

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3670562号

(P3670562)

(45) 発行日 平成17年7月13日(2005. 7. 13)

(24) 登録日 平成17年4月22日(2005. 4. 22)

(51) Int.Cl.⁷

F I

H04S 1/00

H04S 1/00

K

H04S 5/02

H04S 5/02

K

請求項の数 15 (全 15 頁)

(21) 出願番号 特願2000-268442 (P2000-268442)
 (22) 出願日 平成12年9月5日(2000. 9. 5)
 (65) 公開番号 特開2002-78100 (P2002-78100A)
 (43) 公開日 平成14年3月15日(2002. 3. 15)
 審査請求日 平成14年5月24日(2002. 5. 24)

(73) 特許権者 000004226
 日本電信電話株式会社
 東京都千代田区大手町二丁目3番1号
 (74) 代理人 100066153
 弁理士 草野 卓
 (74) 代理人 100100642
 弁理士 稲垣 稔
 (72) 発明者 水島 昌英
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内
 (72) 発明者 青木 真理子
 東京都千代田区大手町二丁目3番1号 日
 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】ステレオ音響信号処理方法及び装置並びにステレオ音響信号処理プログラムを記録した記録媒体

(57) 【特許請求の範囲】

【請求項1】

ステレオ收音、もしくはステレオ再生用に加工された2チャンネル音響信号から中央付近に定位する音源信号を抑圧もしくは強調するステレオ音響信号処理方法において、ステレオ信号を各チャンネルごとに複数の周波数帯域に分割し、各周波数帯域ごとにチャンネル間の類似度を計算し、類似度から中央付近に定位する音源信号を抑圧、もしくは強調するための減衰係数を計算し、その減衰係数を各周波数帯域信号に乗算し、減衰係数を乗じた後の各チャンネルごとの各周波数帯域信号を再合成して出力することを特徴とするステレオ音響信号処理方法。

【請求項2】

請求項1に記載のステレオ音響信号処理方法において、各周波数帯域ごとの類似度は、周波数帯域ごとにチャンネル信号間の大きさの比と位相差によって二つ求め、求めた二つの類似度から中央付近に定位する音源信号を抑圧もしくは強調するための減衰係数を二つ求めることを特徴とするステレオ音響信号処理方法。

【請求項3】

請求項2に記載のステレオ音響信号処理方法において、求めた二つの減衰係数のうち、中央付近に定位する音源信号を強調する時には小さい係数を、抑圧する時には大きい係数を各周波数帯域信号に乗算することを特徴とするステレオ音響信号処理方法。

10

20

【請求項 4】

請求項 1 乃至 3 のいずれか 1 項に記載のステレオ音響信号処理方法において、
チャンネル間のレベル差及び時間差がわずかな音源信号の平均的なパワーとそれ以外の音源信号のパワーとの比を算出し、その比と所望する固定比から抑圧に必要な減衰係数を計算し、チャンネル間のレベル差及び時間差がわずかな音源信号の平均的なパワーとそれ以外の音源信号のパワーの比を所望な一定に保つこと
を特徴とするステレオ音響信号処理方法。

【請求項 5】

請求項 1 乃至 4 のいずれか 1 項に記載のステレオ音響信号処理方法において、
音の立ち上がり時間を周波数帯域ごとに観測し、立ち上がり時間の早さに応じた減衰係数
10
を前記類似度による減衰係数に乗算して新たな減衰係数とすることで、中央付近に音声と共に定位する、音声に比べて立ち上がりが速いと判断された音源信号をさらに抑圧すること
を特徴とするステレオ音響信号処理方法。

【請求項 6】

ステレオ収音、もしくはステレオ再生用に加工された 2 チャンネル音響信号から中央付近に定位する音源信号を抑圧もしくは強調するステレオ音響信号処理装置において、
ステレオ信号を各チャンネルごとに複数の周波数帯域に分割する周波数帯域分割手段と、
各周波数帯域ごとにチャンネル間の類似度を計算する類似度計算手段と、
20
類似度から中央付近に定位する音源信号を抑圧、もしくは強調するための減衰係数を計算する減衰係数計算手段と、
その減衰係数を各周波数帯域信号に乗算する乗算手段と、
減衰係数を乗じた後の各チャンネルごとの各周波数帯域信号を再合成して出力する音源信号合成・出力手段を備えたこと
を特徴とするステレオ音響信号処理装置。

【請求項 7】

請求項 6 に記載のステレオ音響信号処理装置において、
類似度計算手段は、各周波数帯域ごとにチャンネル信号間の大きさの比と位相差によって類似度を二つ求め、
減衰係数計算手段は、類似度計算手段で求めた二つの類似度から中央付近に定位する音源
30
信号を抑圧もしくは強調するための減衰係数を二つ求めること
を特徴とするステレオ音響信号処理装置。

【請求項 8】

請求項 7 に記載のステレオ音響信号処理装置において、
乗算手段は、求めた二つの減衰係数のうち、中央付近に定位する音源信号を強調する時には小さい係数を、抑圧する時には大きい係数を各周波数帯域信号に乗算すること
を特徴とするステレオ音響信号処理装置。

【請求項 9】

請求項 6 乃至 8 のいずれか 1 項に記載のステレオ音響信号処理装置において、
減衰係数計算手段は、チャンネル間のレベル差及び時間差がわずかな音源信号の平均的なパ
40
ワーとそれ以外の音源信号のパワーとの比を算出し、その比と所望する固定比から抑圧に必要な減衰係数を計算し、チャンネル間のレベル差及び時間差がわずかな音源信号の平均的なパワーとそれ以外の音源信号のパワーの比を所望な一定に保つこと
を特徴とするステレオ音響信号処理装置。

【請求項 10】

請求項 6 乃至 9 のいずれか 1 項に記載のステレオ音響信号処理装置において、
減衰係数計算手段は、音の立ち上がり時間を周波数帯域ごとに観測し、立ち上がり時間の早さに応じた減衰係数を前記類似度による減衰係数に乗算して新たな減衰係数とすることで、中央に音声と共に定位する、音声に比べて立ち上がりが速いと判断された音源信号をさらに抑圧すること

を特徴とするステレオ音響信号処理装置。

【請求項 1 1】

ステレオ収音、もしくはステレオ再生用に加工された 2 チャネル音響信号から中央付近に定位する音源信号を抑圧もしくは強調するステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、
ステレオ信号を各チャネルごとに複数の周波数帯域に分割する処理と、
各周波数帯域ごとにチャネル間の類似度を計算する処理と、
類似度から中央付近に定位する音源信号を抑圧、もしくは強調するための減衰係数を計算する処理と、
その減衰係数を各周波数帯域信号に乗算する処理と、
減衰係数を乗じた後の各チャネルごとの各周波数帯域信号を再合成して出力する処理と、
をステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

10

【請求項 1 2】

請求項 1 1 に記載のステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、
類似度を計算する処理は、周波数帯域ごとにチャネル信号間の大きさの比と位相差によって二つ求める処理を有し、
減衰係数を計算する処理は、周波数帯域ごとにチャネル信号間の大きさの比と位相差によって二つ求めた類似度から中央付近に定位する音源信号を抑圧もしくは強調するための減衰係数を二つ求める処理を有すること
を特徴とするステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

20

【請求項 1 3】

請求項 1 2 に記載のステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、
乗算する処理は、求めた二つの減衰係数のうち、中央付近に定位する音源信号を強調する時には小さい係数を、抑圧する時には大きい係数を各周波数帯域信号に乗算する処理を有すること
を特徴とするステレオ音響信号方法を実行させる処理プログラムを記録したコンピュータ読み取り可能な記録媒体。

30

【請求項 1 4】

請求項 1 1 乃至 1 3 のいずれか 1 項に記載のステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、
減衰係数を計算する処理は、チャネル間のレベル差及び時間差がわずかな音源信号の平均的なパワーとそれ以外の音源信号のパワーとの比を算出し、その比と所望する固定比から抑圧に必要な減衰係数を計算し、チャネル間のレベル差及び時間差がわずかな音源信号の平均的なパワーとそれ以外の音源信号のパワーの比を所望な一定に保つ処理を有すること
を特徴とするステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 1 5】

請求項 1 1 乃至 1 4 のいずれか 1 項に記載のステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体において、
減衰係数を計算する処理は、音の立ち上がり時間を周波数帯域ごとに観測し、立ち上がり時間の早さに応じた減衰係数を前記類似度による減衰係数に乗算して新たな減衰係数とすることで、中央に音声と共に定位する、音声に比べて立ち上がりが速いと判断された音源信号をさらに抑圧する処理を有すること
を特徴とするステレオ音響信号処理方法を実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体。

40

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

50

この発明は、音声、楽音、各種環境音源などの複数の音源から発せられた複数の音響信号が混ざった2チャンネルステレオ信号において、中央付近に定位する音源信号を強調もしくは抑圧するステレオ音響信号処理方法及び装置並びにステレオ音響信号処理プログラムを記録した記録媒体に関し、ステレオ音楽ソースの受聴者の好みに応じた再生や、騒音環境下で目的とする音声だけを受聴する時などに適用される。

【0002】

【従来の技術】

二本のマイクロホンで収音されたステレオ音響信号、もしくは人工的にチャンネル間でレベル差や位相差などをつけることで複数音源を複数位置に定位されたステレオ音響信号から、中央付近に定位する音源信号のみを抑圧するには、片側の信号の正負を反転、逆相にしてもう一方の信号に加算すればよい。これは中央に定位する音源の左右の信号の差違が小さいことにより実現される方法である。この方法は、既にでき上がった音楽信号から歌など主旋律のパートを消去し、伴奏だけを取りだすいわゆるボーカルキャンセル技術として利用される。

10

【0003】

しかしこの方法では、ステレオであった伴奏は加算によってモノラルになってしまうという問題があった。加えてこの方法では、中央に定位する音源を抑圧する量を調整することは出来ない。

一方、難聴者は、複数の音源が存在する中から目的とする音源信号を聞き取る能力（いわゆるカクテルパーティー効果と呼ばれる。）が劣っているといわれている。このため、健聴者を対象に作成された音楽信号では、しばしば伴奏が歌より大きく感じられることが指摘されている。この場合にはセンターに定位する歌を強調し、伴奏を抑圧することが望まれるが、これは前述の方法では実現できない。

20

【0004】

複数の音源が混合された信号から目的とする音源信号を抽出、もしくは強調する方法は他にもある。

第1の方法は、周期構造を持つ音源を周波数領域において基本周波数を推定し、調波構造を抜き出すことにより、同一音源と推定する成分を再合成する方法である。

しかしこの第1の方法では、音源は調波構造に限定され、さらに音源の調波構造の推定には必ず誤差が生じるため、それが雑音として知覚されることにより、目的音源信号の抽出精度が悪くなる問題があった。

30

【0005】

第2の方法は、周波数特性の変動が比較的ゆるやかな定常的な雑音源と周波数特性が定常的な音源よりも頻繁に変動する例えば音声のような目的信号音源が重畳された信号から、後者の目的音源信号を抽出、もしくは強調する方法である。これは混合された信号を周波数領域において、まず目的音源信号が重畳されていない部分、すなわち雑音源信号を推定し、雑音源信号の平均的な周波数特性を記憶する。そして、周波数領域において、雑音源信号と目的音源信号が重畳された信号から記憶された雑音源の平均的な周波数構造を減算することで目的音源信号を強調、もしくは抽出する方法である。

【0006】

しかしこの第2の方法では、雑音源信号が定常であることが必要で、歌の伴奏のように非定常な音源の伴奏のみの個所の推定、及び抑圧は困難であった。

40

【0007】

【発明が解決しようとする課題】

本発明の目的は、ステレオ音響信号から中央付近に定位する音源信号を抑圧、もしくは強調する技術において、抑圧、強調する割合の調整を可能とすることである。

本発明の別の目的は原信号の定位を損なわず、中央付近に定位する音源信号だけを強調、もしくは抑圧することである。

【0008】

本発明の別の目的は、目的とする音源信号の調波構造に依存せずに高精度に抑圧、もしくは

50

は強調することである。

本発明の別の目的は、目的外の信号（雑音信号）が非定常な信号であっても高精度に目的音を抑圧、もしくは強調することである。

【0009】

【課題を解決するための手段】

上記目的を達成するために、本発明のステレオ音響信号処理方法及び装置は、ステレオ音響信号を入力し、二つのチャンネル信号を各チャンネルごとに複数の周波数帯域に分割し、各周波数帯域ごとにチャンネル間の類似度を計算し、類似度などから中央に定位する音源信号を抑圧、もしくは強調するための減衰係数を計算し、この減衰係数を各周波数帯域信号に乗じた後の各チャンネルごとの各周波数帯域信号を再合成し、再合成した信号を出力することにより構成される。

10

【0010】

本発明は、入力されたステレオ信号をチャンネルごとに複数の周波数帯域に分割する。そして、各周波数帯域ごとにチャンネル間の信号の類似度をその振幅比や位相差などによって決定する。そして、類似度の高い周波数帯域に比べて類似度の低い周波数帯域に小さな減衰係数を乗算して、各チャンネルごとに再合成して出力すれば、減衰係数の下限値に応じて中央に定位する音源が強調される。反対に類似度の低い周波数帯域に比べて、類似度の高い周波数帯域に小さな減衰係数を乗算して各チャンネルごとに再合成して出力すれば、減衰係数の下限値に応じて中央に定位する音源が抑圧される。

【0011】

20

【発明の実施の形態】

（第1実施例）

図1は本発明の第1の実施例を示すブロック図である。

ステレオ信号入力部102に入力される音響信号は、強調、もしくは抑圧したい目的音源信号が中央付近に定位するように収音されているステレオ信号であれば本発明は有効である。

【0012】

ステレオ信号入力部102に入力されたステレオ信号は左右のチャンネルごとに処理される。

以下にその処理方法の詳細を述べる。

左チャンネルの信号 s_L は、左チャンネル周波数帯域分割部103によって周波数領域に変換される。同様に右チャンネルの信号 s_R は、右チャンネル周波数帯域分割部110によって周波数領域に変換される。ここで帯域分割数を N とする。左チャンネルにおいて帯域分割された信号を低い周波数から順に $f_L(0), f_L(1), f_L(2), \dots, f_L(k), \dots, f_L(N-1)$ とする。右チャンネルにおいて帯域分割された信号を低い周波数から順に $f_R(0), f_R(1), f_R(2), \dots, f_R(k), \dots, f_R(N-1)$ とする。類似度計算部104において、 $f_L(k), f_R(k)$ は、同じ周波数帯域ごとに類似度 $a(0), a(1), a(2), \dots, a(k), \dots, a(N-1)$ が計算される。ステレオ信号において、中央付近に定位する音源信号は左右の信号が一致、もしくはその差違が非常に小さい。これは、周波数領域に変換したのちも全ての周波数領域において、左右の成分の差違は小さいことを意味する。このことから類似度は、 k が等しい、即ち同じ周波数成分間で、 $f_L(k)$ と $f_R(k)$ の差違で決定することが出来る。

30

40

【0013】

そして、各周波数帯域ごとに計算された類似度 $a(k)$ に基づき各周波数帯域ごとに減衰係数計算部105において減衰係数 $g(k)$ ($k=0 \sim N-1$) が算出される。減衰係数は同一周波数帯域において、左右チャンネル間で同一なものが各周波数帯域信号 $f_L(k)$ に乗算器116で乗算される。

つまり、各周波数帯域ごとの左右レベル差、位相差から各周波数帯域ごとに類似度、そして減衰係数を計算し、各帯域に乗じて、左右チャンネル音源信号合成部106, 111で再合成することで、類似度の大きな成分だけの成分集合 s_L', s_R' が出力され、その結果、中央付近に定位する音源信号だけが残る。

類似度の計算方法

50

類似度 $a(k)$ の計算方法について、左右周波数帯域分割部103,110が短時間フーリエ変換（以下、FFTと略する）である場合について述べる。

【0014】

FFTで周波数分割した場合、 $f_L(k)$ および $f_R(k)$ は一般に複素数となり、位相を考慮する必要がある。そこで、各成分の大きさの比と位相差によって二つの類似度を計算する。大きさの比による類似度を $a_i(k)$ 、位相差による類似度 $a_p(k)$ とすると、

【0015】

【数1】

$$\begin{aligned} a_i(k) &= |f_R(k)| / |f_L(k)| \quad (|f_R(k)| \leq |f_L(k)|) \\ &= |f_L(k)| / |f_R(k)| \quad (|f_L(k)| < |f_R(k)|) \end{aligned} \quad (1)$$

ここで、 $|A|$ は A の複素数の大きさを表す。

$$a_p(k) = \cos \theta \quad (2)$$

ここで、 θ は $f_L(k)$ と $f_R(k)$ の位相差を表す。

【0016】

類似度 $a_i(k)$, $a_p(k)$ は減衰係数計算部105に送られ、減衰係数 $g(k)$ が計算される。

減衰係数の計算方法

減衰係数 $g(k)$ の計算方法について説明する。

1. 中央定位音源信号を強調する場合

中央に定位する音源信号を強調する場合について説明する。

▲1▼大きさの比による減衰係数 $g_i(k)$ の計算方法を説明する。

【0017】

(1) 式から明らかなように、類似度 $a_i(k)$ は、 $f_L(k)$ と $f_R(k)$ の大きさが等しい時に1になり、それ以外は1より小さい値となる。したがって、大きさの比による類似度 $a_i(k)$ を引数とする関数において、単調増加の関数の出力を $g_i(k)$ に選ばばよい。

図2にその一例を示す。横軸は $20\log_{10}(a_i(k))$ 、縦軸は $20\log_{10}(g_i(k))$ を示している。

【0018】

ここで、 $A_i(k) = 20\log_{10}(a_i(k))$, $G_i(k) = 20\log_{10}(g_i(k))$ とすると、

$$\begin{aligned} G_i(k) &= 0 \quad (\varepsilon < A_i(k) \leq 0) \\ &= G_{\min} (A_i(k) / (\beta - \varepsilon) - \varepsilon / (\beta - \varepsilon)) \quad (\beta < A_i(k) \leq \varepsilon) \\ &= G_{\min} \quad (A_i(k) \leq \beta) \end{aligned}$$

中央に定位する音源信号だけであるならば、全ての k に対して $a_i(k)$ は1 ($20\log_{10}(a_i(k)) = 0$) になるが、その他に定位する信号が重畳されることにより、中央定位音源信号が支配的な帯域であっても1よりもやや小さくなることもある。よって図2のように適当な幅 ε を持たせることが有効である。この適当な幅 ε は例えば β は左右のレベル差や位相差が僅かで中央に音を知覚させる中央定位音源信号について音質などの変化が無視できる範囲で予め聴感上で決めることが好適である。ただし、 ε を大きくしすぎると、中央付近で左右いずれかの方向にずれて定位した音源信号などを抑圧することが出来なくなる。よって、 ε は誤差による中央定位音源信号の音質などの変化が無視できる範囲で0に近い値にすることが望ましい。

【0019】

$G_{\min}$ は、中央定位音源信号以外の信号の抑圧量に相当する。この値を変化させることで、歌と伴奏に例えるならば、歌の大きさに対する伴奏の大きさを調整することが可能となる。図2において、 β を ε と一致させてもよいし、一致させなくてもよい。 β を ε に近づけると中央定位音源信号以外の信号は等しく $G_{\min}$ の減衰量で減衰されることが期待でき

10

20

30

40

50

るが、中央定位音源信号の支配的な帯域が誤って抑圧された場合の誤差の影響も大きくなる。 β を ε から離すことで中央定位音源信号が支配的な帯域を誤って抑圧した場合の誤差の影響を小さく出来るが、定位する位置によって抑圧量が変わってしまい、歌の伴奏に例えるならば、伴奏楽器間の音量のバランスが変わってしまうことが予想される。よって、中央定位音源信号の音質などの変化が無視できる範囲で β は ε に近い値($0 > \varepsilon > \beta$)にすることが望ましい。

▲2▼位相差による減衰係数 $g_p(k)$ の計算方法を説明する。

【0020】

(2)式から明らかなように、類似度 $ap(k)$ は、 $f_L(k)$ と $f_R(k)$ の位相が一致したときに1になり、それ以外は1より小さい値であり、位相差 θ が $\pi/2$ ラジアンの際に0、 θ が π ラジアンの際、すなわち逆相の際に-1で最小である。一般に位相差による定位知覚は周波数帯域に依存し、大きさの比ほど単純ではない。しかし、少なくとも中央に定位する音源信号に関しては位相差は0に近く、よって $ap(k)$ は1に近い値であることが期待できる。このことから位相差による減衰係数 $g_p(k)$ は例えば図3に示すように計算すればよい。

【0021】

図3にその一例を示す。横軸は $ap(k)$ 、縦軸は $20\log_{10}(g_p(k))$ を表す。

ここで、 $G_p(k) = 20\log_{10}(g_p(k))$ とすると、

$$\begin{aligned} G_p(k) &= 0 & (\zeta < ap(k) \leq 1) \\ &= G_{pmin}(ap(k)/(\alpha - \zeta) - \zeta/(\alpha - \zeta)) & (\alpha < ap(k) \leq \zeta) \\ &= G_{pmin} & (-1 \leq ap(k) \leq \alpha) \end{aligned}$$

中央に定位する音源信号だけであるならば、全ての k に対して $ap(k)$ は1になるが、その他の雑音信号が重畳されることにより、中央定位音源信号が支配的な帯域であっても1よりやや小さくなることがある。よって図2のように適当な幅 ζ を持たせることが有効である。しかし ζ を大きくしすぎると、中央に定位しない他の音源信号の抑圧が不十分になる。よって、 ζ は誤差による中央定位音源信号の変化が無視できる範囲で1に近い値($1 > \zeta$)にすることが望ましい。 G_{pmin} は、中央定位音源信号以外の信号の抑圧量に相当する。この値を変化させることで、歌と伴奏に例えるならば、歌の大きさに対する伴奏の大きさを調整することが可能となる。

【0022】

図3において、 α と ζ とを一致させてもよいし、一致させなくてもよい。 α を ζ に近づけると中央定位音源信号以外の信号は等しく G_{pmin} の減衰量で減衰されることが期待できるが、中央定位音源信号の支配的な帯域が誤って抑圧された場合の誤差も大きくなる。 α を ζ から離すことで中央定位音源信号が支配的な帯域を誤って抑圧された場合の誤差の影響を小さく出来るが、位相差による抑圧量の違いは周波数帯域によってその影響度が異なるため、歌の伴奏に例えるならば、伴奏楽器の音量のバランスだけではなく音色などが変わってしまうことなどが予想される。よって、中央定位音源信号の変化が無視できる範囲で α は ζ に近い値($0 > \zeta > \alpha$)にすることが望ましい。

▲3▼二つの減衰係数 $g_i(k)$ と $g_p(k)$ から実際に $f_L(k)$, $f_R(k)$ に乗算する減衰係数 $g(k)$ の計算方法を説明する。

【0023】

適当な距離を離れた二つのマイクロホンで比較的マイクロホンから距離が近い複数の音源信号を收音したステレオ信号が入力信号である場合には、ステレオ再生における定位は左右のマイクロホンに入ってくる信号の位相差と大きさの比(レベル差)に依存する。低い周波数においてはレベル差はつきにくく、位相差が大きく影響する。高い周波数では、大きさの比が大きく影響する。よって、例えば周波数帯域を二つに分けてそれよりも低い周波数においては $g_i(k)$ を、高い周波数においては $g_p(k)$ を採用することが考えられる。しかしながら、壁に囲まれた残響のある部屋において、マイクロホンから離れた位置に存在する音源からの信号は一般に左右のレベル差はほとんどなく、逆に位相が左右のマイクロホ

10

20

30

40

50

ンでランダムになるため(2)式の値が0に近くなる。この場合は全ての周波数において優先的に $g_p(k)$ を使うことが望ましい。さらにポピュラー音楽等の場合は、直接マイクロホンで収音するだけでなく、左右チャンネル信号に大きさの比や時間差、あるいは位相の時間的な変化を人工的に付加することで自然界には存在しない定位を得ることが普通であり、もっと複雑になる。以上のように様々なステレオ入力信号に応じて、最適な $g(k)$ の選択をすることは非常に困難である。しかしながら、どの場合も少なくとも中央に定位する音源信号の大きさの比と位相差は共に小さい、そこで、 $g(k)$ として、 $g_i(k)$ と $g_p(k)$ の小さいほうを採用することにする。即ち、

$$g(k) = \min(g_i(k), g_p(k)) \quad (3)$$

ここで、 $\min(A, B)$ はAとBのどちらか小さい方を出力することを意味する。つまり、どんなステレオ入力信号であっても、大きさか位相の少なくとも一方が左右で異なる場合は抑圧することになり、その結果、中央に定位する音源信号を強調することが可能となる。

【0024】

上記のように減衰係数計算部105で計算された $g(k)$ は図1にあるように各チャンネル各周波数帯域の $f_L(k)$, $f_R(k)$ に乗算器116で乗算される。同じ帯域 k において左右のチャンネルに同じ $g(k)$ を乗算することで、中央に定位する音源信号以外の音源信号を定位を維持したまま抑圧することが可能となる。 $g(k)$ を乗算した信号は、 $f_L(k)$ は左チャンネル音源信号合成部106で再合成して時間波形 s_L' に変換される。 $f_R(k)$ は右チャンネル音源信号合成部111で再合成して時間波形 s_R' に変換される。 s_L' , s_R' はステレオ信号出力部107から、ラウドスピーカ108やステレオヘッドホン109に送られる。

【0025】

以上の処理により、中央に定位する音源信号を強調、その他の音源信号を抑圧した合成信号をステレオラウドスピーカ108やステレオヘッドホン109等で受聴することが可能となる。

2. 中央定位音源信号を抑圧する場合

中央に定位する音源信号を抑圧し、それ以外の音源信号を強調する場合について説明する。

図1において類似度計算部104で類似度 $a_i(k)$, $a_p(k)$ を計算するところまでは先に述べた中央に定位する音源信号を強調する場合と同じであり、類似度から減衰係数を計算する部分が異なる。中央に定位する音源信号を抑圧するのであるから、大きさによる減衰係数 $g_i(k)$ を図4に示すように計算し、位相による減衰係数を図5に示すように計算すればよい。

【0026】

図4、5にその一例を示す。

ここで、 $A_i(k) = 20 \log_{10}(a_i(k))$, $G_i(k) = 20 \log_{10}(g_i(k))$, $G_p(k) = 20 \log_{10}(g_p(k))$ とすると、

$$\begin{aligned} G_i(k) &= 0 & (A_i(k) \leq \beta) \\ &= G_{i\min}(A_i(k) / (\varepsilon - \beta) - \beta / (\varepsilon - \beta)) & (\beta < A_i(k) \leq \varepsilon) \\ &= G_{i\min} & (\varepsilon < A_i(k) \leq 0) \\ G_p(k) &= 0 & (-1 < a_p(k) \leq \alpha) \\ &= G_{p\min}(a_p(k) / (\zeta - \alpha) - \alpha / (\zeta - \alpha)) & (\alpha < a_p(k) \leq \zeta) \\ &= G_{p\min} & (\zeta < a_p(k) \leq 1) \end{aligned}$$

即ち、左右の類似度が大きいほど減衰係数を小さくすることによって、中央に定位する音源信号を抑圧することが可能となる。 $\alpha, \beta, \varepsilon, \zeta$ の考え方は前述の中央に定位する音源信号を強調する場合と同様であるため割愛する。 $g_i(k)$ と $g_p(k)$ から $g(k)$ を得る方法も強調の場合と同じ考えで、

$$g(k) = \max(g_i(k), g_p(k)) \quad (4)$$

と計算する。ここで、 $\max(A, B)$ はAとBから大きいほうを出力することを意味する。即ち、

大きさによる減衰係数 $g_i(k)$ と位相による減衰係数 $g_p(k)$ の少なくともどちらか一方が大きい場合には、左右チャンネル信号に位相差か大きさの違いがあることを意味し、その信号は中央に定位する音源信号ではないと考えるからである。 $g_i(k)$ と $g_p(k)$ が共に小さい場合のみ、中央に定位する音源信号であり、抑圧の対象となる。

【0027】

減衰係数計算部105で計算された $g(k)$ を各周波数帯域の $f_L(k)$, $f_R(k)$ に乗算するところから先は中央に定位する音源を強調する場合と同じであるので割愛する。上記の方法では、原信号における中央付近に定位する音源信号とそれ以外の音源信号（例えば、歌と伴奏に対応する）の音量差にかかわらず、減衰係数 $g(k)$ により、一律に抑圧される。つまり、原信号において既に適切な音量差であった場合でも、さらに抑圧され、音量差は拡大する。

10

【0028】

そこで、次に、原信号の中央付近に定位する音源信号とそれ以外の音源信号の音量差を推定し、その差を一定にするように減衰係数を決定する方法について説明する。

中央付近に定位する音源信号を強調する場合には、 $g(k)$ の値が大きければ中央定位音源信号成分と推測される。そこで、 $g(k)$ の値に適当なしきい値 g_{th} を設定し、 g_{th} よりも大きな周波数成分の大きさの合計を中央付近定位信号の大きさの推定値 g_c とする。同様に g_{th} よりも小さな周波数成分の大きさの合計を中央以外に定位する音源信号の大きさの推定値 g_b とする。これらの値は瞬時の値であるため、これらの長時間平均をとる、その方法には、例えば、移動平均法などが考えられる。それらの値を g_{ca} , g_{ba} とすると、

$$rcb = g_{ca} / g_{ba} \quad (5)$$

20

rcb は騒音で考えると長時間平均のSN比に相当する。 S が中央付近定位音源信号の大きさ（パワー）で、 N がそれ以外の定位音源信号の大きさに相当する。次にこの rcb を基準に所望のSN比にするように「 N 」を抑圧することを考える。所望のSN比を rd 、必要な抑圧量を g_2 とすると、 rcb を抑圧量 g_2 で割った値が rd になればよい。よって、

$$g_2 = rcb / rd \quad (rcb \leq rd \text{ のとき})$$

$$g_2 = 1.0 \quad (rcb > rd \text{ のとき})$$

(6)

上式では、原信号のSN比が、所望のSN比よりも大きい時には抑圧しない。その際、原信号よりもSN比を強制的に小さくしたい場合には、 rcb と rd の大小関係にかかわらず、(6)式上段のみを使用し、1より大きな g_2 に設定すればよい。

【0029】

30

各周波数帯域において、減衰係数が上述の g_{th} よりも小さな帯域成分に対し、 g_2 を乗算することで、平均的なSN比、すなわち中央定位音源信号とそれ以外の音源信号の音量差を所望量にすることが可能となる。

以上の処理により、中央に定位する音源信号を抑圧、その他の音源信号を強調した合成信号をステレオラウドスピーカ108やステレオヘッドホン109等で受聴することが可能となる。

(第2実施例)

図6は、本発明の第2の実施例を示すブロック図である。

【0030】

減衰係数 $g(k)$ を乗算した後、左右のチャンネル信号を加算器117で加算することでモノラル化する。ステレオの効果はなくなるが、左右のチャンネルの信号を加算することで左右チャンネルで無相関な音源信号成分をより抑圧することが可能である。

40

多くの歌の入ったポピュラー音楽において中央には歌の他にベースドラムやベースの音を定位させる場合が多い。これらの主たる周波数成分は歌の周波数成分よりも低いため、これらを抑圧するには、例えば減衰係数計算部105において、図7に示したような $g_e(k)$ を $g(k)$ に乗算し、新たな $g(k)$ とすることも有効である。図7において、横軸は分割周波数帯域 k 、縦軸は $20 \log_{10} (g_e(k))$ である。 k_L は低音楽器を抑圧するための下限周波数帯域を示す。 k_L 以下の周波数帯域をGemin抑圧する。 m は周波数帯域が低くなるにつれて徐々に抑圧量を増やしていくことで周波数軸上の不連続を押さえるための小さな正の整数である。 k_L や m を大きくしすぎると歌の低域周波数成分を抑圧してしまうため、例えば周波数に換算

50

して k_L は100Hzから200Hzくらいが適当である。

【0031】

逆に中央に定位する歌のみを抑圧し、中央に定位するベースドラムやベースの音を抑圧しないようにするには、 $g(k)$ による中央定位音源帯域の抑圧を、低い周波数帯域（例えば、0～200Hzの帯域）では行わないようにすればよい。

中央に定位するベースドラムなどを抑圧するもう一つの方法を説明する。ベースドラムは、音の立ち上がり時間が音声に比べて速い。そこでベースドラムの主たる周波数帯域において、音の立ち上がり時間を観測し、立ち上がり時間の短さに応じた減衰係数 $gak(k)$ を $g(k)$ に乗算して新たな $g(k)$ とすることで立ち上がりの速いベースドラムだけを抑圧することが可能となる。

10

【0032】

その一手法について説明する。

ある周波数帯域 k の T 時刻の左右チャンネルの平均を取った大きさを $A(k,T) = (|f_L(k,T)| + |f_R(k,T)|) / 2$ とする。1時刻前の $A(k,T-1)$ との比を r とする。ここで、単位時刻は1フレームで通例数十ミリ秒程度である。

$r = A(k,T) / A(k,T-1)$ ($A(k,T) > A(k,T-1)$ の時)

$r = 1.0$ ($A(k,T) < A(k,T-1)$ の時) (7)

r が大きいほど、 k の周波数帯域の立ち上がりが鋭いことを意味するから、立ち上がりの鋭さに対する減衰係数を $gak(k,T)$ を、 r が大きいほど $gak(k,T)$ が小さくなるような関数の出力にすればよい。

20

例えば、

【0033】

【数2】

$$gak(k,T) = r^{\wedge}rt \quad (rt < 0) \quad (8)$$

$\wedge$ は指数演算を表し、

rt は r の値に対してどの程度の割合で減衰させるかを表す負の実数である。

図8は $rt = -3$ の時、横軸に r 、縦軸に $20\log_{10}(gak(k,T))$ を示した例である。 $Gakmin = -50$ は、減衰量の下限值を表している。各時刻において常に(8)式における $gak(k,T)$ を乗算しているとスペクトルの時間変化が不連続になるので、

30

$$gak'(k,T) = gak'(k,T-1) + \delta(gak(k,T) - gak'(k,T-1)) \quad (9)$$

(9)式のようにスムージング処理を施した $gak'(k,T)$ を用いるのがよい。ここで δ はスムージングのための係数で、0より大きく1以下の実数である。 $gak'(k,T-1)$ と $gak(k,T)$ の大小関係で異なる値を用いてもよい。

【0034】

以上の処理により、ステレオ音響信号の中央に定位する音源信号を強調、もしくは抑圧することが可能となる。

本発明のステレオ音声処理装置はCPUやメモリ等を有するコンピュータと、アクセス主体となるユーザが利用するユーザ端末と記録媒体とから構成することができる。記録媒体はCD-ROM、磁気ディスク装置、半導体メモリ等のコンピュータ読み取り可能な記録媒体であり、ここに記録されたステレオ音声処理プログラムはコンピュータに読み取られ、コンピュータの動作を制御し、コンピュータに左右チャンネルごとに複数の周波数帯域に分割する処理、各周波数帯域ごとにチャンネル間の類似度を計算する処理、類似度から中央付近に定位する音源信号を抑圧、もしくは強調するための減衰係数を計算する処理、その減衰係数を各周波数帯域信号に乗算する処理、及び減衰係数を乗じた後の各チャンネルごとの各周波数帯域信号を再合成する処理等のステレオ音響処理方法を実行させる。なお、上記ステレオ音声処理プログラムは通信回線を介して伝送されたものであってもよい。

40

(利用方法)

次に本発明の利用方法について説明する。

50

【0035】

図9は本発明の第1の利用方法を示している。

音楽コンパクトディスク301はステレオ再生用で、その中に中央に定位する主たる音源信号も収録されているものとする。音楽コンパクトディスク301をパーソナルコンピュータ302において、図1、もしくは図4に示した本発明の中央に定位する音源信号を強調する処理と周波数特性等の聴覚補正処理などを施し、出力する。アンプ303で利得を調整した後、ステレオヘッドホン304等で聴取する。これは、例えば聴覚者等が歌に比べて伴奏を小さくして聞きたい場合などに利用できる。

【0036】

図10は本発明の第2の利用方法でミュージックオンデマンドに本発明を利用する例を示している。 10

音楽ソースはネットワーク306に接続されたホストコンピュータ305に多数格納されている。利用者はネットワーク306に接続したパーソナルコンピュータ302から、ホストコンピュータ305に自分の聞きたい音楽ソース名と、信号の処理方法を指定する。信号の処理方法の指定とは、例えば、本発明における中央に定位する歌以外の伴奏をどれだけ小さくするか、あるいは周波数特性を自分の好みに応じてどのように調整するか、などの処理の指定である。ホストコンピュータはその指定に従って音楽ソースを検索、指示通りの信号処理をした後、もしくは処理をしながらネットワークを介して利用者のパーソナルコンピュータ302へ音楽信号を送信する。利用者は、パーソナルコンピュータ302から出力された音楽信号をアンプ303で利得を調整し、ステレオヘッドホン304等で送られて来た音楽信号を聴 20

【0037】

また、パーソナルコンピュータの代わりに、ネットワークに無線で接続できる機能を内蔵した携帯型の音楽再生機でも、同じことが可能である。

また、以上の利用方法において、中央に定位する歌などを抑圧する処理をすれば、例えば携帯型の簡易カラオケなどに利用することも出来る。

図11は本発明の第3の利用方法を示した図である。

正面で話す話者の声を強調することを目的とする。難聴者の例えば頭部左右に配置した単一指向性マイクロホン201,202で收音した音響信号を、難聴者が携帯する小型の筐体に内蔵した本発明処理部で処理することで正面話者の音声以外の騒音を抑圧する。その後、同 30

【0038】

【発明の効果】

以上の説明のように本発明によれば、ステレオ音響信号から原信号の定位を損ねず、中央に定位する音源信号を所望の量だけ強調、もしくは抑圧することが、音源の定位情報のみで可能となり、以下のような効果が期待できる。

(1) 難聴者等が市販の音楽ソースを受聴する際、中央に定位する主たる音源信号である歌とそれ以外の伴奏の音量バランスを、難聴者自身が自由に聞き易いように調整し、音楽をより良く楽しむことが期待できる。 40

(2) 騒音環境下において、正面にいる目的話者の音声のみを強調することが可能となり、快適なコミュニケーションを実現することが期待できる。

(3) 中央に定位する歌などを抑圧することでカラオケ音源などを作成することが可能となる。

【図面の簡単な説明】

【図1】本発明の第1の実施例を示すブロック図。

【図2】中央定位音源を強調する時の $a_i(k)$ と $g_i(k)$ の関係を示す図。

【図3】中央定位音源を強調する時の $a_p(k)$ と $g_p(k)$ の関係を示す図。

【図4】中央定位音源を抑圧する時の $a_i(k)$ と $g_i(k)$ の関係を示す図。

【図5】中央定位音源を抑圧する時の $a_p(k)$ と $g_p(k)$ の関係を示す図。 50

【図 6】 本発明の第 2 の実施例を示すブロック図。

【図 7】 k と $ge(k)$ の関係を示す図。

【図 8】 r と $gak(k, T)$ の関係を示す図。

【図 9】 本発明の第 1 の利用方法を示す図。

【図 10】 本発明の第 2 の利用方法を示す図。

【図 11】 本発明の第 3 の利用方法を示す図。

【符号の説明】

102	ステレオ信号入力部	
103	左チャンネル周波数帯域分割部	
104	類似度計算部	10
105	減衰係数計算部	
106	左チャンネル音源信号合成部	
107	ステレオ信号出力部	
108	ステレオラウドスピーカ	
109	ステレオヘッドホン	
110	右チャンネル周波数帯域分割部	
111	右チャンネル周波数音源信号合成部	
112	音源信号合成部	
113	モノラル信号出力部	
114	ラウドスピーカ	20
115	モノラルイヤホン	
116	乗算器	
117	加算器	
201	左 ch 単一指向性マイクロホン	
202	右 ch 単一指向性マイクロホン	
203	左 ch イヤホン	
204	右 ch イヤホン	
205	本発明処理部、補聴部	
301	音楽コンパクトディスク	
302	パーソナルコンピュータ	30
303	アンプ	
304	ステレオヘッドホン	
305	ホストコンピュータ	
306	ネットワーク	

【図 1】

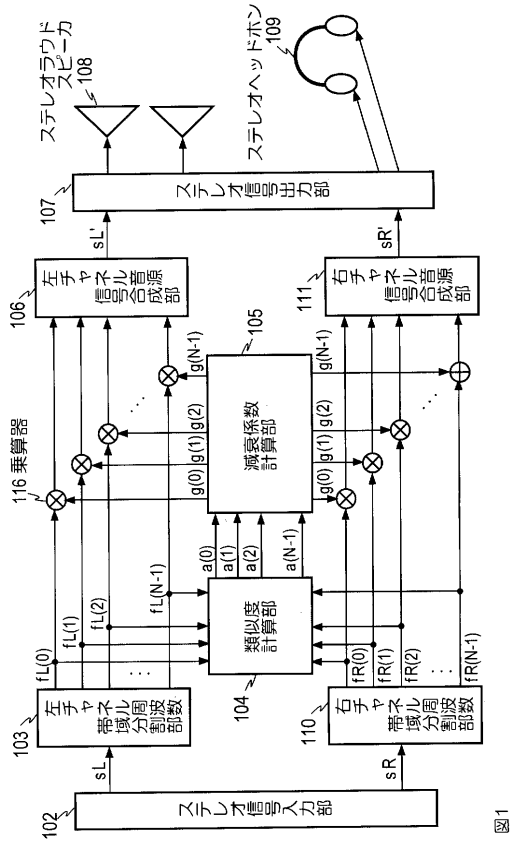


図1

【図 2】

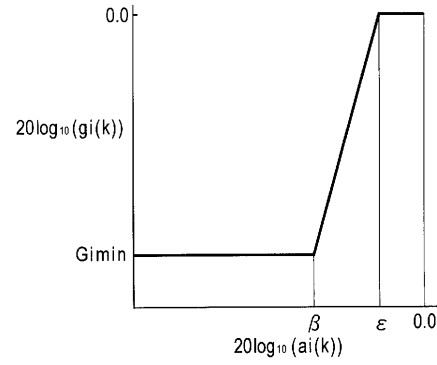


図2

【図 3】

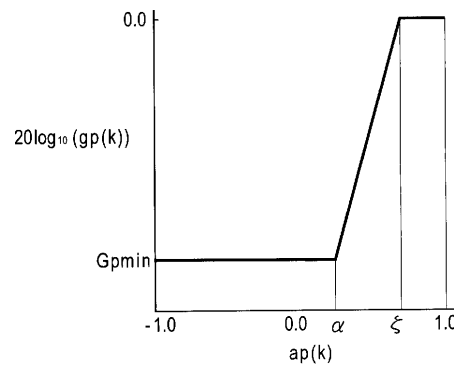


図3

【図 4】

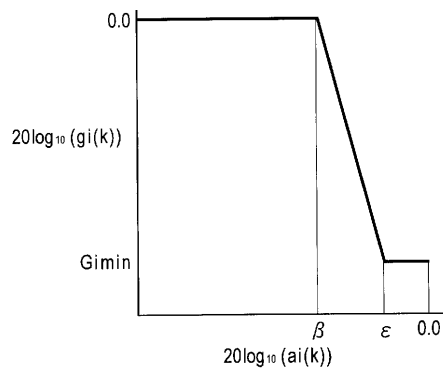


図4

【図 5】

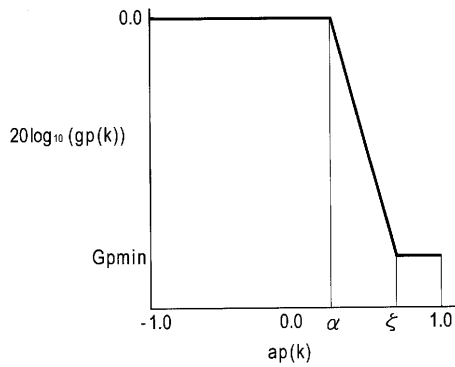


図5

【図 6】

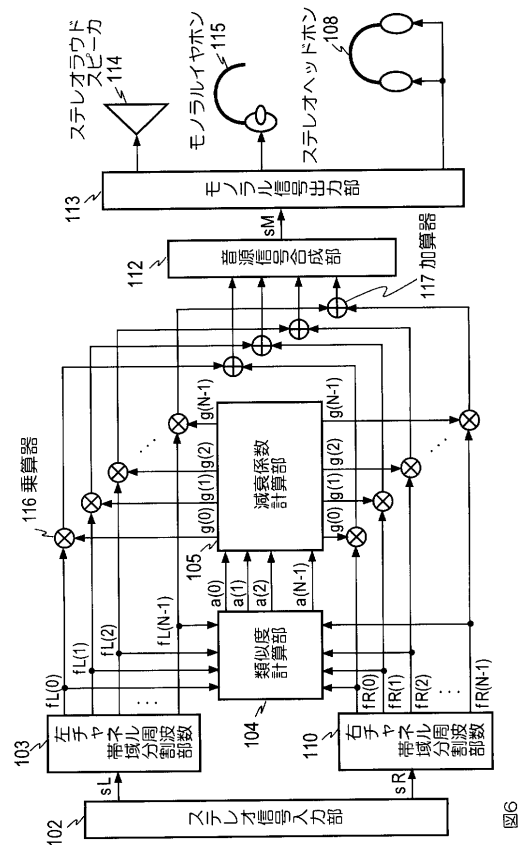


図6

【図 7】

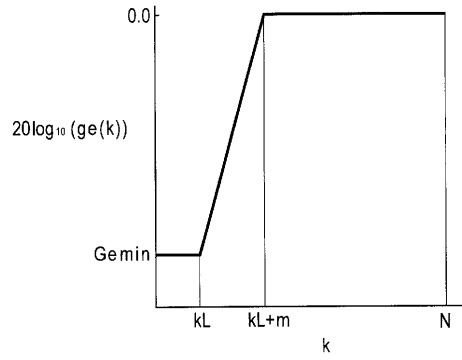


図7

【図 9】

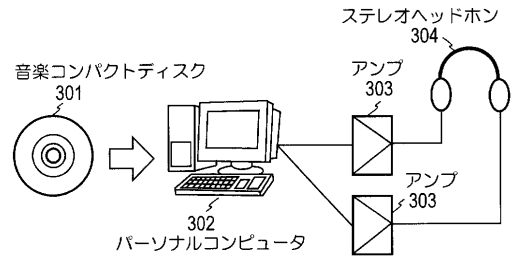


図9

【図 8】

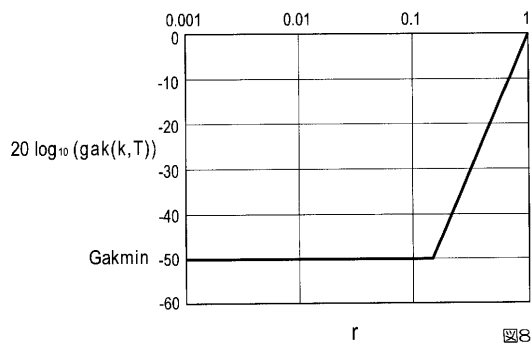


図8

【図 10】

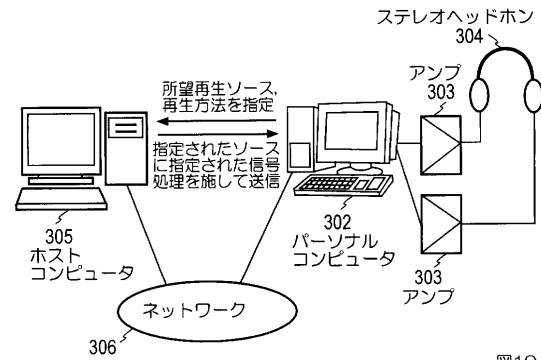


図10

【図 11】

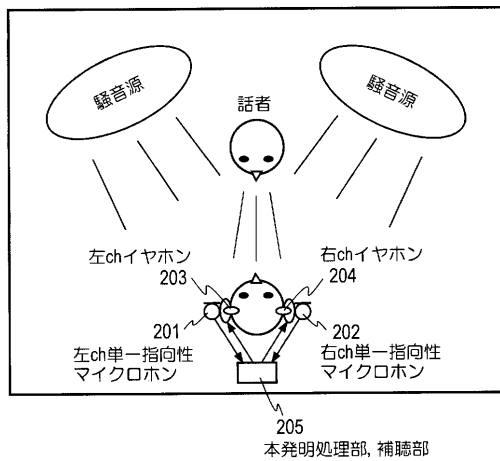


図11

フロントページの続き

(72)発明者 三好 正人

東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 江嶋 清仁

(56)参考文献 特開平6-269096 (JP, A)

特開平5-153698 (JP, A)

(58)調査した分野(Int.Cl.⁷, DB名)

H04S 1/00

H04S 5/00