

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4625084号  
(P4625084)

(45) 発行日 平成23年2月2日(2011.2.2)

(24) 登録日 平成22年11月12日(2010.11.12)

(51) Int.Cl.		F I	
G 1 0 L 19/00	(2006.01)	G 1 0 L 19/00	2 1 3
G 1 0 L 19/02	(2006.01)	G 1 0 L 19/02	1 5 0
H 0 3 M 7/30	(2006.01)	H 0 3 M 7/30	Z

請求項の数 31 (全 31 頁)

(21) 出願番号	特願2007-537134 (P2007-537134)	(73) 特許権者	591037214 フラウンホッフアーゲーゼルシャフト ツ ァ フェルダールング デア アンゲヴァ ンテン フォアシュンク エー. ファオ ドイツ連邦共和国 80686 ミュンヘ ン ハンザシュトラッセ 27ツェー
(86) (22) 出願日	平成17年9月12日 (2005.9.12)	(73) 特許権者	500587067 アギア システムズ インコーポレーテッ ド アメリカ合衆国, 18109 ペンシルヴ アニア, アレントアウン, アメリカン パー クウェイ エヌイー 1110
(65) 公表番号	特表2008-517334 (P2008-517334A)	(74) 代理人	100079577 弁理士 岡田 全啓
(43) 公表日	平成20年5月22日 (2008.5.22)		
(86) 国際出願番号	PCT/EP2005/009784		
(87) 国際公開番号	W02006/045373		
(87) 国際公開日	平成18年5月4日 (2006.5.4)		
審査請求日	平成19年6月20日 (2007.6.20)		
(31) 優先権主張番号	60/620,401		
(32) 優先日	平成16年10月20日 (2004.10.20)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	11/006,492		
(32) 優先日	平成16年12月7日 (2004.12.7)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 バイノーラルキュー符号化方法等のための拡散音の整形

(57) 【特許請求の範囲】

【請求項 1】

入力時間エンベロープを有する入力音声信号を、出力時間エンベロープを有する出力音声信号に変換する方法であって、

前記入力音声信号の入力時間エンベロープの特徴を調べるステップと、

処理は入力信号を非相関処理し、前記処理された音声信号を生成するために、音声信号を処理するステップと、

前記出力時間エンベロープは実質的に前記入力時間エンベロープに整合し、前記出力音声信号を生成するために、特徴を調べられた入力時間エンベロープに基づき処理された音声信号を調整するステップとを含む方法。

【請求項 2】

前記処理は、内部チャンネル間相関である、請求項 1 に記載の発明。

【請求項 3】

前記 I C C 合成は、バイノーラルキュー符号化 ( B C C ) 合成の部分である、請求項 2 に記載の発明。

【請求項 4】

前記 B C C 合成は、さらに、少なくとも、内部チャンネル間レベル差 ( I C L D ) 合成および内部チャンネル間時間差 ( I C T D ) を含む、請求項 3 に記載の発明。

【請求項 5】

前記 I C C 合成は、後期残響 I C C 合成を含む、請求項 2 に記載の発明。

## 【請求項 6】

前記調整するステップは、  
前記処理された音声信号の処理された時間エンベロープの特徴を調べるステップと、  
前記出力音声信号を生成するために、前記特徴を調べられた入力および処理された時間エンベロープに基づき、処理された信号を調整するステップを含む、請求項 1 に記載の発明。

## 【請求項 7】

前記調整するステップは、  
前記特徴を調べられた入力および処理された時間エンベロープに基づきスケーリング関数を生成するステップと、  
前記出力音声信号を生成するために、前記処理された音声信号にスケーリング関数を適用するステップとを含む、請求項 6 に記載の発明。

10

## 【請求項 8】

さらに、平坦化された音声信号を生成するために、前記特徴を調べられた入力時間エンベロープに基づき、前記入力音声信号を調整するステップを含み、前記調整するステップは、前記処理された音声信号を生成するために、前記平坦化された音声信号に適用される、請求項 1 に記載の発明。

## 【請求項 9】

非相関処理信号および相関処理信号を生成する処理ステップと、  
調整された処理信号を生成するために、前記非相関処理信号に適用される調整ステップと、  
前記出力信号は、前記調整された処理された音声および相関処理された信号を合計することによって生成する、請求項 1 に記載の発明。

20

## 【請求項 10】

前記特徴を調べるステップは、前記入力音声信号の明確な周波数のみに適用し、  
前記調整するステップは、前記処理された音声信号の明確な周波数のみに適用する、請求項 1 に記載の発明。

## 【請求項 11】

前記特徴を調べるステップは、明確なカットオフ周波数を超える前記入力音声信号の周波数のみに適用され、  
前記調整するステップは、前記明確なカットオフ周波数を超える前記処理された音声信号の周波数のみに適用される、請求項 10 に記載の発明。

30

## 【請求項 12】

前記特徴を調べるステップ、前記処理するステップ、前記調整するステップの各々が、周波数領域の信号に適用される、請求項 1 に記載の発明。

## 【請求項 13】

前記特徴を調べるステップ、前記処理するステップ、前記調整するステップの各々が、異なる信号サブバンドに個別に適用される、請求項 12 に記載の発明。

## 【請求項 14】

前記周波数領域が、高速フーリエ変換 (FFT) に対応している、請求項 12 に記載の発明。

40

## 【請求項 15】

前記周波数領域が、直交ミラーフィルタ (QMF) に対応している、請求項 12 に記載の発明。

## 【請求項 16】

前記特徴を調べるステップおよび前記調整するステップが、時間領域の信号に適用される、請求項 1 に記載の発明。

## 【請求項 17】

前記処理するステップは、周波数領域の信号に適用される、請求項 16 に記載の発明。

## 【請求項 18】

50

前記周波数領域は、FFTに対応している、請求項17に記載の発明。

【請求項19】

前記周波数領域は、QMFに対応している、請求項17に記載の発明。

【請求項20】

前記特徴を調べるステップおよび前記調整するステップを有効にするか、無効にするか判定するステップをさらに含む、請求項1に記載の発明。

【請求項21】

前記入力音声信号を生成した音声エンコーダによって有効/無効のフラグに基づき判定するステップである、請求項20に記載の発明。

【請求項22】

一時的現象の発生を検出した場合は、前記特徴を調べるステップおよび前記調整するステップを有効にするように前記入力音声信号において一時的現象を検出するために、前記入力信号を分析する分析するステップに基づき判定するステップである、請求項20に記載の発明。

【請求項23】

入力時間エンベロープを有する入力音声信号を、出力時間エンベロープを有する出力音声信号に変換する装置であって、

前記入力音声信号の入力時間エンベロープの特徴を調べる手段と、

処理は、前記入力音声信号の非相関に適應されるものであって、処理された音声信号を生成するために、前記入力信号を処理する手段と、

前記出力時間エンベロープは、実質的に前記入力時間エンベロープに整合し、前記出力音声信号を生成するために、特徴を調べられた入力時間エンベロープに基づき処理された音声信号を調整する手段とを含む装置。

【請求項24】

エンベロープ抽出器を含む前記特徴を調べる手段と、

前記入力音声信号を処理するための合成器を含む前記処理する手段と、

前記処理された音声信号を調整するために適應されるエンベロープ調整器を含む前記調整する手段とを含む、請求項23に記載の装置。

【請求項25】

前記装置は、デジタルビデオプレーヤ、デジタル音声プレーヤ、コンピュータ、衛星受信機、ケーブル受信機、地上波放送受信機、ホームエンターテインメントシステム、および映画館システムからなるグループから選択したシステムであって、

前記システムが、前記エンベロープ抽出器、前記合成器、および前記エンベロープ調整器を備える、請求項24に記載の発明。

【請求項26】

E個の送信音声チャンネルを生成するために、C個の音声チャンネルを符号化する方法であって、

2つ以上の前記C個の入力チャンネルに対する1つ以上のキューコードを生成するステップと、

$C > E$  1である、前記E個の送信チャンネルを生成するために、前記C個の入力チャンネルをダウンミックスするステップと、

前記E個の送信チャンネルを復号化する間に、前記E個の送信チャンネルのデコーダが、エンベロープ整形を行うかどうかを提示するフラグを生成するために、1つ以上のC個の入力チャンネルおよびE個の送信チャンネルを分析するステップと、

デコーダにおいて、一時的現象だけでなく、一時的現象の前後の信号を形成するために先取りの方法における一時的現象の検出を含み、一時的現象が検出されるか、または、検出するためにランダム性検出を含む場合は、前記フラグがセットされ、時間エンベロープが、擬似ランダムの方法において変動している場合は、前記フラグがセットされようとも、前記E個の送信チャンネルが音である場合は、前記フラグがセットされないために、音の検出を含まうとも、時間エンベロープは、擬似ランダムの方法において変動している、前

10

20

30

40

50

記分析するステップとを含む方法。

【請求項 27】

エンベロープ整形は、送信チャネルに対応する時間エンベロープを実質的に整合するためのデコーダによってデコードされたチャネルの時間エンベロープを調整する、請求項 26 に記載の発明。

【請求項 28】

E 個の送信音声チャネルを生成するために、C 個の入力音声チャネルを符号化するための装置であって、

2 つ以上の前記 C 個の入力チャネルに対する 1 つ以上のキューコードを生成する手段と

、  
C > E - 1 である、前記 E 個の送信チャネルを生成するために、前記 C 個の入力チャネルをダウンミックスする手段と、

前記 E 個の送信チャネルを復号化する間に、前記 E 個の送信チャネルのデコーダが、エンベロープ整形を行うかどうかを提示するフラグを生成するために、1 つ以上の C 個の入力チャネルおよび E 個の送信チャネルを分析する手段と、

デコーダにおいて、一時的現象だけでなく、一時的現象の前後の信号を形成するために先取りの方法における一時的現象の検出を含み、一時的現象が検出されるか、または、検出するためにランダム性検出を含む場合は、前記フラグがセットされ、時間エンベロープが、擬似ランダムの方法において変動している場合は、前記フラグがセットされようとも、前記 E 個の送信チャネルが音である場合は、前記フラグがセットされないために、音の検出を含まうとも、時間エンベロープは、擬似ランダムの方法において変動している、前記分析する手段とを含む装置。

【請求項 29】

コード推定器を含む前記生成する手段と、

ダウンミキサを含む前記ダウンミキシングする手段とを含む、請求項 28 に記載の装置

【請求項 30】

前記装置が、デジタルビデオプレーヤ、デジタル音声プレーヤ、コンピュータ、衛星受信機、ケーブル受信機、地上波放送受信機、ホームエンターテインメントシステム、および映画館システムからなるグループから選択したシステムであって、

前記システムが、前記コード推定器および前記ダウンミキサを備える、請求項 27 に記載の発明。

【請求項 31】

コンピュータ上で実行する場合は、請求項 1 に従って、入力音声信号を変換する方法、または請求項 26 に従って、C 個の入力音声チャネルを符号化する方法を実行するための機械読み取り可能な指示を有するコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願のクロスリファレンス

本出願は、2004年10月20日出願の米国特許仮出願第60/620,480号(代理人整理番号第Allamanche2-3-18-4)の優先権を主張するものである。この要旨は、ここに引例として組み込まれている。

【0002】

また、本出願の内容は、次の米国特許出願の内容に関連し、これらの要旨すべては、ここに引例として組み込まれている。

・米国出願第09/848,877号2001年5月4日出願(代理人整理番号第Fallier5)

・米国出願第10/045,458号2001年11月7日出願(代理人整理番号第Baumgarthe1-6-8)。これは、2001年8月10日出願の米国仮出願第6

10

20

30

40

50

0 / 3 1 1 , 5 6 5 号の優先権を主張するものである。

・米国出願第 1 0 / 1 5 5 , 4 3 7 号 2 0 0 2 年 5 月 2 4 日出願 ( 代理人整理番号第 B a u m g a r t e 2 - 1 0 )

・米国出願第 1 0 / 2 4 6 , 5 7 0 号 2 0 0 2 年 9 月 1 8 日出願 ( 代理人整理番号第 B a u m g a r t e 3 - 1 1 )

・米国出願第 1 0 / 8 1 5 , 5 9 1 号 2 0 0 4 年 4 月 1 日出願 ( 代理人整理番号第 B a u m g a r t e 7 - 1 2 )

・米国出願第 1 0 / 9 3 6 , 4 6 4 号 2 0 0 4 年 9 月 8 日出願 ( 代理人整理番号第 B a u m g a r t e 8 - 7 - 1 5 )

・米国出願第 1 0 / 7 6 2 , 1 0 0 号 2 0 0 4 年 1 月 2 0 日出願 ( F a l l e r 1 3 - 1 ) 10

・米国出願第 1 0 / x x x , x x x 号は、本出願 ( 代理人整理番号第 A l l a m a n c h e 2 - 3 - 1 8 - 4 ) と同日に出願したものである。

#### 【 0 0 0 3 】

本出願の内容は、次の論文の要旨にも関連している。これらの要旨すべては、ここに引例として組み込まれている。

・C. フォーラ ( F a l l e r ) 、 F . パウムガルテ ( B a u m g a r t e ) 著、「バイノーラルキュー符号化パート I : 心理音響学基礎および設計原理 ( B i n a u r a l C u e C o d i n g - P a r t I : P s y c h o a c o u s t i c f u n d a m e n t a l s a n d d e s i g n p r i n c i p l e s ) 」 ( I E E E 会報、スピーチおよび音声学会紀要第 1 1 巻第 6 号、2 0 0 3 年 1 1 月) 20

・C. フォーラおよび F . パウムガルテ著「バイノーラル用キュー符号化パート II : 方法および応用例 ( B i n a u r a l C u e C o d i n g - P a r t I I : S c h e m e s a n d a p p l i c a t i o n s ) 」 ( I E E E 会報、オーディオおよびスピーチ学会紀要、1 1 巻、第 6 号、2 0 0 3 年 1 1 月)

・C. フォーラ著、「異なる再生フォーマットと互換性のある空間音声符号化 ( C o d i n g o f s p a t i a l a u d i o c o m p a t i b l e w i t h d i f f e r e n t p l a y b a c k f o r m a t s ) 」 ( 音声工学学会第 1 1 7 回大会予稿、2 0 0 4 年 1 0 月) )

#### 【 0 0 0 4 】

30

本発明は、音声信号を符号化して、次に符号化した音声データから聴覚情景の合成を行うことに関する。

#### 【 背景技術 】

#### 【 0 0 0 5 】

人が特定の音源により生成した音声信号 ( すなわち、音 ) を聞く場合、この音声信号は通常、その人の左右の耳に 2 つの異なる時間で 2 つの異なる音声 ( 例えば、デシベル ) レベルで到達する。それらの異なる時間およびレベルは、音声信号が左右の耳それぞれに到達する経路の差の関数である。人の脳は、聞こえた音声信号が、その人を基準として、特定の位置 ( 例えば、方向および距離 ) にある音源から生成されているという知覚をその人に与えるために、時間およびレベルのこれらの差を解釈する。聴覚情景は、その人を基準として、1 つ以上の異なる位置にある 1 つ以上の異なる音源が生成した音声信号を同時に聞いている人の正味の影響である。 40

#### 【 0 0 0 6 】

脳によるこの処理は、聴覚情景を合成するために用いられる。1 つ以上の異なる音源からの音声信号は、聴取者を基準として、異なる音源が異なる位置にあるという知覚を与える左右の音声信号を生成するために、意図的に変更される。

#### 【 0 0 0 7 】

図 1 は、従来のバイノーラル信号合成装置 1 0 0 の上位ブロック図を示す。この装置は、1 つの音源信号 ( 例えば、モノラル信号 ) をバイノーラル信号の左右の音声信号に変換し、バイノーラル信号は、聴取者の鼓膜で受け取る 2 つの信号であると定義される。音源 50

信号の他に、合成装置100は、聴取者を基準として、所望の位置の音源に対応する空間キューのセットを受信する。典型的な実施例では、空間キューのセットは、チャンネル間レベル差(ICLD)値(左右の音声信号の間の音声レベルの差を、それぞれ左右の耳で聞こえるように特定する値)と、チャンネル間時間差(ICTD)値(左右の音声信号の間の到達時間差を、それぞれ左右の耳で聞こえるように特定する値)とを含んでいる。このほかに、または別のものとして、合成技術の中には、頭部伝達関数(HRTF)とも呼ぶ、信号源から鼓膜への音の方向依存の変換関数モデリングを必要とするものもある。例えば、J.ブラウエルト(Blauert)、「人のサウンドローカリゼーションの精神物理学(psychophysics of Human Sound Localization)」(MIT出版、1983年)を参照のこと、これは、この要旨は、ここに引例として組み込まれている。

10

## 【0008】

ヘッドホンで聞いた場合に、適切な空間キューのセット(例えば、ICLD、ICTD、および/またはHRTF)を適用して、それぞれの耳に対して音声信号を生成することにより、音源を空間的に配置したように、図1のバイノーラル信号合成装置100を用いて、1つの音源が生成したモノラル音声信号を処理することができる。例えば、D.R.ベゴールト(Begault)、「バーチャルリアリティーおよびマルチメディア用3-Dサウンド(3-D Sound for Virtual Reality and Multimedia)」(アカデミックプレス社マサチューセッツ州ケンブリッジ1994年)を参照のこと。

20

## 【0009】

図1のバイノーラル信号合成装置100は、聴取者を基準として配置された1つの音源を有する、最も単純な型の聴覚情景を生成する。基本的に、バイノーラル信号合成装置の複数のインスタンスを用いて実施する聴覚情景合成装置を用いることにより、聴取者を基準として、異なる位置に配置された2つ以上の音源を含むもっと複雑な聴覚情景を生成することができる。ここでは、各バイノーラル信号合成装置インスタンスが、異なる音源に対応するバイノーラル信号を生成する。各異なる音源それぞれが、聴取者を基準として異なる場所にあるので、異なる空間キューのセットが、各異なる音源それぞれに対しバイノーラル音声信号を生成するために、用いられる。

30

## 【0010】

【特許文献1】米国特許出願番号第09/848,877号

【特許文献2】米国特許出願番号第10/045,458号

【特許文献3】米国特許出願番号第10/155,437号

【特許文献4】米国特許出願番号第10/246,570号

【特許文献5】米国特許出願番号第10/815,591号

【特許文献6】米国特許出願番号第10/936,464号

【特許文献7】米国特許出願番号第10/762,100号

【非特許文献1】C.フォーラ(Faller)、F.バウムガルテ(Baumgartel)、「バイノーラルキュー符号化パートI:心理音響学基礎および設計原理(Binaural Cue Coding - Part I: Psychoacoustic fundamentals and design principles)」、IEEE会報、スピーチおよび音声学会紀要第11巻第6号、2003年11月

40

【非特許文献2】C.フォーラおよびF.バウムガルテ、「バイノーラル用キュー符号化パートII:方法および応用例(Binaural Cue Coding - Part II: Schemes and applications)」、IEEE会報、オーディオおよびスピーチ学会紀要、11巻、第6号、2003年11月

【非特許文献3】C.フォーラ、「異なる再生フォーマットと互換性のある空間音声符号化(Coding of spatial audio compatible with different playback formats)」、音声工学学会第117回大会予稿、2004年10月

50

## 【発明の開示】

## 【課題を解決するための手段】

## 【0011】

一実施の形態によれば、本発明は、入力時間エンベロープを有する入力音声信号を、出力時間エンベロープを有する出力音声信号に変換するための方法および装置である。入力音声信号の入力時間エンベロープは、特徴を調べられる。入力音声信号は、処理された音声信号を生成するために、処理される。前記処理は、前記入力音声信号の非相関することである。処理された音声信号は、出力音声信号を生成するために特徴を調べられた入力時間エンベロープに基づき調整される。出力時間エンベロープは、実質的に入力時間エンベロープに整合する。

10

## 【0012】

別の実施の形態によれば、本発明は、E個の送信音声チャンネルを生成するために、C個の入力音声チャンネルを符号化する装置である。1つ以上のキューコードは2つ以上のC個の入力チャンネルに対して生成される。C > E = 1である、E個の送信チャンネルに生成するために、C個の入力チャンネルがダウンミキシングされる。E個の送信チャンネルを復号化する間に、E個の送信チャンネルのデコーダが、エンベロープ整形をすることがを提示するフラグを生成するために、1つ以上のC個の入力チャンネルおよびE個の送信チャンネルは、分析される。

## 【0013】

別の実施の形態によれば、本発明は、上述の段落の方法により、符号化音声ビットストリームである。

20

## 【0014】

別の実施の形態によれば、本発明は、E個の送信チャンネル、1つ以上のキューコード、およびフラグを含む符合化音声ビットストリームである。1つ以上のキューコードは、2つ以上のC個の入力チャンネルに対して1つ以上のキューコードを生成することにより生成される。C > E = 1である、E個の送信チャンネルは、C個の入力チャンネルをダウンミキシングすることにより生成される。E個の送信チャンネルを復号化する間に、フラグは、E個の送信チャンネルのデコーダが、エンベロープ整形をすることがを提示し、前記フラグが、1つ以上のC個のチャンネルおよびE個のチャンネルを分析することにより、生成される。

## 【発明を実施するための最良の形態】

30

## 【0015】

以下の詳細な説明、特許請求の範囲、添付の図面から、本発明の他の側面、特徴および利点について、より完全に明らかになるであろう。類似の、または全く同じ構成要素には、同じ参照番号が付されている。

図1は、従来のバイノーラル信号合成装置の上位ブロック図を示す。

図2は、一般バイノーラルキュー符号化(BCC)音声処理システムを示すブロック図である。

図3は、図2のダウンミキサに用いることができるダウンミキサのブロック図を示す。

図4は、図2のデコーダに用いることができるBCC合成装置のブロック図を示す。

図5は、本発明の一実施の形態による、図2のBCC推定器のブロック図を示す。

40

図6は、5チャンネル音声用ICTDおよびICLDデータ生成を説明する図である。

図7Aは、5チャンネル音声用ICCデータを説明する図である。

図7Bは、5チャンネル音声用ICCデータを説明する図である。

図8は、BCCデコーダに用いることができ、1つの送信した和信号s(n)プラス空間キューが与えられたステレオ音声信号またはマルチチャンネル音声信号を生成する、図4のBCC合成装置の一実施例のブロック図を示す。

図9は、周波数関数として、サブバンド内でICTDおよびICLDがどのように変化するかを説明する図である。

図10は、本発明の一実施の形態による、BCCデコーダの少なくとも部分を示すブロック図を示す。

50

図 1 1 A は、図 4 の B C C 合成装置の条件での図 1 0 のエンベロープ整形方法の典型的なアプリケーションを示す。

図 1 1 B は、図 4 の B C C 合成装置の条件での図 1 0 のエンベロープ整形方法の典型的なアプリケーションを示す。

図 1 1 C は、図 4 の B C C 合成装置の条件での図 1 0 のエンベロープ整形方法の典型的なアプリケーションを示す。

図 1 2 A は、エンベロープ整形は、時間領域で適用され、図 4 の B C C 合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

図 1 2 B は、エンベロープ整形は、時間領域で適用され、図 4 の B C C 合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

図 1 2 C は、エンベロープ整形は、時間領域で適用され、図 4 の B C C 合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

図 1 3 は、カットオフ周波数  $f_{tp}$  より高い周波数にだけエンベロープ整形が行われる、図 1 2 の T P A および T P の考えられる実施例である。

図 1 4 は、代理人整理番号第 B a u m g a r t e 7 - 2 として 2 0 0 4 年 4 月 1 日に出願の米国特許出願番号第 1 0 / 8 1 5 , 5 9 1 に記載の I C C 合成方法に基づく後期残響の条件での図 1 0 のエンベロープ整形方法の典型的なアプリケーションを示す。

図 1 5 は、図 1 0 に示される方法に代わりに、本発明の実施例により、B C C デコーダの少なくとも部分を示すブロック図を示す。

図 1 6 は、図 1 0 および図 1 5 に示される方法の代わりに、本発明の実施例により、B C C デコーダの少なくとも部分を示すブロック図を示す。

図 1 7 は、図 4 の B C C 合成装置の条件での図 1 5 のエンベロープ整形の典型的なアプリケーションを示す。

図 1 8 A は、図 1 5 および図 1 6 の T P A と、図 1 6 の I T P および T P との考えられる実施例のブロック図を示す。

図 1 8 B は、図 1 5 および図 1 6 の T P A と、図 1 6 の I T P および T P との考えられる実施例のブロック図を示す。

図 1 8 C は、図 1 5 および図 1 6 の T P A と、図 1 6 の I T P および T P との考えられる実施例のブロック図を示す。

#### 【 0 0 1 6 】

バイノーラルキュー符号化 ( B C C ) は、 $C > E = 1$  である、音声チャンネルを生成するために、エンコーダが  $C$  個の入力音声チャンネルを符号化される。特に、2 つ以上の  $C$  個の入力チャンネルが周波数領域に提供され、周波数領域の 2 つ以上の入力チャンネルで、1 つ以上の異なる周波数帯域それぞれに 1 つ以上のキューコードが生成される。また、 $C$  個の入力チャンネルは、 $E$  個の送信チャンネルを生成するために、ダウンミキシングされる。ダウンミキシング実施例の中には、 $E$  個の送信チャンネルのうちの少なくとも 1 つのチャンネルが 2 つ以上の  $C$  個の入力チャンネルに基づいていて、 $E$  個の送信チャンネルのうちの少なくとも 1 つのチャンネルが  $C$  個の入力チャンネルのうちの 1 つのチャンネルだけに基づいている場合もある。

#### 【 0 0 1 7 】

一実施の形態では、B C C コーダは、2 つ以上のフィルタバンク、コード推定器、およびダウンミキサを備えている。2 つ以上のフィルタバンクは、2 つ以上の  $C$  個の入力チャンネルを時間領域から周波数領域に変換する。コード推定器は、2 つ以上の変換した入力チャンネルで 1 つ以上の異なる周波数帯域それぞれに 1 つ以上のキューコードを生成する。ダウンミキサは、 $C > E = 1$  である、 $E$  個の送信チャンネルを生成するために、 $C$  個の入力チャンネルをダウンミキシングする。

#### 【 0 0 1 8 】

B C C 復号化では、 $C$  個の再生音声チャンネルを生成するために、 $E$  個の送信音声チャンネルが復号化される。特に、1 つ以上の異なる周波数帯域それぞれに対して、周波数領域で 1 つ以上の  $E$  個の送信チャンネルが、 $C > E = 1$  である、周波数領域で 2 つ以上の  $C$  個の再

10

20

30

40

50



生チャンネルを生成するために、アップミキシングされる。周波数領域の2つ以上の再生チャンネルで、1つ以上のキューコードが、2つ以上の変更チャンネルを生成するために、1つ以上の異なる周波数帯域それぞれに適用され、2つ以上の変更チャンネルが、周波数領域から時間領域に変換される。アップミキシングの実施例の中には、少なくとも1つのC個の再生チャンネルが、E個の送信チャンネルのうち少なくとも1つのチャンネルと、少なくとも1つのキューコードとに基づいていて、少なくとも1つのC個の再生チャンネルが、キューコードとは無関係に、E個の送信チャンネルのうち1つのチャンネルだけに基づいている場合もある。

#### 【0019】

一実施の形態では、BCCデコーダは、アップミキサ、合成装置、および1つ以上の逆フィルタバンクを備える。1つ以上の異なる周波数帯域それぞれに対して、アップミキサは、 $C > E = 1$ である、周波数領域で2つ以上のC個の再生チャンネルを生成するために、周波数領域で1つ以上のE個の送信チャンネルをアップミキシングする。合成装置は、2つ以上の変更チャンネルを生成するために、周波数領域の2つ以上の再生チャンネルで、1つ以上のキューコードを1つ以上の異なる周波数帯域それぞれに適用する。1つ以上の逆フィルタバンクは、2つ以上の変更チャンネルを周波数領域から時間領域に変換する。

#### 【0020】

特定の実施例によるが、任意の再生チャンネルは、2つ以上の送信チャンネルの組み合わせよりむしろ、1つの送信チャンネルに基づいている場合もある。例えば、送信チャンネルが1つだけある場合、C個の再生チャンネルのそれぞれは、その1つの送信チャンネルに基づいている。このような状況では、アップミキシングは、送信チャンネルをコピーすることに対応している。従って、送信チャンネルが1つだけある適用例では、各再生チャンネルに対して送信チャンネルをコピーするレプリケータを用いて、アップミキサを実施してもよい。

#### 【0021】

BCCエンコーダおよび/またはデコーダを、例えば、デジタルビデオレコーダ/プレーヤ、デジタル音声レコーダ/プレーヤ、コンピュータ、衛星送信機/受信機、ケーブル送信機/受信機、地上波放送送信機/受信機、ホームエンターテインメントシステム、および映画館システム等の、多数のシステムまたは適用例に組み込むこともできる。

#### 【0022】

##### 一般BCC処理

図2は、エンコーダ202とデコーダ204とを備える一般バイノーラルキュー符号化(BCC)音声処理システム200を示すブロック図である。エンコーダ202は、ダウンミキサ206とBCC推定器208とを含んでいる。

#### 【0023】

ダウンミキサ206は、 $C > E = 1$ である、C個の入力音声チャンネル $x_i(n)$ をE個の送信音声チャンネル $y_i(n)$ に変換する。この明細書では、変数 $n$ を用いて表す信号は時間領域信号であって、変数 $k$ を用いて表す信号は周波数領域信号である。特定の実施例によるが、時間領域または周波数領域のいずれかでダウンミキシングを行うことが可能である。BCC推定器208は、BCCコードをC個の入力音声チャンネルから生成して、E個の送信音声チャンネルを基準として、帯域内または帯域外サイド情報のいずれかとしてそれらのBCCコードを送信する。典型的なBCCコードは、1つ以上のチャンネル間時間差(ICTD)と、チャンネル間レベル差(ICLD)と、周波数および時間の関数として、特定の対の入力チャンネル間で推定した、チャンネル間相関(ICC)データとを含んでいる。特定の実施例では、どの特定の対の入力チャンネルでBCCコードを推定するか要求している。

#### 【0024】

ICCデータは、音源の知覚した幅に関するバイノーラル信号のコヒーレンスに対応している。音源がより広くなると、得られるバイノーラル信号の左右のチャンネル間のコヒーレンスがより低下する。例えば、ホールのステージいっばいに広がるオーケストラに対応するバイノーラル信号のコヒーレンスは、通常、独奏している1つのバイオリンに対応す

10

20

30

40

50

るバイノーラル信号のコヒーレンスより低い。一般に、コヒーレンスが低い音声信号は通常、聴覚空間ではより広がって知覚される。従って、ＩＣＣデータは通常、明白な音源の幅と、聴取者が包み込まれた状態の度合いに関係する。すなわち、J. ブラウエルト、「人のサウンドローカリゼーションの精神物理学」(MIT出版、1983年)を参照のこと。

#### 【0025】

特定の適用例によるが、E個の送信音声チャンネルおよび対応するＢＣＣコードをデコーダ204に直接送信することもできるし、ある適したタイプの記憶装置に記憶して、デコーダ204が続いてアクセスすることもできる。場合によるが、“送信”という言葉は、デコーダに直接送信すること、または、デコーダが続いて利用することができるように記憶することのいずれかを言う場合もある。いずれの場合でも、デコーダ204は送信音声チャンネルとサイド情報とを受信して、ＢＣＣコードを用いてアップミキシングとＢＣＣ合成とを行って、E個の送信音声チャンネルをE個より多い(通常C個であるが、必ずしもC個でなくてもよい)再生音声チャンネル $\hat{x}_i(n)$ に変換して、音声再生を行う。特定の実施例によるが、時間領域または周波数領域のいずれかでアップミキシングを行うことができる。

10

#### 【0026】

図2に示すＢＣＣ処理の他に、一般ＢＣＣ音声処理システムは、さらに符号化段、復号化段をさらに備え、エンコーダで音声信号を圧縮して、デコーダで音声信号を復元することもできる。これらの音声コーデックは、パルスコード変調(PCM)、差動PCM(DPCM)、または適応DPCM(ADPCM)に基づいた、従来の音声圧縮/復元技術に基づくものであってもよい。

20

#### 【0027】

ダウンミキサ206が1つの和信号(すなわち、 $E=1$ )を生成する場合は、ＢＣＣ符号化により、モノラル音声信号を表すのに必要なものよりも若干高いビットレートでマルチチャンネル音声信号を表すことが可能である。これは、チャンネル対の間の推定したICTD、ICLD、およびＩＣＣデータが、音声波形よりも約2桁小さい情報を含んでいるからである。

#### 【0028】

ＢＣＣ符号化は、ビットレートが低いことばかりでなく、その後方互換性の点でも関心を集めている。1つの送信した和信号は、元のステレオまたはマルチチャンネル信号のモノラルダウンミキシングに対応している。ステレオまたはマルチチャンネル音響再生をサポートしていない受信機にとって、送信した和信号を聴取することは、薄型モノラル再生装置で音声素材を表す有効な方法である。従って、ＢＣＣ符号化を用いて、モノラル音声素材をマルチチャンネル音声にすることが必要な既存のサービスを向上させることもできる。例えば、ＢＣＣサイド情報が、既存の伝送チャンネルに埋め込むことができれば、ステレオまたはマルチチャンネル再生が行えるように、既存のモノラル音声無線放送システムを向上させることができる。マルチチャンネル音声を、ステレオ音声に対応する2つの和信号にダウンミキシングする際に、類似の能力が存在する。

30

40

#### 【0029】

ＢＣＣにより、ある時間および周波数分解能で音声信号を処理する。用いられる周波数分解能は、人間の聴覚システムの周波数分解能が主な動機となっている。心理音響学では、空間知覚は、音響入力信号の重要な帯域表現に基づいている可能性が一番高いことを示唆している。人間の聴覚システムの重要な帯域幅と同じか、または比例する帯域幅のサブバンドを有する可逆フィルタバンク(例えば、高速フーリエ変換(FFT)または直交ミラーフィルタ(QMF)に基づくもの)を用いて、この周波数分解能について考える。

#### 【0030】

一般ダウンミキシング

好適な実施例では、送信した和信号は、入力音声信号の全信号成分を含んでいる。目的

50

は、各信号成分を完全に維持することである。音声入力チャンネルを単純に加算することは、信号成分が増幅したり、減衰したりすることがよくある。言い換えれば、“単純に”加算した信号成分のパワーが、各チャンネルの対応する信号成分のパワーの合計よりも大きかったり、小さかったりすることがよくある。和信号の信号成分のパワーが全入力チャンネルの対応するパワーとほぼ同じになるように、和信号を等しくするダウンミキシング技術を用いることができる。

【 0 0 3 1 】

図 3 は、B C C システム 2 0 0 のある実施例による、図 2 のダウンミキサ 2 0 6 に用いることができるダウンミキサ 3 0 0 を示すブロック図である。ダウンミキサ 3 0 0 は、各入力チャンネル  $x_i(n)$  のフィルタバンク (F B) 3 0 2 と、ダウンミキシングブロック 3 0 4 と、オプションのスケーリング/遅延ブロック 3 0 6 と、各符号化したチャンネル  $y_i(n)$  の逆 F B (I F B) 3 0 8 とを備える。

【 0 0 3 2 】

各フィルタバンク 3 0 2 は、時間領域の対応するデジタル入力チャンネル  $x_i(n)$  の各フレーム (例えば、2 0 ミリ秒) を周波数領域の入力係数のセット  $\tilde{x}_i(k)$  に変換する。ダウンミキシングブロック 3 0 4 は、各サブバンドの C 個の対応する入力係数を、対応するサブバンドの E 個のダウンミキシング周波数領域係数にダウンミキシングする。次の式 (1) は、k 番目のサブバンドの入力係数  $(\tilde{x}_1(k), \tilde{x}_2(k), \dots, \tilde{x}_C(k))$  をダウンミキシングして、k 番目のサブバンドのダウンミキシング係数  $(\tilde{y}_1(k), \tilde{y}_2(k), \dots, \tilde{y}_E(k))$  を生成することを表す。D<sub>CE</sub> は、実数値の C × E ダウンミキシング行列である。

$$\begin{bmatrix} \hat{y}_1(k) \\ \hat{y}_2(k) \\ \vdots \\ \hat{y}_E(k) \end{bmatrix} = \mathbf{D}_{CE} \begin{bmatrix} \tilde{x}_1(k) \\ \tilde{x}_2(k) \\ \vdots \\ \tilde{x}_C(k) \end{bmatrix}, \quad (1)$$

【 0 0 3 3 】

オプションのスケーリング/遅延ブロック 3 0 6 は、乗算器 3 1 0 のセットを含んでいる。それぞれ、対応するスケーリング係数  $\tilde{y}_i(k)$  を生成するために、対応するダウンミキシング係数  $\hat{y}_i(k)$  に倍率  $e_i(k)$  を乗算する。スケーリング演算の目的は、各チャンネルの任意の重み係数でダウンミキシングするように一般化した、等化に相当する。入力チャンネルが独立していれば、次の式 (2) により、各サブバンドのダウンミキシング信号のパワー  $P_{\tilde{y}_i(k)}$  が得られる。

$$\begin{bmatrix} P_{\tilde{y}_1(k)} \\ P_{\tilde{y}_2(k)} \\ \vdots \\ P_{\tilde{y}_E(k)} \end{bmatrix} = \overline{\mathbf{D}}_{CE} \begin{bmatrix} P_{\tilde{x}_1(k)} \\ P_{\tilde{x}_2(k)} \\ \vdots \\ P_{\tilde{x}_C(k)} \end{bmatrix}, \quad (2)$$

ここで、 $\overline{\mathbf{D}}_{CE}$  は、C × E ダウンミキシング行列 D<sub>CE</sub> の各行列要素を二乗することにより導出したもので、 $P_{\tilde{x}_i(k)}$  は、入力チャンネル i のサブバンド k のパワーである。

【 0 0 3 4 】

サブバンドが独立していなければ、ダウンミキシング信号のパワー値  $P_{\tilde{y}_i(k)}$  は、信号成分が同相または位相のずれている場合は、信号増幅または相殺により、式(2)を用いて算出したものよりも、大きいまたは小さい。これを避けるために、式(1)のダウンミキシング演算をサブバンドに適用して、次に、乗算器310のスケーリング演算を行う。以下の式(3)を用いて、倍率  $e_i(k)$  ( $1 \cdot i \cdot E$ ) を導出することができる。

$$e_i(k) = \sqrt{\frac{P_{\tilde{y}_i(k)}}{P_{\hat{y}_i(k)}}}, \quad (3)$$

10

ここで、 $P_{\tilde{y}_i(k)}$  は、式(2)により算出したサブバンドパワーで、 $P_{\hat{y}_i(k)}$  は、対応するダウンミキシングサブバンド信号  $\hat{y}_i(k)$  のパワーである。

【0035】

オプションのスケーリングを行う他に、またはこの代わりに、スケーリング/遅延ブロック306は、オプションで信号を遅延してもよい。

【0036】

各逆フィルタバンク308は、周波数領域の対応するスケーリング係数のセットを、対応するデジタル送信チャネル  $y_i(n)$  のフレームに変換する。

20

【0037】

図3は、C個の入力チャネルすべてを周波数領域に変換して、続いてダウンミキシングを行っているが、別の実施例として、1つ以上(しかし、C-1よりも小さい数)のC個の入力チャネルについて、図3に示す処理の一部またはすべてを省略して、同等の数の変更していない音声チャネルとして送信してもよい。特定の実施例によるが、送信BCCコードの生成に、図2のBCC推定器208がこれらの変更していない音声チャネルを用いても、用いなくてもよい。

【0038】

1つの和信号  $y(n)$  を生成するダウンミキサ300の実施例において、次の式(4)により、 $E=1$ 、各入力チャネル  $c$  の各サブバンドの信号  $\tilde{x}_c(k)$  を加算して、係数  $e(k)$  を乗算する。

30

$$\tilde{y}(k) = e(k) \sum_{c=1}^C \tilde{x}_c(k). \quad (4)$$

【0039】

係数  $e(k)$  は、次の式(5)から得られる。

$$e(k) = \sqrt{\frac{\sum_{c=1}^C p_{\tilde{x}_c}(k)}{p_{\tilde{y}}(k)}}, \quad (5)$$

40

ここで、 $P_{\tilde{x}_c}(k)$  は、時間インデックス  $k$  での  $\tilde{x}_c(k)$  のパワーの短時間推定値で、 $P_{\tilde{y}}(k)$  は、 $\sum_{c=1}^C \tilde{x}_c(k)$  のパワーの短時間推定値である。等化サブバンドを、時間領域に逆に変換して、和信号  $y(n)$  を生成して、BCCデコーダに送信する。

【0040】

### 一般BCC合成

図4は、BCCシステム200のある実施例による、図2のデコーダ204に用いられるBCC合成装置400を示すブロック図である。BCC合成装置400は、各送信チャネル $y_i(n)$ 用のフィルタバンク402と、アップミキシングブロック404と、遅延406と、乗算器408と、相関ブロック410と、各再生チャネル $\hat{x}_i(n)$ 用の逆フィル

【0041】

各フィルタバンク402は、時間領域の対応するデジタル送信チャネル $y_i(n)$ の各フレームを、周波数領域の入力係数のセット $\tilde{y}_i(k)$ に変換する。アップミキシングブロック404は、各サブバンドのE個の対応する送信チャネル係数を、対応するサブバンドのC個のアップミキシング周波数領域係数にアップミキシングする。次の式(4)は、k番目のサブバンドのアップミキシング係数 $(\tilde{s}_1(k), \tilde{s}_2(k), \dots, \tilde{s}_C(k))$ を生成するために、サブバンドの送信チャネル係数 $(\tilde{y}_1(k), \tilde{y}_2(k), \dots, \tilde{y}_E(k))$ のアップミキシングを行うことを表す。

$$\begin{bmatrix} \tilde{s}_1(k) \\ \tilde{s}_2(k) \\ \vdots \\ \tilde{s}_C(k) \end{bmatrix} = \mathbf{U}_{EC} \begin{bmatrix} \tilde{y}_1(k) \\ \tilde{y}_2(k) \\ \vdots \\ \tilde{y}_E(k) \end{bmatrix}, \quad (6)$$

ここで、 $\mathbf{U}_{EC}$ は、実数値の $E \times C$ アップミキシング行列である。周波数領域でアップミキシングを行うことにより、異なるサブバンドそれぞれに個別にアップミキシングを行えるようになる。

### 【0042】

各遅延406は、ICTDデータの対応するBCCコードに基づいて、遅延値 $d_i(k)$ を適用して、再生チャネルの特定の対の間で確実に所望のICTD値が現れるようにする。各乗算器408は、ICLDデータの対応するBCCコードに基づいて、倍率 $a_i(k)$ を適用して、特定の対の再生チャネルの間で確実に所望のICLD値が現れるようにする。相関ブロック410は、ICCデータの対応するBCCコードに基づいて、非相関演算Aを行って、特定の対の再生チャネルの間で確実に所望のICC値が現れるようにする。さらに相関ブロック410の演算の記載は、米国出願第10/155,437号2002年5月24日出願(代理人整理番号第Baumgartel-10)に記載されている。

### 【0043】

ICLD合成を行うには、サブバンド信号のスケーリングを行うだけでよいので、ICLD値の合成は、ICTD値およびICC値の合成よりも煩わしくない。ICLDキューは最も一般的に用いられる方向キューなので、ICLD値により元の音声信号のICLDキューを近似することは、通常、さらに重要なことである。従って、ICLDデータを、全チャネル対の間で推定する場合もある。好ましくは、各再生チャネルのサブバンドパワーが、元の入力音声チャネルの対応するパワーを近似するように、各サブバンドの倍率 $a_i(k)$ ( $1 \cdot i \cdot C$ )が選択される。

### 【0044】

目的の1つは、ICTD値およびICC値を合成するために、比較的少ない回数の信号変更を適用することである。従って、BCCデータは、全チャネル対のICTD値およびICC値を含んでいなくてもよい。その場合は、BCC合成装置400は、あるチャネル対の間だけでICTD値およびICC値を合成する。

### 【0045】

10

20

30

40

各逆フィルタバンク 4 1 2 は、周波数領域の対応する合成した係数のセット  $\tilde{\hat{x}}_i(k)$  を、  
 対応するデジタル再生チャネル  $\hat{x}_i(k)$  のフレームに変換する。

【 0 0 4 6 】

図 4 は、続いてアップミキシングおよび B C C 処理を行うために、E 個の送信チャネルをすべて周波数領域に変換することを示しているが、別の実施例では、(すべてではないが) 1 つ以上の E 個の送信チャネルについて、図 4 に示す処理の一部またはすべてを回避してもよい。例えば、1 つ以上の送信チャネルが、アップミキシングを行っていない、変更していないチャネルであってもよい。1 つ以上の C 個の再生チャネルの他に、必ずしも  
 10 行う必要はないが、これらの変更していないチャネルを順に、基準チャネルとして用いて、B C C 処理を行って、1 つ以上の他の再生チャネルを合成してもよい。いずれの場合でも、このような変更していないチャネルを遅延して残りの再生チャネルの生成に用いられるアップミキシングおよび / または B C C 処理に必要な処理時間を補償することもできる。

【 0 0 4 7 】

図 4 は、C は元の入力チャネルの数である、C 個の再生チャネルを E 個の送信チャネルから合成することを示しているが、B C C 合成は、再生チャネルの数に限られるわけではないことに留意されたい。一般に、再生チャネルの数を任意の数のチャネルとすることができ、C より多い、または少ない数や、再生チャネルの数が送信チャネルの数以下である  
 20 場合も考えられる。

【 0 0 4 8 】

音声チャネル間の“知覚的関連差”

1 つの和信号を仮定し、I C T D、I C L D、および I C C が元の音声信号の対応するキューを近似するように、B C C は、ステレオ音声信号またはマルチチャネル音声信号を合成する。以下では、聴覚空間イメージ属性に関連した I C T D、I C L D、および I C C の役割が説明される。

【 0 0 4 9 】

空間聴力の知識は、1 つの聴覚イベントについて、I C T D および I C L D は、知覚した方向に関係していることを意味している。1 つの音源のバイノーラル室内インパルス応答 (B R I R) を考える場合、聴覚イベントの幅、聴取者が包み込まれた状態、B R I R のはじめの部分および後の部分について推定した I C C データの間に、関係がある。しかしながら、I C C と、(単に B R I R ばかりでなく) 一般的な信号のこれらの特性との間の関係は、直接的なものではない。

【 0 0 5 0 】

ステレオ音声信号およびマルチチャネル音声信号は通常、アクティブな音源信号を、同時に、閉鎖空間でのレコーディングから得られる反射信号成分と重畳したものの、または、レコーディングエンジニアによって、人工的に生成した空間印象を加えられたものの、複雑な混合物を含んでいる。異なる源信号およびそれらの残響は、時間周波数平面で異なる領域を占めている。このことは、I C T D、I C L D、および I C C に反映され、時間および周波数の関数として変化する。この場合は、瞬時 I C T D、I C L D、I C C、聴覚イベント方向、空間印象の間の関係は、明白でない。B C C を行うある実施の形態の手法は、元の音声信号の対応するキューを近似するように、これらのキューを盲目的に合成することである。

【 0 0 5 1 】

時間等価矩形帯域幅 (E R B) の 2 倍と等しいサブバンドの帯域幅を持つフィルタバンクを用いる。非公式な聴き取りでは、より高い周波数分解能を選択した場合は、B C C の音声品質があまり向上しないことが明らかになっている。デコーダに送信する必要がある I C T D、I C L D、および I C C 値が小さくなり、従ってビットレートが低くなるので、より低い周波数分解能が望ましい。

10

20

30

40

50

## 【 0 0 5 2 】

時間分解能については、ICTD、ICLD、およびICCは、通常、通常の時間間隔で考えられている。ICTD、ICLD、およびICCを約4～16ミリ秒毎に考える場合に、高い音効果が得られる。キューを非常に短時間の間隔で考える場合を除いて、先行音効果については直接考えないことに留意されたい。進みおよび遅れが時間間隔になり、1セットのキューが合成され、進みの局所的な優越については考えない場合は、音刺激の従来の遅れ進み対を想定する。このことにもかかわらず、BCCにより、平均で、平均MUSHRASコアで約87（すなわち、“非常によい”音声品質）の音声品質になり、ある音声信号については100近くにまでなる。

## 【 0 0 5 3 】

基準信号と合成した信号との間の、しばしば見受けられる知覚的に小さな差は、通常の時間間隔でICTD、ICLD、およびICCを合成することにより、幅の広い聴覚空間イメージ属性の関するキューを暗黙的に考えることを意味している。以下では、ICTD、ICLD、およびICCが、聴覚空間イメージ属性の幅とどのように関連しているかについて、説明する。

## 【 0 0 5 4 】

空間キューの推定

以下では、ICTD、ICLD、およびICCをどのように推定するか、説明する。これらの（量子化および符号化）空間キューを伝送するビットレートは、わずか数キロビット/秒なので、BCCにより、1つの音声チャンネルに必要なものに近いビットレートで、ステレオ音声信号およびマルチチャンネル音声信号を送信することが可能である。

## 【 0 0 5 5 】

図5は、本発明の一実施の形態による、図2のBCC推定器208を示すブロック図である。BCC推定器208は、図3のフィルタバンク302と同じであってもよいフィルタバンク（FB）502と、フィルタバンク502によって生成されるそれぞれ異なる周波数サブバンドに対するICTD、ICLD、およびICC空間キューを生成する推定ブロック504とを備える。

## 【 0 0 5 6 】

ステレオ信号に対するICTDICLDおよびICCの推定

次の手順は、2つの（例えば、ステレオ）音声チャンネルの対応するサブバンド信号 $\tilde{x}_1(k)$ および $\tilde{x}_2(k)$ のICTD、ICLD、およびICCに用いる。

## 【 0 0 5 7 】

・ ICTD [ サンプル ]

$$\tau_{12}(k) = \arg \max_d \{ \Phi_{12}(d, k) \}, \quad (7)$$

## 【 0 0 5 8 】

次の式（8）から得られる正規化相互相関関数の短時間推定値による。

$$\Phi_{12}(d, k) = \frac{p_{\tilde{x}_1 \tilde{x}_2}(d, k)}{\sqrt{p_{\tilde{x}_1}(k-d_1) p_{\tilde{x}_2}(k-d_2)}}, \quad (8)$$

ここで、

$$\begin{aligned} d_1 &= \max \{ -d, 0 \} \\ d_2 &= \max \{ d, 0 \} \end{aligned} \quad (9)$$

また、 $P_{\tilde{x}_1 \tilde{x}_2}$  は、 $\tilde{x}_1(k-d_1) \tilde{x}_2(k-d_2)$  の平均の短時間推定値である。

## 【 0 0 5 9 】

・ ICLD [ dB ] :

$$\Delta L_{12}(k) = 10 \log_{10} \left( \frac{p_{\hat{x}_2}(k)}{p_{\hat{x}_1}(k)} \right). \quad (10)$$

【 0 0 6 0 】

・ I C C :

$$c_{12}(k) = \max_d |\Phi_{12}(d, k)|. \quad (11)$$

【 0 0 6 1 】

正規化相互相関の絶対値について考慮し、 $c_{12}(k)$ の範囲は $[0, 1]$ であることに留意されたい。 10

【 0 0 6 2 】

マルチチャネル音声信号に対する I C T D I C L D、および I C C の推定

入力チャネルが3つ以上ある場合は、通常、基準チャネル（例えば、チャネルナンバー1）と他のチャネルとの間で I C T D および I C L D を定義することで十分である。図6には、 $C = 5$ チャネルの場合を示している。 $c_{1c}(k)$ および $c_{12}(k)$ はそれぞれ、基準チャネル1とチャネル $c$ との間の I C T D および I C L D を表す。

【 0 0 6 3 】

I C T D および I C L D とは反対に、I C C は通常、より自由度がある。定義されている I C C は、考えられるすべての入力チャネル対の間で異なる値をとることができる。 $C$  20  
個のチャネルの場合、 $C(C-1)/2$ 個の考えられるチャネル対がある。例えば、5チャネルの場合は、図7(a)に示すように10個のチャネル対がある。しかしながら、このような方法では、各時間インデックスでの各サブバンドに対し、 $C(C-1)/2$ 個の I C C 値を推定して送信することが必要になり、計算量が大きくなり、ビットレートが高くなってしまふ。

【 0 0 6 4 】

また、各サブバンドに対して、I C T D および I C L D が、サブバンドの対応する信号成分の聴覚イベントを表現する方向を求める。次に、サブバンド1つ毎に1つの I C C パラメータを用いて、全音声チャネル間の全体的なコヒーレンスを記述してもよい。各時間インデックスで、各サブバンドで最もエネルギーが大きい2つのチャネルの間だけで、I 30  
C C キューを推定して送信することにより、良い結果を得ることができる。このことを、図7(b)に示す。時刻 $k-1$ および $k$ で、チャネル対(3, 4)および(1, 2)それぞれが最も強い。発見的規則を用いて、他のチャネル対の間の I C C を求めてもよい。

【 0 0 6 5 】

空間キューの合成

図8は、図4の B C C 合成装置 4 0 0 の実施例を示すブロック図である。この装置を B C C デコーダで用いることができ、1つの送信した和信号 $s(n)$ プラス空間キューが与えられたステレオ音声信号またはマルチチャネル音声信号を生成する。和信号 $s(n)$ をサブバンドに分解する。 $\tilde{s}(k)$ は、このようなサブバンドの1つを表す。出力チャネルそ 40  
れぞれの対応するサブバンドを生成するために、遅延 $d_c$ 、倍率 $a_c$ 、およびフィルタ $h_c$ を、対応するサブバンドの和信号に適用する。（表記を簡単にするために、遅延、倍率、およびフィルタの時間インデックス $k$ を無視する。）I C T D は遅延を行い、I C L D をスケーリングし、I C C に非相関フィルタを適用してそれぞれ合成する。図8に示す処理は、別々に各サブバンドに適用される。

【 0 0 6 6 】

I C T D 合成

次の式(12)により、I C T D  $c_{1c}(k)$ から遅延 $d_c$ を求める。遅延 $d_c$ の最大大きさを最小にするように、基準チャネル $d_1$ の遅延が算出される。変更するサブバンド信号が少なくなるほど、アーティファクトが発生する危険性がより少なくなる。サブバンドの 50



サンプリングレートが、ICTD合成に対する時間分解能が十分高くない場合は、適したオールパスフィルタを用いることにより、より正確に遅延を行うようにする。

$$d_c = \begin{cases} -\frac{1}{2}(\max_{2 \leq l \leq C} \tau_{1l}(k) + \min_{2 \leq l \leq C} \tau_{1l}(k)), & c=1 \\ \tau_{1l}(k) + d_1, & 2 \leq c \leq C. \end{cases} \quad (12)$$

【0067】

ICLD合成

チャンネルcおよび基準チャンネル1の間で出力サブバンド信号が所望のICLD<sub>12</sub>(k)を有するようにするために、利得係数a<sub>c</sub>は、次の式(13)を満たす必要がある。

10

$$\frac{a_c}{a_1} = 10^{\frac{\Delta_{1c}(k)}{20}}. \quad (13)$$

また、好ましくは、全出力チャンネルのパワーの合計が、入力信号のパワーと等しくなるように、出力サブバンドを正規化する。各サブバンドの元の信号パワーのすべてが和信号に保たれているので、この正規化は、各出力チャンネルの絶対サブバンドパワーにおいて、元のエンコーダ入力音声信号の対応するパワーを近似することになる。これらの制約条件から、次の式(14)により、倍率a<sub>c</sub>が得られる。

$$a_c = \begin{cases} 1/\sqrt{1 + \sum_{i=2}^C 10^{\Delta_{1i}/10}}, & c=1 \\ 10^{\Delta_{1c}/20} a_1, & \text{otherwise.} \end{cases} \quad (14) \quad 20$$

【0068】

ICC合成

ある実施の形態では、ICC合成の目的は、ICTDおよびICLDに影響を与えることなく、遅延およびスケリングを行った後で、サブバンド間の相関を低減することである。平均変動が各サブバンド(聴覚的に重要な帯域)でゼロになるように、周波数の関数としてICTDおよびICLDが効果的に変化するように、図8のフィルタh<sub>c</sub>を設計することにより、このことを行うことが可能である。

【0069】

30

図9は、周波数の関数として、ICTDおよびICLDがサブバンド内でどのように変化するかを示している。ICTDおよびICLD変動の振幅が、非相関の度合いを求め、ICCの関数として制御する。ICTDは滑らかに変化し(図9(a)に示す)、ICLDはランダムに変化する(図9(b)に示す)ことに留意されたい。ICLDを、ICTDのように滑らかに変化させることもできるが、このことにより、得られる音声信号をさらに特徴付けることになる。

【0070】

マルチチャンネルICC合成を行うのに特に適した、ICCを合成する別の方法は、次の文献に詳細に記載されている。C.フォーラ、「パラメトリックマルチチャンネル音声符号化: コヒーレンスキューの合成 (Parametric multi-channel audio coding: Synthesis of coherence cues)」(IEEE会報、スピーチおよびオーディオ学会、2003年)」(IEEE会報、スピーチおよびオーディオ学会、2003年)。この要旨は、ここに引例として組み込まれている。時間および周波数の関数として、一定量のアーティフィシアルな後期残響を出力チャンネルそれぞれに加算して、所望のICCを得る。また、得られる信号のスペクトルエンベロープが元々の音声信号のスペクトルエンベロープに近づくように、スペクトル変更を行うこともできる。

40

【0071】

ステレオ信号(または音声チャンネル対)に対するICC合成技術に関係する、または関係しない他の技術が、次の文献に記載されている。E.シュイエールス(Schuijse

50

rs)、W.オーメン(Oomen)、B.デン・ブリンカー(den Brinker)、J.ブレーバールト(Breebaart)、「高品質音声のためのパラメトリック符号化の進歩(Advances in parametric coding for high-quality audio)」(音声工学学会第114回大会予稿集、2003年3月)、J.エンゲデガルド(Engdegard)、H.ブルンハーゲン(Purnhagen)、J.ローデン(Roden)、L.リルジェリド(Liljeryd)、「パラメトリックステレオ符号化における合成環境(Synthetic ambience in parametric stereo coding)」(音声工学学会第117回大会予稿集2004年5月)。これらの要旨は、ここに引例として組み込まれている。

10

## 【0072】

CチャンネルからEチャンネルへのBCC

前述のように、2つ以上の伝送チャンネルでBCCを行うことができる。BCCのバリエーションが、1つの(送信した)チャンネルだけでなく、E個のチャンネルとしてもC個の音声チャンネルを表すことについて、説明してきたが、これを、CチャンネルからEチャンネルへのBCCと記載する。CチャンネルからEチャンネルへのBCCを行う、(少なくとも)2つの目的がある。

- ・1つの伝送チャンネルで、BCCは、既存のモノラルシステムをグレードアップして、ステレオ音声再生またはマルチチャンネル音声再生を行う後方互換性経路を提供する。グレードアップしたシステムは、既存のモノラルインフラを介してBCCダウンミキシング和信号を送信し、さらにBCCサイド情報を送信する。CチャンネルからEチャンネルへのBCCを、Cチャンネル音声のEチャンネル後方互換性符号化に適用することができる。

20

- ・CチャンネルからEチャンネルへのBCCは、送信チャンネルの数を異なる度合いで低減するという意味で、拡張性を導入する。送信する音声チャンネルの音声品質がさらに向上することが期待できる。

ICTD、ICLD、およびICCキューをどのように定義するかというような、CチャンネルからEチャンネルへのBCCの信号処理の詳細は、米国出願第10/762,100号01/20/042004年1月20日出願(Faller 13-1)に記載されている。

## 【0073】

拡散音波形整形

ある実施の形態では、BCC符号化は、ICTD、ICLDおよびICC合成のためのアルゴリズムを含む。ICCキューは、対応するサブバンドの信号成分を非相関させることによって合成されることができる。これは、ICLD、ICTDの周波数に依存するバリエーションおよびICLD(全通過のフィルタリング)の、または、残響アルゴリズムに関連した考えを有する周波数に依存するバリエーションによってされることができる。

30

## 【0074】

これらの技術が音声信号に適用される場合に、信号の時間エンベロープ特性は保存されない。具体的には、一時的現象に適用される場合に、瞬間的な信号エネルギーは特定の期間広がりそうである。これは、プレエコーまたはくたびれた一時的現象のアーティファクトである。

40

## 【0075】

本発明のある実施の形態の一般的な原理は、BCCデコーダによって合成される音が元のそれと類似しているスペクトル特性を有しなければならないだけでなく、類似の知覚的な特性を有するために全く密接に元の時間エンベロープにも似ていなければならないという観察に関する。通常、これは、各信号チャンネルの時間エンベロープに近づくために時間様々なスケール動作を適用する動的なICLD合成を含むことによって、BCCのような方式において成し遂げられる。しかしながら、一時的な信号(音の立ち上がり、打楽器、その他)という場合は、このプロセスの時間軸分解度は、十分に密接に元の時間エンベロープに近い合成された信号を生成するために、充分ではなくてもよい。このセクショ

50

ンは、十分に微細な時間分解能によってこれをするために、多くの方法を記載する。

【 0 0 7 6 】

さらに、元の信号の時間エンベロープにアクセスしない B C C デコーダのために、アイデアは、その代わりに近似として送信された「合計信号 ( s )」の時間エンベロープをとることである。このように、このようなエンベロープ情報を伝達するために、 B C C エンコーダから B C C デコーダまで送信されるのに必要なサイド情報がない。要約すると、本発明は、以下の原理に依存する：

- ・送信された音声チャンネル (すなわち、「合計チャンネル」) または、これらのチャンネルの線形結合に基づく B C C 合成は、(例えば、 B C C ブロックサイズより著しく微細な) 高時間分解を有するそれらの時間エンベロープに対する時間エンベロープ抽出器によって分析される。

10

- ・各出力チャンネルに対する次の合成された音は、 I C C 合成の後でさえ、できるだけ密接に抽出器により求められた時間エンベロープに整合するように、形成される。これは、一時的な信号の場合さえ、合成された出力音が I C C 合成 / 信号非相関性処理によって著しく劣化しないことを確実にする。

【 0 0 7 7 】

本発明の一実施の形態によれば、図 1 0 は、少なくとも一部の B C C デコーダ 1 0 0 0 で、ブロック図表示式を示す。図 1 0 において、ブロック 1 0 0 2 は、少なくとも、 I C C 合成を含む B C C 合成処理を表す。 B C C 合成ブロック 1 0 0 2 は、ベースチャンネル 1 0 0 1 を受信して、合成されたチャンネル 1 0 0 3 を生成する。特定の実施例において、ブロック 1 0 0 2 は図 4 のブロック 4 0 6、4 0 8 および 4 1 0 の処理を表す。そこにおいて、ベースチャンネル 1 0 0 1 はブロック 4 0 4 をアップミキシングすることによって発生する信号であり、そして、合成されたチャンネル 1 0 0 3 は相関ブロック 4 1 0 によって生成する信号である。図 1 0 は、処理が 1 つのベースチャンネル 1 0 0 1 ' およびその対応する合成されたチャンネルのために行うことを表す。類似の処理は、各他のベースチャンネルおよびその対応する合成されたチャンネルにも適用される。

20

【 0 0 7 8 】

エンベロープ抽出器 1 0 0 4 は、ベースチャンネル 1 0 0 1 ' の微細な時間エンベロープ a を決定し、そして、エンベロープ抽出器 1 0 0 6 は、合成されたチャンネル 1 0 0 3 ' の微細な時間エンベロープ b を決定する。逆エンベロープ調整装置 1 0 0 8 は、フラット (例えば、一様な) 時間エンベロープを有する平坦化された信号 1 0 0 5 ' を生成するために、合成されたチャンネル 1 0 0 3 ' のエンベロープ (すなわち、時間的微細なエンベロープの「平坦化」) を規格化するために、エンベロープ抽出器 1 0 0 6 から時間エンベロープ b を使用する。特定の実施例に応じて、平坦化は、アップミキシングする、前または後ろに適用されることができる。エンベロープ調整装置 1 0 1 0 は、ベースチャンネル 1 0 0 1 の時間エンベロープと実質的に等しい時間エンベロープを有している出力信号 1 0 0 7 ' を生成するために、平坦化された信号 1 0 0 5 ' における元の信号エンベロープを再び課すためのエンベロープ抽出器 1 0 0 4 から時間エンベロープ a が使用される。

30

【 0 0 7 9 】

実施の形態に応じて、この時間エンベロープ処理 (また、本明細書において、「エンベロープ整形」と称される) は、全ての合成されたチャンネル (示すように) に、または、合成されたチャンネル (その後記載されているように) の直交化された一部 (例えば後期残響パート、非相関している一部) だけに適用されることができる。さらに、実施例に応じて、エンベロープ整形は、時間領域信号に、または、周波数に依存する方法 (例えば、時間エンベロープは、異なる周波数で個々に推定されて、課される) で適用されることができる。

40

【 0 0 8 0 】

逆エンベロープ調整装置 1 0 0 8 およびエンベロープ調整装置 1 0 1 0 は、異なる方法で行うことができる。 1 つの実施の形態の形式において、信号のエンベロープは、時間可変振幅変更関数 (例えば、逆エンベロープ調整装置 1 0 0 8 に対する  $1 / b$  およびエンベ

50

ロープ調整装置 1010 に対する a) を有する信号の時間領域のサンプル (または、スペクトル/サブバンドサンプル) の乗算によって乗算される。あるいは、周波数の上の信号のスペクトル表現の畳込み/フィルタリングが、低いビットレートな音声符合化の量子化雑音を成形するために既知の発明において使われて、それに類似した方法で使われることができる。同様に、信号の時間エンベロープは、信号の時間が構築する分析によって、または、周波数の上の信号スペクトルの自己相関を調べることによって直接抽出されることもできる。

#### 【0081】

図 11 は、図 4 の BCC 合成装置 400 の条件で、図 10 のエンベロープ整形方式の典型的なアプリケーションを例示する。本実施の形態において、一つの送信された合計信号  $s(n)$  があり、C 個のベース信号が、その合計信号を複製することにより生成され、そして、エンベロープ整形が、個々に異なるサブバンドに適用される。別の実施例では、遅延、スケールングおよび他の処理の順序は、異なってもよい。さらに、別の実施例では、エンベロープ整形は、それぞれに各サブバンドを処理することに制限されない。これは、信号の時間微細構造に関する情報を引き出すために周波数帯の上の共分散を利用する実施例に基づく、特に畳込み/フィルタリングのために当てはまる。

10

#### 【0082】

図 11 (a) において、時間処理分析 1104 (TPA) を処理することは、図 10 のエンベロープ抽出器 1004 に類似している、そして、各時間処理、1106 はエンベロープ抽出器 1006、逆エンベロープ調整装置 1008 および図 10 のエンベロープ調整装置 1010 の組合せに類似している (TP)。

20

#### 【0083】

図 11 (b) は、TPA 1104 の一つの時間領域に基づく実施可能なブロック図を示す。ベース信号サンプルは、二乗され (1110)、次に、ベース信号の時間エンベロープ a を特徴づけるためにローパスフィルタされる (1112)。

#### 【0084】

図 11 (c) は、TP 1106 の一つの時間領域に基づく実施可能なブロック図を示す。合成された信号サンプルは、二乗され (1114)、次に、合成された信号の時間エンベロープ b を特徴づけるためにローパスフィルタされる (1116)。倍率 (例えば、平方根 ( $a/b$ )) が生成されて (1118)、次に、元のベースチャンネルのそれに実質的に等しい時間エンベロープを有する出力信号を生成するために、合成された信号に適用される (1120)。

30

#### 【0085】

TPA 1104 および TP 1106 の他の実施例において、時間エンベロープは、信号サンプルを二乗することによってよりむしろ大きさの演算を使用して特徴づけられる。このような実施の形態では、比率  $a/b$  が、平方根演算を適用することなく、倍率として使用してもよい。

#### 【0086】

図 11 (c) のスケールング演算が、TP 処理の時間領域ベースの実施例に対応しているが、(以下に説明する) 図 17 18 の実施の形態のような、周波数領域信号を使用して、TP 処理 (TPA および逆 TP (ITP) 処理とともに) は行うこともできる。従って、この明細書の目的においては、用語「スケールング関数」は、図 18 b および c のフィルタリング演算のように、時間領域または周波数領域のいずれもカバーするように解釈する必要がある。

40

#### 【0087】

一般に、好ましくは、それらが信号パワー (すなわちエネルギー) を変更しないように、TPA 1104 および TP 1106 は設計される。ある実施の形態に応じて、この信号パワーは、例えば、合成ウィンドウまたはパワーの若干の他の適した計測によって定義される期間のチャンネル当たりの全体の信号パワーに基づく各チャンネルにおける短時間平均信号パワーとしてもよい。従って、エンベロープ整形の前か後で、ICLD 合成 (例えば、

50

乗算器 408 を使用して) のためのスケーリングは、適用されることことができる。

【0088】

なお、各チャネルのために、図 11 (a) で、2 つの出力がある。TP 処理は、それらの中のわずか 1 つに適用されることに留意されたい。これは、2 つの信号成分を混合する ICC 合成方法を反映する：修正されないおよび直交化された信号、修正されないおよび直交化された信号の成分の比率は、ICC を決定する。図 11 (a) の図示した実施例において、TP は、直交された信号成分だけ適用される。和ノード 1108 は、対応する時間的に形作られ、直交化された信号の成分を有する修正されない信号成分に再結合する。

【0089】

図 12 は図 4 の BCC 合成装置 400 の条件で、図 10 のエンベロープ整形方法の代わりの典型的なアプリケーションを例示する。エンベロープ整形は時間領域において適用される。そのような実施の形態は、ICTD、ICLD および ICC 合成が、実行されるスペクトル表現の時間分解能が、所望の時間エンベロープを課すことにより、「プレエコー」を効果的に妨げることが十分に出来ない場合、正当化されてもよい。例えば、BCC が短いフーリエ変換 (STFT) によって行う場合に、このようなケースでもよい。

【0090】

図 12 (a) に示すように、TPA1204 および各 TP1206 は時間領域において行う。それが所望の時間エンベロープ (例えば、送信された合計信号から推定されるものとしてのエンベロープ) を有するように、全帯域信号はスケーリングされる。図 12 (b) および (c) は、図 11 (b) および (c) に示されるそれらに類似する TPA1204 および TP1206 の可能な実施例を示す。

【0091】

本実施の形態では、直交化された信号成分だけでなく、TP 処理は、出力信号に適用される。別の実施の形態では、時間領域ベールの TP 処理は、所望の場合は、直交化された信号成分にちょうど適用されることが出来る。修正されないおよび直交化されたサブバンドのケースは、別々の逆フィルタバンクを有する時間領域に変換される。

【0092】

BCC 出力信号の全帯域スケーリングがアーティファクトとして発生する場合があるので、エンベロープ整形は指定された周波数、例えば、特定のカットオフ周波数  $f_{tp}$  (例えば、500 Hz) より大きい周波数だけに適用されることが出来る。分析のための周波数範囲 (TPA) が、合成 (TP) のための周波数範囲と異なる場合もあることに留意されたい。図 13 (a) および (b) は、TPA1204 および TA1206 の可能な実施例を示す。エンベロープ整形は、カットオフ周波数  $f_{tp}$  より高い周波数でのみ適用される。特に、図 13 (a) はハイパスフィルタ 1302 のさらに示されており、これは、時間エンベロープ特徴を調べる前に、 $f_{tp}$  より低い周波数をフィルタする。図 13 (b) は、2 つのサブバンドの間の  $f_{tp}$  のカットオフ周波数を有する 2 帯域フィルタバンクがさらに示されており、高い周波数部分のみ、時間的に整形される。次に、2 帯域の逆フィルタバンク 1306 は、出力信号を生成するために、低い周波数部分を時間的に整形される高い周波数部分と再合成する。

【0093】

図 14 は、代理人明細書、no. Baumgarte 7-12 04/01/04 に出願された米国特許番号 10/815, 591 に記載された後期残響に基づく ICC 合成方法の条件で、図 10 のエンベロープ整形方法の典型的なアプリケーションを例示する。本実施の形態において、TPA1404 および各 TP1406 は、図 12 または図 13 として時間領域において適用される。しかし、各 TP1406 は、異なる後期残響 (LR) ブロック 1402 から出力に適用される。

【0094】

図 10 に示される方法の変形例である本発明の一実施例によれば、図 15 は、少なくとも一部の BCC デコーダ 1500 で、ブロック図表示式を示す。図 15 において、BCC

10

20

30

40

50

合成ブロック 1502、エンベロープ抽出器 1504 およびエンベロープ調整装置 1510 は、図 10 の BCC 合成ブロック 1002、エンベロープ抽出器 1004 およびエンベロープ調整装置 1010 に類似している。図 15 において、しかしながら、逆エンベロープ調整装置 1508 は、BCC 合成の後よりはむしろ、図 10 のような BCC 合成の前に使用される。このようにして、BCC 合成が適用される前に、逆エンベロープ調整装置 1508 はベースチャンネルを平坦化する。

【0095】

図 10 および図 15 に示される方法の変形例である本発明の一実施例によれば、図 16 は、少なくとも一部の BCC デコーダ 1600 で、ブロック図表示式を示す。図 16 において、エンベロープ抽出器 1604 およびエンベロープ調整装置 1610 は、図 15 のエンベロープ抽出器 1504 およびエンベロープ調整装置 1510 に類似している。図 15 の実施例において、しかしながら、合成ブロック 1602 は、それと類似の残響ベースの ICC 合成が図 16 に示されることを表す。この場合、エンベロープ整形は無関係の後期残響信号だけに適用され、そして、和ノード 1612 は時間的に整形された、後期残響信号を最初のベースチャンネル（それは、すでに所望の時間エンベロープを有する）に加える。なお、この場合、逆エンベロープ調整装置は、使用される必要はない、なぜなら、後期残響信号がブロック 1602 のその生成プロセスのためおよそ平坦な時間エンベロープを有するからである。

【0096】

図 17 は、図 4 の BCC 合成装置 400 の条件で、図 15 のエンベロープ整形方式の典型的なアプリケーションを例示する。図 17 において、TPA 1704、逆 TP (ITP) 1708 および TP 1710 は、図 15 におけるエンベロープ抽出器 1504、逆エンベロープ調整装置 1508 およびエンベロープ調整装置 1510 に類似している。

【0097】

この周波数ベースの実施例において、拡散音のエンベロープ整形は、周波数軸に沿って畳み込みを（例えば STF T）フィルタバンク 402 の周波数ビンに適用することによって行う。米国特許 5,781,888 (Herre) および米国を参照する、内容の特許 5,812,971 (Herre)（その教示は本願明細書に引用したものとする）は、この技術に関した。

【0098】

図 18 (a) は、図 17 の TPA 1704 の 1 つの実施可能なブロック図を示す。この実施の形態において、TPA 1704 は、周波数の上のスペクトル係数の直列に最適の予測係数を決定する線形予測分析 (LPC) 分析動作として行う。このような LPC 分析技術は、例えば、周知である LPC 係数の効果的な算出のための音声符号化および多くのアルゴリズムから、自己相関方法（信号の自己相関関数および次のレビンソン ダービン再帰の算出を含む）は、公知である。この計算の結果、一組の LPC 係数は、信号の時間エンベロープを表す出力で利用できる。

【0099】

図 18 (b) および (c) は図 17 の ITP 1708 および TP 1710 の実施可能なブロック図を示す。両方の実施の形態において、処理される信号のスペクトル係数は、（増減するまたは減少する）周波数の順に処理される。それは、スイッチ回路の丸められ、これらの係数を予測フィルタ処理（およびこの処理の後に元のところへ）により処理されるシリアルオーダーに変換されることによって象徴化される。ITP 1708 の場合、予測するフィルタリングは、残余の予測を算出して、このようにして時間的信号エンベロープを「平坦化」する。TP 1710 の場合、逆フィルタは、TPA 1704 から LPC 係数によって表される時間エンベロープを再導入する。

【0100】

TPA 1704 による信号の時間エンベロープの算出のために、このようなウィンドウが使われる場合、フィルタバンク 402 の分析ウィンドウの影響を除去することは重要である。これは（知られている）分析ウィンドウ形状によって、または、分析ウィンドウを

10

20

30

40

50

使用しない別々の分析フィルタバンクを用いることによって、結果として得られるエンベロープを規格化することによって、達成されることも出来る。

【0101】

図17の畳み込み/フィルタリングに基づく技術は、図16のエンベロープ整形方法の条件でも適用される。エンベロープ抽出器1604およびエンベロープ調整装置1610は、それぞれ、図18(a)のTPAおよび図18(c)のTPに基づいている。

【0102】

更なる別の実施例

BCCデコーダは、選択的にエンベロープ整形を可能にして/使用不能にするように設計されることができる。例えば、エンベロープ整形の利点が、エンベロープ整形が生成することが出来るいかなるアーティファクトを支配するように、合成された信号の時間エンベロープが十分に変動する場合に、BCCデコーダは、従来のBCC合成方法を適用することが可能であり、エンベロープ整形を可能にすることができる。この有効な/無効制御は、以下によって成し遂げられることができる：

(1) 一時的な検出：一時的現象が検出される場合、TP処理は使用可である。一時的現象の検知は、一時的現象だけでなく、一時的現象の前後の短時間にも効果的に形作るために先取りの方法で行うことができる。一時的現象を検出する考えられる方法は、以下を含む：

- ・一時的現象の発生を示しているパワーの急増があるときに送信されたBCCの時間エンベロープが決定する信号を合計するのを観察すること；そして、

- ・予測(LPC)フィルタの利得を検証すること。LPC予測利得が指定された閾値を上回る場合、信号が一時的であるかまたは非常に変動していると仮定されることができる。LPC分析は、スペクトルの自己相関により算出される。

(2) ランダム性検出：時間エンベロープが疑似ランダム的に変動するときに、シナリオがある。このようなシナリオにおいて、一時的現象は検出されないかもしれない、しかし、TP処理はまだ適用されることができる(例えば、密度の高い拍手信号は、このようなシナリオに対応する)。

【0103】

加えて、ある実施の形態では、音の信号の可能なアーティファクトを防止するために、送信された合計信号の調性が高いときに、TP処理は適用されない。

【0104】

さらにまた、いつのTP処理が作動中でなければならぬかについて検出するために、類似の方法が、BCCエンコーダで用いられることができる。エンコーダがすべての元の入力信号にアクセスするので、TP処理が可能な場合の決定をさせるより高度なアルゴリズム(例えば、推定ブロック208の部分)を使用することができる。この決定(TPが作動中の場合、フラグシグナリング)の結果は、BCCデコーダ(例えば、図2のサイド情報の一部として)に送信されることができる。

【0105】

本発明が、一つの合計信号があるBCC符号化方法の条件で、記載されているにもかかわらず、本発明は、2以上の合計信号を有するBCC符号化方法の条件で、行うことができる。この場合、各異なる「ベース」合計信号のための時間エンベロープは、BCC合成を適用する前に推定されることができる、そして、異なるBCC出力チャネルは異なる時間エンベロープに基づいて発生することができる。そして、それに応じて、合計信号は異なる出力チャネルを合成するために用いられた。2つ以上の異なる合計チャネルから合成される出力チャネルは、成分和チャネルの相対的な効果を考慮する(例えば、加重平均算出を介して)効果的な時間エンベロープに基づいて生成することができる。

【0106】

本発明がICTD、ICLDおよびICCコードを含んでいるBCC符号化方法の条件で、記載されていたが、これら3つのタイプのコードのうちの1つまたは2つだけを用いる他のBCC符号化方法(例えば、ICTDではなく、ICLDおよびICC)および/

10

20

30

40

50

または1つ以上の別のタイプのコードをさらに用いる他のBCC符号化方法の条件で、本発明を実施することもできる。さらに、BCC合成処理およびエンベロープ整形のシーケンスは、異なる実施例において変形することができる。例えば、図14および16に示すように、周波数領域信号に適用される場合に、ICTD合成（ICTD合成を使用するこれらの実施例において）の後、ICLD合成の前であるが、エンベロープ整形を行うこともできる。他の実施の形態において、任意の他のBCC合成が適用される前に、エンベロープ整形はアップミキシングされた信号に適用されることことができる。

【0107】

BCC符号化方法の条件で、本発明について説明してきたが、音声信号を非相関にする他の音声処理システム、または信号を非相関にする必要がある他の音声処理の条件で、本発明を実施することもできる。

10

【0108】

エンコーダが時間領域で入力音声信号を受信し、時間領域で送信音声信号を生成し、デコーダが時間領域で送信音声信号を受信し、時間領域で再生音声信号を生成する実施例の条件で、本発明について説明してきたが、本発明はこれに限定されない。例えば、他の実施の形態において、任意の1つ以上の入力し、送信し、再生した音声信号を、周波数領域で表現することができる。

【0109】

BCCエンコーダおよび/またはデコーダを、テレビまたは電子音楽配信、映画館、放送、ストリーミング、および/または受信システム等の、様々な異なる適用例またはシステムとともに用いたり、これらに組み込んで用いたりすることもできる。これらは、例えば、地上波、衛星、ケーブル、インターネット、イントラネット、または物理媒体（例えば、コンパクトディスク、デジタルバーサタイルディスク、半導体チップ、ハードドライブ、メモリカード等）を介して、符号化/復号化伝送を行うシステムを含む。BCCエンコーダおよび/またはデコーダを、ゲームおよびゲームシステムも用いることもできる。これらは、例えば、ユーザインタラクティブな娯楽用（アクションゲーム、ロールプレイングゲーム、戦略ゲーム、アドベンチャーゲーム、シミュレーションゲーム、レーシングゲーム、スポーツゲーム、ゲームセンター、カードゲーム、およびボードゲーム）および/または複数のマシン、プラットフォーム、またはメディア等に発行した教育向けの、インタラクティブなソフトウェア製品を含む。さらに、BCCエンコーダおよび/またはデコーダを、音声レコーダ/プレーヤまたはCD-ROM/DVDシステムに組み込んでよい。BCCエンコーダおよび/またはデコーダを、デジタル復号化（例えば、プレーヤ、デコーダ）を組み込んだPCソフトウェアアプリケーション、デジタル符号化する能力（例えば、エンコーダ、リッパ、レコーダ、およびジュークボックス）を組み込んだソフトウェアアプリケーションに組み込むこともできる。

20

30

【0110】

本発明を、回路ベースの処理として実施することもできる。1つの集積回路（ASICまたはFPGA等）、マルチチップモジュール、シングルカード、またはマルチカード回路パッケージ等の考えられる実施例が挙げられる。回路素子の各種の関数を、ソフトウェアプログラムの処理工程として実施できることも、当業者にとって明らかになるであろう。このようなソフトウェアを、例えば、デジタル信号プロセッサ、マイクロコントローラ、または汎用コンピュータに用いることもできる。

40

【0111】

それらの方法を行う方法および装置の形態で、本発明を実施することができる。本発明を、フロッピー（登録商標）ディスク、CD-ROM、ハードドライブ、または任意の他の機械読み取り可能記憶媒体等の、有形媒体で、プログラムコードの形態で実施することもできる。プログラムコードをコンピュータ等のマシンにロードして実行する場合は、このマシンが本発明を実施する装置となる。本発明を、プログラムコードの形態で実施することもできる。例えば、記憶媒体に記録したり、マシンにロードしたり、マシンで実行したり、マシンにロードしてマシンで実行したり、および/またはマシンで実行したり、電

50



子ワイヤまたはケーブル、光ファイバ、または電磁放射等の、伝送媒体またはキャリアで送信したりする。プログラムコードをコンピュータ等のマシンにロードして実行する場合は、そのマシンが本発明を実施する装置となる。汎用プロセッサ上で実施する場合は、プログラムコードセグメントをプロセッサと組み合わせて、一意のデバイスに送って、特定の論理回路と同様に動作させる。

【0112】

以下の特許請求の範囲で述べる本発明の範囲を逸脱することなく、この本発明の本質を説明するために、述べ、示してきた、詳細、素材、構成について、当業者が様々に変更できることが、さらに理解できるであろう。

【0113】

クレームを反復することにより、それらのステップの一部またはすべてを実行する特定のシーケンスを示さない限り、必要の場合は、次の方法クレームのステップを、対応するラベル構成で特定のシーケンスで反復するが、それらのステップを、その特定のシーケンスで実行することを、必ずしも必要としているものではない。

【図面の簡単な説明】

【0114】

【図1】従来のバイノーラル信号合成装置の上位ブロック図を示す。

【図2】一般バイノーラルキュー符号化(BCC)音声処理システムを示すブロック図である。

【図3】図2のダウンミキサに用いることができるダウンミキサのブロック図を示す。

【図4】図2のデコーダに用いることができるBCC合成装置のブロック図を示す。

【図5】本発明の一実施の形態による、図2のBCC推定器のブロック図を示す。

【図6】5チャンネル音声用ICTDおよびICLDデータ生成を説明する図である。

【図7A】5チャンネル音声用ICCデータを説明する図である。

【図7B】5チャンネル音声用ICCデータを説明する図である。

【図8】BCCデコーダに用いることができ、1つの送信した和信号 $s(n)$ プラス空間キューが与えられたステレオ音声信号またはマルチチャンネル音声信号を生成する、図4のBCC合成装置の一実施例のブロック図を示す。

【図9】周波数関数として、サブバンド内でICTDおよびICLDがどのように変化するかを説明する図である。

【図10】本発明の一実施の形態による、BCCデコーダの少なくとも部分を示すブロック図を示す。

【図11A】図4のBCC合成装置の条件での図10のエンベロープ整形方法の典型的なアプリケーションを示す。

【図11B】図4のBCC合成装置の条件での図10のエンベロープ整形方法の典型的なアプリケーションを示す。

【図11C】図4のBCC合成装置の条件での図10のエンベロープ整形方法の典型的なアプリケーションを示す。

【図12A】エンベロープ整形は、時間領域で適用され、図4のBCC合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

【図12B】エンベロープ整形は、時間領域で適用され、図4のBCC合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

【図12C】エンベロープ整形は、時間領域で適用され、図4のBCC合成装置の条件でのエンベロープ整形方法の代わりとなる典型的なアプリケーションを示す。

【図13】カットオフ周波数 $f_{tp}$ より高い周波数にだけエンベロープ整形が行われる、図12のTPAおよびTPの考えられる実施例である。

【図14】代理人整理番号第Baumgart 7-2として2004年4月1日出願の米国特許出願番号第10/815,591に記載のICC合成方法に基づく後期残響の条件での図10のエンベロープ整形方法の典型的なアプリケーションを示す。

【図15】図10に示される方法に代わりに、本発明の実施例により、BCCデコーダの

10

20

30

40

50

少なくとも部分を示すブロック図を示す。

【図16】図10および図15に示される方法の代わりに、本発明の実施例により、BCCデコーダの少なくとも部分を示すブロック図を示す。

【図17】図4のBCC合成装置の条件での図15のエンベロープ整形の典型的なアプリケーションを示す。

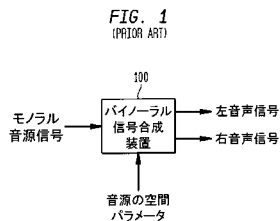
【図18A】図15および図16のTPAと、図16のITPおよびTPとの考えられる実施例のブロック図を示す。

【図18B】図15および図16のTPAと、図16のITPおよびTPとの考えられる実施例のブロック図を示す。

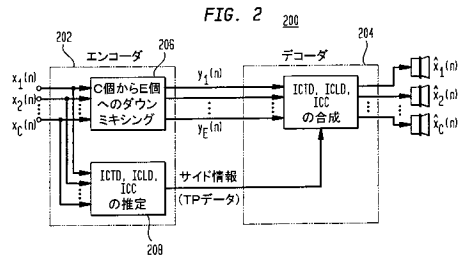
【図18C】図15および図16のTPAと、図16のITPおよびTPとの考えられる実施例のブロック図を示す。

10

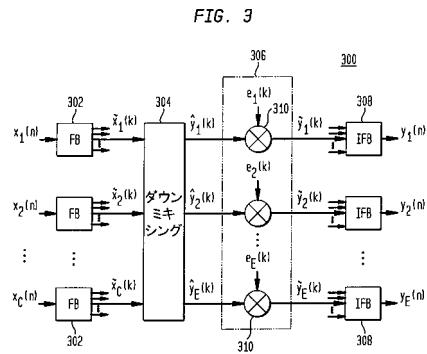
【図1】



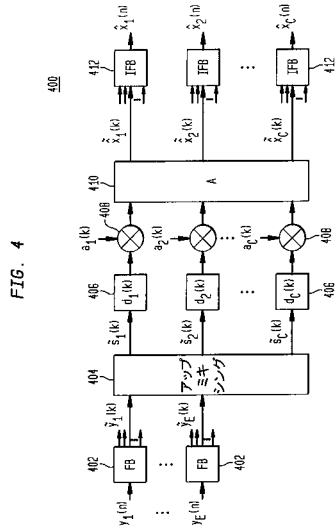
【図2】



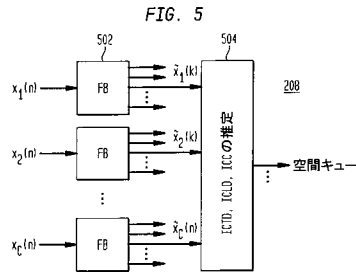
【図3】



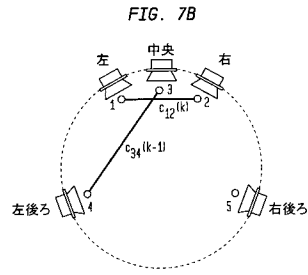
【 図 4 】



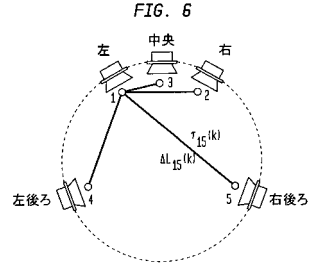
【 図 5 】



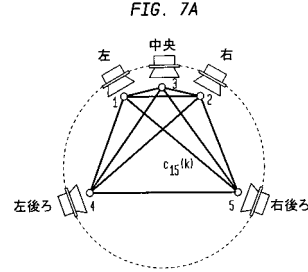
【 図 7 B 】



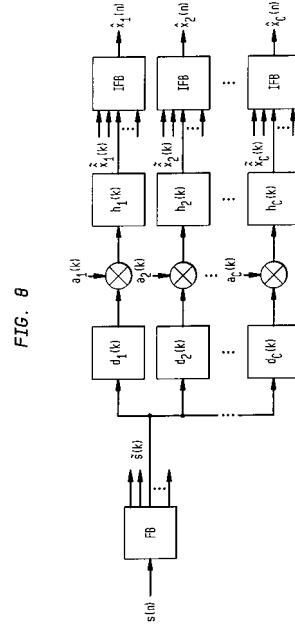
【 図 6 】



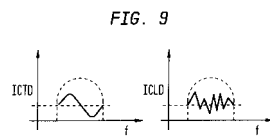
【 図 7 A 】



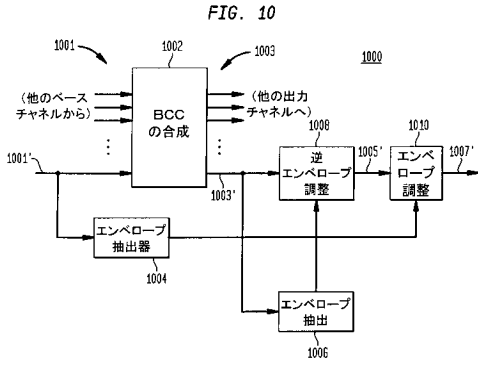
【 図 8 】



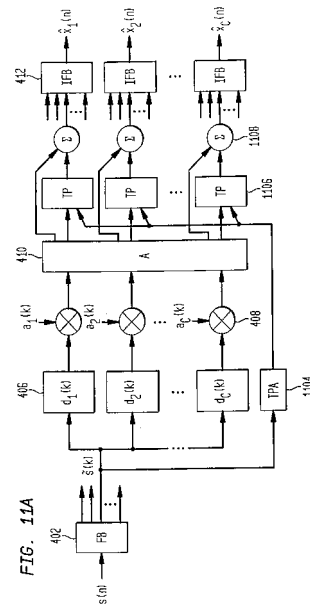
【 図 9 】



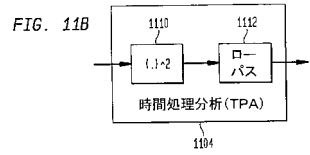
【図10】



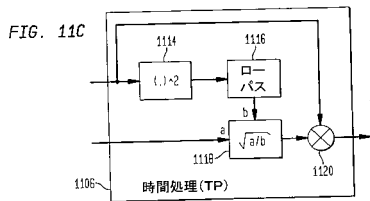
【図11A】



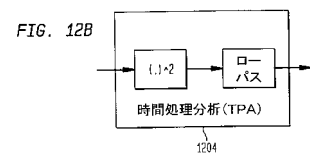
【図11B】



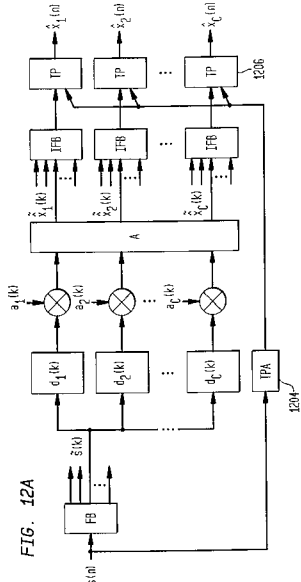
【図11C】



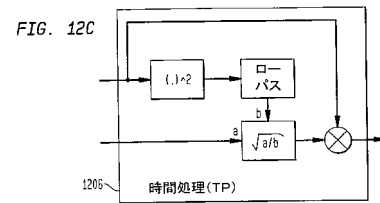
【図12B】



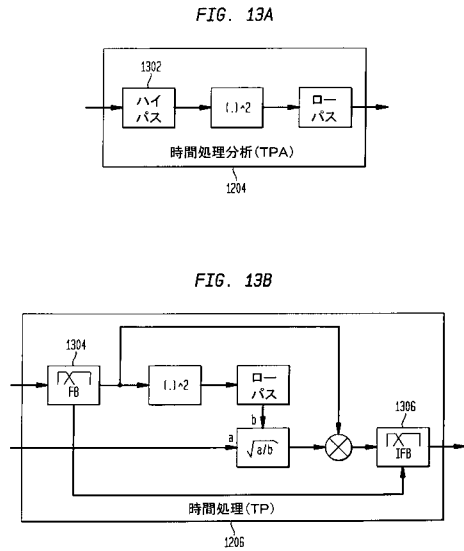
【図12A】



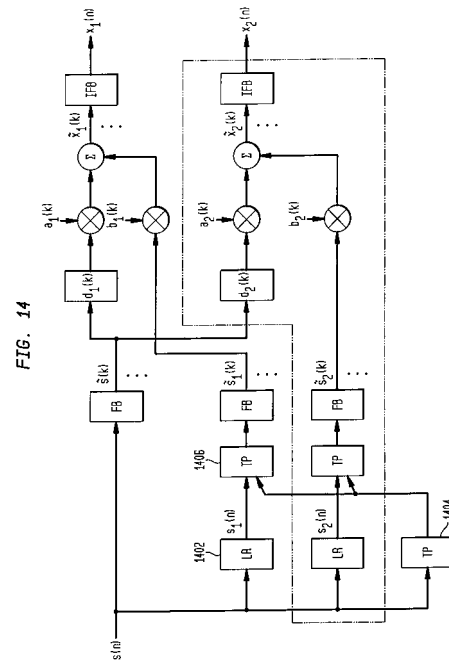
【図12C】



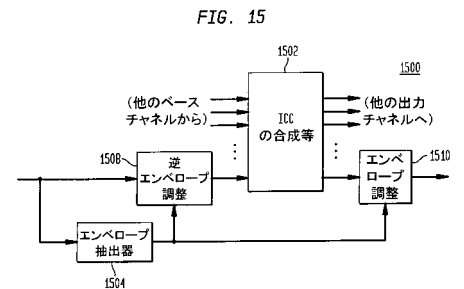
【 図 1 3 】



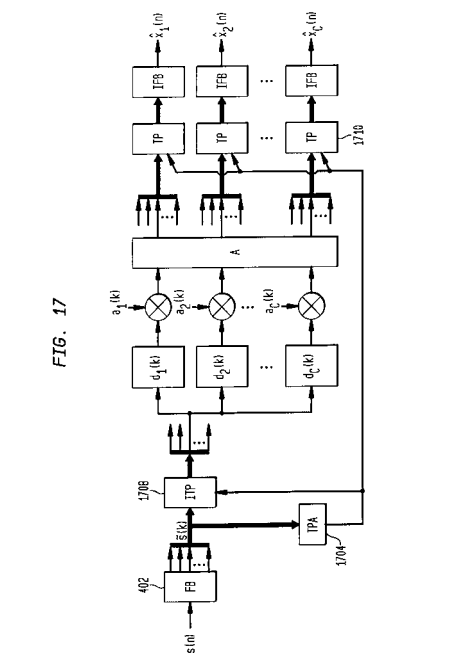
【 図 1 4 】



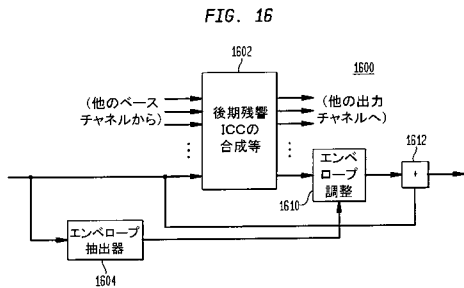
【 図 1 5 】



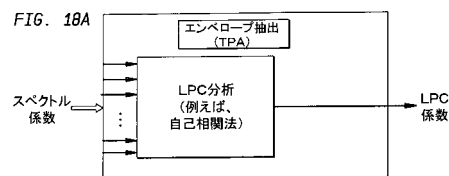
【 図 1 7 】



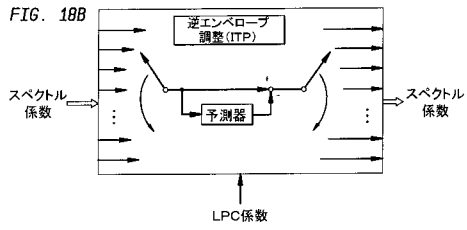
【 図 1 6 】



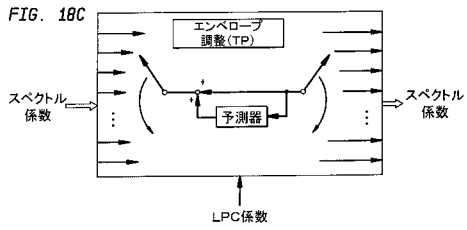
【 図 1 8 A 】



【図 18B】



【図 18C】



## フロントページの続き

- (72)発明者 アラマンヒェ エリック  
アメリカ合衆国 94608 カリフォルニア州 エメリービル 6400 クリスティ アベニ  
ュー アパートメント 3101
- (72)発明者 ディスヒ サッシャ  
ドイツ連邦共和国 90763 フルス トゥルンシュトラッセ 7
- (72)発明者 フォーラー クリストフ  
スイス国 8274 テゲルヴィレン ギュートレイン 1
- (72)発明者 ヘレ ユールゲン  
ドイツ連邦共和国 91054 ブッケンホフ ハーラーシュトラッセ 24

審査官 井出 和水

- (56)参考文献 特開平10-051313(JP,A)  
特表2004-535145(JP,A)  
国際公開第2004/008806(WO,A1)  
国際公開第2004/072956(WO,A1)

## (58)調査した分野(Int.Cl., DB名)

G10L 19/00  
G10L 19/02  
H03M 7/30