

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6231762号
(P6231762)

(45) 発行日 平成29年11月15日(2017.11.15)

(24) 登録日 平成29年10月27日(2017.10.27)

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

請求項の数 7 (全 12 頁)

(21) 出願番号	特願2013-82330 (P2013-82330)	(73) 特許権者	000004352
(22) 出願日	平成25年4月10日 (2013.4.10)		日本放送協会
(65) 公開番号	特開2014-206559 (P2014-206559A)		東京都渋谷区神南2丁目2番1号
(43) 公開日	平成26年10月30日 (2014.10.30)	(73) 特許権者	591053926
審査請求日	平成28年3月1日 (2016.3.1)		一般財団法人NHKエンジニアリングシステム
			東京都世田谷区砧一丁目10番11号
		(74) 代理人	100147485
			弁理士 杉村 憲司
		(74) 代理人	100161148
			弁理士 福尾 誠
		(72) 発明者	中山 靖茂
			東京都世田谷区砧一丁目10番11号 一般財団法人NHKエンジニアリングシステム内

最終頁に続く

(54) 【発明の名称】 受信装置及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

放送された番組音声を受信する第1受信部と、
 インターネット通信により前記番組音声のダイアログ音声を受信する第2受信部と、
 前記番組音声及び前記ダイアログ音声に基づいて、各フレームの有効性を判断する判断部と、

前記判断部で有効と判断されたフレームの前記番組音声及び前記ダイアログ音声に基づいて、前記番組音声に含まれるダイアログ音声のゲイン量を推定する推定部と、

前記ゲイン量に基づいて前記ダイアログ音声を調整する調整部と、

前記調整部で調整された前記ダイアログ音声と前記番組音声との合成を行う合成部と、
 を備える受信装置。

10

【請求項 2】

前記判断部が、前記ダイアログ音声のフレームエネルギー、及び前記番組音声と前記ダイアログ音声との類似性に基づいて、前記各フレームの有効性を判断する、請求項1に記載の受信装置。

【請求項 3】

前記判断部が、前記番組音声及び前記ダイアログ音声の帯域を複数に分割して、少なくとも1つの帯域で類似性を有するフレームを有効と判断する、請求項2に記載の受信装置。

【請求項 4】

20

前記推定部が、有効と判断されなかったフレームの前記ゲイン量を、1つ前のフレームのゲイン量と同一とする、請求項1～3の何れか一項に記載の受信装置。

【請求項5】

前記調整部が、前記ダイアログ音声を逆位相化して前記ゲイン量を乗算することにより、前記ダイアログ音声を調整する、請求項1～4の何れか一項に記載の受信装置。

【請求項6】

前記合成部で合成した結果と、第2受信部によって受信したダイアログ音声のうち、前記ダイアログ音声とは異なるダイアログ音声とを更に合成する第2合成部を備える、請求項1～5の何れか一項に記載の受信装置。

【請求項7】

受信装置のコンピュータに、
放送された番組音声を取得するステップと、
インターネット通信により前記番組音声のダイアログ音声を取得するステップと、
前記番組音声と前記ダイアログ音声に基づいて、各フレームの有効性を判断するステップと、

有効と判断されたフレームの前記番組音声及び前記ダイアログ音声に基づいて、前記番組音声に含まれるダイアログ音声のゲイン量を推定するステップと、

前記ゲイン量に基づいて前記ダイアログ音声を調整するステップと、

前記調整するステップで調整された前記ダイアログ音声及び前記番組音声の合成を行うステップと、を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、受信装置及びプログラムに関するものであり、特に、放送された情報及びインターネット通信において送信された情報の双方を受信できる受信装置及びプログラムに関するものである。

【背景技術】

【0002】

放送番組の音声に対する要望は、多言語放送や解説放送の充実化、さらに聞き取りやすいダイアログ音声レベルの設定など様々である。中でも、聞き取りやすい音声を求める要望は大きく、放送局は、例えば、番組制作にラウドネスメータを導入することにより、聞き取りやすい番組音声実現に向けた取り組みを始めている。しかし、音声が聞き取りやすいように番組制作が行われても、音声の聞き取りやすさは、放送信号の受信側の環境や聴取者（聴取者の聴覚能力）によっても変化してしまうものである。

【0003】

そのため、既に放送局側で制作された番組音声から所望の音声のみを抽出することは困難であるものの、聞き取りやすさを追求するためには、ダイアログ音声のレベルを受信装置側で調整できることが望まれる。従来、受信装置が、チャンネル間の相関を利用して、ダイアログ音声（例えば、ナレーション音声）とそれ以外の音声（例えば、背景音や効果音）との音量差を制御する手法が提案されている（例えば、特許文献1参照）。特許文献1に記載の手法では、受信装置が、まず、放送信号に含まれる2チャンネルの音声信号中から、相関が高い同相成分であるダイアログ音声信号を抽出し、このダイアログ音声信号に対して増幅・減衰・フィルタ処理などの処理を施す。そして、受信装置が、同相成分値を抽出した2チャンネルの音声信号に対して、処理済みのダイアログ音声信号を加算する。これにより、音声信号中のダイアログ音声信号に対して、種々の処理が施されることになる。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特許第4970174号公報

10

20

30

40

50

【発明の概要】**【発明が解決しようとする課題】****【0005】**

しかし、ダイアログ音声信号の相関のみが高いとは限らず、特許文献1に記載の従来手法では、例えば、背景音などのモノ信号は、相関が高いとみなされ、ダイアログ音声信号と共に抽出されてしまう。この場合、受信装置は、ダイアログ音声信号のみを増幅したり、減衰したりできなくなる。

【0006】

また、従来手法は、単に、ダイアログ音声を強調したり、弱めたりすることを目的にするものであり、受信装置側でダイアログ音声の差し替えを可能にするものではない。つまり、従来の受信装置は、放送信号以外を受信するものではないため、ダイアログ音声に対して処理が施された放送信号に、放送信号以外の別の信号を合成することはできない。そのため、従来手法では、放送信号中の元のダイアログ音声（例えば、日本語）を別のダイアログ音声（例えば、英語）に置き換えるために、元のダイアログ音声を削除する、つまり元のダイアログ音声のレベルをゼロにすることについては何ら想定されていない。仮に、従来手法によってダイアログ音声の削除が試されても、上記理由により、相関値の利用ではダイアログ音声のみの削除は困難であり、効果が不十分である。

【0007】

従って、上記のような問題点に鑑みてなされた本発明の目的は、受信した放送信号に含まれるダイアログ音声のみのレベルを調整できる受信装置及びプログラムを提供すること

【課題を解決するための手段】**【0008】**

上記課題を解決するために、本発明に係る受信装置は、
放送された番組音声を受信する第1受信部と、
インターネット通信により前記番組音声のダイアログ音声を受信する第2受信部と、
前記番組音声及び前記ダイアログ音声に基づいて、各フレームの有効性を判断する判断部と、
前記判断部で有効と判断されたフレームの前記番組音声及び前記ダイアログ音声に基づいて、前記番組音声に含まれるダイアログ音声のゲイン量を推定する推定部と、
前記ゲイン量に基づいて前記ダイアログ音声を調整する調整部と、
前記調整部で調整された前記ダイアログ音声と前記番組音声との合成を行う合成部と、
を備える。

【0009】

また、本発明に係る受信装置は、
前記判断部が、前記ダイアログ音声のフレームエネルギー、及び前記番組音声と前記ダイアログ音声との類似性に基づいて、前記各フレームの有効性を判断する。

【0010】

また、本発明に係る受信装置は、
前記判断部が、前記番組音声及び前記ダイアログ音声の帯域を複数に分割して、少なくとも1つの帯域で類似性を有するフレームを有効と判断する。

【0011】

また、本発明に係る受信装置は、
前記推定部が、有効と判断されなかったフレームの前記ゲイン量を、1つ前のフレームのゲイン量と同一とする。

【0012】

また、本発明に係る受信装置は、
前記調整部が、前記ダイアログ音声を逆位相化して前記ゲイン量を乗算することにより、前記ダイアログ音声を調整する。

【0013】

また本発明に係る受信装置は、
前記合成部で合成した結果と、第2受信部によって受信したダイアログ音声のうち、前記ダイアログ音声とは異なるダイアログ音声とを更に合成する第2合成部を備える。

【0014】

また、本発明に係るプログラムは、
受信装置のコンピュータに、
放送された番組音声を取得するステップと、
インターネット通信により前記番組音声のダイアログ音声を取得するステップと、
前記番組音声と前記ダイアログ音声に基づいて、フレームの有効性を判断するステップと、

有効と判断されたフレームの前記番組音声及び前記ダイアログ音声に基づいて、前記番組音声に含まれるダイアログ音声のゲイン量を推定するステップと、

前記ゲイン量に基づいて前記ダイアログ音声を調整するステップと、

前記調整するステップで調整された前記ダイアログ音声及び前記番組音声の合成を行うステップと、を実行させる。

【発明の効果】

【0015】

本発明によれば、受信した放送信号に含まれるダイアログ音声のみのレベルを調整できる受信装置及びプログラムを提供することができる。

【図面の簡単な説明】

【0016】

【図1】図1は、本発明の第1実施形態に係る受信装置の概略構成を示す機能ブロック図である。

【図2】図2は、本発明の第1実施形態に係る受信装置の処理を示すフローチャートである。

【図3】図3は、本発明の第2実施形態に係る受信装置の概略構成を示す機能ブロック図である。

【図4】図4は、本発明の第2実施形態に係る受信装置の処理を示すフローチャートである。

【発明を実施するための形態】

【0017】

以下、本発明の実施形態に係る受信装置について、図面を参照して説明する。受信装置は、放送信号を受信できるものであり、例えば、テレビ受像機等の放送受信端末である。受信装置は、以下に説明される計算処理を行うコンピュータを含むものである。なお、受信装置は、放送信号を受信できればよく、受信された放送信号の情報を表示する表示部（ディスプレイ）の有無に限定されるものではない。そのため、受信装置には、チューナ搭載のレコーダ、単体チューナなどが含まれる。

【0018】

放送とは、放送局から公衆（不特定多数）に向けた放送信号の送信であり、例えば、テレビジョン放送である。放送信号は、放送局が提供する番組の映像（番組映像）と音声（番組音声）の2つの信号が合成されたものである。番組音声には、人声に関するナレーション、アナウンサーコメント、ヴォーカルなどのダイアログ音声と、それ以外の音声（背景音や効果音など）とが含まれる。また、番組音声に含まれるダイアログ音声は、番組の制作の過程でレベル変化処理されることにより、時刻とともにそのレベルが変化するものとなっている。

【0019】

そして、本受信装置は、放送信号だけでなく、インターネットを介した通信信号を放送局から受信するものである。そのため、受信装置と放送局とは、有線接続又は無線接続されている。このような放送信号と通信信号の双方を受信する受信装置は、例えば、放送通信連携サービスの基盤システムであるHybridcast（ハイブリッドキャスト）（登録商標）

10

20

30

40

50

における受信装置である。

【0020】

(第1実施形態)

図1は、本発明の第1実施形態に係る受信装置1の概略構成を示す機能ブロック図である。本実施形態の受信装置1は、第1受信部3と、第2受信部5と、判断部7と、推定部9と、調整部11と、合成部13とを備える。第1受信部3は、判断部7と合成部13とに接続され、第2受信部5は判断部7と調整部11に接続されている。なお、図示は省略するが、受信装置1は、各機能を実現する処理内容を記述したプログラムを実行する制御部、例えば、CPU(中央処理装置)やDSP(デジタルシグナルプロセッサ)等を備えることができる。また、受信装置1は、各機能を実現する処理内容を記述したプログラムや各種情報を記憶し、また、バッファとしても機能可能な記憶部、例えば、RAM(Random Access Memory)等の揮発性の記憶媒体やROM(Read Only Memory)等の不揮発性の記憶媒体等を備えることができる。

10

【0021】

第1受信部3は、放送局によって放送された放送信号を受信し、放送信号を構成する番組映像と番組音声とを分離する。そして、第1受信部3は、番組音声をフレームごとに判断部7及び合成部13に送る。

【0022】

第2受信部5は、放送局からインターネットを介して通信信号を受信するものである。本実施形態において通信信号は、レベル変化処理される前のダイアログ音声である。すなわち、上記の第1受信部3を介した番組音声に含まれるダイアログ音声と、第2受信部5を介したダイアログ音声とは、レベルが異なるものであり、また、レベルの差が時刻とともに変化する。なお、上記の第1受信部3から取得した番組音声に含まれるダイアログ音声と、第2受信部5から取得したダイアログ音声とは、レベルを除いては、時間軸上で同期する。第2受信部5は、受信したダイアログ音声をフレームごとに判断部7及び調整部11に送る。

20

【0023】

判断部7は、番組音声及びダイアログ音声用の2つの入力を有し、それぞれ第1受信部3と第2受信部5とに接続されている。また、判断部7は番組音声及びダイアログ音声用の2つの出力を有し、それぞれ推定部9に接続されている。

30

【0024】

判断部7が、第1受信部3から取得した番組音声及び第2受信部5から取得したダイアログ音声に基づいて、各フレームの有効性を判断するための処理について、以下に説明する。なお、本実施形態において、番組音声及びダイアログ音声はデジタル信号として扱い、それらの信号は時間軸において適切なサンプリング周波数で標本化(サンプリング)されているものとする。また、以下の処理は、すべて所定の単位時間における、所定の周波数サンプル数を塊としたフレーム処理を前提とする。また、本実施形態において、番組音声を1chのモノ信号として説明するが、これに限られず、番組音声は2ch以上の場合にも、本発明は適用可能である。

【0025】

判断部7は、第2受信部5から取得したダイアログ音声のフレームエネルギーEdを算出する。ダイアログ音声のフレームエネルギーEd[dB]は、例えば以下の式により与えられる。

40

【0026】

【数1】

$$Ed = 20 \log_{10}(\max(\text{abs}(d(n)+m))) \quad (1)$$

ここで、d(n)はダイアログ音声の時間表現であり、abs()は絶対値、maxは最大値をとる関数である。またmは、log(0)を防ぐための、d(n)に比べて非常に小さな値である。フレームエネルギーEdは、フレーム内のダイアログ音声のエネルギーの最大値を表し、判断

50

部7は、E dの値が所定の値よりも大きいフレームを、有意なエネルギーを有するフレームと判断する。本実施形態では、E d > - 20 [dB]を満たしたフレームを、有意なエネルギーを有すると判断する。これにより、ダイアログ音声に含まれ得る、無音の時間(“ま”)からなるフレーム又は十分なエネルギーを有しないフレームを用いてゲイン量が推定されることを防ぐことができる。

【0027】

さらに判断部7は、同一フレームの番組音声とダイアログ音声との類似性を求める。本実施形態において、判断部7は、フィルタバンクとして、離散フーリエ変換(FFT)を用いて番組音声及びダイアログ音声それぞれの振幅周波数特性を求める。ダイアログ音声の振幅周波数特性D(k)及び番組音声の振幅周波数特性B(k)は、以下の式により与えられる。

【0028】

【数2】

$$D(k) = \text{abs}(FFT(d(n))) \quad (2)$$

$$B(k) = \text{abs}(FFT(b(n))) \quad (3)$$

ここで、d(n)はダイアログ音声の時間表現、b(n)は番組音声の時間表現であり、kは周波数上のサンプリングインデックスを表す。またFFT()は離散フーリエ変換を表し、abs()は絶対値をとる関数である。

【0029】

さらに、本実施形態において、判断部7は、番組音声の振幅周波数特性B(k)及びダイアログ音声の振幅周波数特性D(k)をそれぞれ複数の帯域に分割して、それぞれ同じ帯域成分同士の間係数を計算する。なお、ダイアログ音声は、一般の音声信号と比較して、低い帯域にエネルギーが集中しているため、本実施形態においては、伝送された番組音声における一部の帯域のみを計算する。ここで、各フレームのサンプル数を2048とすると、離散フーリエ変換により周波数表現されたものは1024サンプルで表現することが可能である。本実施形態においては、時間サンプリング周波数を48kHzとし、簡易化のために計算の上限を5kHzと設定することにより、各フレームの1024周波数サンプルのうち、およそ20周波数サンプルについて計算する。また、本実施形態においては、200周波数サンプル以下の帯域を4分割する例を示すが、これに限られず、4分割よりも多い分割数又は少ない分割数で分割してもよく、また、帯域を分割せずに計算することも可能である。ダイアログ音声の振幅周波数特性D(k)及び番組音声の振幅周波数特性B(k)は、以下の式により帯域ごとに4分割される。

【0030】

【数3】

$$D(k) = \{D1, D2, D3, D4\} \quad k = 0 \sim 199$$

$$D1 = D(k) \quad k = 0 \sim 49$$

$$D2 = D(k) \quad k = 50 \sim 99$$

$$D3 = D(k) \quad k = 100 \sim 149$$

$$D4 = D(k) \quad k = 150 \sim 199$$

(4)

$$B(k) = \{B1, B2, B3, B4\} \quad k = 0 \sim 199$$

$$B1 = B(k) \quad k = 0 \sim 49$$

$$B2 = B(k) \quad k = 50 \sim 99$$

$$B3 = B(k) \quad k = 100 \sim 149$$

$$B4 = B(k) \quad k = 150 \sim 199$$

10

20

30

40

50

さらに判断部 7 は、 $D(k)$ 及び $B(k)$ の帯域成分ごとの相関関数の最大値 MC を以下の式により求める。

【 0 0 3 1 】

【 数 4 】

$$MC = \max(\text{cor}(B1, D1), \text{cor}(B2, D2), \text{cor}(B3, D3), \text{cor}(B4, D4)) \quad (5)$$

上式 (5) の MC が 1 に近いほど、番組音声の帯域成分はダイアログ音声の同じ帯域成分と類似していることを示す。本実施形態においては、 $MC > 0.9$ を満たしたフレームを類似度が高いと判断する。

【 0 0 3 2 】

判断部 7 は、ダイアログ音声のフレームエネルギー E_d 、ならびに番組音声とダイアログ音声との類似性に基づいて、各フレームの有効性を判断する。すなわち、判断部 7 は、上述したフレームエネルギー E_d [dB] が、 $E_d > -20$ を満たし、且つ、帯域成分ごとの相関関数の最大値 MC が、 $MC > 0.9$ を満たすフレームを有効と判断する。そして、判断部 7 は、有効と判断されたフレームの番組音声及びダイアログ音声を推定部 9 に送る。

【 0 0 3 3 】

なお、本実施形態においては、判断部 7 での処理を、番組音声及びダイアログ音声の時間表現 $b(n)$ 、 $d(n)$ を用いて行っているが、第 1 受信部 3 及び第 2 受信部 5 でデコードされる前の番組音声及びダイアログ音声は、周波数領域で表現されている場合には、そのまま周波数領域で表現された番組音声及びダイアログ音声をを用いて判断部 7 での処理を行うことも可能である。

【 0 0 3 4 】

推定部 9 は、判断部 7 で有効と判断されたフレームの番組音声及びダイアログ音声に基づいて、番組音声に含まれるダイアログ音声のゲイン量 G を推定する。ここで、ゲイン量 G とは、あるフレームのダイアログ音声にゲイン量 G を乗算することで、当該フレームの番組音声に含まれるダイアログ音声と同一のレベルが得られる値である。番組音声に含まれるダイアログ音声のゲイン量 G は、以下の式により与えられる。

【 0 0 3 5 】

【 数 5 】

$$G = \frac{\sum_{n=0}^{2048} (b(n) \times d(n))}{\sum_{n=0}^{2048} d(n)^2} \quad (6)$$

【 0 0 3 6 】

また、推定部 9 は、判断部 7 で有効と判断されなかったフレームのゲイン量 G を、1 つ前のフレームのゲイン量 G と同一とする。なお、本実施形態において推定部 9 は、判断部 7 で最初のフレームが有効と判断されなかった場合には、当該フレームのゲイン量 G を 1 とする。推定部 9 は、各フレームのゲイン量 G を、調整部 1 1 に送る。

【 0 0 3 7 】

調整部 1 1 は、推定部 9 から取得したゲイン量 G に基づいて、第 2 受信部 5 から取得したダイアログ音声を調整する。すなわち、本実施形態において調整部 1 1 は、第 2 受信部 5 から取得したダイアログ音声を逆位相化するとともにゲイン量 G を乗算することにより、ダイアログ音声のレベルを調整する。したがって、本実施形態において、調整部 1 1 で調整されたダイアログ音声は、番組音声に含まれるダイアログ音声の逆位相信号となる。調整部 1 1 は、調整されたダイアログ音声を、合成部 1 3 に送る。

【 0 0 3 8 】

合成部 1 3 は、調整部 1 1 から取得した調整されたダイアログ音声と、第 1 受信部 3 から取得した番組音声とを合成する。本実施形態では、調整部 1 1 で調整されたダイアログ音声は、番組音声に含まれるダイアログ音声の逆位相信号であるため、合成部 1 3 の処理により、調整されたダイアログ音声と、番組音声に含まれるダイアログ音声とは打ち消し

10

20

30

40

50

あうことになる。よって、合成部13は、番組音声からダイアログ音声除去されたダイアログ音声除去番組音声（つまり、ダイアログ音声以外の音声）を、例えば音出力部（図示せず）に出力することができる。なお、番組音声は2ch以上の場合には、それぞれのチャンネルで上記手法を適用することによりダイアログ音声を取り去ることが可能である。

【0039】

続いて、図2を用いて、受信装置1の処理について説明する。図2は、本発明の第1実施形態に係る受信装置1の処理を示すフローチャートである。

【0040】

まず、第1受信部3は、放送局により放送された放送信号を受信し、放送信号から番組音声を取り出し、フレームごとに判断部7及び合成部13に送る（ステップS101）。そして、第2受信部5は、インターネット通信により、番組音声に含まれるダイアログ音声（通信信号）を放送局から受信し、フレームごとに判断部7及び調整部11に送る（ステップS102）。

【0041】

そして、判断部7は、番組音声及びダイアログ音声の各フレームの有効性を、ダイアログ音声のフレームエネルギー、ならびに番組音声とダイアログ音声との類似性に基づいて判断し、有効と判断されたフレームの番組音声及びダイアログ音声を推定部9に送る（ステップS103）。

【0042】

そして、推定部9は、有効と判断されたフレームの番組音声及びダイアログ音声に基づいて、番組音声に含まれるダイアログ音声のゲイン量Gを推定し、有効と判断されなかったフレームのゲイン量Gは、1つ前のフレームのゲイン量Gと同一として、各フレームのゲイン量Gを調整部11に送る（ステップS104）。

【0043】

そして、調整部11は、第2受信部5から取得したダイアログ音声を逆位相化するとともにゲイン量Gを乗算して、ダイアログ音声を調整し、合成部13に送る（ステップS105）。

【0044】

そして、合成部13は、第1受信部3から取得した番組音声と、調整部11から取得した調整されたダイアログ音声とを合成する（ステップS106）。合成部13は、ダイアログ音声除去番組音声を出力することになる。

【0045】

このように本実施形態では、判断部7は、番組音声及びダイアログ音声に基づいて、各フレームの有効性を判断し、推定部9は、有効と判断されたフレームの番組音声及びダイアログ音声に基づいて、番組音声に含まれるダイアログ音声のゲイン量Gを推定する。つまり、本実施形態における受信装置1は、有効なフレームに限定してゲイン量Gの推定を行うため、信頼性の高いゲイン量Gを得ることができる。その結果、受信した放送信号に含まれるダイアログ音声のみのレベルを調整することが可能となる。

【0046】

また、本実施形態における受信装置1は、判断部7が、ダイアログ音声のフレームエネルギー、及び番組音声とダイアログ音声との類似性に基づいて、各フレームの有効性を判断する。つまり、レベル調整の対象であるダイアログ音声のうち、無音又は微小レベルの音声のフレームに対してはゲイン量Gの推定を行わず、ダイアログ音声が十分なエネルギーを有し、番組音声とダイアログ音声とが十分に類似したフレームに限定して、ゲイン量Gの推定を行うため、より信頼性の高いゲイン量Gを得ることができ、放送信号の番組音声に含まれるダイアログ音声のレベルを、より厳密に調整することが可能となる。また、本実施形態における受信装置1は、番組音声及びダイアログ音声の帯域を複数に分割して、少なくとも1つの帯域で類似性を有するフレームを有効と判断することにより、より信頼性の高いゲイン量Gを得ることができ、放送信号の番組音声に含まれるダイアログ音声のレベルを、より厳密に調整することが可能となる。また、本実施形態における受信装置

10

20

30

40

50

1 は、推定部 109 が、有効と判断されなかったフレームのゲイン量 G を、1 つ前のフレームのゲイン量 G と同一とすることで、より信頼性の高いゲイン量 G を用いてダイアログ音声のレベルを調整することが可能となる。さらに、本実施形態における受信装置 1 は、調整部 11 がダイアログ音声を逆位相化してゲイン量 G を乗算することによりダイアログ音声を調整し、合成部 13 が調整したダイアログ音声と番組音声とを合成することにより、時刻とともにレベルが変化する番組音声に含まれるダイアログ音声のみを適切に消去することが可能となる。

【0047】

(第2実施形態)

第1実施形態では、受信装置 1 の出力信号の 1 つとしてダイアログ音声除去番組音声を出力する場合について説明したが、第2実施形態では、ダイアログ音声除去番組音声に更なる信号を付加する場合について説明する。

【0048】

図3は、本発明の第2実施形態に係る受信装置 101 の概略構成を示す機能ブロック図である。本実施形態の受信装置 101 は、第1受信部 103 と、第2受信部 105 と、判断部 107 と、推定部 109 と、第1調整部 111 と、第1合成部 113 と、第2調整部 112 と、第2合成部 114 とを備える。第1受信部 103、判断部 107、推定部 109、第1調整部 111、及び第1合成部 113 の機能は、それぞれ対応する第1実施形態の第1受信部 3、判断部 7、推定部 9、調整部 11、及び合成部 13 の機能と同一であるため、説明は省略する。

【0049】

本実施形態における第2受信部 105 は、放送局からインターネットを介して通信信号を受信し、受信したダイアログ音声を判断部 107、第1調整部 111、及び第2調整部 112 に送る。

【0050】

第2調整部 112 は、第2受信部 105 から取得したダイアログ音声に所定の値を乗算して調整し、調整されたダイアログ音声を第2合成部 114 に送る。

【0051】

第2合成部 114 は、第2調整部 112 から取得した調整されたダイアログ音声と、第1合成部 113 から取得したダイアログ音声除去番組音声とを合成する。これにより、番組音声に含まれるレベル変化処理されたダイアログ音声を、第2調整部 112 で適切に調整されたダイアログ音声に置き換えた番組音声を出力することができる。なお、第2調整部 112 においてダイアログ音声に乘算する所定の値は、予め受信装置 101 で設定された値としてもよいし、あるいは、例えば受信装置 101 がボリューム用つまみ等の入力部を備え、当該入力部の操作によりユーザが適宜入力した値とすることもできる。

【0052】

続いて、図4を用いて、受信装置 101 の処理について説明する。図4は、本発明の第2実施形態に係る受信装置 101 の処理を示すフローチャートである。

【0053】

まず、第1実施形態におけるステップ S101 と同様、第1受信部 103 は、放送局により放送された放送信号を受信し、放送信号から番組音声を取り出し、フレームごとに判断部 107 及び第1合成部 113 に送る(ステップ S201)。そして、第2受信部 105 は、インターネット通信により、番組音声に含まれるダイアログ音声(通信信号)を放送局から受信し、フレームごとに判断部 107、第1調整部 111、及び第2調整部 102 に送る(ステップ S202)。

【0054】

そして、判断部 107 は、第1実施形態におけるステップ S103 と同様、番組音声及びダイアログ音声の各フレームの有効性を、ダイアログ音声のフレームエネルギー E_d、ならびに番組音声とダイアログ音声との類似性に基づいてから判断し、番組音声及びダイアログ音声の有効と判断されたフレームを推定部 109 に送る(ステップ S203)。

10

20

30

40

50

【 0 0 5 5 】

そして、推定部 1 0 9 は、第 1 実施形態におけるステップ S 1 0 4 と同様、有効と判断されたフレームの番組音声及びダイアログ音声に基づいて、番組音声に含まれるダイアログ音声のゲイン量 G を推定し、有効と判断されなかったフレームのゲイン量 G は、1 つ前のフレームのゲイン量 G と同一として、各フレームのゲイン量 G を第 1 調整部 1 1 1 に送る（ステップ S 2 0 4）。

【 0 0 5 6 】

そして、第 1 調整部 1 1 1 は、第 1 実施形態におけるステップ S 1 0 5 と同様、ダイアログ音声を逆位相化するとともにゲイン量 G を乗算することによりダイアログ音声を調整し、第 1 合成部 1 1 3 に送る（ステップ S 2 0 5）。

10

【 0 0 5 7 】

そして、第 1 合成部 1 1 3 は、第 1 実施形態におけるステップ S 1 0 6 と同様、第 1 受信部 1 0 3 から取得した番組音声と、第 1 調整部 1 1 1 から取得した調整されたダイアログ音声とを合成する（ステップ S 2 0 6）。

【 0 0 5 8 】

そして、第 2 調整部 1 1 2 は、ダイアログ音声に所定の値を乗算して、ダイアログ音声を調整し、第 2 合成部 1 1 4 に送る（ステップ S 2 0 7）。

【 0 0 5 9 】

そして、第 2 合成部 1 1 4 は、第 1 合成部 1 1 3 で合成された、ダイアログ音声除去番組音声と、第 2 調整部 1 1 2 から取得した調整されたダイアログ音声とを合成する（ステップ S 2 0 8）。第 2 合成部 1 1 4 は、番組音声に含まれたレベル変化処理されたダイアログ音声を、第 2 調整部 1 1 2 で適切に調整されたダイアログ音声に置き換えた番組音声を出力することになる。

20

【 0 0 6 0 】

このように本実施形態における受信装置 1 0 1 は、受信した放送信号に含まれるダイアログ音声を消去し、さらに第 2 調整部 1 1 2 で適切に調整されたダイアログ音声を合成することにより、所望のレベルのダイアログ音声を含む番組音声を出力することが可能となる。

【 産業上の利用可能性 】

【 0 0 6 1 】

かくして本発明によって、受信した放送信号に含まれるダイアログ音声のみのレベルを調整できる受信装置及びプログラムを提供することが可能となった。

30

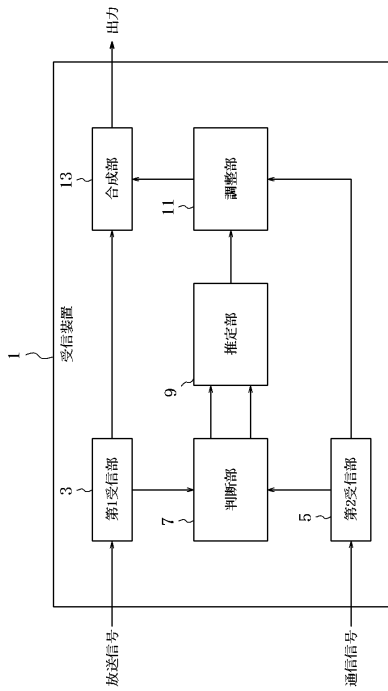
【 符号の説明 】

【 0 0 6 2 】

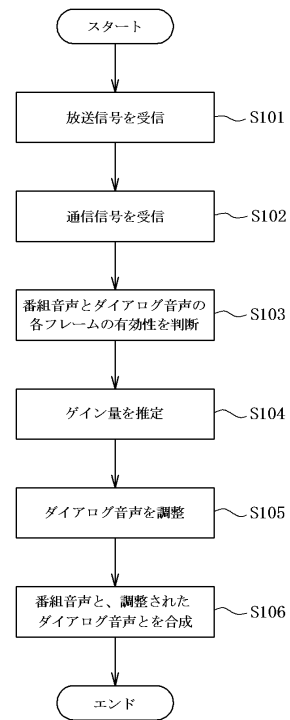
- 1、1 0 1 受信装置
- 3、1 0 3 第 1 受信部
- 5、1 0 5 第 2 受信部
- 7、1 0 7 判断部
- 9、1 0 9 推定部
- 1 1、1 1 1 調整部（第 1 調整部）
- 1 3、1 1 3 合成部（第 1 合成部）
- 1 1 2 第 2 調整部
- 1 1 4 第 2 合成部

40

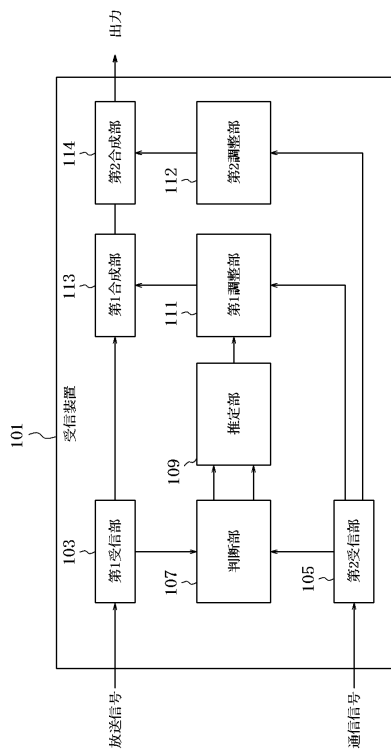
【図1】



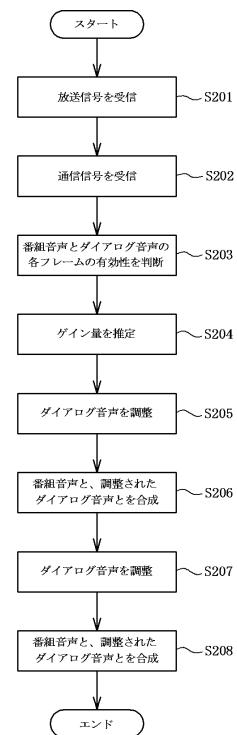
【図2】



【図3】



【図4】



フロントページの続き

- (72)発明者 杉本 岳大
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 今井 篤
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内
- (72)発明者 大竹 剛
東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

審査官 菊池 智紀

- (56)参考文献 特開2012-205075(JP,A)
特表2010-521867(JP,A)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|------|---------------------------------|
| G10L | 21/02 - 21/0364 |
| H04N | 7/10 - 7/56 ,
21/00 - 21/858 |