_2 が L_4 より 20 dB 以上低い場合は、前後の音声信号と比べてパワーが小さく、無理に増幅しても雑音との判別が難しくなるため、徐々に増幅度を低下させる。例えば、図 7 (c) のように、増幅度 G を $L_2 - L_4$ が 10 dB 下がるごとに 3.6 dB 下げ、 $L_2 - L_3$ が -60 dB のときに増幅度 G を 0 とするものである。しかし、図 7 (c) は一例としてあげたもので、ステップ状に低下させると、増幅度が不連続に変化するところで違和感のある音声

50

となるので、より滑らかな一点鎖線のような低減の仕方をするのが好ましい。

【0093】

このように実施例3は、とくに日本語やイタリア語のように重要な情報が音節の始まりに存在することが多いCV型の言語に対して、きわめて簡単な構成で容易に子音強調が可能になる。なお、音環境、使用目的に応じて、時間フレーム1の抽出幅、偏りや、最大ゲインなどのパラメータを変えることができる。

【0094】

さらに、実施例1, 2のフレーム分割は日本語でも外国語でも子音強調を行える汎用性のあるものであるが、実施例3のフレーム分割は日本語等のCV型の言語の子音強調を効果的に行えるものである。従って、実施例3のフレーム分割を単独で使用しても、実施例1, 2のフレーム分割と組合せて使用することもできる。このとき、2つの処理を並行して行い、増幅度の大きい方を選ぶようにすればよい。

【0095】

なお、実施例3の子音加工装置と音声情報伝達装置は、子音と母音の判断を逆にするだけで、VC (Vowel-Consonant) 型の子音強調処理に応用することができる。子音判定部4が子音または音節の端点でないと判断した場合には、音声信号の増幅対象点または増幅対象幅を増幅せず、子音または音節の端点と判断した場合に音声信号の増幅対象点または増幅対象幅を増幅すればよい。この場合、音節の終端部が強調され、音節の終端部が強調され、CV型言語以外の外国語の子音強調が効果的に行える。日本語においても、撥音「ん」、発声のさいに母音が脱落し無声化した音節などに対して効果がある。

【0096】

また、増幅度を負にした場合、言い換えると子音の大きさを抑制する方向の増幅になり、 音声を子音または音節の端点が聞き取り難い音声に加工することができ、聴力検査、聞き取り訓練等に利用できる。 本発明の、リアルタイムに近い時間内に音声信号の分析、加工(音声情報伝達)が行えるという目的はこのような環境の用途においても利用できる。

【0097】

以上説明したように、実施例3の子音加工装置、音声情報伝達装置及び子音加工方法は、フレーム信号の平均パワーの差を比較するだけで子音強調が行えるから、並列的に様々な処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、論理判断が少なく信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、境騒音が増加するのを防ぐことができる。また、日本語等のCV型の言語の子音強調に好適で、増幅度を簡単に調整できるため構成が簡単で、安価に製造できる音強調処理装置等の子音加工装置、音声情報伝達装置を提供することができる。そして、子音または音節の端点でないと判断した場合には、音声信号の増幅対象点または増幅対象幅を増幅せず、子音または音節の端点と判断した場合に音声信号の増幅対象点または増幅対象幅を増幅すると、音節の終端部が強調され、CV型言語以外の外国語などの子音強調が効果的に行える。

【0098】

(実施例4)

本発明の実施例4における子音加工装置と音声情報伝達装置、子音加工方法について説明する。図8は本発明の実施例4における子音加工装置の増幅特性の説明図、図9は音声刺激のパターン説明図、図10は音声刺激ごとの子音強調処理前後の正答率の比較図である。

【0099】

実施例3の比較部3はデシベル表示した平均パワー L_i ($i = 1, 2, 4$)の差を計算して増幅度を計算したが、実施例4は各時間フレームの平均パワー P_i ($i = 1, 2, 4$)の比率を計算して増幅度を計算するものである。従って、実施例4と実施例3とで同一符号は同様の構成であり、比較部3の計算方法が異なるだけで、その余の点は実施例3と

10

20

30

40

50

同様である。これらの詳細な説明は実施例 3 に譲ってここでは省略する。従って、図 5、図 6 を参照する。

【 0 1 0 0 】

実施例 4 においては、比較部 3 が各フレーム信号の平均パワー P_i ($i = 1, 2, 4$) の比率 $K_{ij} = P_i / P_j$ ($i, j = 1, 2, 4; i < j$) を計算し、増幅度決定部 7 で増幅度を算出する。 L_1 と L_2 の関係は実施例 3 と同様に比率で雑音を判別できればよい。そこで、以下 L_2 と L_4 の関係を説明する。

【 0 1 0 1 】

比較部 3 は、第 2 パワー算出部 2 b からの出力であるデシベル表示した L_2 、第 4 パワー算出部 2 d からの出力であるデシベル表示した L_4 の差 $L_2 - L_4$ を算出し、 $L_2 - L_4 > 0$ であれば増幅度決定部 7 は増幅を行わない。比率 $K_{24} = P_2 / P_4$ で判定する場合、 $K_{24} > 1$ となる。これは図 10 において A 点より L が大きい場合である。

【 0 1 0 2 】

これに対し、 $L_2 - L_4 = 0$ 、あるいは比率 $K_{24} = 1$ の場合、増幅度決定部 7 は増幅を行う。この場合、増幅度 $= c \cdot (L_2 - L_4)$ となる。図 8 においては、この c は (線分) / (線分) で表される比で表される。 c を増やすほど増幅度が大きくなり、 c

が 0 のときには音声信号が増幅されない。破線上の点の入力があつたとき、 $L_2 < L_4$ であれば、出力は線分 分持ち上げられ、点にまで増幅されることを意味する。

【 0 1 0 3 】

図 8 において B 点はニーポイント (増幅度の切り換わり点) であつて、これ以下のレベルの入出力信号はノイズと判別が難しくなるので、増幅度を下げています。図 8 の場合、ニーポイント B 点を -20 dB とし、ニーポイント B 点より小さな入出力信号に対しては、増幅度を徐々に下げ、ニーポイント B 点で増幅度が最大となる。

【 0 1 0 4 】

また、実施例 4 の子音加工装置と音声情報伝達装置は、デシベル表示した平均パワーの差 $L_2 - L_4$ が $L_2 - L_4 > 0$ 、 $L_2 - L_4 = 0$ 、あるいは平均パワー P_2, P_4 の比率 $K_{24} > 1$ と $K_{24} = 1$ における判断を逆にすることなどで、V C 型の子音強調処理に応用することができる。すなわち、増幅対象点または増幅対象幅の増幅度の判断を逆にすることで、音節の終端部が強調され、C V 型言語以外の外国語の子音強調が効果的に行える。そして、このフレーム分割を実施例 1, 2, 3 のフレーム分割と組合せて使用することもできる。組合せる場合、2 つの処理を並行して行い、増幅度の大きい方を選ぶようにすればよい。これにより、音節の最終端と判断される場合に比率 K_{24} が 1 以下であれば音声信号を増幅するので、音節の最後を明瞭にすることができる。

【 0 1 0 5 】

さて、実施例 4 の子音加工装置の有効性を確認するために、明瞭度を検証した。音声刺激としては、「人工内耳装用による語音聴取評価検査 (C I 2 0 0 4)」(日本人工内耳研究会編) に収録されている成人用子音検査の音源を用いた。この音源には、「a b a」, 「a d a」, 「a g a」, 「a h a」, 「a k a」, 「a m a」, 「a n a」, 「a p a」, 「a r a」, 「a s a」, 「a t a」, 「a w a」, 「a y a」, 「a z a」の 14 種類の V C V 音節が設けられている。この音源を 44.1 kHz で、子音強調処理を施したものと処理しないものを用意し、図 9 に示すように上限、下限周波数が 8000 Hz、 50 Hz の背景雑音を加えて、音声刺激とした。背景雑音の継続時間は 5000 ms、 500 ms の立ち上がり及び立ち下りを設け、 5000 ms の継続時間の中央に子音強調処理を施した音声刺激を配した。次の音声刺激までの時間間隔は 2000 ms とした。

【 0 1 0 6 】

この音声刺激を正常な聴力をもつ 14 人の実験参加者に与え、子音強調処理を施したものと処理しないものとの正答率を比較した。図 10 は音声刺激ごとの子音強調処理前後の正答率を比較したものである。図 10 で両者の全体の平均値を比較すると、子音強調処理を施したもののほうが処理しないものより高いことが分かる。実施例 1 の子音加工装置が有

10

20

30

40

50

効に機能していることが分かる。

【0107】

この中で、とくに「asa」、「aza」の正答率が高いのは、「s」や「z」のエネルギーの大半が8000Hz以上で背景雑音によってマスクされなかったからと考えられるし、摩擦音は摩擦性の雑音、及び前後の母音との遷移部（VOTや無音区間）に特徴があるため、「s」や「z」はこの遷移部より雑音部（子音部）に多くの音声情報をもっているとみられることから、実施例1の増幅部6による子音強調処理が有効に機能し、明瞭度を増したと考えられる。

【0108】

これに対し、「aba」、「ada」、「aga」は有声閉鎖子音であり、第2ホルマント遷移の形状が音声の識別に大きな影響を与える。有声破裂子音はこの第2ホルマント遷移に多くの音声情報を有しているとみられ、また、第2ホルマント遷移部の振幅は母音に対して大きい値を示すために、実施例1の増幅部6による子音強調処理を行ったものも行わなかったものとして、明瞭度にはそれほど差が出なかったものと考えられる。

【0109】

このように実施例4の子音加工装置、音声情報伝達装置及び子音加工方法は、フレーム信号の平均パワーの比率を比較するだけで子音強調が行えるから、並列的に様々の処理を行う必要がなく、リアルタイムに近い時間内に音声情報伝達が行え、信号処理が簡単で、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも子音が聞き取り易くなり、これにより音声の明瞭さを損なうことなく音声全体の強さを減らすことができ、環境騒音が増加するのを防ぐことができる。また、日本語等のCV型の言語の子音強調に好適で、増幅度を簡単に調整できるため構成が簡単で、安価に製造できる子音強調処理装置等の子音加工装置、音声情報伝達装置を提供することができる。

【0110】

また、実施例3と同様に増幅度を負にした場合、子音抑制処理装置等として音声を子音または音節の端点が聞き取り難い音声に加工することができ、聴力検査、聞き取り訓練等に利用できる。

【0111】

（実施例5）

以下、本発明の実施例5における子音加工装置と音声情報伝達装置、子音加工方法について説明する。図11は本発明の実施例5における子音加工装置とこれを搭載した音声情報伝達装置の構成図である。

【0112】

実施例5における子音加工装置は、音声信号の子音あるいは音節の境界をより明瞭に検出するために、予め音声信号を処理して時間フレーム1に入力するものである。

【0113】

図11に示す8はフレーム分割部1の直前に置かれたフィルタ部である。フィルタ部8は、3000Hz以下の周波数成分を通過させ1000Hz近辺にピークがあるような特性を有しており、これによって子音または音節の境界をより適切に検出することが可能になる。なお、実施例5は、実施例1の子音加工装置10と音声情報伝達装置20にフィルタ部8を設けたものを示しているが、フィルタ部8を実施例2～4の子音加工装置10と音声情報伝達装置20に設けるのでも同様である。これらは図示しない。

【0114】

このように実施例5の子音加工装置、音声情報伝達装置及び子音加工方法は、簡単に子音若しくは音節の境界を明瞭に検出することができ、騒音下、あるいは音声が他の音響信号と競合する状況であっても、また、難聴者、高齢者でも聞き取り易くなる。

【産業上の利用可能性】

【0115】

本発明は、アナウンス放送装置や携帯電話等、補聴器等の音声情報伝達装置に適用できる。

10

20

30

40

50

【図面の簡単な説明】

【0116】

【図1】本発明の実施例1における子音加工装置とこれを搭載した音声情報伝達装置の構成図

【図2】本発明の実施例1における子音加工装置の処理の説明図

【図3】本発明の実施例2における子音加工装置とこれを搭載した音声情報伝達装置の構成図

【図4】(a)補充現象の説明図、(b)静寂な環境での音と雑音中での音の比較図

【図5】本発明の実施例3における子音加工装置とこれを搭載した音声情報伝達装置の構成図

10

【図6】本発明の実施例3における子音加工装置の処理の説明図

【図7】(a)本発明の実施例1における増幅時の増幅度の説明図、(b)本発明の実施例2における増幅時の増幅度の説明図、(c)本発明の実施例3における増幅時の増幅度の説明図

【図8】本発明の実施例4における子音加工装置の増幅特性の説明図

【図9】音声刺激のパターン説明図

【図10】音声刺激ごとの子音強調処理前後の正答率の比較図

【図11】本発明の実施例5における子音加工装置とこれを搭載した音声情報伝達装置の構成図

【符号の説明】

20

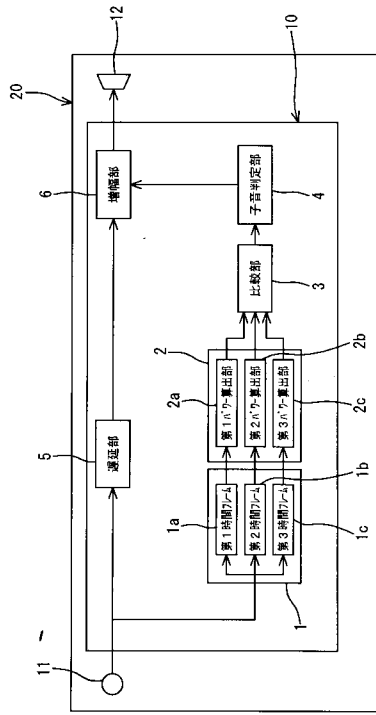
【0117】

- 1 フレーム分割部
- 1 a 第1時間フレーム
- 1 b 第2時間フレーム
- 1 c 第3時間フレーム
- 1 d 第4時間フレーム
- 2 パワー算出部
- 2 a 第1パワー算出部
- 2 b 第2パワー算出部
- 2 c 第3パワー算出部
- 2 d 第4パワー算出部
- 3 比較部
- 4 子音判定部
- 5 遅延部
- 6 増幅部
- 7 増幅度決定部
- 8 フィルタ部
- 10 子音加工装置
- 11 マイク
- 12 スピーカ
- 20 音声情報伝達装置

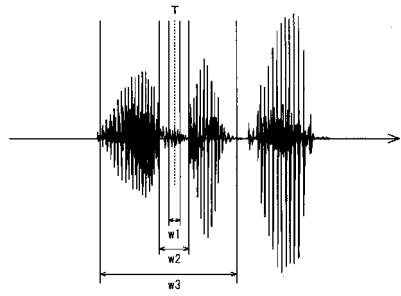
30

40

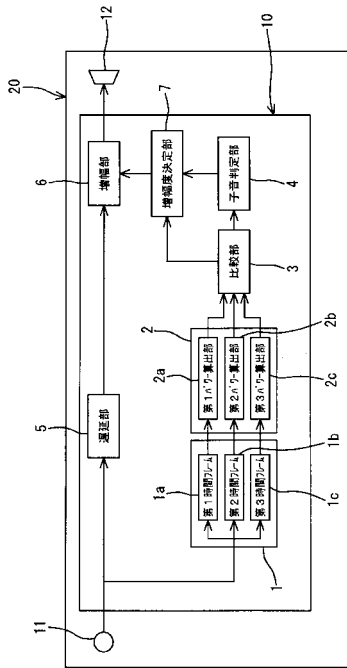
【図1】



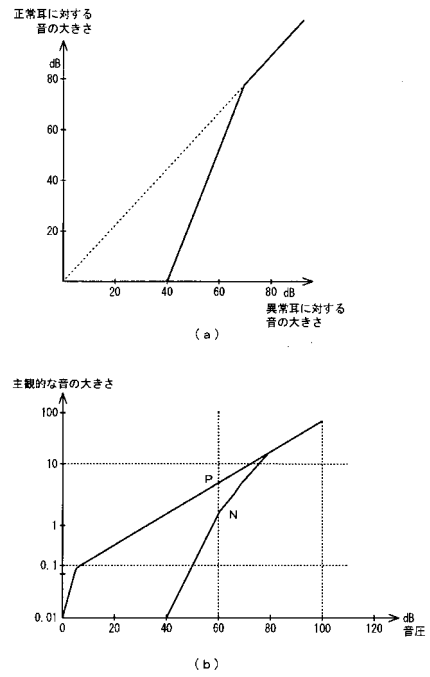
【図2】



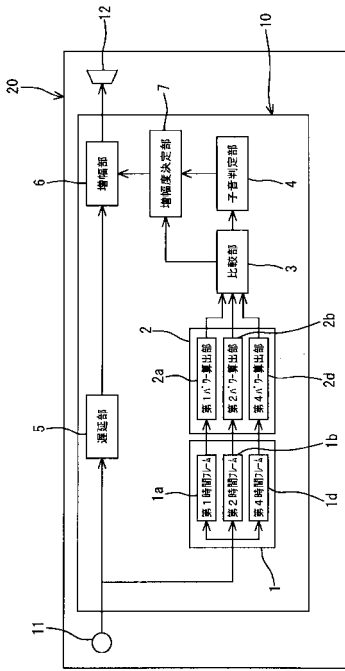
【図3】



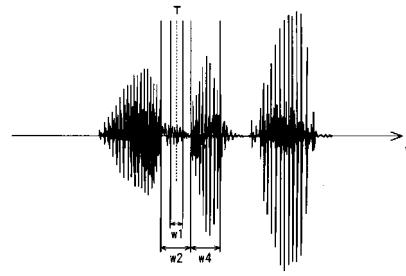
【図4】



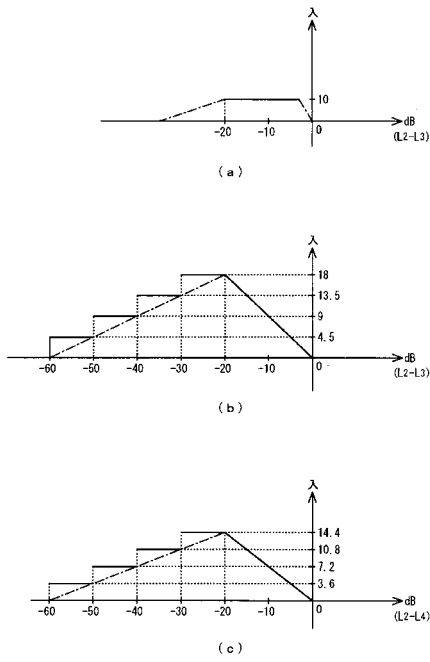
【 図 5 】



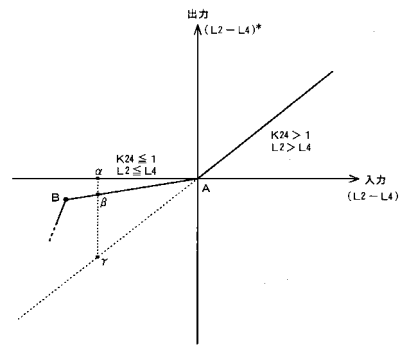
【 図 6 】



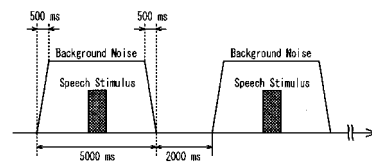
【 図 7 】



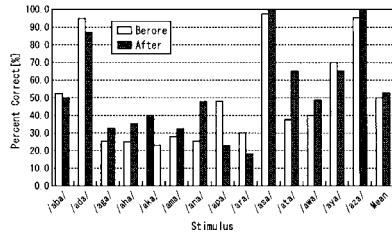
【 図 8 】



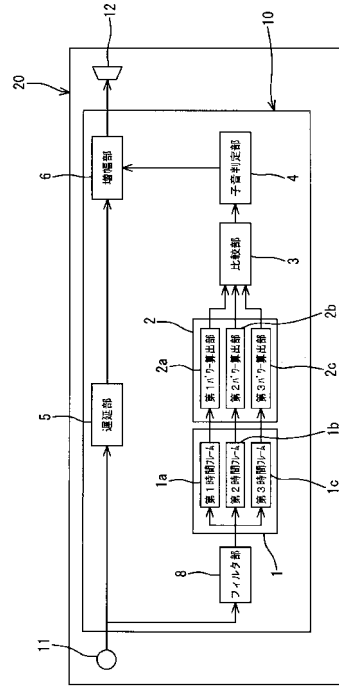
【 図 9 】



【図10】



【図11】



フロントページの続き

- (56)参考文献 特開平09 - 016193 (JP, A)
特開平02 - 083595 (JP, A)
特開平07 - 129190 (JP, A)
特開平01 - 260499 (JP, A)
特開昭60 - 147797 (JP, A)
国際公開第2004 / 111996 (WO, A1)

- (58)調査した分野(Int.Cl., DB名)
G10L 11/00 - 21/06