

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2016-134706

(P2016-134706A)

(43) 公開日 平成28年7月25日(2016.7.25)

(51) Int.Cl.		F I		テーマコード (参考)
HO4R 3/00 (2006.01)		HO4R	3/00	5D220
HO3G 9/00 (2006.01)		HO3G	9/00	Z 5J030
HO3F 3/181 (2006.01)		HO3F	3/181	B 5J500

審査請求 未請求 請求項の数 9 O L (全 28 頁)

(21) 出願番号 特願2015-7380 (P2015-7380)
 (22) 出願日 平成27年1月19日 (2015.1.19)

(71) 出願人 504133110
 国立大学法人電気通信大学
 東京都調布市調布ヶ丘一丁目5番地1
 (74) 代理人 100107766
 弁理士 伊東 忠重
 (74) 代理人 100070150
 弁理士 伊東 忠彦
 (72) 発明者 高橋 弘太
 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
 Fターム(参考) 5D220 EE21
 5J030 AA05 AA10 AB01 AC10
 5J500 AA02 AA57 AC00 AF00 AF20
 AS05 AT03 LV04

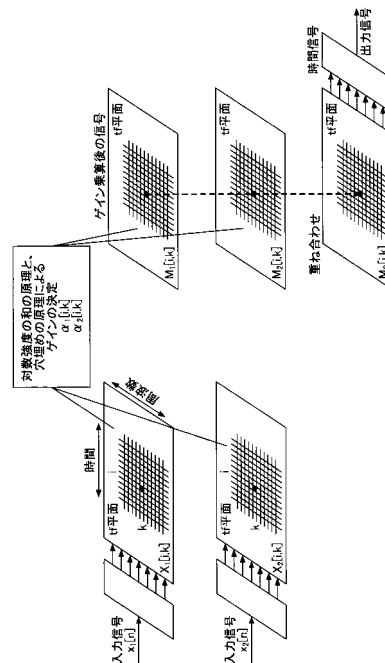
(54) 【発明の名称】 ミキシング装置、信号ミキシング方法、及びミキシングプログラム

(57) 【要約】

【課題】 入力信号のミキシング時に適切なゲインを設定し、ミキシング動作の向上と安定を図る。

【解決手段】 ミキシング装置は、時間領域の第1入力信号と第2入力信号をそれぞれ時間周波数平面上の第1信号と第2信号に展開する周波数解析部と、前記第1信号と前記第2信号を混合した混合信号を生成する信号処理部と、前記混合信号を時間領域の信号に変換する周波数時間変換部と、前記変換された信号を出力する信号出力部と、を有し、前記信号処理部は、前記時間周波数平面上の各点ごとに、出力信号の対数強度が前記第1信号の対数強度と前記第2信号の対数強度の和を超えない範囲内で前記第1信号のパワーを調整する第1ゲインと、前記第1信号のパワーの増加分を超えない範囲内で前記第2信号のパワーを減少させる第2ゲインとを決定するゲイン決定部と、前記第1ゲインで調整された前記第1信号と前記第2ゲインで調整された前記第2信号を加算する加算部とを有する。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

時間領域の第 1 入力信号と第 2 入力信号をそれぞれ時間周波数平面上の第 1 信号と第 2 信号に展開する周波数解析部と、

前記第 1 信号と前記第 2 信号を混合した混合信号を生成する信号処理部と、

前記混合信号を時間領域の信号に変換する周波数時間変換部と、

前記変換された信号を出力する信号出力部と、

を有し、

前記信号処理部は、

前記時間周波数平面の各点ごとに、出力信号の対数強度が前記第 1 信号の対数強度と前記第 2 信号の対数強度の和を超えない範囲内で前記第 1 信号のパワーを調整する第 1 ゲインと、前記第 1 信号のパワーの増加分を超えない範囲内で前記第 2 信号のパワーを減少させる第 2 ゲインとを決定するゲイン決定部と、

前記第 1 ゲインで調整された前記第 1 信号と前記第 2 ゲインで調整された前記第 2 信号を加算する加算部と、

を有することを特徴とするミキシング装置。

10

【請求項 2】

前記ゲイン決定部は、(a) 前記第 1 信号と前記第 2 信号を単純加算したパワーの一定倍を超えないように前記第 1 ゲインの調整に第 1 上限を設定する、(b) 前記第 1 ゲインに固定的な第 2 上限を設ける、または (c) 前記第 2 ゲインに固定的な下限を設ける、の少なくとも 1 つの条件を追加し、前記追加された条件を満たす範囲で前記第 1 ゲインと前記第 2 ゲインを決定することを特徴とする請求項 1 に記載のミキシング装置。

20

【請求項 3】

前記ゲイン決定部は、前記第 2 信号のパワーに対する前記第 1 信号のパワーの比率が所定の比率以下である場合に、前記少なくとも 1 つの条件を緩和することを特徴とする請求項 2 に記載のミキシング装置。

【請求項 4】

前記ゲイン決定部は、前記第 1 信号のパワーと前記第 2 信号のパワーを、それぞれ聴力限界レベルを基準とする第 1 聴感補正パワーと第 2 聴感補正パワーに補正し、前記出力信号の対数強度が、前記第 1 聴感補正パワーの対数強度と前記第 2 聴感補正パワーの対数強度の和を超えない範囲で前記第 1 ゲイン及び前記第 2 ゲインを決定することを特徴とする請求項 1 に記載のミキシング装置。

30

【請求項 5】

前記ゲイン決定部は、前記第 1 ゲインを平滑化した第 3 ゲインと、前記第 2 ゲインを平滑化した第 4 ゲインを生成し、

前記加算部は、前記第 3 ゲインで調整された前記第 1 信号と、前記第 4 ゲインで調整された前記第 2 信号を加算することを特徴とする請求項 1 に記載のミキシング装置。

【請求項 6】

前記ゲイン決定部は、前記第 1 ゲインと前記第 2 ゲインを前記時間周波数平面上の各点について逐次更新することを特徴とする、請求項 1 ~ 5 のいずれか 1 項に記載のミキシング装置。

40

【請求項 7】

前記第 1 入力信号はミキシング処理で優先的に明瞭化される優先音であり、前記第 2 入力信号は前記優先音以外の非優先音であることを特徴とする請求項 1 ~ 6 のいずれか 1 項に記載のミキシング装置。

【請求項 8】

時間領域の第 1 入力信号と第 2 入力信号を受け取り、

前記第 1 入力信号と前記第 2 入力信号を、それぞれ時間周波数平面上の第 1 信号と第 2 信号に展開し、

前記時間周波数平面の各点ごとに、出力信号の対数強度が前記第 1 信号の対数強度と前

50

記第 2 信号の対数強度の和を超えない範囲内で前記第 1 信号のパワーを調整する第 1 ゲインと、前記第 1 信号のパワーの増加分を超えない範囲内で前記第 2 信号のパワーを減少させる第 2 ゲインとを決定し、

前記第 1 信号を前記第 1 ゲインで乗算した第 1 乗算結果と、前記第 2 信号を前記第 2 ゲインで乗算した第 2 乗算結果を加算して混合信号を生成し、

前記混合信号を時間領域の信号に変換して出力する、
ことを特徴とする信号ミキシング方法。

【請求項 9】

信号ミキシング処理をコンピュータに実行させるミキシングプログラムであって、前記コンピュータに、

時間領域の第 1 入力信号と第 2 入力信号を受け取らせる手順と、

前記第 1 入力信号と前記第 2 入力信号を、それぞれ時間周波数平面上の第 1 信号と第 2 信号に展開させる手順と、

前記時間周波数平面の各点ごとに、出力信号の対数強度が前記第 1 信号の対数強度と前記第 2 信号の対数強度の和を超えない範囲内で前記第 1 信号のパワーを調整する第 1 ゲインと、前記第 1 信号のパワーの増加分を超えない範囲内で前記第 2 信号のパワーを減少させる第 2 ゲインとを決定させる手順と、

前記第 1 信号を前記第 1 ゲインで乗算した第 1 乗算結果と、前記第 2 信号を前記第 2 ゲインで乗算した第 2 乗算結果を加算して混合信号を生成させる手順と、

前記混合信号を時間領域の信号に変換して出力させる手順と、
を実行させることを特徴とするミキシングプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、2 以上の入力信号のミキシング技術に関する。

【背景技術】

【0002】

ミキシングの基本構成は、複数の入力信号の加算である。レコーディングや放送の現場では、ミキシングされた各種楽器や音声をバランスよく聴かせるために、加算器の前後にイコライザが配置されている。イコライザを用いることで、入力信号ごとに所望の周波数帯域を強調したり、バックグラウンドの中の重要でない周波数帯域の音量を下げて混合音中の優先音を引き立てる等の処理が行われる。

【0003】

一方、入力信号を時間周波数平面上に展開し、時間周波数平面上の点ごとに振幅調整と位相調整を行った後に加算して、時間軸上の信号に戻す「スマートミキシング」の手法が提案されている（たとえば、特許文献 1 参照）。音声や楽器音は、時間周波数平面上できめ細かな構造を持っている。各種入力信号を、時間周波数平面上での構造に合わせてミキシング処理を行うことで、よりきめ細かなミキシング処理が可能になる。特許文献 1 では、時間周波数平面上での処理対象点と所定の関係を有する他の点の入力データを用いて処理対象点の信号特性を判断し、その信号特性に応じて優先音の明りょう度を決定している。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特許第 5 0 5 7 5 3 5 号

【発明の概要】

【発明が解決しようとする課題】

【0005】

特許文献 1 では、入力信号に与えられるゲインを最適化する合理的な方法が規定されていない。特許文献 1 のミキシング方法を現実に応用する場合、周波数ごとのゲインは音源

10

20

30

40

50

にあわせて試行錯誤で決定されることになり、適切なゲインが設定されない場合がある。また、従来からの一般的なミキシング技術は経験と勘に頼っており、ゲイン設定やイコライザの特性設定についての合理的な基準が確立されていない。

【0006】

ミキシングで優先的に明瞭化されるべき音（以下、「優先音」と称する）の周波数ごとのゲインが適切に設定されないと、以下の問題が生じる。まず、優先音のゲインの変化が強すぎると、出力（混合音）中の優先音として不自然な感じになり、音として聴こえても音声として内容を聴き分けることができない。また、優先音の音量変化や音質変化が強すぎて不快感を生じさせることがある。逆にゲインの変化が弱すぎると、優先音が十分に聴き取れない。

10

【0007】

優先音以外の音（以下、「非優先音」と称する）のゲインが適切に設定できないと、以下の問題が生じる。ゲインの変化が強すぎると、出力（混合音）において非優先音の欠落や音質急変が目立ってしまい、違和感を感じる。違和感に気をとられることで優先音の聴き取りが阻害されてしまう。逆に、ゲインの変化が弱すぎると、優先音を十分に引き立てることができない。

【0008】

そこで、オーディオデータミキシング時の合理的なゲイン設定方法を確立して、ミキシング装置の動作の向上と安定を図ることを課題とする。

【課題を解決するための手段】

20

【0009】

上記課題を解決するために、本発明では、次の2つの原理を基本原理として用いる。

(1) 出力信号の対数強度を入力信号の対数強度の和を超えない範囲に限定する。これを「対数強度の和の原理」と称する。「対数強度の和の原理」によって、優先音が増強されすぎて混合音に違和感が生じることを抑制する。

(2) 非優先音のパワーの減少を、優先音のパワー増加分を超えない範囲に制限する。これを「穴埋めの原理」と称する。「穴埋めの原理」によって、混合音において非優先音が抑制されすぎて違和感が生じることを抑制する。

【0010】

具体的には、本発明の一側面において、ミキシング装置は、
時間領域の第1入力信号と第2入力信号をそれぞれ時間周波数平面上の第1信号と第2信号に展開する周波数解析部と、

30

前記第1信号と前記第2信号を混合した混合信号を生成する信号処理部と、

前記混合信号を時間領域の信号に変換する周波数時間変換部と、

前記変換された信号を出力する信号出力部と、

を有し、

前記信号処理部は、

前記時間周波数平面の各点ごとに、出力信号の対数強度が前記第1信号の対数強度と前記第2信号の対数強度の和を超えない範囲内で前記第1信号のパワーを調整する第1ゲインと、前記第1信号のパワーの増加分を超えない範囲内で前記第2信号のパワーを減少させる第2ゲインとを決定するゲイン決定部と、

40

前記第1ゲインで調整された前記第1信号と前記第2ゲインで調整された前記第2信号を加算する加算部と、

を有する。

【発明の効果】

【0011】

入力信号のミキシング時に適切なゲインが設定され、ミキシング装置の動作の向上と安定を図ることができる。

【図面の簡単な説明】

【0012】

50

【図 1】本発明の基本原理を説明する図である。

【図 2】実施形態のミキシング装置の概略構成図である。

【図 3】優先音と非優先音のパワーをリスナーの可聴レベルとともに示す図である。

【図 4】図 3 の優先音と非優先音のそれぞれから可聴レベルを減算して、リスナーが感じる音強度（聴感補正パワー）としてプロットした図である。

【図 5】図 4 の優先音と非優先音の対数尺度での和を示す図である。

【図 6】図 4 の優先音と非優先音の単純加算和を一定倍した結果を示す図である。

【図 7】図 5（対数強度和）と図 6（単純加算和の一定倍）の双方を満たす範囲を加算（ミキシング）時の上限とする図である。

【図 8】図 7 の範囲内でゲイン調整された優先音と非優先音を示す図である。

10

【図 9】最小可聴パワーの選択に用いるラウドネス曲線の例を示す図である。

【図 10】実施形態のミキシング装置の動作例を示す図である。

【図 11】ゲインを平滑化した場合のミキシング装置の動作状態を示す図である。

【図 12】補正係数 $B[k]$ を直接設定する場合の例を示す図である。

【図 13 A】実施形態のミキシング信号処理方法の一例を示すフローチャートである。

【図 13 B】実施形態のミキシング信号処理方法の一例を示すフローチャートであり図 13 A に続くフローチャートである。

【図 14】実施形態のミキシング信号処理方法の別の例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】

【0013】

20

図 1 は、本発明の基本原理を説明する図である。以下の説明では、入力信号 $x_1[n]$ と $x_2[n]$ をミキシングする場合を例にとる。入力信号 $x_1[n]$ は、たとえば音声等の優先信号とする。入力信号 $x_2[n]$ は、バックグラウンド音等の非優先信号とする。

【0014】

入力信号 $x_1[n]$ と $x_2[n]$ は、周波数解析によって、それぞれ時間周波数平面（図中「 $t-f$ 平面」と表記されている）上に展開される。周波数解析は、短時間 FFT（Fast Fourier Transform）、短時間フーリエ変換、ウェーブレット変換、フィルタバンクによる変換、ウィグナー分布などの時間周波数分布への変換等、任意の手法を用いることができる。時間周波数平面上に展開された各信号を、 $X_1[i, k]$ 、 $X_2[i, k]$ とする。 $X_1[i, k]$ 、 $X_2[i, k]$ は、時間方向の座標値 i と、周波数方向の座標値 k で表される時間周波数平面上での入力信号の点である。

30

【0015】

時間周波数平面に展開された入力信号の各点のパワーに基づき、「対数強度の和の原理」と「穴埋めの原理」を用いて、各点での優先音と非優先音のゲインを決定する。「対数強度の和の原理」は、上述のように出力信号のパワーを入力信号の対数強度の和を超えない範囲に限定する処理である。「穴埋めの原理」は、非優先音のパワーの低減を、優先音のパワー増加分を超えない範囲に限定する処理である。これらの原理の具体的な処理方法については後述する。

【0016】

なお、実施形態では最適なゲインの決定のために、対数強度の和の原理に基づく処理（1）と、穴埋めの原理に基づく処理（2）に加えて、オプションとしてさらに以下の処理（3）～（5）を導入する。

40

（3）ゲイン決定に際して、(a)対数強度の和の原理により決定されるパワーの増加率に対して、入力音の単純加算値を一定倍したレベルを超えないように上限を設ける、(b)優先音のゲインに固定的な上限を設ける、(c)非優先音のゲインに固定的な下限を設ける、という3つの条件のうちの少なくとも1つを加える。これにより、混合音をさらに自然で穏やかにすることができる。

（4）極端に信号対雑音比が低下している時間区間では、上記（3）の上限や下限を緩和する。これにより、信号対雑音比が低い時間区間でも優先音を目立たせてリスナーにとって聞き取りやすい音にする。

50

(5) ミキシング処理におけるパラメータは、最適化問題の解として算出するのではなく逐次更新するように構成する。逐次更新の採用によって、「方程式を解く」ことを「不等式の真偽の判定」に置き換えることができ、指数関数、対数関数、乗算などの演算を排除して、乗算と加減算だけの高速アルゴリズムを構成することができる。これにより、FPGA (Field-programmable gate array) 等のプログラマブルロジックデバイスへの実装や、DAW (digital audio workstation) 用のプラグインとしての実装が容易になり、リアルタイム処理が実現する。

【0017】

時間周波数平面上の各点で、 $X_1[i, k]$ 、 $X_2[i, k]$ にそれぞれ対応するゲインを乗算する。ゲイン乗算後の信号を、 $M_1[i, k]$ 、 $M_2[i, k]$ とする。ゲイン調整された信号 $M_1[i, k]$ 、 $M_2[i, k]$ を加算して時間周波数平面上で2つの信号を重ね合わせる。その後、時間領域の信号に戻して、混合音を出力する。

10

【0018】

このように、時間周波数平面上の各点で入力信号に対するゲインを決定して乗算することで、自然な混合音を生成することができる。

【0019】

図2は、実施形態のミキシング装置1の概略図である。ミキシング装置1は、信号入力部11、周波数解析部12、信号処理部15、周波数時間変換部16、及び信号出力部17を有する。信号入力部11は、ミキシングの対象となる複数の入力信号を入力する。入力信号はたとえばオーディオ信号であり、音声等の優先信号と、バックグラウンド音等の非優先信号を含む。

20

【0020】

周波数解析部12はたとえば、短時間FFTにより入力信号を時間周波数平面上に展開する。信号処理部15は、時間周波数平面上の各点で、入力信号のパワーを算出し、パワーを平滑化した後、ゲイン決定部151で優先信号と非優先信号に対するゲインをそれぞれ算出する。そして、優先信号と非優先信号に対して算出されたそれぞれのゲインを乗算した後、加算して加算結果を出力する。周波数時間変換部16は、信号処理部15からの出力信号を時間領域の信号に変換する。信号出力部17は、時間領域に復元された信号を出力する。

【0021】

図3～図8を参照して、信号処理部15での基本処理を説明する。以下の説明で用いられる記号を、表1と表2に示す。表1は定数の記号の一覧であり、表2は変数の記号の一覧である。

30

【0022】

【表 1】

定数の意味	記号	デフォルト値
サンプリング周波数	F_S	44.1 kHz
信号 $x_j[n]$ の絶対値の最大値	x_{max}	32768
FFT点数	N_F	256
F_S の1/2の周波数に相当するビン番号	N_H	128
解析時の窓関数	$h[n]$	ガウス型
解析時の窓関数の広がり	σ	$\frac{1}{3\sqrt{2}}$
窓関数の半値幅	N_h	128 点
フレームシフト	N_d	1 点
等ラウドネス曲線から選ぶ最小可聴曲線	L_p	30 phon
フルスケール信号をSPLで表現	L_f	106 phon
パワー平滑化の時定数	τ_s	20 ms
ゲイン平滑化の時定数	τ_a	5ms
有音判定のパラメータ	T_e	2.0
低SNR判定のパラメータ	T_{SN}	10.0
混合パワーの増幅限界(振幅比に換算)	T_G	4.0
優先音に対するゲインの上限	T_{1H}	4.0
非優先音に対するゲインの下限	T_{2L}	0.001
優先音に対するゲインの増減刻み	Δ_1	0.001
(優先音のゲインを平滑化する場合)		0.01
非優先音に対するゲインの増減刻み	Δ_2	0.001
(非優先音のゲインを平滑化する場合)		0.01
ゲイン調整する最低周波数(優先音)	f_{1L}	350 Hz
ゲイン調整する最低周波数(非優先音)	f_{2L}	350 Hz
ゲイン調整する最高周波数(優先音)	f_{1H}	20000 Hz
ゲイン調整する最高周波数(非優先音)	f_{2H}	20000 Hz
ゲイン調整する最低ビン(優先音)	k_{1L}	$\text{rd}(N_F f_{1L}/F_S)$
ゲイン調整する最低ビン(非優先音)	k_{2L}	$\text{rd}(N_F f_{2L}/F_S)$
ゲイン調整する最高ビン(優先音)	k_{1H}	$\text{rd}(N_F f_{1H}/F_S)$
ゲイン調整する最高ビン(非優先音)	k_{2H}	$\text{rd}(N_F f_{2H}/F_S)$
等ラウドネス曲線(20 phon)	$C_{20}[k]$	図9(B)
等ラウドネス曲線(70 phon)	$C_{70}[k]$	図9(B)
最小可聴パワー	$A[k]$	式(11)
最小可聴パワーの逆数	$B[k]$	式(13)

10

20

30

40

【表 2】

変数の意味	優先音	非優先音	型
時間周波数領域での入力信号	$X_1[i,k]$	$X_2[i,k]$	複素数
入出力間のゲイン	$\alpha_1[i,k]$	$\alpha_2[i,k]$	正の実数
平滑化した入出力間のゲイン	$\beta_1[k,i]$	$\beta_2[i]$	正の実数
時間周波数領域での出力信号	$Y[i,k]$		複素数
平滑化パワー	$E_1[i,k]$	$E_2[i,k]$	複素数
聴感補正パワー	$P_1[i,k]$	$P_2[i,k]$	正の実数
聴感補正総パワー	$Q_1[i]$	$Q_2[i]$	正の実数
修正前の α_j での聴感補正パワー	$L_1[i,k]$	$L_2[i,k]$	正の実数
$L_1[i,k]+L_2[i,k]$	$L[i,k]$		正の実数
有音判定の結果	$e[i]$		boolean
低SN判定の結果	$l[i]$		boolean
ブースト判定の結果	$b[i]$		boolean
ブーストレシオの分子	$b_n[i]$		正の実数
ブーストレシオの分母	$b_d[i]$		正の実数

10

20

< 対数強度の和の原理 >

図 3 は、時間周波数平面のある時刻における優先音（太線）と非優先音（実線）のパワー [dB] を周波数の関数として模式的に示したものである。このパワーは、信号処理部 15 で平滑化されたパワー値 E である。点線は、このレベル以上であればリスナーが音を検知できる聴力限界を示す。

30

【 0 0 2 4 】

人間の聴覚は、パワーの対数で音の強さを感じると言われている。この考え方に基けば、点線の可聴限界よりも 10 dB 高い音成分同士はほぼ同じ強さを感じられ、聴力限界よりも 20 dB 高い音成分同士もほぼ同じ強さを感じられる。また、可聴限界よりも 10 dB 高い音成分と、20 dB 高い音成分の音量差は、可聴限界よりも 20 dB 高い音成分と、30 dB 高い音成分の音量差と同じであるとみなすことができる。

【 0 0 2 5 】

図 4 は、図 3 の優先音と非優先音のそれぞれから、点線の聴力限界可聴レベルを減算してプロットしたものである。図 4 で得られる優先信号と非優先信号のパワー値は、人間の聴力限界をゼロ dB として補正された「聴感補正パワー P」である。図 4 の縦軸の大きさが、リスナーが感じる音強度となる。

40

【 0 0 2 6 】

図 5 は、図 4 で聴感補正された優先音と非優先音の対数尺度での和を点線 A で示す。上述した人間の聴覚についての知見によれば、リスナーに聴感補正された優先音と非優先音の和（対数強度の和）を混合音として感じさせることが妥当であるという原理に行き着く。すなわち、混合音のパワーを図 4 の点線 A とする。これが、「対数強度の和の原理」である。人間の聴力限界と等しい音の大きさを 1 倍とすると、20 dB は 10 倍であり、その対数は 1 である。40 dB は 100 倍であり、その対数は 2 である。対数の尺度で表現

50

すると、人間の聴力限界の10倍の音と100倍の音の和は、 10^3 すなわち1000倍のパワーとなる。

【0027】

ただし、この原理をそのまま使うと、聴感補正後(図4)の優先音のパワー P_1 と非優先音のパワー P_2 の和で表される混合音のパワーは $P_1 \times P_2$ になり、場合によってはパワーが大きくなりすぎる。たとえば、 $P_1 = P_2 = 10^5$ であれば、混合音のパワーは 10^{10} になり、多くの処理系でオーバーフローが起きる可能性がある。そこで、「対数強度の和の原理」で導かれる混合音のパワーをミキシングの上限値として用いる。

<追加条件>

対数強度の和の原理で求められる混合音は場合によって音量が強くなりすぎるという問題を解決するために、処理(3)の3つの条件(a)~(c)の少なくともひとつを追加する。

【0028】

図6は、条件(a)を示す図である。条件(a)は、混合音のパワー増加率を、2つの入力音のパワーの単純加算値の一定倍率(比率)に制限するものである。自然界において、人間は2つの音の加算(混合音)を聴くときに、単純加算での混合を聴いている。たとえば、聴力限界の10倍の入力音と、100倍の入力音の単純加算値は110倍である。これに対して、対数尺度での加算値は1000倍となる。

【0029】

そこで、混合パワーの増幅限界 T_G を設定する。混合パワーの増幅限界 T_G は単純加算値の振幅比で表わし、たとえば $T_G = 4.0$ と設定する。この場合、単純加算で求められる振幅(たとえば110)の4.0倍が混合パワーの増幅限界となる。図6の点線Bが、優先音と非優先音の和(混合)のパワーを単純加算値の所定倍率 T_G に制限する増幅限界である。

【0030】

図7は、図5の対数強度の和の原理で求められる混合音パワーの上限と、図6の単純加算の一定倍で設定される増幅限界の双方を満たすレベル、すなわちいずれか低い方のパワーレベルを信号加算の上限として設定する処理を示す。図7で、点線A(対数強度の和の原理で求められる上限)と、点線B(単純加算に基づく増幅限界)のパワーの低い方を選択した実線が、信号加算時の上限となる。

<穴埋めの原理>

図8は、図7の信号加算の上限の範囲内でのゲインの設定を示す図である。優先音の聴き取りを向上させるためには、時間周波数平面上の必要な部分において、非優先音を抑制する必要がある。抑制量は多ければ多いほどよいというものではない。非優先音を無条件に抑制してしまうと、非優先音の音量変化が刺激的になりすぎて、違和感が増すだけでなく、優先音の聴き取りが妨害され得るからである。したがって、非優先音の抑制に対しても合理的な基準が必要になる。

【0031】

実施形態では、優先音のゲインを増加させたことによってパワーが増大した分を超えない範囲で、非優先音のパワーが低減される。つまり、非優先音の抑制によって作られた穴を、優先音の増大によって埋める。この処理により、非優先音に対する違和感の発生を回避できる。

【0032】

図8において、優先音は、図7の信号加算の上限の範囲内で増幅される。点線Cがゲイン調整後の優先音のパワーである。他方、非優先音は、優先音の変化量(すなわちパワー増大分)を超えない範囲で低減される。点線Dが、ゲイン調整後の非優先音のパワーである。

【0033】

このように、信号処理部15は、優先音パワーの増大に対する制限と、非優先音パワーの減少に対する制限に基づいて、それぞれのゲインを決定する。時間周波数平面の各点での優先音の増大と非優先音の減少(凹凸)を決定したものがゲインマスクである。

10

20

30

40

50

【 0 0 3 4 】

なお、処理(3)の条件(a)に替えて、あるいは条件(a)に加えて、優先音のゲインに固定的な上限を設定する条件(b)や、非優先音のゲインに固定的な下限を設定する条件(c)を追加してもよい。対数強度の和の原理と、穴埋めの原理にこれらの条件を追加することで自然な混合音を生成することができる。

< 短時間 F F T >

次に、周波数解析部 1 2 の処理の詳細を説明する。実施形態では、周波数解析部 1 2 は F F T 点数として 2 5 6 程度の短時間 F F T を実施する。この短時間 F F T は、一次元の入力信号を、2次元の時間周波数 (t f) 平面上に展開する処理である。

【 0 0 3 5 】

サンプリング周波数 F_s で採取された信号 $x_1[n]$ と $x_2[n]$ を、それぞれ優先音と非優先音とする。両信号 $x_j[n]$ ($j = 1, 2$) を N_d 点シフトで N_F 点の短時間フーリエ変換する。ブロック番号 i 、周波数ビン番号 k における変換結果を $X_j[i, k]$ とすると、 $X_j[i, k]$ は式(1)で表される。

【 0 0 3 6 】

【数 1】

$$X_j[i, k] = \sum_{n=-N_h+1}^{N_h-1} h[n] x_j[iN_d+n] \exp\left(-\frac{2\pi i k n}{N_F}\right) \quad (j=1, 2) \quad (1)$$

ここで、 $h[n]$ は窓関数である。 N_h は窓関数の幅を決めるパラメータであり、 $|n| \leq N_h$ である n に対して、 $h[n] = 0$ とする。窓関数としては、ハン窓、ハニング窓、ガウス窓等、任意の窓関数を用いることができる。実施形態では、式(2)のガウス窓を使用する。

【 0 0 3 7 】

【数 2】

$$h[n] = \begin{cases} \exp\left(-\frac{n^2}{2\sigma^2 N_h^2}\right), & |n| < N_h \\ 0, & |n| \geq N_h \end{cases} \quad (2)$$

ここで、 σ は窓関数の幅を調整するパラメータである。

【 0 0 3 8 】

実信号の F F T 結果は、正負の周波数出力が互いに複素共役の関係にあるため、負の周波数を扱う必要はない。 $N_H = N_F / 2$ として、 $0 \leq k < N_H$ の範囲の周波数ビンのみを扱えばよい。また、 $N_d = 1$ のときの逆 F F T を、加算と加算後の定数倍だけで済ませるために、式(3)の変換を行っておく。加算と乗算のみの演算により計算量を軽減する。

【 0 0 3 9 】

10

20

30

40

【数 3】

$$X_j[i, k] \leftarrow 2X_j[i, k] \quad (0 < k < N_H) \quad (3)$$

逆 F F T は、ミキシング装置 1 の周波数時間変換部 1 6 で行われる。実施形態の信号処理部 1 5 は位相処理を行わず、振幅処理だけで入力信号を混合する。これは F F T の点数 N_F が少ないためである。一例として、 $N_F = 256$ 、サンプリング周波数 $F_S = 44.1$ kHz とする。これらの条件では音声の線スペクトル構造を分解するには不十分であり、ひとつの周波数ビンに複数の高調波成分が混在し位相の利用が困難になる。

10

【0040】

実施形態では振幅処理のみを行うので、ミキシング出力 $Y[i, k]$ は、 $X_1[i, k]$ と $X_2[i, k]$ に、それぞれゲイン $\alpha_1[i, k]$ と $\alpha_2[i, k]$ を乗算して加算することで生成される。

【0041】

【数 4】

$$Y[i, k] = \alpha_1[i, k] X_1[i, k] + \alpha_2[i, k] X_2[i, k] \quad (4)$$

20

時間領域での出力 y は、式 (4) の $Y[i, k]$ を逆 F F T して得られる。

【0042】

【数 5】

$$y[iN_d + n] = \frac{1}{N_F} \sum_{k=0}^{N_F} Y[i, k] \exp\left(\frac{2\pi i k n}{N_F}\right) \quad (5)$$

30

ここで、1 サンプルシフト ($N_d = 1$) の場合は、 n をゼロに固定しても y を生成できるので、式 (6) のように簡単な処理になる。

【0043】

【数 6】

$$y[i] = \frac{1}{N_F} \sum_{k=0}^{N_F-1} Y[i, k] \quad (6)$$

40

さらに、 $X[i, k]$ について、式 (3) の変換をしておくことで、加算する周波数ビンを、式 (7) のようにほぼ半分に減らすことができる。

【0044】

【数 7】

$$y[i] = \frac{1}{N_F} \sum_{k=0}^{N_H} Y[i, k] \quad (7)$$

< 平滑化パワーの計算 >

次に、信号処理部 15 による平滑化パワーの計算を説明する。パワーの平滑化に先立って、時間周波数領域の信号 $X_j [i, k]$ の絶対値の 2 乗 ($|X_j [i, k]|^2$) を計算し、これを平滑化する。平滑化として、たとえば式 (8) で示される指数平滑化を用いる。指数平滑化法は、計算量と必要なメモリ量が少ないので、FPGA 化に適している。

【0045】

【数 8】

$$E_j [i, k] = \mu E_j [i-1, k] + (1 - \mu) |X_j [i, k]|^2 \quad (j=1, 2) \quad (8)$$

20

ここで、 μ は指数平滑化法の係数であり、平滑の時定数 τ_s から式 (9) で導出する。

【0046】

【数 9】

$$\mu = \exp\left(-\frac{N_d}{\tau_s F_s}\right) \quad (9)$$

30

式 (8) を IIR (Infinite Impulse Response) 型デジタルフィルタとみたとき、そのインパルス応答がピーク値の $1/e$ に減衰する時間が τ_s である。実施形態では、平滑化に指数平滑化を用いるが、FIR (Finite Impulse Response) フィルタ、IIR フィルタ等、任意の平滑化法を用いることができる。

< 最小可聴パワーの計算 >

入力信号のミキシングには、時間周波数平面上での各点の成分について、それが聴こえる成分なのか、聴こえない成分なのかを判定する必要がある。そのために、それぞれの音源 j の各周波数ビン k について、その成分が可聴であるための最小のパワー $A [k]$ を定義する。

40

【0047】

図 9 (A) は、国際標準化規格 ISO 226 : 2003 で規定された等ラウドネス曲線のうち、20 phon と 70 phon の曲線から主要部分を抽出してサンプリングしたものである。これらをそれぞれ $C_{20} [k]$ と $C_{70} [k]$ と呼ぶ。

【0048】

本来であれば、0 phon の曲線が最小可聴パワーである。しかし、リスナーにどのような音量で音が提示されるかは電気音響装置のボリューム設定によってその都度違うので、実施形態のミキシング装置 1 の信号処理部 15 は、ラウドネスレベルが指定された値になったときに可聴であると判断する。ミキシング装置 1 のユーザが最小可聴パワーとして、等ラウドネス曲線の中から L_p phon の曲線を選択できるように設計してもよい。 L_p ph

50

onの曲線は、 $C_{20}[k]$ と $C_{70}[k]$ を補間または補外した近似値として、式(10)で得ることができる。

【0049】

【数10】

$$C_{L_p}[k] = \frac{(L_p - 20)C_{70}[k] + (70 - L_p)C_{20}[k]}{70 - 20} \quad (10)$$

10

なお、平滑化されたパワーレベル $E_j[i, k]$ が可聴であるか否かを判断するとき、 $E_j[i, k]$ を $C_{L_p}[k]$ と比較することはできず、信号 $x_j[n]$ の絶対値の最大値 x_{max} や窓関数 $h[n]$ を勘案する必要がある。そこで、 $C_{L_p}[k]$ を式(11)のように変換し、最小可聴パワー $A[k]$ を導出する。

【0050】

【数11】

$$A[k] = \left(x_{max} \sum_n h[n] \right)^2 \exp\left(\frac{\log(10)}{10} (C_{L_p}[i] - L_f) \right) \quad (11)$$

20

ここで、定数 L_f は、 $x_j[n]$ がフルスケールの信号であったときに、それを図9(A)の縦軸の音圧レベル(SPL: Sound Pressure Level)の何dBに相当させるかを自由に設定するための定数である。

【0051】

ミキシング装置1の動作を自由に設定するという観点に立てば、 $C_{20}[k]$ と $C_{70}[k]$ をISO 226:2003に準拠させる必然性はなく、図9(B)のような等ラウドネス曲線を生成してもよい。図9(B)の曲線を用いると、8kHz程度の高い周波数の音は可聴とみなされやすくなるので、優先音においてこの帯域付近の成分は尊重されることになる。結果として、優先音にメリハリがつくので、実際に聴いた感じとして高評価が得られやすい。後述する実験結果は、図9(B)の $C_{20}[k]$ と $C_{70}[k]$ を用いている。

30

<聴感補正パワーの計算>

ゲインを決定するための聴感補正パワーの計算について説明する。聴感補正パワーの計算は図4の処理に該当する。平滑化後のパワー $E_j[i, k]$ を最小可聴パワー $A[k]$ で除算した結果が1より大きければ可聴であり、その可聴のレベルは、 $E_j[i, k] / A[k]$ で表現される。たとえば、 $E_j[i, k] / A[k] = 100$ であれば、最小可聴の音に比べて100倍のパワーを持っている。

40

【0052】

この評価法では除算が生じるが、FPGAは除算が苦手である。そこで、最小可聴パワー $A[k]$ は事前に決定されているので、あらかじめその逆数 $B[k]$ を作っておくことで除算を回避する。

【0053】

【数 1 2】

$$B[k] = \frac{1}{A[k]} \quad (12)$$

$$= \frac{1}{(x_{\max} \sum_n h[n])^2} \exp\left(-\frac{\log(10)}{10}(C_{L_p}[k] - L_f)\right) \quad (13)$$

10

この補正係数 $B[k]$ を用いて、平滑化パワー $E_j[i, k]$ から聴感補正パワー $P_j[i, k]$ を式 (14) の乗算により生成する。

【0054】

【数 1 3】

$$P_j[i, k] = B[k] E_j[i, k] \quad (14)$$

20

聴感補正パワー $P_j[i, k]$ は、時間周波数平面の 1 点ごとに値が決まる量である。各点での聴感補正パワー $P_j[i, k]$ から、式 (15) で定義する聴感補正総パワー $Q_j[i]$ を算出する。

【0055】

【数 1 4】

$$Q_j[i] = \sum_{k=0}^{N_H} P_j[i, k] \quad (15)$$

30

聴感補正総パワー $Q_j[i]$ は、各点のパワーを周波数方向に積算した量であり、リスナーが感じることのできる音のエネルギーの簡略化された推定値である。聴感補正総パワー $Q_j[i]$ は、以下で説明する時区間の属性判定に用いられる。

< 時区間の属性判定計算 >

信号処理部 15 は、ミキシング処理を行う際に各時間区間において、有音判定、低 SNR (Signal to Noise Ratio: 信号対雑音比) 判定、及びブースト判定を行う。これらの判定は、上述した処理 (4) と関連する。

【0056】

40

まず、有音判定について説明する。有音でない部分でミキシング処理を行うと、優先信号に含まれるわずかな音、たとえばナレーションの合間の風の音などが増強され、好ましくない混合音が生成される。これを防ぐために、優先音の中でこの時間区間は聴き落してはならないという部分を有音部としてあらかじめ設定しておく。

【0057】

有音部の判定は、有音時に 1 となる関数 $e[i]$ を式 (16) により定義する。

【0058】

【数 15】

$$e[i] = \begin{cases} 1 & Q_1[i] > (N_H + 1) T_e \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

ここで、 T_e は有音判定のためのパラメータである。たとえば、 $T_e = 1.0$ とすれば、全ピンが可聴判定ぎりぎりであるときに有音と判定される。

10

【0059】

次に、低SNR判定について説明する。図5～図7を参照して説明したように、ミキシング装置1では、優先音のゲインに上限を設ける。このため、優先音が非優先音に比べて極端にレベルが低い場合は、ゲインの上限値を使っても、優先音の聴き取りが困難になる場合がある。これを防ぐため、低SNRか否かを判定し、低SNRと判定された時間区間で上限の引き上げを行う。

【0060】

低SNRの判定は、低SNR時に1となる関数 $\ell[i]$ を式(17)で定義することができる。

【0061】

20

【数16】

$$\ell[i] = \begin{cases} 1 & T_{SN}^2 Q_1[i] < Q_2[i] \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

ここで、 T_{SN} は低SNR判定のためのパラメータである。たとえば、 $T_{SN} = 10.0$ とすれば、聴感補正総パワーについて、優先音と非優先音の間に、振幅比で10倍（パワー比で100倍）の開きがあるときに低SNRと判定される。

30

【0062】

最後に、ブースト判定について説明する。ブースト判定は、優先音が有音であり、かつ低SNRであるときに行われる。ブースト時に1となる $b[i]$ を、式(18)で定義する。

【0063】

【数17】

$$b[i] = \begin{cases} 1 & e[i] = 1 \wedge \ell[i] = 1 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

40

ブースト判定が真となったときに、除算なしでブースト動作を行うために、ブーストレシオを分数表示したときの分子 b_n と分母 b_d を、それぞれ式(19)と式(20)で求めておく。これらを用いて、各種の評価基準に対して b_n / b_d のブーストが行われる。

【0064】

【数 18】

$$b_n[i] = \begin{cases} Q_2[i] & b[i] = 1 \\ 1 & \text{otherwise} \end{cases} \quad (19)$$

$$b_a[i] = \begin{cases} T_{SN}^2 Q_1[i] & b[i] = 1 \\ 1 & \text{otherwise} \end{cases} \quad (20)$$

10

< ゲインの生成 >

ゲインの生成は実施形態のミキシング処理の核心である。優先音のためのゲイン $g_1[i, k]$ と、非優先音のためのゲイン $g_2[i, k]$ を生成する。ミキシング装置 1 の動作開始時は、両ゲインを 1 に初期化しておく。すべての k について、 $g_1[0, k] = g_2[0, k] = 1$ である。

【0065】

今、時間ブロック i に関する処理を始めたところであるとする。このとき、すべての k について $g_1[i-1, k]$ と $g_2[i-1, k]$ がすでに決定している。 $g_1[i, k]$ は $g_1[i-1, k]$ に g_1 を使った増減を行うことで更新される。 $g_2[i, k]$ は $g_2[i-1, k]$ に g_2 を使った増減を行うことで更新される。

20

【0066】

$g_1[i, k]$ の増減は $g_1[i-1, k]$ に対して $(1 + g_1)$ の乗算、もしくは $(1 + g_1)^{-1}$ の乗算を行うことで実現する。一方、 $g_2[i, k]$ の増減は、 $g_2[i-1, k]$ に $\pm g_2$ を加算することで行う。

【0067】

このように異なる更新方法を採用する理由を説明する。優先音のためのゲイン $g_1[i, k]$ は、条件によっては 10 以上の値にすることがある。特に、 $g_1[i, k]$ が大きいときには変化の差分を大きくする必要があり、乗算的更新が適している。一方、非優先音のためのゲイン $g_2[i, k]$ は、0 から 1 の範囲に限定されているため、一定刻みで十分であるし、一定刻みのほうが低レベルになったときの信号の抑圧をシャープに行うことができる。

30

【0068】

ゲイン $g_1[i, k]$ 、 $g_2[i, k]$ の更新を加減算と乗算のみにしたのは、処理 (5) で説明したとおり、演算を軽くするためである。方程式を解いて次のゲインを決めるという方法では、多くの場合、除算や平方根等が発生する。また、ゲインが大きく変動して出力波形に不連続が生じることも懸念される。

【0069】

これに対し、実施形態では微量の増減に限定することで、ゲインは滑らかに変化し、出力に段差が生じることを抑止できる。

40

(A) ゲイン調整信号の聴感補正パワーの計算

もし、ゲインの増減を行わず、ひとつ前のフレームのゲイン $g_j[i-1, k]$ をそのまま用いた場合、すなわち、 $g_j[i, k] = g_j[i-1, k]$ とした場合、音源 j に関する優先音と非優先音の聴感補正パワーは、それぞれ式 (21) と式 (22) で表される。

【0070】

【数 19】

$$L_1[i, k] = \alpha_1[i-1, k]^2 P_1[i, k] \quad (21)$$

$$L_2[i, k] = \alpha_2[i-1, k]^2 P_2[i, k] \quad (22)$$

このとき、ミキシング出力の聴感補正パワー $L[i, k]$ は、両音源の寄与の和として式 (23) で表される。

10

【0071】

【数 20】

$$L[i, k] = L_1[i, k] + L_2[i, k] \quad (23)$$

優先音のゲインを増加させた場合の聴感補正パワーを $L_{1p}[i, k]$ と定義しておく。

20

【0072】

【数 21】

$$L_{1p}[i, k] = ((1 + \Delta_1) \alpha_1[i-1, k])^2 P_1[i, k] \quad (24)$$

増加時のミキシング出力の聴感補正パワーを $L_p[i, k]$ とする。

30

【0073】

【数 22】

$$L_p[i, k] = L_{1p}[i, k] + L_2[i, k] \quad (25)$$

非優先音のゲインを α_2 だけ減少させた増加させた場合の聴感補正パワーを $L_{2m}[i, k]$ と定義しておく。

40

【0074】

【数 23】

$$L_{2m}[i, k] = (\alpha_2[i-1, k] - \Delta_2)^2 P_2[i, k] \quad (26)$$

調整後のゲイン $\alpha_1[i, k]$ を用いた場合の優先音に関する聴感補正パワーを $L_{1a}[i, k]$ と定義しておく。

【0075】

50

【数 2 4】

$$L_{1a}[i, k] = \alpha_1[i, k]^2 P_1[i, k] \quad (27)$$

(B) 操作する帯域の制限

次に、ゲイン調整する帯域の制限について説明する。0 Hz に相当する周波数ビンの信号ゲインを操作すると、音の自然感が損なわれる場合がある。また、高い周波数の信号ゲインを操作すると、聴き取り易さ向上のメリットよりも耳障りな音の付加というデメリットが大きくなる場合がある。

【0076】

そこで、優先音に対しては、 f_{1L} f_{1H} の範囲にある周波数 f でのみ $1[i, k]$ を更新する。この範囲は、周波数ビン k の範囲で、 k_{1L} k_{1H} の範囲に相当する。ただし、

$$k_{1L} = \text{rd} (N_F f_{1L} / F_s)$$

$$k_{1H} = \text{rd} (N_F f_{1H} / F_s)$$

である。ここで、「 $\text{rd}()$ 」は最も近い整数への丸め関数（四捨五入関数）を意味する。

【0077】

非優先音に対しても、同様に、 f_{2L} f_{2H} の範囲に限定してゲイン調整を行い、 k_{2L} k_{2H} を満たす $2[i, k]$ だけを増減させる。

(C) 1 を増加するための条件

1 の増加、すなわち $1[i, k] = (1 + 1) \times 1[i - 1, k]$ の演算を行うのは、式 (28) ~ (32) の条件がすべて満たされるときである。

【0078】

【数 2 5】

$$P_1[i, k] \geq 1 \quad (28)$$

$$P_2[i, k] \geq 1 \quad (29)$$

$$b_d L_p[i, k] \leq b_n P_1 P_2 \quad (30)$$

$$b_d (\alpha_1[i - 1, k] (1 + \Delta_1))^2 \leq b_n T_{1H}^2 \quad (31)$$

$$b_d L_p[i, k] < T_G^2 (b_n P_1[i, k] + b_d P_2[i, k]) \quad (32)$$

式 (28) と式 (29) は、優先音と非優先音の双方が可聴であるときにのみ増加を行うことを規定している。式 (30) は、混合音の対数強度（パワー）が優先音と非優先音の対数強度の和を上回らないように働く（対数強度の和の原理）。式 (31) は、優先音に対するゲインを一定値（ T_{1H} ）以下に抑えるように働く。式 (32) は、単純加算の場合の混合と比較して、時間周波数平面の局所であってもパワーの上昇を一定限界（振幅比で T_G 倍）以下に抑えるように働く（処理 (3) の条件 (a)）。

【0079】

式 (30) ~ (32) に対しては、低 SNR 判定時には補正をかけるのが望ましい。こ

の補正は、 P_1 を $(b_n / b_d) P_1$ に置き換えることによって優先音のレベルを上昇させたとみなすことによつて行われる。

(D) P_1 を減少するための条件

P_1 の減少、すなわち $P_1[i, k] = (1 + \alpha_1)^{-1} \times P_1[i-1, k]$ の演算を行うのは、式(33)~(37)のいずれかが成り立ち、かつ式(38)が成り立つときである。

【0080】

【数26】

$$P_1[i, k] < 1 \quad (33)$$

$$P_2[i, k] < 1 \quad (34)$$

$$b_d L[i, k] > b_n P_1 P_2 \quad (35)$$

$$b_d (\alpha_1 [i-1, k])^2 > b_n T_{1H}^2 \quad (36)$$

$$b_d L[i, k] > T_G^2 (b_n P_1[i, k] + b_d P_2[i, k]) \quad (37)$$

$$\alpha_1 [i-1, k] > 1 \quad (38)$$

10

20

30

40

50

式(33)と式(34)は、時間周波数平面上の点 (i, k) において、優先音と非優先音の少なくとも一方が可聴レベルを満たさない場合は、優先音のゲインを戻すことを意図する。式(35)は、混合音の対数強度が優先音と非優先音の対数強度の和を上回っている場合に、優先音のゲインを戻すように働く。式(36)は、優先音に対するゲイン α_1 があらかじめ設定された上限 T_{1H} を超えていたとき、その超過を解消する方向に働く(処理(3)の条件(b))。式(37)は、単純加算による混合音に所定の倍率(比率) T_G を乗算したレベル(図6参照)を超える場合に優先音のゲインを戻す方向に働く。式(38)は、優先音のゲイン値が1よりも大きいときにのみ減少させることを示す。

【0081】

式(33)~(36)は、式(28)~(31)の否定である。一方、式(37)は式(32)の否定になっていない。式(37)は、修正前に対する条件式であり、式(32)は修正後に対する条件式であるという差異がある。この差異により、ゲインが振動することを抑制している。

【0082】

このような減少操作によつて、 P_1 は増加の必要がないときには1に戻っていく。減少操作によつて $P_1[i, k] < 1$ となつてしまった場合は、 P_1 を強制代入することで、 $P_1[i, k] = 1$ を回復させる。この回復操作がある場合は、式(38)の条件は必ずしも必要ではないが、ソフトウェア実装の場合は、無駄な乗算時間の増大を防止するため、FPGA実装の場合は消費電力抑制のために、式(38)の判定があつたほうがよい。

【0083】

P_1 の増加と減少の条件がどちらも満たされない場合は、値の保持、すなわち $P_1[i, k] = P_1[i-1, k]$ を行う。

(E) P_2 を減少するための条件

P_2 の減少、すなわち $P_2[i, k] = P_2[i-1, k] - \alpha_2$ の演算を行うのは、式(39)と式(40)の双方が満たされる場合である。

【0084】

【数 27】

$$L_{1a}[i, k] - P_1[i, k] > P_2[i, k] - L_{2m}[i, k] \quad (39)$$

$$\alpha_2[i - 1, k] - \Delta_2 \geq T_{2L} \quad (40)$$

式(39)は、優先音のパワー増加分を超えない量であれば、非優先音のパワーを減少させてもよいことを示す。式(40)は、非優先音に対するゲインを一定値(T_{2L})以上に保つように働く。

(F) 2を増加するための条件

2の増加、すなわち $2[i, k] = 2[i - 1, k] + 2$ の演算を行うのは、式(41)と式(42)の双方が満たされる場合である。

【0085】

【数 28】

$$L_{1a}[i, k] - P_1[i, k] < P_2[i, k] - L_2[i, k] \quad (41)$$

$$\alpha_2[i - 1, k] < 1 \quad (42)$$

式(41)は、この時点までに決定されたゲイン $1[i, k]$ 、 $2[i - 1, k]$ を用いると、優先音のパワー増加分よりも非優先音のパワー減少のほうがおおきくなってしまふことを示している。式(41)は式(39)の否定に近いが、式(41)は修正前に対する条件式であるのに対し、式(39)は修正後に対する条件式であるという差異がある。この差異によって、ゲインが振動することを防止する。

【0086】

この操作により、2は減少させる必要がないときは1に戻っていく。2の増加により $2[i, k] > 1$ となった場合は、1を強制代入することで、 $2[i, k] = 1$ を回復する。

【0087】

2の増加と減少の条件がどちらも満たされない場合は、値の保持、すなわち $2[i, k] = 2[i - 1, k]$ を行う。

<動作例>

図10は、実施形態のミキシング装置1の動作例を示す図である。2つの音源セット(セット1、セット2)を用意し、各音源セットで音声を優先音とし、楽器音を非優先音とした。図10(A)はブーストが効いていない場合の例、図10(B)はブーストが効いている場合の例であり、ともに音源セット1を対象にしたものである。上述のように、優先音が有音であって、かつ低SNRのときにブースト処理が行われる。図10(A)と図10(B)はともに、修正前の各種変数をプロットしている。

【0088】

図中のMULは、 $b_n P_1 \cdot P_2 / b_d$ (b_n / b_d はブーストレシオ)であり、PLUSは、 $T_g^2 (b_n P_1 + b_d P_2) / b_d$ である。図中のLは、式(23)で定義したミキシング出力の聴感補正パワーLである。LをMULを超えない範囲でできるだけ大きくするというのが式(30)の条件であり、LをPLUSを超えない範囲でできるだけ大きくするというのが式(32)の条件である。

【 0 0 8 9 】

図 1 0 (A) と図 1 0 (B) の双方で、M U L と P L U S の大小関係は周波数に依存しており、常にどちらかが高いということはない。このことから、式 (3 0) と式 (3 2) の条件は両方とも効いており、併用すべきであることがわかる。

< 発展例 1 >

発展例 1 として、ゲインの平滑化による改良例を示す。上述した方法で、2つの音源セットのいずれに対しても良好な結果を得ることができたが、入力のス N R が低い部分で混合音がやや聴き取りにくくなることがわかった。

【 0 0 9 0 】

その原因を探ったところ、優先音のゲイン 1 の上昇が穏やかすぎて必要な値が確保できていないためであるとわかった。これに対処するためにはゲイン増加のステップサイズ 1 を大きくすればよいが、1 を大きくすると、ゲインの推移やゲインの差分の推移に大きな不連続が生じるおそれがある。この場合、スペクトルの散逸 (ノイズの発生) が起こってしまう。

【 0 0 9 1 】

そこで、発展例 1 では、以下のようにして 1、2 を平滑化し、平滑化されたゲイン 1、2 を用いる。これによってゲイン調整のステップサイズ 1、2 を 1 0 倍以上に大きくしても、スペクトル散逸の問題を回避することができる。

【 0 0 9 2 】

【 数 2 9 】

$$\beta_j[i, k] = \eta \beta_j[i-1, k] + (1 - \eta) \alpha_j[i, k] \quad (j=1, 2) \quad (43)$$

ここで、 η は指数平滑化法の係数であり、平滑の時定数 τ_α から式 (4 4) で導出する。

【 0 0 9 3 】

【 数 3 0 】

$$\eta = \exp\left(-\frac{N_d}{\tau_\alpha F_s}\right) \quad (44)$$

このようにして生成した 1、2 は、上述した音源セット 1、セット 2 のいずれに対しても良好なミキシング結果をもたらす。計算負荷や回路規模に支障がなければ、発展例 1 のゲインの平滑化を行うのが望ましい。

【 0 0 9 4 】

図 1 1 は、音源セット 2 を用い、平滑化されたゲイン 1、2 でミキシングしたときのミキシング装置 1 の動作状態を示す。図 1 1 で、横軸は時間、縦軸は周波数である。図 1 1 (A) は優先信号 X 1 としての音声、図 1 1 (B) は非優先信号 X 2 としての音楽、図 1 1 (C) は従来の単純加算によるミキシング結果 (X 1 + X 2)、図 1 1 (D) は実施形態のミキシング結果である。図 1 1 (E) は優先信号の平滑化後のパワー E 1、図 1 1 (F) は非優先信号の平滑化後のパワー E 2 である (図 3 参照)。図 1 1 (G) は聴感補正後のパワー P 1、図 1 1 (H) は聴感補正後のパワー P 2 (図 4 参照) を諧調表示したものである。

【 0 0 9 5 】

図 1 1 (G) と図 1 1 (H) において、淡い灰色の領域が 0 d B 以上 2 0 d B 未満、黒

10

20

30

40

50

色の領域が 20 dB 以上 40 dB 未満、濃い灰色の領域が 40 dB 以上の領域である。すなわち、聴感補正により可聴として取り扱われたのは、白色以外の領域である。図 11 (G) で、線分 e で示される領域が有音判定 ($e[i]$) された領域、線分 l で示される領域が低 SNR 判定 ($l[i]$) された領域、線分 b で示される領域がブースト判定 ($b[i]$) された領域である。用いた音源セット 2 は、優先音の SNR が低い音源セットであり、5 秒以降の時間区間では、有音区間はすべてブースト処理の対象となっている。

【0096】

図 11 (I) は、発展例 1 で平滑化されたゲイン G_1 に基づいて作成された優先音のゲインマスクであり、 G_1 の対数を濃淡表示した図である。白色が 0 dB、黒色が 35 dB に相当する。図 11 (J) は、発展例 1 で平滑化されたゲイン G_2 に基づいて作成された非優先音のゲインマスクであり、 G_2 の値を濃淡表示した図である。白色が 1.0、黒色が 0.0 に相当する。

10

【0097】

図 11 (I) 及び図 11 (J) のゲインマスクを用いてゲイン調整した後に加算することによって、時間周波数平面でのきめ細かなミキシングが可能になる。従来法による図 11 (C) では、低周波領域で非優先音 (ギター) の成分しか見えないのに対し、図 11 (D) では、優先音 (音声) の成分が混ざりあっている。

< 発展例 2 >

上述した実施形態では、演算量を低減するために、方程式を解くのではなく、不等式の実偽判定による逐次更新を行っている (処理 (5))。特に、FPGA の実装に際しては、極力処理を簡略化したい。そこで、70 phon と 20 phon のラウドネス曲線を信号処理部 15 にセットして、式 (10) ~ (12) により順次 $C_{Lp}[k]$ 、 $A[k]$ 、 $B[k]$ を導出する方法に替えて、最初から $B[k]$ を与える。たとえば出荷時に補正係数 $B[k]$ (最小可聴パワー $A[i]$ の逆数) を定数テーブルとして与えておく。動作中に一時的に合理性を無視しても特に強い優先感を与えなくなったりしたなどの場合、 $B[k]$ に強制的に任意の値を代入して自由に好みの特性を持たせることも可能である。

20

【0098】

図 12 は、図 9 (B) の等ラウドネス曲線を設定し、発展例 2 を適用したときの補正係数 $B[i]$ の具体例を示す。図 12 の場合、 $B[i]$ をテーブルとして記憶する替わりに関数としてあらかじめ記憶しておいてもよい。

30

< 処理フロー >

図 13 A 及び図 13 B は、ミキシング装置 1 の信号処理部 15 のゲイン決定部 151 で実行されるゲイン決定の処理の一例を示すフローである。この処理フローは、ゲインを平滑化してゲイン G を生成する発展例 1 に対応する。

【0099】

まず、 $G_1[k]$ 、 $G_2[k]$ 、 $A[k]$ 、 $B[k]$ をすべての周波数ビン k について「1」に初期化し、聴感補正総パワー $Q_1 = 0$ 、 $Q_2 = 0$ 、 $i = 0$ に設定して係数 $B[k]$ を読み込む (S11)。 $k = 0$ から処理を開始し (S12)、平滑化パワー $E_1[i, k]$ 、 $E_2[i, k]$ を読み込んで (S13)、聴感補正パワー $P_1[i, k]$ 、 $P_2[i, k]$ を求め (S14)、聴感補正総パワー $Q_1[i]$ と $Q_2[i]$ を計算する (S15)。 k の値をインクリメントして (S16)、 k が周波数ビン数 N_H に達するまで (S17 で NO)、S13 ~ S16 を繰り返す。これは、周波数ビン k についてのループの 1 回目のパスである。

40

【0100】

k が N_H を超えると (S17 で YES)、有音判定結果 $e[i]$ 、低 SNR 判定結果 $l[i]$ 、ブースト判定結果 $b[i]$ 、ブーストレシオの分子 $b_n[i]$ 、ブーストレシオの分母 $b_d[i]$ を求めて (S18)、 k についてのループの 2 回目のパスの処理を開始する (S19)。周波数ビン k について、優先音の k がゲイン調整する最低ビン k_{1L} と最高ビン k_{1H} の範囲内にあるか否かを判断する (S20)。範囲内にある場合に、平滑化パワー $E_1[i, k]$ 、 $E_2[i, k]$ を読み込み (S21)、 P_1 (優先音の聴感補

50

正パワー)、 P_2 (非優先音の聴感補正パワー)、 L_1 (更新前のゲイン G_1 での優先音の聴感補正パワー)、 L_{1p} (優先音のゲインを増加させたときの聴感補正パワー)、 L_2 (更新前のゲイン G_2 での非優先音の聴感補正パワー)、 L_{2m} (非優先音のゲインを G_2 減少させたときの聴感補正パワー)、 L (式 (23))、 L_p (式 (25)) を求める (S22)。

【0101】

求めた値を用いて、式 (28) ~ 式 (32) のすべてが成立するか否か、すなわち G_1 を増加するか否かを判断する (S23)。成立する場合に (S23でYES)、 G_1 を増加し (S24)、成立しない場合は (S23でNO)、 G_1 を維持する。

【0102】

次に、式 (33) ~ 式 (37) のいずれかが成立し、かつ式 (38) が成立するか否か、すなわち G_1 を減少するか否かを判断する (S25)。S25の条件が満たされない場合は G_1 を維持し、満たされる場合に G_1 を減少する (S26)。さらに、減少後の G_1 が1未満か否かを判断する (S27)。 G_1 が1未満になった場合は G_1 を1に戻し (S28)、 G_1 が1以上であれば更新された G_1 を維持する。

【0103】

続いて、非優先音の k がゲイン調整する最低ビン k_{2L} と最高ビン k_{2H} の範囲内にあるか否かを判断する (S29)。範囲内にある場合に、 L_{1a} (調整後のゲイン G_1 を用いた場合の優先音に関する聴感補正パワー) を求め (S30)、式 (39) と式 (40) が成立するか否か、すなわち G_2 を減少するか否かを判断する (S31)。成立する場合に (S31でYES)、 G_2 を減少し (S32)、成立しない場合は (S31でNO)、 G_2 を維持する。

【0104】

次に、式 (41) と式 (42) が成立するか否か、すなわち G_2 を増加するか否かを判断する (S33)。S33の条件が満たされない場合は G_2 を維持し、満たされる場合に G_2 を増加する (S34)。さらに、増加後の G_2 が1を超えるか否かを判断し (S35)、1以下であれば増加後の G_2 を維持し、1を超える場合は G_2 を1に戻す (S36)。

【0105】

次に、式 (43) に基づき、 G_2 と G_1 をそれぞれ平滑化して G_2 と G_1 を生成し (S37及びS38)、 G_1 と G_2 を出力する (S39)。その後、 k をインクリメントして (S40)、 k が N_H に達するまで (S41でNO)、S20 ~ S40を繰り返す。

【0106】

k が N_H を超えると (S41でYES)、時間ブロック i をインクリメントして (S42)、最後の時間ブロック i になるまで (S43でNO)、S12 ~ S42を繰り返し、最後の時間ブロック i の処理が終わったらプロセスを終了する。

【0107】

図13A及び図13Bの処理では、プーストレシオの分子 $b_n[i]$ と分母 $b_d[i]$ を決定するために (S18)、すべての周波数ビン k について $X_1[i, k]$ と $X_2[i, k]$ を参照しなければならない。一方、ゲイン $G_1[k]$ と $G_2[k]$ の更新には、 $X_1[i, k]$ と $X_2[i, k]$ だけでなく、 $b_n[i]$ と $b_d[i]$ が必要である。このため周波数ビン k についてのループ処理を2回行わないと、ゲインを1回更新することができない。

【0108】

そこで、図14では、プーストレシオだけは前回のサンプルで計算しておき、これを流用することによって、2パスの処理から1パスの処理へ軽減する。これにより、回路の簡素化や高速化が実現する。

【0109】

図14では、図13A及び図13Bとの差分のみを説明する。S11 ~ S15で、初期化、必要なパラメータの読み込みと算出を行った後、S20に飛んで、S21 ~ S28に

10

20

30

40

50

よる 1 の調整と、S 2 9 ~ S 3 6 による 2 の調整を行う。その後、調整後の 1、2 を平滑化して 1、2 を求め、着目している時間ブロック i について処理範囲のすべての周波数 k についてゲインを求める (S 1 3 ~ S 4 1 の繰り返し)。その後、時間ブロック i をインクリメントし (S 4 2)、 $e [i]$ 、 $l [i]$ 、 $b [i]$ 、 $b_n [i]$ 、 $b_d [i]$ を求め (S 5 1)、これらのパラメータを次の時間ブロック $i = i + 1$ の処理に用いる。

【 0 1 1 0 】

この簡略化について、優先音の S N R が低い音源セット 2 について調べたところ、出力値は完全に同一ではないが、その差分は聴き取ることのできない微小なレベルであった。

【 0 1 1 1 】

以上述べた方法により、入力信号 1 と入力信号 2 のそれぞれにつき、合理的な判定基準に基づいて最適なゲイン値を決定することができる。また、加減算と乗算のみの演算による逐次更新により、演算量を大幅に低減することができる。

【 0 1 1 2 】

上述した信号処理部 1 5 の処理は、ハードウェアによってもソフトウェアによっても実現することができる。信号処理部 1 5 の処理により、入力信号をそのまま用いてミキシング処理を行い、自然な聴感で優先音を目立たせ (処理 (1) の対数強度の和の原理)、非優先音を違和感なく抑制することができる (処理 (2) の穴埋めの原理)。優先音と非優先音のゲインを決める規範が合理的に決められているため (処理 (1) ~ 処理 (3))、ユーザが音を聴きながらパラメータの調整をする必要はない。

【 0 1 1 3 】

非優先音に対して優先音のパワーが極端に小さい場合でも、優先音を目立たせることができる (処理 (4))。たとえば、音楽にナレーションを重ねる場合に、ささやき声でも音楽に埋もれることなく聴き取らせることができる。

【 0 1 1 4 】

また、乗算と加減算だけでゲインを更新する手法により、F P G A 上へのハードウェア実装が容易になる (処理 (5))。また、D A W 上のプラグインとして実装し、リアルタイムに動作するミキシング装置が実現される。これは、逐次更新によって計算負荷を軽くしたことと、ゲインだけの調整で高性能のミキシングが可能になったため F F T の点数を 2 5 6 点程度まで小さくできることから生じる効果である。

【 0 1 1 5 】

なお、優先音と非優先音のパワーの平滑化は必須ではなく、時間周波数平面上に展開された入力信号のパワーから直接ゲイン 1、2 を求めてもよい。

【 0 1 1 6 】

上述したミキシング装置は、入力されたオーディオ信号を、合理的に決定されるゲインを用いて高速に自動合成できるため、レコーディングだけではなく、ニュース速報、カーナビゲーション、ディスクジョッキー、会議、カラオケ装置などに広く適用することができる。たとえば、番組を邪魔せずに緊急速報を報道する、カーステレオで音楽を流している場合でもカーナビゲーションの音声聴こえやすい、ディスクジョッキーが音楽の音量を下げずにトークできる、会議時に司会者の声を際立たせる、伴奏に対してボーカル音を自動調整する、などである。

【 0 1 1 7 】

また、パソコンやスマートフォン等のユーザ端末装置にミキシングプログラムをインストールすることで、ユーザが所望の音楽をミキシングし、あるいは所望の画像に所望の音楽を重ねて、通信相手に送信することができる。受信側では、受け取ったデータに対してさらに音声を重ねて保存、あるいは返信することができる。

【 符号の説明 】

【 0 1 1 8 】

1 ミキシング装置

1 1 信号入力部

10

20

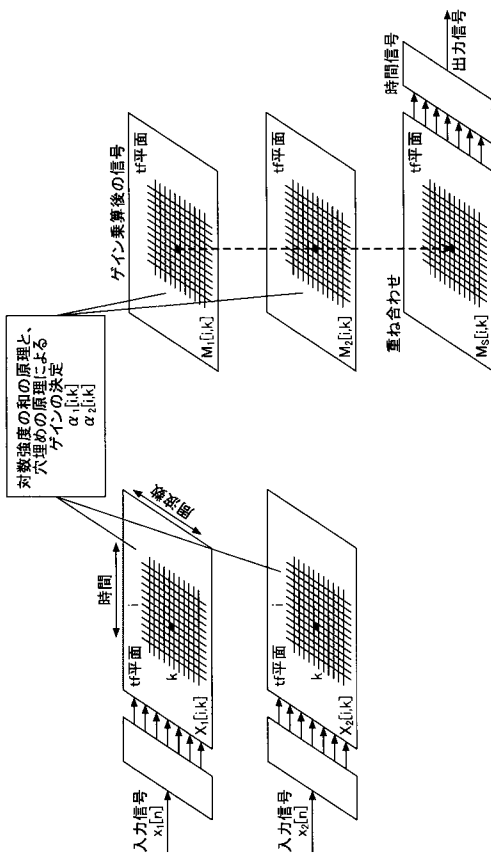
30

40

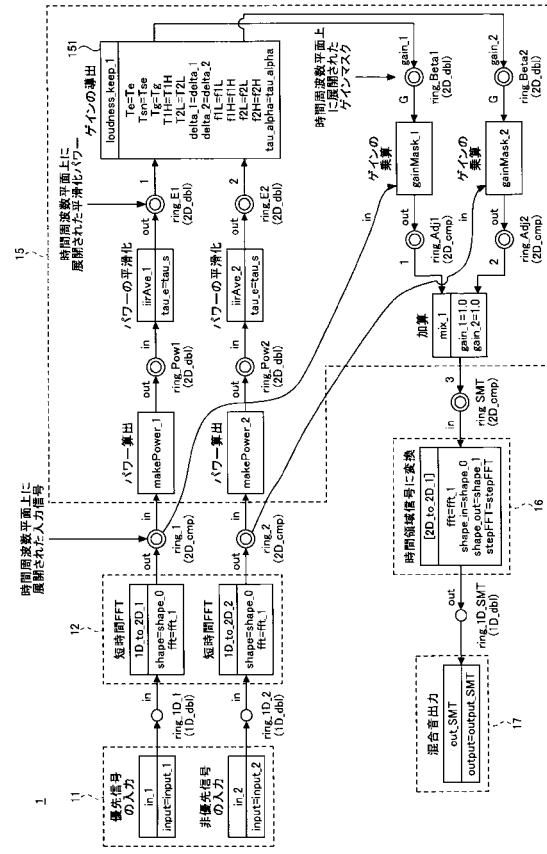
50

- 1 2 周波数解析部
- 1 5 信号処理部
- 1 6 周波数時間変換部
- 1 7 信号出力部
- 1 5 1 ゲイン決定部

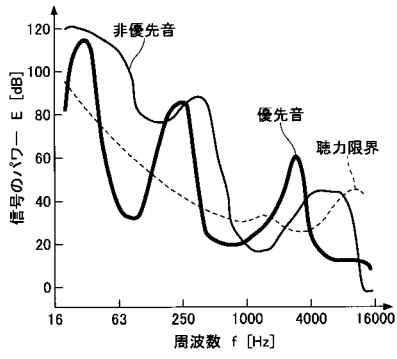
【 図 1 】



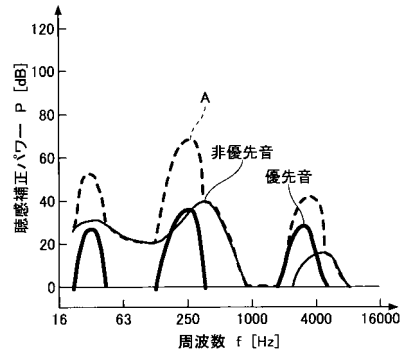
【 図 2 】



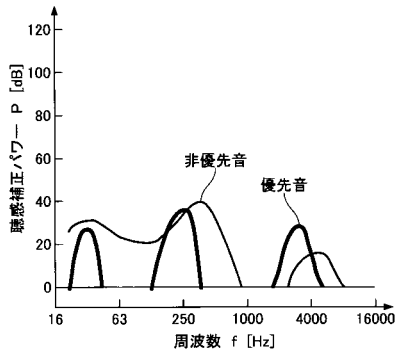
【 図 3 】



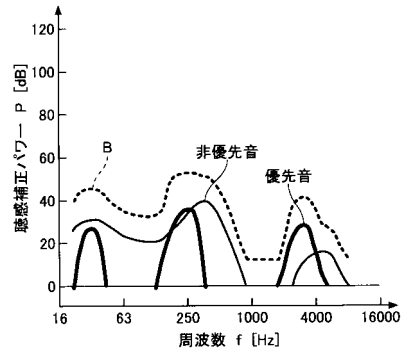
【 図 5 】



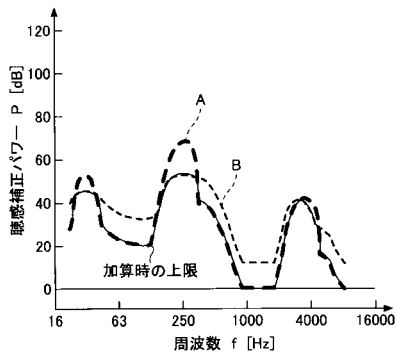
【 図 4 】



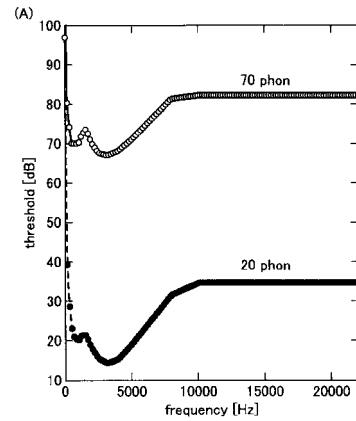
【 図 6 】



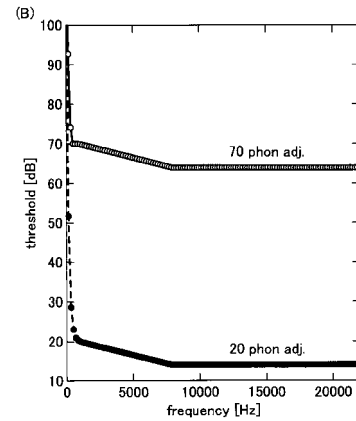
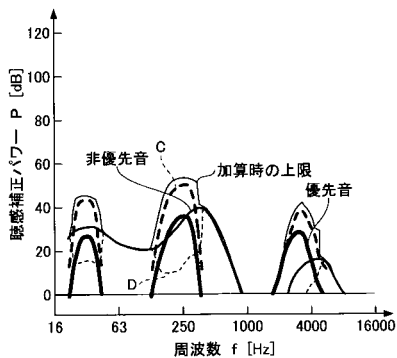
【 図 7 】



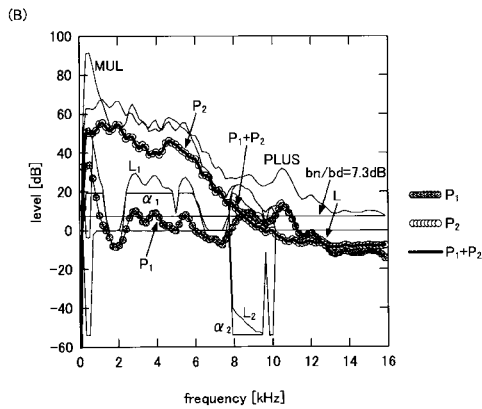
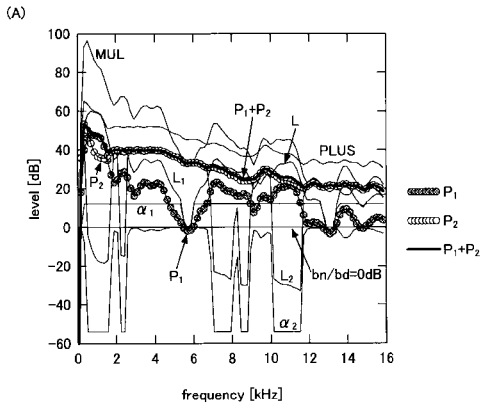
【 図 9 】



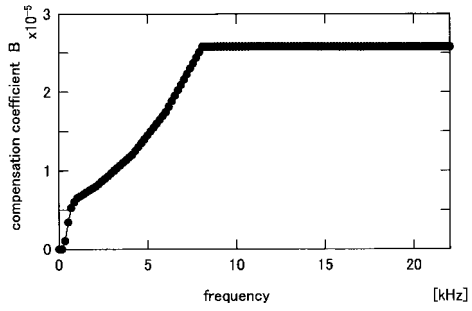
【 図 8 】



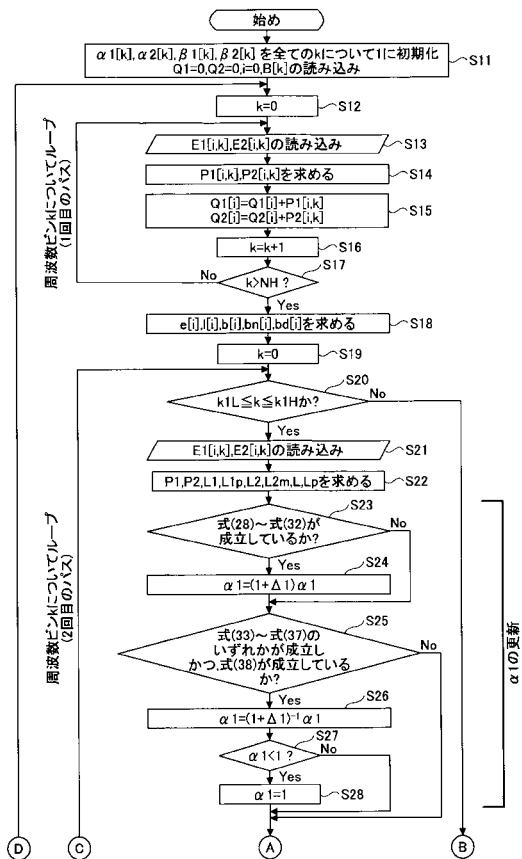
【図10】



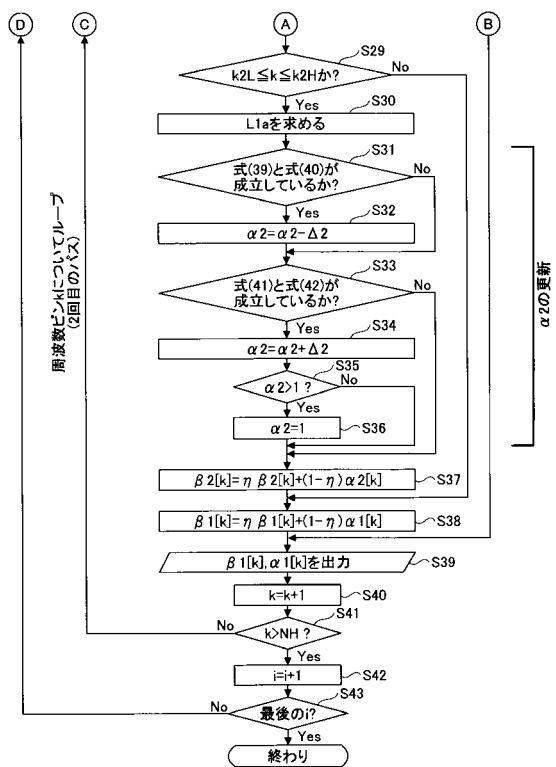
【図12】



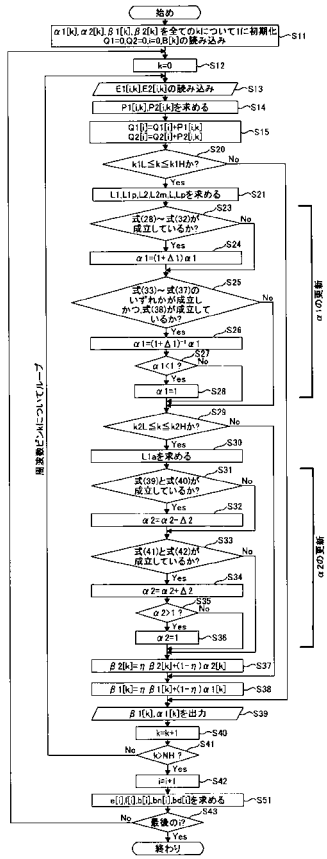
【図13A】



【図13B】



【図14】



【図11】

