

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-528596  
(P2004-528596A)

(43) 公表日 平成16年9月16日(2004.9.16)

(51) Int. Cl. <sup>7</sup>	F I	テーマコード (参考)
G 1 0 G 3/04	G 1 0 G 3/04	5 D 0 8 2
G 1 0 H 1/057	G 1 0 H 1/057	Z 5 D 3 7 8
G 1 0 H 1/40	G 1 0 H 1/40	

審査請求 有 予備審査請求 有 (全 47 頁)

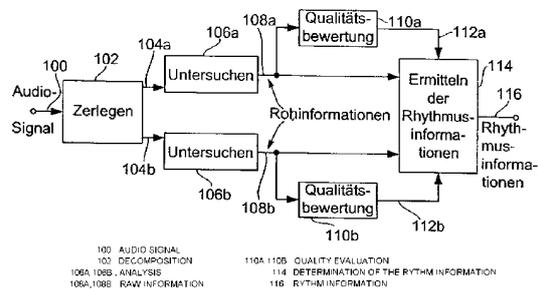
(21) 出願番号	特願2002-590149 (P2002-590149)	(71) 出願人	500341779 フラウンホーファー・ゲゼルシャフト・ツール・フェルデルング・デル・アンゲヴァンテン・フォルシュング・アインゲトラージェネル・フェライン ドイツ連邦共和国, 80636 ミュンヘン, ハンザシュトラッセ 27 ツェー
(86) (22) 出願日	平成14年4月25日 (2002. 4. 25)	(74) 代理人	100091432 弁理士 森下 武一
(85) 翻訳文提出日	平成15年7月29日 (2003. 7. 29)	(72) 発明者	ヘレ、ユルゲン ドイツ国、D-91054 ブッケンホフ、アム アイヘンガルテン 11
(86) 国際出願番号	PCT/EP2002/004618	(72) 発明者	ローデン、ジャン ドイツ国、D-98693 イルメナオ、ヴァイマーラー ストラッセ 55
(87) 国際公開番号	W02002/093557		最終頁に続く
(87) 国際公開日	平成14年11月21日 (2002. 11. 21)		
(31) 優先権主張番号	101 23 366.3		
(32) 優先日	平成13年5月14日 (2001. 5. 14)		
(33) 優先権主張国	ドイツ (DE)		
(81) 指定国	EP (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), JP, US		

(54) 【発明の名称】 リズム情報に関するオーディオ信号を解析するための装置及び方法

(57) 【要約】

オーディオ信号をそのリズム情報に関して解析するための装置であって、オーディオ信号を少なくとも二つの副帯信号に分けるためのフィルタバンク(102)を有している。各副帯信号のリズム生情報を得るために、各副帯信号は副帯信号(104a, 104b)の周期性に関して識別される。各副帯信号の重要度を得るために、リズム生情報は質の評価(110a, 110b)をされる。最終的に、副帯信号の重要度及びリズム生情報(108a, 108b)を考慮して、オーディオ信号のリズム情報が決定される。このような装置はオーディオ信号のより確実な解析を可能にする。リズム情報を決定する際に、重要なリズム情報が存在する副帯信号が、あまり重要でないリズム情報が存在する副帯信号よりも、重要視されるからである。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

オーディオ信号を、そのオーディオ信号のリズム情報に関して解析するための装置であり、以下のものを備えている、

オーディオ信号を少なくとも二つの副帯信号(104a, 104b)に分ける手段(102)、

前記副帯信号のリズム生情報(108a, 108b)を得るために、前記副帯信号の周期性に関して該副帯信号を識別する手段(106a, 106b)、

前記副帯信号の重要度(112a)を得るために、副帯信号(104a)のリズム生情報(108a)の周期性の質に関して評価する手段(110a, 110b)、及び、

前記副帯信号の重要度(112a)と少なくとも一つの副帯信号のリズム生情報(108a, 108b)を考慮して、オーディオ信号のリズム情報を決定する手段(114)。

10

**【請求項 2】**

請求項 1 に記載の装置において、前記識別手段(106a, 106b)は、前記少なくとも二つの副帯信号のそれぞれに対する自動相関関数を算出するためのものである。

**【請求項 3】**

請求項 1 又は請求項 2 に記載の装置において、前記識別手段(106a, 106b)は以下のものを含む、

副帯信号のエンベロープを形成する手段、

平滑なエンベロープを得るために前記副帯信号のエンベロープを平滑化する手段、

微分されたエンベロープを得るために平滑化されたエンベロープを微分する手段、

限定されたエンベロープを得るために微分されたエンベロープを正の値に限定する手段、及び、

リズム生情報(108a, 108b)を得るために、限定されたエンベロープの自動相関関数を作成する手段。

20

**【請求項 4】**

請求項 2 又は請求項 3 に記載の装置において、前記質の評価手段(110a, 110b)は、自動相関関数の最大値の自動相関関数の平均値に対する比を重要度として用いるものである。

**【請求項 5】**

請求項 2 又は請求項 3 に記載の装置において、前記質の評価手段(110a, 110b)は、リズム生情報の算術的平均値のリズム生情報の幾何学的平均値に対する比を重要度として用いるものである。

30

**【請求項 6】**

請求項 4 又は請求項 5 に記載の装置において、前記質の評価手段(110a, 110b)は、最大テンポを得るための最小遅延から最小テンポを得るための最大遅延に至るテンポの範囲内でのみ、自動相関関数を評価するものである。

**【請求項 7】**

請求項 1、請求項 2、請求項 3、請求項 4、請求項 5 又は請求項 6 に記載の装置において、前記決定手段(114)は以下のものを含む、

副帯信号に対する重要度を用いて、副帯信号に対する重み付け係数を引き出す手段(114a)、及び

副帯信号に対する重み付けされたリズム生情報を得るために、該副帯信号の重み付け係数を用いて副帯信号のリズム生情報を重み付けし、オーディオ信号のリズム情報を得るために、重み付けされた該副帯信号のリズム生情報を、重み付けされた又は重み付けされていない他の副帯信号のリズム生情報と統合する手段(114b)。

40

**【請求項 8】**

請求項 7 に記載の装置において、前記重み付け係数を引き出す手段(114a)は、それぞれの副帯信号に対して相関的な重み付け係数を算出するものであり、全ての副帯信号の重み付け係数の合計は 1 となる。

50

## 【請求項 9】

請求項 8 に記載の装置において、前記重み付け係数を引き出す手段 ( 1 1 4 a ) は、重み付け係数を、副帯信号の重要度の全ての副帯信号の重要度の合計に対する比として算出するものである。

## 【請求項 10】

請求項 9 に記載の装置において、前記副帯信号識別手段 ( 1 0 6 a , 1 0 6 b ) は、その長さが 1 0 秒より長い副帯信号を識別する。

## 【請求項 11】

オーディオ信号を、該オーディオ信号のリズム情報に関して解析する方法であり、以下のステップを備えている、

10

オーディオ信号を少なくとも二つの副帯信号 ( 1 0 4 a , 1 0 4 b ) に分ける、

前記副帯信号のリズム生情報 ( 1 0 8 a , 1 0 8 b ) を得るために、前記副帯信号の周期性に関して該副帯信号を識別する ( 1 0 6 a , 1 0 6 b )、

前記副帯信号の重要度 ( 1 1 2 a ) を得るために、副帯信号 ( 1 0 4 a ) のリズム生情報 ( 1 0 8 a ) の周期性の質を評価する ( 1 1 0 a , 1 1 0 b )、及び、

前記副帯信号の重要度 ( 1 1 2 a ) と少なくとも一つの副帯信号のリズム生情報 ( 1 0 8 a , 1 0 8 b ) を考慮して、オーディオ信号のリズム情報を決定する ( 1 1 4 )。

## 【発明の詳細な説明】

## 【0001】

## 【発明の属する技術分野】

20

本発明は、信号処理概念、特にリズム情報に関するオーディオ信号の解析に関する。

## 【0002】

## 【従来の技術及び課題】

近年、オーディオデータやビデオデータのようなマルチメディアデータ資料の入手可能性が著しく進歩してきた。これは、特にインターネット、能率的なコンピュータのハードウェア及びソフトウェア、能率的なデータ圧縮つまりソース符号化方法やオーディオ及びビデオ方法の幅広い有用度に基づく一連の技術的要因による。

## 【0003】

例えば、インターネットにより世界的に入手可能な大容量の音声視覚信号は、内容の基準によってこれらのデータを判断したり分類することが可能になる概念を必要とする。有効な基準を指定することによって、所定の計算方法でマルチメディアを検索し、見つけ出せることが求められている。

30

## 【0004】

この要求は「コンテンツベース」技術と呼ばれ、音声視覚データから、その信号の重要な特性を表すいわゆる特徴点を抜粋する。このような特徴点及びこれらの特徴点の組合せに基づき、オーディオ信号やビデオ信号それぞれの間で、類似のものや共通の特徴点が引き出される。これは、単純に「作品」とも称される異なる信号から抜粋された特徴値をそれぞれ比較及び関連付けることによって実行される。

## 【0005】

信号論理的だけでなく直接意味論的である意義を有する、つまりリスナーによって直接受信された特徴点を表すような特徴点の決定と抜粋のそれぞれが、特に注目されている。

40

## 【0006】

これにより、ユーザーは、オーディオ信号データバンクの一覧にある全てのデータからいくつかの作品を見つけ出すために、簡単で直感的な方法で検索要求を表現することができる。同様に、意味論的に関連する特徴点により、人間の感覚に近い方法で、作品の間の類似関係を類型化することができる。意味論的意義を有する特徴点の使用は、また、ユーザーの好み明らかである場合には、例えばそのユーザーのために作品を自動的に提案することも可能にする。

## 【0007】

音楽解析の分野において、テンポは重要な音楽パラメータであり、これは意味論的意義を

50

有するものである。テンポは、通常、分単位のビート (b p m) で測定される。テンポだけでなく「ビート」の小節ごとの強勢の自動的な抜粋、つまり概してリズム情報の自動的な抜粋は、一つの音楽作品の意味論的に重要な特徴点をとらえるための一つの例である。

【0008】

さらに、特徴点の抜粋、つまり、オーディオ信号からリズム情報を抜粋することは、确实でありコンピュータ的に効率的な方法で行われることが望まれている。确实であるというのは、その作品がソース符号化され、再び復号化されたものかどうか、その作品はスピーカを介して演奏されたりマイクから受信されるものかどうか、大きくあるいは穏やかに演奏されているのかどうか、一つの楽器あるいは複数の楽器で演奏されているのかどうかは関係ないということである。

10

【0009】

小節の強勢及びこれによりテンポを決定するために、つまりリズム情報を決定するために、「ビートトラッキング」という言葉が専門家の間で確立してきた。先行技術より、楽譜のような書き写された信号表記に基づき、つまりミディフォーマットで、ビートトラッキングを行うことが知られている。しかし、このようなメタ表記を必要とするのではなく、例えば、PCM符号化された又は概してデジタルオーディオ信号を直接解析することが目標である。

【0010】

ジャーナル音響社会学アメリカ103:1(1998年1月)588~601ページ、Eric D. Scheirerによる“Tempo and Beat Analysis of Acoustic Musical Signals”に、音楽的な抜粋からリズムパルスを自動的に抜粋する方法が開示されている。入力信号は、フィルターバンクを介して一連の副帯、例えば、200Hz、400Hz、800Hz、1600Hz、3200Hzという遷移周波数を有する六つの副帯に分割される。低域フィルタリングが最初の副帯のために行われる。最後の副帯のためには高域フィルタリングが行われ、他の中間副帯のためには帯域フィルタリングが行われる。各副帯は以下のように処理される。まず、副帯信号を整流する。それとは別にサンプルの絶対値を決定する。そして、エンベロープ信号を得るために、決定した値nを、例えば適当なウィンドウで平均化することによって平滑にする。コンピュータ演算の複雑さを緩和するために、エンベロープ信号を二段抽出することが可能である。エンベロープ信号は微分され、つまり、好ましくは微分フィルタによって信号の振幅の突然の変化にも対応できる。結果は負ではない値に制限される。各エンベロープ信号は共振フィルター、つまり振動子のバンクに格納される。各振動子は各テンポ領域のためのフィルターを備えており、従ってその音楽のテンポに合うフィルターが最も刺激される。入力信号のテンポをそのフィルターに属するテンポに合わせるための手段として、各フィルターの出力信号のエネルギーが算出される。各テンポのためのエネルギーは全ての副帯において合計され、最も大きなエネルギー合計が結果として与えられるテンポ、つまりリズム情報を特徴づける。

20

30

【0011】

この方法の重要な不利点は、特に、多数の振動子を並行して共振させ、最終的にはそのうちのただ一つを選択するということを実現するために行われる膨大な計算量とメモリの複雑さにある。このことが、リアルタイムな応用というような効率的な実行をほとんど不可能にしている。

40

【0012】

Eric D. Scheirerによる“Pulse Tracking with a Pitch Tracker”(音声及び音波信号処理の応用に関するワークショップ1997年会報、モーホック、ニューヨーク、1997年10月)では、前述の振動子概念と別の概念との比較を検討している。別の概念とは、オーディオ信号からの周期性の抜粋、つまり信号のリズム情報抜粋のための、自動相関関数の使用に基づくものである。ビートトラッキングのために、人間のピッチ感覚の変調のためのアルゴリズムが使用される。

50

## 【0013】

図3は、公知のアルゴリズムを示すブロック図である。オーディオ信号は、オーディオ入力300を介して解析フィルターバンク302に送り込まれる。解析フィルターバンクは、そのオーディオ入力からn個のチャンネル、つまり個々の副帯信号を生成する。各副帯信号は、オーディオ信号のある特定の周波数領域を有している。解析フィルターバンクのフィルターは、人間の内耳の選択特性に近くなるように選択される。このような解析フィルターバンクは、トーンフィルターバンクとも称される。

## 【0014】

各副帯のリズム情報は手段304a~304cで評価される。各入力信号に対して、まず、エンベロープ状の出力信号が（耳の中のいわゆる内部繊毛細胞処理に関して）計算され、二段抽出される。この結果から自動相関関数（ACF）が算出され、信号の周期性を遅延の関数として得る。 10

## 【0015】

手段304a~304cの出力側において、自動相関関数が各副帯信号に対して与えられ、これは各副帯信号のリズム情報の様相を表す。

## 【0016】

副帯信号のそれぞれの自動相関関数は、手段306で合計されることで組み合わせられ、総自動相関関数（SACF）を得る。総自動相関関数は、オーディオ入力300で信号のリズム情報を再生する。この情報はテンポ出力308で出力できる。総自動相関関数において高い値は、小節の始まりの高い周期性が総自動相関関数のピークのずれとして現れることを示している。従って、例えば総自動相関関数のピーク値は、音楽的に有用な遅延内において探し出される。 20

## 【0017】

音楽的に有用な遅延とは、例えば、60~200bpmの範囲のテンポである。手段306は、遅延時間をテンポ情報に変換させるために設けられている。従って、1秒間のピークの遅延は、例えば毎分60ビートというテンポに相当する。遅延がそれより小さい場合はテンポがそれよりも速いことを示し、遅延がそれより大きい場合はテンポがそれよりも遅いことを示している。

## 【0018】

いずれの振動子をも高い計算及び記憶能力を必要としないので、この方法は最初に述べた方法に比べて有利である。しかし他方では、結果の質はオーディオ信号のタイプにかなり左右されるという点で、この概念は不利である。例えば、オーディオ信号から主要なリズム楽器の音が聞こえる場合には、図3に示した概念は良好に作用するであろう。しかし、声が主で、明白なリズム情報が与えられない場合には、リズム決定は曖昧なものになってしまう。しかし、オーディオ信号中に、単にリズム情報を含むだけの帯域が存在している可能性がある。例えば、ドラムのハイハットが周波数上位置している高周波帯域、又は大きなドラムが周波数上位置している低周波帯域である。それぞれの情報の組み合わせにより、このような特定の副帯に関するかなり明確な情報は、他の副帯に関する曖昧な情報に重畳され、「減殺」される。 30

## 【0019】

副帯信号の周期性を引き出すために自動相関関数を用いる際の他の問題点は、手段306によって得られる総自動相関関数が曖昧であるということである。出力306における総自動相関関数は、自動相関関数のピーク値は複数の遅延期間を持って形成されるという点で曖昧である。これは、自動相関関数処理される際、 $t_0$ の期間の湾曲成分が、 $t_0$ での望ましい最高値及び複数の遅延期間をおいた時点、即ち、 $2t_0$ 、 $3t_0$ 等の時点での最高値とは無関係に形成されるという事実から納得できる。 40

## 【0020】

TolonenとKarjalainenによる“A Computationally Efficient Multipitch Analysis Model”（スピーチ及びオーディオ処理に関するIEEE転送、第8巻、2000年11月）は、複雑な 50

オーディオ信号の周期性解析のための時間効率的な計算モデルを開示している。この計算モデルでは、信号を二つのチャンネル、つまり、1000Hz未満のチャンネルと1000Hz以上のチャンネルに分ける。これらから、低周波チャンネルの自動相関関数及び高周波チャンネルのエンベロープの自動相関関数が算出される。最終的に、これら二つの自動相関関数が合計される。総自動相関関数の曖昧さを除去するために、この総自動相関関数はさらに処理され、いわゆるエンハンストサマリー自動相関関数(ESACF)が得られる。この総自動相関関数の後処理は、負でない値に制限して、整数の係数で拡張された様々な変形自動相関関数を総自動相関関数から繰り返し引くという処理を含む。

【0021】

本発明の目的は、リズム情報を有するオーディオ信号を解析するための計算時間に関して効率的で確実な装置、及び、効率的で確実な方法を提供することである。 10

【0022】

【発明の構成、作用及び効果】

前記目的は、請求項1に係るオーディオ信号を解析するための装置、又は、請求項11に係るオーディオ信号を解析するための方法によって達成される。

【0023】

本発明は、それぞれの周波数帯域、つまり副帯に、リズムカルな周期性を探し出すための様々な好ましい条件がしばしば存在するという知識に基づいている。例えば、ポップミュージックでは、しばしば1kHz周辺の中央領域では、ビートに呼応しない声が主であり、高周波領域では主にドラムにハイハットのようなパーカッションの音が存在し、これによりリズムの周期性を非常によく抜粋することができる。また、異なる周波数帯域には、オーディオ信号によって異なる量のリズム情報が存在し、これらの周波数帯域はそのオーディオ信号にとって異なる質又は重要性をそれぞれ有している。 20

【0024】

従って、本発明によると、オーディオ信号はまず副帯信号に分けられる。各副帯信号はその周期性に関して識別され、各副帯信号のリズム生情報が得られる。本発明によると、その後、各副帯信号の重要度を得るために、各副帯信号の周期性の質の評価が行われる。高い重要度は、その副帯信号に明白なリズム情報が存在することを示し、低い重要度は、その副帯にはあまり明白でないリズム情報が存在することを示している。

【0025】

本発明の好ましい一実施形態によると、副帯信号をその周期性に関して識別する際、まず、副帯信号の変形エンベロープが計算され、そのエンベロープの自動相関関数が算出される。このエンベロープの自動相関関数はリズムの生情報を示している。自動相関関数が明らかな最高値を示す場合には明白なリズム情報が存在し、一方、その副帯信号のエンベロープの自動相関関数が重要な信号ピーク値を有さないか又は全くピーク値を有さない場合には、明白なリズム情報がほとんど存在しない。従って、明白な信号ピークを有する自動相関関数は重要度が高く、比較的平坦な信号波形を有する自動相関関数は重要度が低い。 30

【0026】

本発明によると、それぞれの副帯信号のそれぞれのリズム生情報は「盲目的に」のみ組み合わせられるのではなく、オーディオ信号のリズム情報を得るための各副帯信号の重要度を考慮して組み合わせられる。ある副帯信号が高い重要度を有している場合、リズム情報を決定する際、これを十分に考慮することが好ましく、一方、ある副帯信号が低い重要度を有している場合、つまり、リズム情報に関して低い質しか有さない場合には、リズム情報を決定する際、これをほとんどあるいは極端な場合には全く考慮しない。 40

【0027】

これは、重要度に基づく重み付け係数により、良好な方法で計算時間的に効率的に実施することができる。リズム情報にとって良質な、つまり高い重要度を有する副帯信号には1の重み付け係数を与えることも可能であり、重要度の低い他の副帯信号には1よりも小さい重み付け係数を与えられる。極端な場合、全く平坦な自動相関関数を有する副帯信号は0の重み付け係数を有することになる。重み付けされた自動相関関数、つまり重み付けさ 50

れた生リズム情報は、その後単純に合計される。全ての副帯信号のうちの一つの副帯信号のみが良質なリズム情報を提供し、他の副帯信号は平坦な信号波形の自動相関関数を有している場合、この重み付けは、極端な場合、その唯一の副帯信号を除く他の全ての副帯信号には重み付け係数0が与えられ、リズム情報決定の際には全く考慮されないことになる。そして、このオーディオ信号のリズム情報は単に一つの副帯信号からのみ決定されることになる。

#### 【0028】

本発明は、リズム情報の確実な決定が可能になることに利点がある。明白でなくさらにそれぞれ異なるリズム情報を有する副帯信号、例えば、声はその作品の実際のビートとは異なるリズムを有している場合、それらがオーディオ信号のリズム情報を弱めたり崩壊させたりしないからである。さらに、非常にノイズに近い副帯信号はシステム自動相関関数に全く平坦な信号波形を与えるが、これらの信号は、リズム情報を決定する際、信号雑音比を低下させない。しかし、先行技術のように、全ての副帯信号の自動相関関数を同じ重みで単に合計すれば、信号雑音比を低下させることになる。

10

#### 【0029】

本発明の方法の他の利点は、わずかな付加的な演算処理で重要度が決定できること、及びその重要度を有するリズム生情報の評価とそれに続く合計が膨大な記憶量や計算時間を費やさずに効率的に行えることであり、従って、本発明は特にリアルタイムな応用に適している。

#### 【0030】

20

#### 【発明の実施の形態】

本発明の好ましい実施形態を添付図面を参照して以下に説明する。

#### 【0031】

図1は、オーディオ信号をリズム情報に関して解析するための装置のブロック図を示す。オーディオ信号は入力100を介して手段102に送られ、少なくとも二つの副帯信号104a, 104bに分けられる。各副帯信号104a, 104bはそれぞれ手段106a, 106bに送られ、その副帯信号中の周期性に関して識別され、それぞれの副帯信号に対してリズム生情報108a, 108bが得られる。このリズム生情報108a, 108bは手段110a, 110bに送られ、その少なくとも二つの副帯信号のそれぞれの周期性の質が評価され、それぞれの副帯信号に対して重要度112a, 112bが与えられる。これらのリズム生情報108a, 108b及び重要度112a, 112bの両方が手段114に送られ、このオーディオ信号のリズム情報を決定する。オーディオ信号を決定する際、手段114は、各副帯信号の重要度112a, 112b及び少なくとも一つの副帯信号のリズム生情報108a, 108bを考慮する。

30

#### 【0032】

例えば、質評価手段110aが副帯信号104a中には特に周期性はないと判断した場合、重要度112aは非常に小さくほぼ0に等しいものとなる。この場合、リズム情報を決定する手段114は、重要度112aが0であるので、オーディオ信号のリズム情報を決定する際には、副帯信号104aのリズム生情報108aを全く考慮する必要はないと判断する。このオーディオ信号のリズム情報は、副帯信号104bのリズム生情報108b

40

#### 【0033】

図2を参照して、図1の装置の具体的な実施形態を以下に説明する。オーディオ信号を分割する手段102として、共通解析フィルタバンクが使用でき、この場合、出力側に与えられる副帯信号の数はユーザーにより選択可能となる。各副帯信号はそれぞれ手段106a, 106b, 106cの処理を受け、各リズム生情報の重要度が手段110a~110cにより決定される。図2に示す好ましい実施形態において、手段114は、各副帯信号の重み付け係数をその副帯信号の重要度及び任意に他の副帯信号の重要度に基づき算出する手段114aを含んでいる。そして、手段114bにおいて、リズム生情報108a~108cの重み付けがその副帯信号の重み付け係数をもって行われる。その後、手段1

50

1 4 bでは、重み付けされたリズム生情報が例えば合計されて組み合わせられ、テンポ出力 1 1 6でオーディオ信号のリズム情報が得られる。

【0034】

従って、本発明の概念は以下のようなものである。例えば、エンベロープ作成、平滑化、微分、正の値への限定及び自動相関関数の作成(手段106a~106c)を行うことによってそれぞれの帯域のリズム情報を評価した後、これらの中間結果の重要度及び質の評価が手段110a~110cにおいて行われる。これは、それぞれの結果の信用性を重要度をもって評価する評価関数の助けを借りて行われる。リズム情報の抜粋のために、重み付け係数が各帯域の副帯信号全ての重要度から引き出される。そして、それぞれの帯域幅の結果をそれぞれの重み付け係数を考慮して組み合わせることにより、リズム抜粋の最終的な結果が手段114bで得られる。

10

【0035】

結果的に、このように実行されたリズム解析のアルゴリズムは、不都合な条件下にあっても、信号内のリズム情報を確実に見つけ出すことができる。従って、本発明の概念は高い確実性によって区別される。

【0036】

好ましい実施形態において、各副帯信号の周期性を示すリズム生情報108a, 108b, 108cは自動相関関数によって決定される。この場合、自動相関関数の最大値をその自動相関関数の平均で割り、そして1を引くことにより、重要度を決定することが好ましい。ここで明記すべき点は、全ての自動相関関数は常に遅延0で信号のエネルギーを示す局部的最大値を与えるということである。質の決定を台無しにしないように、この最大値は考慮すべきではない。

20

【0037】

さらに、自動相関関数はある特定のテンポの範囲内でのみ、例えば、最大遅延つまり関連する最小テンポから最小遅延つまり関連する最大テンポまででのみ、考慮すべきである。典型的なテンポの範囲は60~200bpmである。

【0038】

あるいは、関連するテンポ範囲内の自動相関関数の算術的平均とそのテンポ範囲内の自動相関関数の幾何学的平均との間の関係を重要度として決定することもできる。周知のように、自動相関関数の全ての値が等しい場合、つまり、自動相関関数が平坦な信号波形を有する場合、自動相関関数の幾何学的平均と自動相関関数の算術的平均は等しい。この場合、重要度は1に等しい値であり、これはそのリズム生情報は重要でないことを意味している。

30

【0039】

強いピークを有するシステム自動相関関数の場合、算術的平均の幾何学的平均に対する比は1以上となり、これはこの自動相関関数が好ましいリズム情報を有していることを意味している。算術的平均と幾何学的平均の間の比が小さければ小さいほど、その自動相関関数は平坦で、低い周期性であり、その副帯信号のリズム情報はさほど重要でない、つまり低い又は0の重み付け係数で表されるような質の低いものであるということの意味している。

40

【0040】

重み付け係数に関して、いくつかの可能性が存在する。比較重み付けが好ましく、全ての副帯信号の重み付け係数の合計が1になるようにする。例えば、ある帯域の重み付け係数は、全ての重要度の値の合計で割られた重要度の値として決定される。この場合、比較重み付けは、オーディオ信号のリズム情報を得るための重み付けされたリズム生情報の合計の前に行われる。

【0041】

前述したように、自動相関関数を用いてリズム情報の評価を行うことが好ましい。図4は、この場合の説明図である。オーディオ信号は、オーディオ信号を副帯信号104a, 104bに分割する手段102に、オーディオ信号入力100を介して送り込まれる。前述

50

したように、各副帯信号は、自動相関関数を使用して、それぞれ手段106a, 106bにおいて識別される。そして、リズム生情報108a, 108bが手段106a, 106bの出力側に与えられる。これらは手段118a, 118bにそれぞれ送られ、手段116aで自動相関関数を介してリズム生情報の後処理が行われる。これにより、自動相関関数の曖昧さ、つまり信号ピークが整数回の遅延をもって起こることが副帯ごとに必ず除去され、後処理されたリズム生情報120a, 120bがそれぞれ得られる。

#### 【0042】

従来技術のように、それぞれの自動相関関数の合計の後に曖昧さが除去されるだけでなく、自動相関関数、つまりリズム生情報108a, 108bの曖昧さは、すでに副帯ごとに除去されているという利点を有している。その上、手段118a, 118bにより自動相関関数の曖昧さを各帯域に関して除去することで、副帯信号のリズム生情報を互いに独立して扱うことが可能になる。例えば、リズム生情報108aの手段110aを介して又はリズム生情報108bの手段110bを介して質の評価を行うことが可能になる。

10

#### 【0043】

図4に点線で示されているように、質の評価は後処理されたリズム生情報に関するものを行うことができ、この最後の可能性が好ましい。後処理されたリズム生情報に基づき質の評価を行うことは、もはや曖昧でない情報の質を評価することになるからである。

#### 【0044】

手段114によるリズム情報の決定は、一つのチャンネルの後処理されたリズム情報に基づき行われ、好ましくは、このチャンネルの重要度にも基づいて行われる。

20

#### 【0045】

質の評価が手段118aの前段の信号を意味するリズム生情報に基づき行われる場合、重要度が0である、即ち、自動相関関数が平坦な信号波形であると判断されると、計算時間を節約するために手段118aを介した後処理は全く削除され得るといような点で有利である。

#### 【0046】

図5を参照して、リズム生情報の後処理を行う手段118a, 118bの構造をより詳細に以下に説明する。まず、104aのような副帯信号が、自動相関関数を介して副帯信号の周期性を識別する手段106aに送り込まれ、リズム生情報108aが得られる。曖昧さを副帯ごとに除去するために、先行技術にあるように、手段121によって拡張自動相関関数を算出することも可能である。その際、自動相関関数が整数回の遅延をもって拡張される拡張自動相関関数を算出するために、手段122が設置されている。この場合、手段122は、この拡張自動相関関数を元の自動相関関数、つまりリズム生情報108aから引くために設置されている。特に、自動相関関数をまず2倍のサイズに拡張し、そしてそれをリズム生情報108aから引くことが好ましい。次のステップでは、係数3で拡張された自動相関関数が手段121で算出され、前段階の引き算の結果から再びそれを引く。このようにして、リズム生情報から、全ての曖昧さが徐々に除去されていく。

30

#### 【0047】

あるいは又は付加的に、整数の係数によって作り出される、例えば1よりも小さい係数で拡張された自動相関関数を計算するために、手段121を設けてもよい。これは、遅延時間 $t_0/2$ ,  $t_0/3$ 等の部分をも形成するために、手段122によってリズム生情報に付加される。

40

#### 【0048】

さらに、拡張され、作成されたリズム生情報の変形108aに対して、確実性という意味でここで融通性を得るために、加算及び減算の前に重み付けをしてもよい。

#### 【0049】

自動相関関数に基づく副帯信号の周期性を識別する方法によって、自動相関関数の特性が取り込まれ、手段118a, 118bによって後処理が行われる際、さらなる向上が可能となる。このように、距離 $t_0$ を有する一連の周期的な小節の始まりは、遅延時間 $t_0$ だけでなく $2t_0$ ,  $3t_0$ でも自動相関関数のピークを形成する。このことは、テンポ

50

検出、つまり自動相関関数の重要な最大値を探し出す際に曖昧さをもたらすことになる。整数の係数によって拡張された自動相関関数の変形が出力値から副帯ごとに（重み付けされて）引かれることにより、この曖昧さが除去される。

【0050】

さらに自動相関関数に関して、「基本のテンポ」の2倍又は3倍を意味する  $t_0/2$  ,  $t_0/3$  等で、何の情報も与えられないという問題がある。特に、異なる副帯に存在する二つの楽器がその信号のリズムを共に決めている場合には、間違った結果につながるおそれがある。この問題は整数の係数によって作成された自動相関関数の変形が算出され、重み付けされるか又は重み付けされないでリズム生情報に付加されるという事実によって考慮される。

10

【0051】

このように、ACF後処理は副帯ごとに行われ、これにより自動相関関数が少なくとも一つの副帯信号に対して算出され、そしてこの関数の拡大変形と組み合わせられる。

【図面の簡単な説明】

【図1】

リズム生情報の質的評価に関してオーディオ信号を解析するための装置のブロック図である。

【図2】

重要度に基づく重み付け係数を使用して、オーディオ信号を解析する装置のブロック図である。

20

【図3】

リズム情報に関してオーディオ情報を解析する従来の装置のブロック図である。

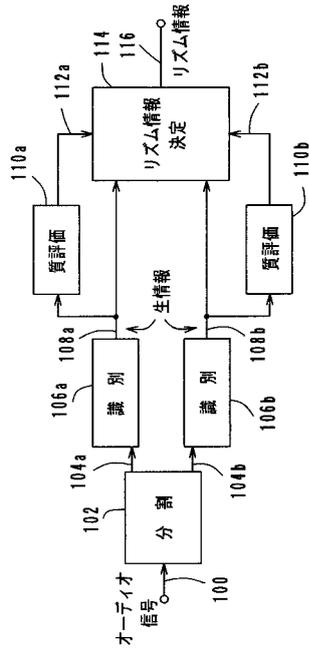
【図4】

リズム生情報の副帯ごとの後処理をした自動相関関数を使用して、オーディオ信号をリズム情報に関して解析する装置のブロック図である。

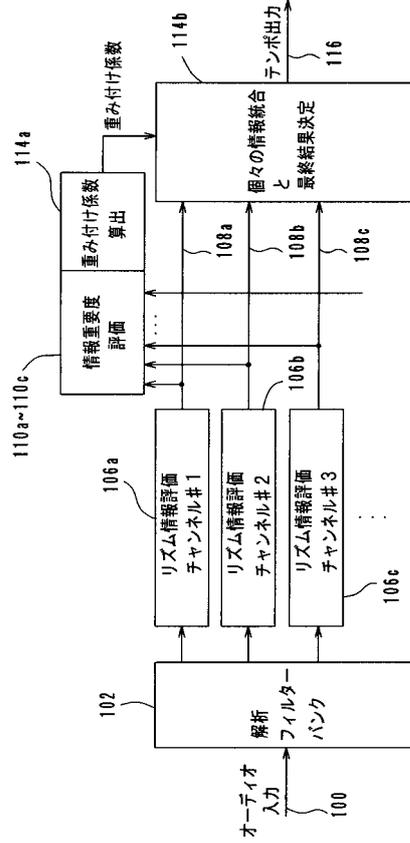
【図5】

図4の後処理手段の詳細なブロック図である。

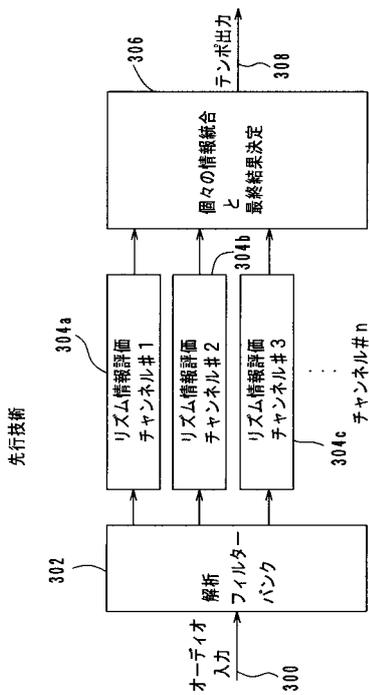
【図 1】



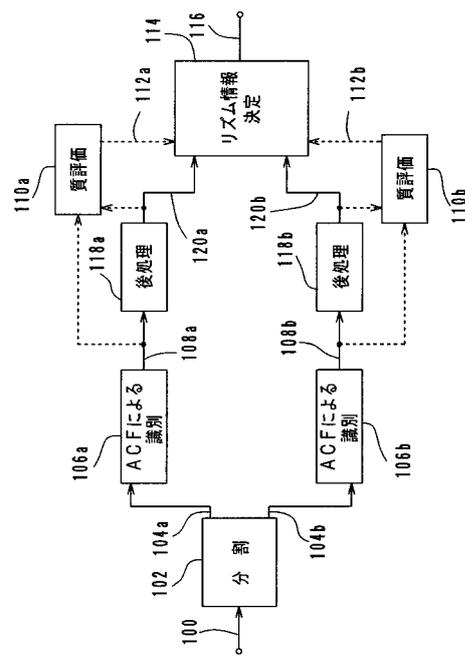
【図 2】



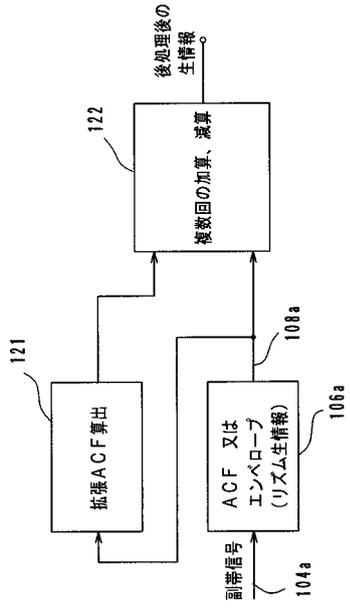
【図 3】



【図 4】



【 図 5 】



【国際公開パンフレット】

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
21. November 2002 (21.11.2002)

PCT

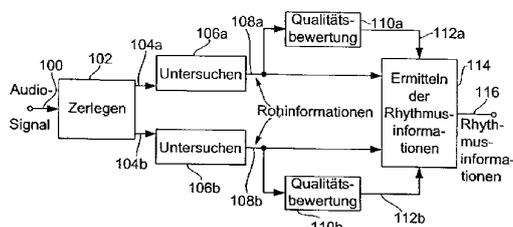
(10) Internationale Veröffentlichungsnummer  
WO 02/093557 A1

- (51) Internationale Patentklassifikation: G10L 19/02 ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG [DE/DE]; Leonrodstrasse 54, 80636 München (DE).
- (21) Internationales Aktenzeichen: PCT/EP02/04618
- (22) Internationales Anmeldedatum: 25. April 2002 (25.04.2002) (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HERRE, Jürgen [DE/DE]; Am Lischengarten 11, 91054 Buckenhof (DE); ROHDEN, Jan [DE/DE]; Weimarer Strasse 55, 98693 Ilmenau (DE); UHLE, Christian [DE/DE]; Schwanzitzstrasse 3, 98693 Ilmenau (DE); CREMER, Markus [DE/DE]; Herderstrasse 4, 98693 Ilmenau (DE).
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität: 101 23 366.3 14. Mai 2001 (14.05.2001) DE (74) Anwälte: SCHOPPE, Fritz usw.; SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER, POSTFACH 71 08 67, 81458 München (DE).
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): FRAUNHOFER-GESELLSCHAFT (81) Bestimmungsstaaten (national): JP, US.

[Fortsetzung auf der nächsten Seite]

(54) Title: DEVICE AND METHOD FOR ANALYSING AN AUDIO SIGNAL IN VIEW OF OBTAINING RHYTHM INFORMATION

(54) Bezeichnung: VORRICHTUNG UND VERFAHREN ZUM ANALYSIEREN EINES AUDIOSIGNALS HINSICHTLICH VON RHYTHMUSINFORMATIONEN



100...AUDIO SIGNAL  
 102...DECOMPOSITION  
 106A,106B...ANALYSIS  
 106A,106B...RAW INFORMATION  
 110A,110B...QUALITY EVALUATION  
 114...DETERMINATION OF THE RHYTHM INFORMATION  
 116...RHYTHM INFORMATION

(57) Abstract: The invention relates to a device for analysing an audio signal in view of obtaining rhythm information of said audio signal. Said device comprises a filter bank for decomposing (102) the audio signal into at least two partial band signals. Each partial band signal is analysed (106a, 106b) in terms of its periodicity, in order to obtain raw rhythm information for each partial band signal. Raw rhythm information is subjected to a quality evaluation (110a, 110b), in order to obtain a significance measure for each partial band signal. The rhythm information of the audio signal is determined, taking into account the significance measure of the partial band signal and the raw rhythm information (114). This enables a more robust analysis of the audio signal, as partial band signals containing clear rhythm information are preferred to partial band signals containing less clear rhythm information, for the determination of rhythm information.

[Fortsetzung auf der nächsten Seite]



WO 02/093557 A1

WO 02/093557 A1 

(84) **Bestimmungsstaaten (regional):** europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR). *Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.*

**Veröffentlicht:**

— mit internationalem Recherchenbericht  
vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen einreffen

(57) **Zusammenfassung:** Eine Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen des Audiosignals umfassend eine Filterbank zum Zerlegen (102) des Audiosignals in zumindest zwei Teilbandsignale. Jedes Teilbandsignal wird hinsichtlich einer Periodizität des Teilbandsignals untersucht (106a, 106b), um Rhythmus-Rohinformationen jedes Teilbandsignals zu erhalten. Die Rhythmus-Rohinformationen werden einer Qualitätsbeurteilung (110a, 110b) unterzogen, um ein Signifikanzmaß für jedes Teilbandsignal zu erhalten. Die Rhythmusinformationen des Audiosignals werden schließlich unter Berücksichtigung des Signifikanzmaßes des Teilbandsignals und der Rhythmus-Rohinformationen ermittelt (114). Dies ermöglicht eine robustere Analyse des Audiosignals, da Teilbandsignale, in denen deutliche Rhythmusinformationen vorliegen, gegenüber Teilbandsignalen, in denen weniger deutliche Rhythmusinformationen vorliegen, beim Ermitteln der Rhythmusinformationen bevorzugt werden.

WO 02/093557

PCT/EP02/04618

**Vorrichtung und Verfahren zum Analysieren eines Audiosignals  
hinsichtlich von Rhythmusinformationen**

Beschreibung

Die vorliegende Erfindung bezieht sich auf Signalverarbeitungskonzepte und insbesondere auf die Analyse von Audiosignalen hinsichtlich von Rhythmusinformationen.

In den letzten Jahren ist die Verfügbarkeit multimedialer Datenmaterials, wie z. B. Audio- oder Video-Daten, stark gestiegen. Dies ist auf eine Reihe von technischen Faktoren zurückzuführen, welche sich insbesondere auf die breite Verfügbarkeit des Internets, von leistungsfähiger Rechnerhardware und Software sowie von leistungsfähigen Verfahren zur Datenkompression, d. h. Quellcodierung, von Audio- und Videoverfahren stützen.

Die riesigen Mengen audiovisueller Daten, die beispielsweise auf dem Internet weltweit verfügbar sind, verlangen nach Konzepten, die es ermöglichen, diese Daten nach inhaltlichen Kriterien beurteilen, katalogisieren usw. zu können. Es besteht der Wunsch, in der Lage zu sein, multimediale Daten gezielt durch Angabe sinnvoller Kriterien zu suchen und aufzufinden.

Dies erfordert sogenannte „inhaltsbasierter“ Techniken, die aus den audiovisuellen Daten sogenannte Merkmale, die in Fachkreisen auch als „Features“ bezeichnet werden, extrahieren, welche wichtige charakteristische Eigenschaften des Signals repräsentieren. Basierend auf solchen Merkmalen bzw. Kombinationen dieser Merkmale können Ähnlichkeitsbeziehungen bzw. Gemeinsamkeiten zwischen Audio- oder Videosignalen hergeleitet werden. Dieser Vorgang erfolgt durch Vergleich bzw. In-

Beziehung-Setzen der extrahierten Merkmalswerte aus den verschiedenen Signalen, welche auch einfach als „Stücke“ bezeichnet werden.

Von besonderem Interesse ist die Bestimmung bzw. Extraktion von Merkmalen, die nicht nur signaltheoretische, sondern möglichst unmittelbare semantische Bedeutung haben, d. h. vom Hörer unmittelbar empfundene Eigenschaften repräsentieren.

Dies ermöglicht es dem Benutzer, auf einfache und intuitive Weise Suchanfragen zu formulieren, um Stücke aus dem gesamten vorhandenen Datenbestand einer Audiosignal-Datenbank zu finden. Ebenso erlauben es semantisch relevante Merkmale, Ähnlichkeitsbeziehungen zwischen Stücken zu modellieren, die der menschlichen Empfindung nahe kommen. Die Verwendung von Merkmalen, die semantische Bedeutung haben, ermöglicht auch beispielsweise ein automatisches Vorschlagen von für einen bestimmten Benutzer interessanten Stücken, wenn seine Vorlieben bekannt sind.

Im Bereich der Musikanalyse ist das Tempo ein wichtiger musikalischer Parameter, der semantische Bedeutung hat. Das Tempo wird üblicherweise in „Beats per Minute“ (BPM) gemessen. Die automatische Extraktion des Tempos sowie der Taktschwerpunkte des „Beats“ bzw. allgemein gesagt die automatische Extraktion von Rhythmusinformationen, ist ein Beispiel für die Gewinnung eines semantisch wichtigen Merkmals eines Musikstückes.

Weiterhin besteht ein Wunsch dahingehend, dass die Merkmalsextraktion, d. h. das Extrahieren von Rhythmusinformationen aus einem Audiosignal, robust und recheneffizient stattfinden kann. Robustheit bedeutet, dass es keine Rolle spielen darf, ob das Stück quellcodiert und wieder decodiert worden ist, ob das Stück über einen Lautsprecher abgespielt und von einem Mi-

Mikrofon empfangen worden ist, ob es laut oder leise gespielt wird, oder ob es von einem Instrument oder einer Mehrzahl von Instrumenten gespielt wird.

Für die Bestimmung der Taktschwerpunkte und damit auch des Tempos, d. h. für die Bestimmung von Rhythmusinformationen, hat sich in den Fachkreisen auch der Begriff „Beat Tracking“ etabliert. Aus dem Stand der Technik ist es bereits bekannt, ein Beat Tracking auf der Basis einer notenartigen bzw. transkribierten Signaldarstellung, z. B. im Midi-Format, durchzuführen. Ziel ist es jedoch, keine solchen Metadarstellungen zu benötigen, sondern eine Analyse direkt mit einem z. B. PCM-codierten oder allgemein gesagt digital vorliegenden Audiosignal vorzunehmen.

Die Fachveröffentlichung „Tempo and Beat Analysis of Acoustic Musical Signals“ von Eric D. Scheirer, J. Acoust. Soc. Am. 103:1, (Jan 1998), Seiten 588 - 601, offenbart ein Verfahren zur automatischen Extraktion eines rhythmischen Pulses aus musikalischen Exzerpten. Das Eingangssignal wird mittels einer Filterbank in eine Reihe von Teilbändern aufgespalten, beispielsweise in 6 Teilbänder mit Übergangsfrequenzen von 200 Hz, 400 Hz, 800 Hz, 1600 Hz und 3200 Hz. Für das erste Teilband wird eine Tiefpassfilterung durchgeführt. Für das letzte Teilband wird eine Hochpassfilterung durchgeführt, für die restlichen, dazwischen liegenden Teilbänder wird eine Bandpassfilterung beschrieben. Jedes Teilband wird folgendermaßen verarbeitet. Das Teilbandsignal wird zunächst gleichgerichtet. In anderen Worten ausgedrückt wird der Absolutbetrag der Abtastwerte bestimmt. Die resultierenden n Werte werden dann geglättet, beispielsweise mit einer Mittelung über ein geeignetes Fenster, um ein Hüllkurvensignal zu erhalten. Zur Senkung der Rechenkomplexität kann das Hüllkurvensignal unterabtastet werden. Die Hüllkurvensignale werden differenziert, d. h.

WO 02/093557

PCT/EP02/04618

4

plötzliche Änderungen der Signalamplitude werden durch das Differenzierungsfilter bevorzugt weitergeleitet. Das Resultat wird dann auf nicht-negative Werte begrenzt. Jedes Hüllkurvensignal wird dann in eine Bank resonanter Filter, d. h. Oszillatoren, gegeben, die jeweils ein Filter für jeden Tempobereich enthalten, so dass das zum musikalischen Tempo passende Filter am stärksten angeregt wird. Für jedes Filter wird die Energie des Ausgangssignals als Maß für die Übereinstimmung des Tempos des Eingangssignals mit dem zum Filter gehörigen Tempo berechnet. Die Energien für jedes Tempo werden schließlich über alle Teilbänder aufsummiert, wobei die größte Energiesumme das als Resultat gelieferte Tempo, d. h. die Rhythmusinformationen, kennzeichnet.

Ein wesentlicher Nachteil dieses Verfahrens besteht in der großen Rechen- und Speicherkomplexität insbesondere zur Realisierung der großen Zahl von parallelschwingenden „Oszillatoren“, von denen letztlich nur ein einziger ausgewählt wird. Dies macht eine effiziente Implementierung beispielsweise für Echtzeitanwendungen nahezu unmöglich.

Die Fachveröffentlichung „Pulse Tracking with a Pitch Tracker“ von Eric D. Scheirer, Proc. 1997 Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, NY, Oct 1997, beschreibt einen Vergleich des oben beschriebenen „Oszillatorkonzeptes“ mit einem Alternativkonzept, das auf der Verwendung von Autokorrelationsfunktionen zur Extraktion der Periodizität aus einem Audiosignal, d. h. der Rhythmusinformationen eines Signals, aufbaut. Ein Algorithmus zur Modulierung der menschlichen Tonhöhenwahrnehmung, d. h. des Pitch, wird für das „Beat Tracking“ eingesetzt.

Der bekannte Algorithmus ist in Fig. 3 als Blockdiagramm dargestellt. Das Audiosignal wird über einen Audioeingang 300 ei-

ner Analyse-Filterbank 302 zugeführt. Die Analyse-Filterbank erzeugt aus dem Audioeingang eine Anzahl  $n$  von Kanälen, d. h. von einzelnen Teilbandsignalen. Jedes Teilbandsignal enthält einen bestimmten Bereich an Frequenzen des Audiosignals. Die Filter der Analyse-Filterbank sind so ausgewählt, dass sie die Selektionscharakteristik des menschlichen Innenohrs annähern. Eine solche Analyse-Filterbank wird auch als Gamma-Ton-Filterbank bezeichnet.

In den Einrichtungen 304a bis 304c werden die Rhythmus-Informationen jedes Teilbandsignals ausgewertet. Für jedes Eingangssignal wird zunächst ein hüllkurvenartiges Ausgangssignal berechnet (entsprechend einer sogenannten „Inner Hair Cell“-Verarbeitung im Ohr) und unterabtastet. Aus diesem Resultat wird eine Autokorrelationsfunktion (AKF) berechnet, um die Periodizität des Signals als Funktion der Verzögerung, d. h. des „Lag“ zu erhalten.

Am Ausgang der Einrichtungen 304a bis 304c liegt dann für jedes Teilbandsignal eine Autokorrelationsfunktion vor, welche Aspekte der Rhythmus-Informationen jedes Teilbandsignals darstellt.

Die einzelnen Autokorrelationsfunktionen der Teilbandsignale werden dann in einer Einrichtung 306 durch Summation kombiniert, um eine Summenautokorrelationsfunktion (SAKF) zu erhalten, welche die Rhythmus-Informationen des Signals am Audio-Eingang 300 wiedergibt. Diese Informationen können an einem Tempo-Ausgang 308 ausgegeben werden. Große Werte in der Summenautokorrelation zeigen an, dass für eine einer Spitze der SAKF zugeordneten Verzögerung (Lag) eine hohe Periodizität der Notenanfänge vorliegt. Daher wird beispielsweise der größte Wert der Summenautokorrelationsfunktion innerhalb der musikalisch sinnvollen Verzögerungen gesucht.

Musikalisch sinnvolle Verzögerungen sind beispielsweise der Tempobereich zwischen 60 bpm und 200 bpm. Die Einrichtung 306 kann ferner angeordnet sein, um eine Verzögerungszeit in Tempoinformationen umzusetzen. So entspricht beispielsweise eine Spitze einer Verzögerung von einer Sekunde einem Tempo von 60 Schlägen pro Minute. Kleinere Verzögerungen deuten auf höhere Tempos hin, während größere Verzögerungen auf kleinere Tempos als 60 bpm hinweisen.

Dieses Verfahren hat gegenüber dem zuerst genannten Verfahren einen Vorteil dahingehend, dass keine Oszillatoren mit großem Rechen- und Speicheraufwand implementiert werden müssen. Andererseits ist das Konzept dahingehend nachteilhaft, dass die Qualität der Ergebnisse sehr stark von der Art des Audiosignals abhängt. Ist aus einem Audiosignal beispielsweise ein dominantes Rhythmusinstrument herauszuhören, so wird das in Fig. 3 beschriebene Konzept gut funktionieren. Ist dagegen die Stimme dominant, welche keine besonders eindeutigen Rhythmusinformationen liefern wird, so wird die Rhythmusbestimmung mehrdeutig sein. In dem Audiosignal könnte durchaus auch ein Band vorkommen, das lediglich Rhythmusinformationen enthält, wie z. B. ein höheres Frequenzband, in dem beispielsweise ein Hi-hat eines Schlagzeugs positioniert ist, oder ein niedriges Frequenzband, in dem die große Trommel eines Schlagzeugs auf der Frequenzskala positioniert ist. Aufgrund der Kombination der Einzelinformationen werden jedoch die einigermaßen eindeutigen Informationen dieser speziellen Teilbänder von den mehrdeutigen Informationen der anderen Teilbänder überlagert bzw. „verwässert“.

Ein weiteres Problem bei der Verwendung von Autokorrelationsfunktionen zum Extrahieren der Periodizität eines Teilbandsignals besteht darin, dass die Summen-Autokorrelationsfunktion,

WO 02/093557

7

PCT/EP02/04618

die durch die Einrichtung 306 gewonnen wird, mehrdeutig ist. Die Summen-Autokorrelationsfunktion am Ausgang 306 ist dahingehend mehrdeutig, dass auch beim Vielfachen einer Verzögerung eine Autokorrelationsfunktions-Spitze erzeugt wird. Dies ist daraus verständlich, dass eine Sinuskomponente mit einer Periode von  $t_0$ , wenn sie einer Autokorrelationsfunktionverarbeitung unterzogen wird, neben dem gewünschten Maximum bei  $t_0$  auch Maxima bei Vielfachen der Verzögerungen, d. h. bei  $2t_0$ ,  $3t_0$ , usw. erzeugt.

Die Fachveröffentlichung „A Computationally Efficient Multipitch Analysis Model“, von Tolonen und Karjalainen, IEEE Transactions on Speech and Audio Processing, Band 8, Nr. 6, Nov. 2000, offenbart ein rechenzeiteffizientes Modell für eine Periodizitätsanalyse von komplexen Audiosignalen. Das Rechenmodell teilt das Signal in zwei Kanäle auf, und zwar in einen Kanal unter 1000 Hz und einen Kanal über 1000 Hz. Hieraus wird eine Autokorrelation des unteren Kanals und eine Autokorrelation der Hüllkurve des oberen Kanals berechnet. Schließlich werden die beiden Autokorrelationsfunktionen summiert. Um die Mehrdeutigkeiten der Summenautokorrelationsfunktion zu eliminieren, wird die Summenautokorrelationsfunktion weiterverarbeitet, um eine sogenannte Enhanced Summary Autocorrelation Function (ESACF) (Weiterentwickelte Summenautokorrelationsfunktion) zu erhalten. Diese Nachbearbeitung der Summenautokorrelationsfunktion beinhaltet ein wiederholtes Abziehen von mit ganzzahligen Faktoren gespreizten Versionen der Autokorrelationsfunktion von der Summenautokorrelationsfunktion mit nachfolgender Begrenzung auf nicht-negative Werte.

Die Aufgabe der vorliegenden Erfindung besteht darin, eine rechenzeiteffiziente und robuste Vorrichtung und ein rechenzeiteffizientes und robustes Verfahren zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen zu schaffen.

Diese Aufgabe wird durch eine Vorrichtung zum Analysieren eines Audiosignals nach Patentanspruch 1 oder durch ein Verfahren zum Analysieren eines Audiosignals nach Patentanspruch 11 gelöst.

Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, dass in den einzelnen Frequenzbändern, d. h. den Teilbändern, oftmals unterschiedlich günstige Bedingungen für das Auffinden von rhythmischen Periodizitäten herrschen. Während beispielsweise bei Popmusik oftmals im Bereich der Mitte, beispielsweise um 1 kHz, das Signal von nicht mit dem Beat korrespondierenden Gesang dominiert wird, sind in den höheren Frequenzbereichen oftmals vor allem Perkussionsklänge präsent, wie z. B. das Hihat des Schlagzeugs, welche eine sehr gute Extraktion rhythmischer Regelmäßigkeiten erlauben. Anders ausgedrückt beinhalten unterschiedliche Frequenzbänder je nach Audiosignal eine unterschiedliche Menge an rhythmischer Information bzw. haben eine unterschiedliche Qualität oder Signifikanz für die Rhythmusinformationen des Audiosignals.

Erfindungsgemäß wird daher das Audiosignal zunächst in Teilbandsignale zerlegt. Jedes Teilbandsignal wird hinsichtlich seiner Periodizität untersucht, um Rhythmus-Rohinformationen für jedes Teilbandsignal zu erhalten. Hierauf wird gemäß der vorliegenden Erfindung eine Bewertung der Qualität der Periodizität jedes Teilbandsignals durchgeführt, um ein Signifikanzmaß für jedes Teilbandsignal zu erhalten. Ein hohes Signifikanzmaß deutet darauf hin, dass in diesem Teilbandsignal deutliche Rhythmusinformationen vorliegen, während ein niedriges Signifikanzmaß darauf hindeutet, dass in diesem Teilbandsignal weniger eindeutige Rhythmusinformationen vorliegen.

Gemäß einem bevorzugten Ausführungsbeispiel der vorliegenden Erfindung wird bei der Untersuchung eines Teilbandsignals hinsichtlich seiner Periodizitäten zunächst eine modifizierte Hüllkurve des Teilbandsignals berechnet und dann eine Autokorrelationsfunktion der Hüllkurve berechnet. Die Autokorrelationsfunktion der Hüllkurve stellt die Rhythmus-Rohinformationen dar. Eindeutige Rhythmusinformationen liegen vor, wenn die Autokorrelationsfunktion deutliche Maxima aufweist, während weniger eindeutige Rhythmusinformationen vorliegen, wenn die Autokorrelationsfunktion der Hüllkurve des Teilbandsignals weniger ausgeprägte Signalspitzen oder überhaupt keine Signalspitzen hat. Eine Autokorrelationsfunktion, die deutliche Signalspitzen hat, wird daher ein hohes Signifikanzmaß erhalten, während eine Autokorrelationsfunktion, die einen relativ flachen Verlauf hat, ein niedriges Signifikanzmaß erhalten wird.

Erfindungsgemäß werden die einzelnen Rhythmus-Rohinformationen der einzelnen Teilbandsignale nicht einfach „blind“ kombiniert, sondern unter Berücksichtigung des Signifikanzmaßes für jedes Teilbandsignal verwendet, um die Rhythmusinformationen des Audiosignals zu erhalten. Hat ein Teilbandsignal ein hohes Signifikanzmaß, so wird es bei der Ermittlung der Rhythmusinformationen bevorzugt, während ein Teilbandsignal, das ein niedriges Signifikanzmaß hat, d. h. das eine niedrige Qualität hinsichtlich der Rhythmusinformationen aufweist, bei der Ermittlung der Rhythmusinformationen des Audiosignals kaum oder im Extremfall überhaupt nicht berücksichtigt wird.

Dies kann rechenzeiteffizient gut durch einen Gewichtungsfaktor implementiert werden, der von dem Signifikanzmaß abhängt. Während ein Teilbandsignal, das eine gute Qualität für die Rhythmusinformationen hat, d. h. das ein hohes Signifikanzmaß hat, einen Gewichtungsfaktor von 1 erhalten könnte, wird ein anderes Teilbandsignal, das ein kleineres Signifikanzmaß hat,

einen Gewichtungsfaktor kleiner als 1 erhalten. Im Extremfall wird ein Teilbandsignal, das eine vollkommen flache Autokorrelationsfunktion hat, einen Gewichtungsfaktor von 0 haben. Die gewichteten Autokorrelationsfunktionen, d. h. die gewichteten Rhythmus-Rohinformationen werden dann einfach aufsummiert. Wenn lediglich ein Teilbandsignal sämtlicher Teilbandsignale gute Rhythmusinformationen liefert, während die anderen Teilbandsignale Autokorrelationsfunktionen mit einem flachen Verlauf aufweisen, kann diese Gewichtung im Extremfall dazu führen, dass alle Teilbandsignale außer dem einen Teilbandsignal einen Gewichtungsfaktor von 0 erhalten, d. h. bei der Ermittlung der Rhythmusinformationen überhaupt nicht berücksichtigt werden, so dass die Rhythmusinformationen des Audiosignals lediglich aus einem einzigen Teilbandsignal ermittelt werden.

Das erfindungsgemäße Konzept ist dahingehend vorteilhaft, dass es eine robuste Bestimmung der Rhythmusinformationen ermöglicht, da Teilbandsignale mit keinen eindeutigen bzw. sogar abweichenden Rhythmusinformationen, d. h. wenn der Gesang einen anderen Rhythmus aufweist als der eigentliche Beat des Stücks, die Rhythmusinformationen des Audiosignals nicht „verwässern“ bzw. „verfälschen“. Darüber hinaus werden sehr rauschartige Teilbandsignale, welche eine Systemautokorrelationsfunktion mit vollständig flachem Verlauf liefern, das Signal/Rausch-Verhältnis bei der Bestimmung der Rhythmusinformationen nicht verschlechtern. Genau dies würde jedoch auftreten, wenn, wie im Stand der Technik, einfach alle Autokorrelationsfunktionen der Teilbandsignale mit gleichem Gewicht aufsummiert werden.

Ein weiterer Vorteil des erfindungsgemäßen Verfahrens besteht darin, dass mit einem kleinen zusätzlichen Rechenaufwand ein Signifikanzmaß bestimmt werden kann, und dass die Bewertung der Rhythmus-Rohinformationen mit dem Signifikanzmaß und die

anschließende Summierung ohne großen Speicher- und Rechenzeitaufwand effizient durchgeführt werden können, was das erfindungsgemäße Konzept insbesondere auch für Echtzeitanwendungen empfiehlt.

Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeichnungen detailliert erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals mit einer Qualitätsbewertung der Rhythmus-Rohinformationen;

Fig. 2 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals unter Verwendung von Gewichtungsfaktoren auf der Basis der Signifikanzmaße;

Fig. 3 ein Blockschaltbild einer bekannten Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen;

Fig. 4 ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen unter Verwendung einer Autokorrelationsfunktion mit einer teilbandweisen Nachbearbeitung der Rhythmus-Rohinformationen; und

Fig. 5 ein detailliertes Blockschaltbild der Einrichtung zum Nachbearbeiten von Fig. 4.

Fig. 1 zeigt ein Blockschaltbild einer Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen. Das Audiosignal wird über einen Eingang 100 einer Einrichtung 102 zum Zerlegen des Audiosignals in zumindest zwei

Teilbandsignale 104a und 104b zugeführt. Jedes Teilbandsignal 104a, 104b wird in eine Einrichtung 106a bzw. 106b zum Untersuchen desselben hinsichtlich von Periodizitäten in dem Teilbandsignal zugeführt, um Rhythmus-Rohinformationen 108a bzw. 108b für jedes Teilbandsignal zu erhalten. Die Rhythmus-Rohinformationen werden dann einer Einrichtung 110a bzw. 110b zum Bewerten einer Qualität der Periodizität jedes der zumindest zwei Teilbandsignale zugeführt, um ein Signifikanzmaß 112a, 112b für jedes der zumindest zwei Teilbandsignale zu erhalten. Sowohl die Rhythmus-Rohinformationen 108a, 108b als auch die Signifikanzmaße 112a, 112b werden einer Einrichtung 114 zum Ermitteln der Rhythmusinformationen des Audiosignals zugeführt. Die Einrichtung 114 berücksichtigt beim Ermitteln der Rhythmusinformationen des Audiosignals die Signifikanzmaße 112a, 112b für die Teilbandsignale sowie die Rhythmus-Rohinformationen 108a, 108b von zumindest einem Teilbandsignal.

Hat die Einrichtung 110a zur Qualitätsbewertung beispielsweise festgestellt, dass in dem Teilbandsignal 104a keine besondere Periodizität vorliegt, so wird das Signifikanzmaß 112a sehr klein, bzw. gleich 0 sein. In diesem Fall stellt die Einrichtung 114 zum Ermitteln der Rhythmusinformationen fest, dass das Signifikanzmaß 112a gleich Null ist, so dass die Rhythmus-Rohinformationen 108a des Teilbandsignals 104a bei der Ermittlung der Rhythmusinformationen des Audiosignals überhaupt nicht mehr berücksichtigt werden müssen. Die Rhythmusinformationen des Audiosignals werden dann allein und ausschließlich auf der Basis der Rhythmus-Rohinformationen 108b des Teilbandsignals 104b bestimmt.

Im nachfolgenden wird auf Fig. 2 hinsichtlich einer speziellen Ausführungsform der Vorrichtung von Fig. 1 eingegangen. Als Einrichtung 102 zum Zerlegen des Audiosignals kann eine übli-

che Analyse-Filterbank verwendet werden, die ausgangsseitig eine von einem Benutzer wählbare Anzahl von Teilbandsignalen liefert. Jedes Teilbandsignal wird dann der Verarbeitung der Einrichtungen 106a, 106b bzw. 106c unterzogen, woraufhin dann durch die Einrichtungen 110a bis 110c von jedem Rhythmus-Rohinformationen Signifikanzmaße ermittelt werden. Die Einrichtung 114 umfaßt bei der in Fig. 2 dargestellten bevorzugten Ausführungsform eine Einrichtung 114a zum Berechnen von Gewichtungsfaktoren für jedes Teilbandsignal auf der Basis des Signifikanzmaßes für dieses Teilbandsignal und optional auch der anderen Teilbandsignale. In der Einrichtung 114b findet dann eine Gewichtung der Rhythmus-Rohinformationen 108a bis 108c mit dem Gewichtungsfaktor für dieses Teilbandsignal statt, woraufhin dann, ebenfalls in der Einrichtung 114b, die gewichteten Rhythmus-Rohinformationen kombiniert, z. B. aufsummiert, werden, um an dem Tempo-Ausgang 116 die Rhythmusinformationen des Audiosignals zu erhalten.

Das erfindungsgemäße Konzept stellt sich somit folgendermaßen dar. Nach der Evaluierung der rhythmischen Informationen der Einzelbänder, welche beispielsweise durch Hüllkurvenbildung, Glättung, Differenzieren, Begrenzen auf positive Werte und Bilden der Autokorrelationsfunktion stattfinden, kann (Einrichtungen 106a bis 106c), findet eine Bewertung der Wertigkeit bzw. der Qualität dieser Zwischenergebnisse in den Einrichtungen 110a bis 110c statt. Dies wird mit Hilfe einer Bewertungsfunktion erreicht, welche die Verlässlichkeit der jeweiligen Einzelergebnisse mit einem Signifikanzmaß bewertet. Aus den Signifikanzmaßen aller Teilbandsignale wird ein Gewichtungsfaktor für jedes Band für die Extraktion der Rhythmusinformationen hergeleitet. Das Gesamtergebnis der Rhythmusextraktion wird dann in der Einrichtung 114b durch Kombination der bandweisen Einzelergebnisse unter Berücksichtigung ihrer jeweiligen Gewichtungsfaktoren erreicht.

Als Resultat zeigt ein solchermaßen implementierter Algorithmus zur Rhythmusanalyse eine gute Fähigkeit, rhythmische Informationen in einem Signal auch unter ungünstigen Bedingungen zuverlässig zu finden. Das erfindungsgemäße Konzept zeichnet sich daher durch eine hohe Robustheit aus.

Bei einer bevorzugten Ausführungsform werden die Rhythmus-Rohinformationen 108a, 108b, 108c, die die Periodizität des jeweiligen Teilbandsignals darstellen, mittels einer Autokorrelationsfunktion bestimmt. In diesem Fall wird es bevorzugt, das Signifikanzmaß zu bestimmen, indem ein Maximum der Autokorrelationsfunktion durch einen Mittelwert der Autokorrelationsfunktion geteilt wird, und dann der Wert 1 subtrahiert wird. Es sei darauf hingewiesen, daß jede Autokorrelationsfunktion immer bei einer Verzögerung von 0 ein lokales Maximum liefert, das die Energie des Signals repräsentiert. Dieses Maximum sollte unberücksichtigt bleiben, damit die Qualitätsbestimmung nicht verfälscht wird.

Ferner soll die Autokorrelationsfunktion lediglich in einem speziellen Tempobereich betrachtet werden, d. h. von einer maximalen Verzögerung, die dem kleinsten interessierenden Tempo entspricht, zu einer minimalen Verzögerung, die dem höchsten interessierenden Tempo entspricht. Ein typischer Tempobereich liegt zwischen 60 bpm und 200 bpm.

Alternativ kann als Signifikanzmaß das Verhältnis zwischen dem arithmetischen Mittelwert der Autokorrelationsfunktion im interessierenden Tempobereich und dem geometrischen Mittelwert der Autokorrelationsfunktion im interessierenden Tempobereich bestimmt werden. Es ist bekannt, dass wenn alle Werte der Autokorrelationsfunktion gleich sind, d. h. wenn die Autokorrelationsfunktion einen flachen Verlauf hat, der geometrische

Mittelwert der Autokorrelationsfunktion und der arithmetische Mittelwert der Autokorrelationsfunktion gleich sind. In diesem Fall hätte das Signifikanzmaß einen Wert gleich 1, was bedeutet, dass die Rhythmus-Rohinformationen nicht signifikant sind.

Im Falle einer Systemautokorrelationsfunktion mit starken Spitzen würde das Verhältnis aus arithmetischem Mittelwert zu geometrischem Mittelwert größer 1 sein, was bedeutet, dass die Autokorrelationsfunktion gute Rhythmusinformationen aufweist. Je kleiner jedoch das Verhältnis zwischen arithmetischem Mittelwert und geometrischem Mittelwert wird, desto flacher ist die Autokorrelationsfunktion und desto weniger Periodizitäten enthält sie, was wiederum bedeutet, dass die Rhythmusinformationen dieses Teilbandsignals weniger signifikant, d. h. eine geringere Qualität haben, was sich in einem geringen bzw. einem Gewichtungsfaktor von 0 äußern wird.

Bezüglich der Gewichtungsfaktoren existieren verschiedene Möglichkeiten. Bevorzugt wird eine relative Gewichtung, derart, dass sich alle Gewichtungsfaktoren sämtlicher Teilbandsignale zu 1 aufaddieren, d. h. dass der Gewichtungsfaktor eines Bandes bestimmt wird als der Signifikanzwert dieses Bandes geteilt durch die Summe aller Signifikanzwerte. In diesem Fall wird eine relative Gewichtung vor der Aufsummation der gewichteten Rhythmus-Rohinformationen durchgeführt, um die Rhythmusinformationen des Audiosignals zu erhalten.

Wie bereits ausgeführt worden ist, wird es bevorzugt, die Auswertung der Rhythmus-Informationen unter Verwendung einer Autokorrelationsfunktion durchzuführen. Dieser Fall ist in Fig. 4 dargestellt. Das Audiosignal wird über den Audiosignaleingang 100 in die Einrichtung 102 zum Zerlegen des Audiosignals in Teilbandsignale 104a und 104b eingespeist. Jedes Teilband-

signal wird dann in der Einrichtung 106a bzw. 106b, wie es ausgeführt worden ist, unter Verwendung einer Autokorrelationsfunktion untersucht, um die Periodizität des Teilbandsignals zu ermitteln. Am Ausgang der Einrichtung 106a bzw. 106b liegen dann die Rhythmus-Rohinformationen 108a, 108b vor. Diese werden in eine Einrichtung 118a bzw. 118b eingespeist, um die mittels der Autokorrelationsfunktion von der Einrichtung 116a ausgegebenen Rhythmus-Rohinformationen nachzubearbeiten. Damit wird u. a. sichergestellt, daß die Mehrdeutigkeiten der Autokorrelationsfunktion, d. h. daß bei ganzzahligen Vielfachen der Verzögerungen ebenfalls Signalspitzen auftreten, teilbandweise eliminiert werden, um nachbearbeitete Rhythmus-Rohinformationen 120a bzw. 120b zu erhalten.

Dies hat den Vorteil, dass die Mehrdeutigkeiten der Autokorrelationsfunktionen, d. h. der Rhythmus-Rohinformationen 108a, 108b, bereits teilbandweise eliminiert werden, und nicht erst, wie im Stand der Technik, nach der Summation der einzelnen Autokorrelationsfunktionen. Darüber hinaus ermöglicht die einzelbandweise Eliminierung der Mehrdeutigkeiten in den Autokorrelationsfunktionen durch die Einrichtungen 118a, 118b, dass die Rhythmus-Rohinformationen der Teilbandsignale unabhängig voneinander gehandhabt werden können. Sie können beispielsweise einer Qualitätsbeurteilung mittels der Einrichtung 110a für die Rhythmus-Rohinformationen 108a oder mittels der Einrichtung 110b für die Rhythmus-Rohinformationen 108b unterzogen werden.

Wie es durch die gestrichelten Linien in Fig. 4 dargestellt ist, kann die Qualitätsbeurteilung jedoch auch anhand der nachbearbeiteten Rhythmus-Rohinformationen stattfinden, wobei diese letztere Möglichkeit bevorzugt wird, da die Qualitätsbeurteilung auf der Basis der nachbearbeiteten Rhythmus-

WO 02/093557

17

PCT/EP02/04618

Rohinformationen sicherstellt, dass die Qualität einer Information beurteilt wird, die nicht mehr vieldeutig ist.

Die Ermittlung der Rhythmus-Informationen durch die Einrichtung 114 findet dann auf der Basis von nachbearbeiteten Rhythmus-Informationen eines Kanals und vorzugsweise auch auf der Basis des Signifikanzmaßes für diesen Kanal statt.

Wenn eine Qualitätsbeurteilung auf der Basis der Rhythmus-Rohinformationen, also des Signals vor der Einrichtung 118a durchgeführt wird, ist dies dahingehend vorteilhaft, dass, wenn bestimmt wird, dass das Signifikanzmaß gleich 0 ist, d. h. dass die Autokorrelationsfunktion einen flachen Verlauf hat, auf die Nachbearbeitung mittels der Einrichtung 118a gänzlich verzichtet werden kann, um Rechenzeitressourcen einzusparen.

Im nachfolgenden wird auf Fig. 5 eingegangen, um einen detaillierteren Aufbau einer Einrichtung 118a oder 118b zur Nachbearbeitung der Rhythmus-Rohinformationen darzustellen. Zunächst wird das Teilband-Signal beispielsweise 104a, in die Einrichtung 106a zur Untersuchung der Periodizität des Teilbandsignals mittels einer Autokorrelationsfunktion eingespeist, um Rhythmus-Rohinformationen 108a zu erhalten. Um die Mehrdeutigkeiten teilbandweise zu eliminieren, kann genauso wie im Stand der Technik, eine gespreizte Autokorrelationsfunktion mittels einer Einrichtung 121 berechnet werden, wobei die Einrichtung 121 angeordnet ist, um die gespreizte Autokorrelationsfunktion so zu berechnen, dass sie um ein ganzzahliges Vielfaches einer Verzögerung gespreizt ist. Eine Einrichtung 122 ist in diesem Fall angeordnet, um die gespreizte Autokorrelationsfunktion von der ursprünglichen Autokorrelationsfunktion, d. h. den Rhythmus-Rohinformationen 108a zu subtrahieren. Insbesondere wird es bevorzugt, zunächst eine auf das Doppelte gespreizte

Autokorrelationsfunktion in der Einrichtung 121 zu berechnen und dann von den Rhythmus-Rohinformationen 108a zu subtrahieren. Dann, in dem nächsten Schritt, wird eine um den Faktor 3 gespreizte Autokorrelationsfunktion in der Einrichtung 121 berechnet und von dem Ergebnis der vorherigen Subtraktion wieder abgezogen, so dass nach und nach sämtliche Mehrdeutigkeiten aus den Rhythmus-Rohinformationen eliminiert werden.

Alternativ oder zusätzlich kann die Einrichtung 121 angeordnet sein, um eine um einen ganzzahligen Faktor gestauchte Autokorrelationsfunktion zu berechnen, wobei diese dann von der Einrichtung 122 zu den Rhythmus-Rohinformationen hinzu addiert wird, um auch Anteile für Verzögerungen  $t_0/2$ ,  $t_0/3$  usw. zu erzeugen.

Darüber hinaus können die gespreizten bzw. gestauchten Versionen der Rhythmus-Rohinformationen 108a vor dem Addieren bzw. Subtrahieren gewichtet werden, um auch hier eine Flexibilität im Sinne einer hohen Robustheit zu erreichen.

Durch das Verfahren, die Periodizität eines Teilbandsignals auf der Basis einer Autokorrelationsfunktion zu untersuchen, kann also eine weitere Verbesserung erzielt werden, wenn die Eigenschaften der Autokorrelationsfunktion mit einbezogen werden und die Nachbearbeitung unter Verwendung der Einrichtung 118a oder 118b durchgeführt wird. So erzeugt eine periodische Folge von Notenanfängen mit einem Abstand  $t_0$  nicht nur eine AKF-Spitze bei einer Verzögerung  $t_0$  sondern auch bei  $2t_0$ ,  $3t_0$ , etc. Dies wird zu einer Vieldeutigkeit in der Tempodetektion, d. h. dem Suchen signifikanter Maxima in der Autokorrelationsfunktion, führen. Die Mehrdeutigkeiten können dadurch eliminiert werden, wenn um ganzzahlige Faktoren gespreizte Versionen der AKF vom Ausgangswert teilbandweise (gewichtet) abgezogen werden.

Ferner besteht bei der Autokorrelationsfunktion das Problem, daß sie keine Informationen bei  $t_0/2$ ,  $t_0/3$  ... usw., also beim Doppelten, Dreifachen, etc. des „Grundtempos“ liefert, was besonders dann zu falschen Ergebnissen führen kann, wenn zwei Instrumente, die in unterschiedlichen Teilbändern liegen, zusammen den Rhythmus des Signals definieren. Diese Sache wird dadurch berücksichtigt, dass um ganzzahlige Faktoren gestauchte Versionen der Autokorrelationsfunktion berechnet werden und diese dann zu den Rhythmus-Rohinformationen gewichtet oder ungewichtet hinzuaddiert werden.

Die AKF-Nachverarbeitung findet somit teilbandweise statt, wobei für mindestens ein Teilbandsignal eine Autokorrelationsfunktion errechnet wird und diese mit gedehnten oder gespreizten Versionen dieser Funktion kombiniert wird.

Patentansprüche

1. Vorrichtung zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen des Audiosignals, mit folgenden Merkmalen:

einer Einrichtung (102) zum Zerlegen des Audiosignals in zumindest zwei Teilbandsignale (104a, 104b);

einer Einrichtung zum Untersuchen (106a, 106b) eines Teilbandsignals hinsichtlich einer Periodizität in dem Teilbandsignal, um Rhythmus-Rohinformationen (108a, 108b) für das Teilbandsignal zu erhalten;

einer Einrichtung zum Bewerten (110a, 110b) einer Qualität der Periodizität der Rhythmus-Rohinformationen (108a) des Teilbandsignals (104a), um ein Signifikanzmaß (112a) für das Teilbandsignal zu erhalten; und

einer Einrichtung (114) zum Ermitteln der Rhythmusinformationen des Audiosignals unter Berücksichtigung des Signifikanzmaßes (112a) des Teilbandsignals und der Rhythmus-Rohinformationen (108a, 108b) von zumindest einem Teilbandsignal.

2. Vorrichtung nach Anspruch 1, bei der die Einrichtung zum Untersuchen (106a, 106b) ausgebildet ist, um eine Autokorrelationsfunktion für jedes der zumindest zwei Teilbandsignale zu berechnen.

3. Vorrichtung nach Anspruch 1 oder 2, bei der die Einrichtung zum Untersuchen (106a, 106b) folgende Merkmale aufweist:

eine Einrichtung zum Bilden einer Hüllkurve eines Teilbandsignals;

eine Einrichtung zum Glätten der Hüllkurve des Teilbandsignals, um eine geglättete Hüllkurve zu erhalten;

eine Einrichtung zum Differenzieren der geglätteten Hüllkurve, um eine differenzierte Hüllkurve zu erhalten;

eine Einrichtung zum Begrenzen der differenzierten Hüllkurve auf positive Werte, um eine begrenzte Hüllkurve zu erhalten; und

eine Einrichtung zum Bilden einer Autokorrelationsfunktion der begrenzten Hüllkurve, um die Rhythmus-Rohinformationen (108a, 108b) zu erhalten.

4. Vorrichtung nach Anspruch 2 oder 3, bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um als Signifikanzmaß ein Verhältnis eines Maximum der Autokorrelationsfunktion zu einem Mittelwert der Autokorrelationsfunktion zu verwenden.

5. Vorrichtung nach Anspruch 2 oder 3, bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um als Signifikanzmaß ein Verhältnis aus einem arithmetischen Mittelwert der Rhythmus-Rohinformationen zu einem geometrischen Mittelwert der Rhythmus-Rohinformationen zu verwenden.

6. Vorrichtung nach Anspruch 4 oder 5,

bei der die Einrichtung zum Bewerten (110a, 110b) der Qualität ausgebildet ist, um die Autokorrelationsfunktion lediglich in-

nerhalb eines Tempobereichs auszuwerten, der sich von einer minimalen Verzögerung, um ein maximales Tempo zu erhalten, zu einer maximalen Verzögerung, um ein minimales Tempo zu erhalten, erstreckt.

7. Vorrichtung nach einem der vorhergehenden Ansprüche, bei der die Einrichtung zum Ermitteln (114) folgende Merkmale aufweist:

eine Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors für ein Teilband unter Verwendung des Signifikanzmaßes für das Teilband;

eine Einrichtung (114b) zum Gewichten der Rhythmus-Rohinformationen des Teilbands unter Verwendung des Gewichtungsfaktors für das Teilband, um gewichtete Rhythmus-Rohinformationen für das Teilband zu erhalten, und zum Zusammenfassen der gewichteten Rhythmus-Rohinformationen des Teilbands mit gewichteten oder ungewichteten Rhythmus-Rohinformationen des anderen Teilbands, um die Rhythmusinformationen des Audiosignals zu erhalten.

8. Vorrichtung nach Anspruch 7, bei der die Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors angeordnet ist, um für jedes Teilbandsignal einen relativen Gewichtungsfaktor abzuleiten, wobei eine Summe aus den Gewichtungsfaktoren für alle Teilbandsignale 1 ergibt.

9. Vorrichtung nach Anspruch 8, bei der die Einrichtung (114a) zum Ableiten eines Gewichtungsfaktors angeordnet ist, um einen Gewichtungsfaktor als Verhältnis aus dem Signifikanzmaß eines Teilbandsignals zu der Summe der Signifikanzmaße aller Teilbandsignale abzuleiten.

WO 02/093557

23

PCT/EP02/04618

10. Vorrichtung nach Anspruch 9, bei der die Einrichtung (106a, 106b) zum Untersuchen eines Teilbandsignals angeordnet ist, um ein Teilbandsignal zu untersuchen, dessen Länge größer als 10 Sekunden ist.

11. Verfahren zum Analysieren eines Audiosignals hinsichtlich von Rhythmusinformationen des Audiosignals, mit folgenden Schritten:

Zerlegen des Audiosignals in zumindest zwei Teilbandsignale (104a, 104b);

Untersuchen (106a, 106b) eines Teilbandsignals hinsichtlich einer Periodizität in dem Teilbandsignal, um Rhythmus-Rohinformationen (108a, 108b) für das Teilbandsignal zu erhalten;

Bewerten (110a, 110b) einer Qualität der Periodizität der Rhythmus-Rohinformationen (108a) des Teilbandsignals (104a), um ein Signifikanzmaß (112a) für das Teilbandsignal zu erhalten; und

Ermitteln der Rhythmusinformationen des Audiosignals unter Berücksichtigung des Signifikanzmaßes (112a) des Teilbandsignals und der Rhythmus-Rohinformationen (108a, 108b) von zumindest einem Teilbandsignal.

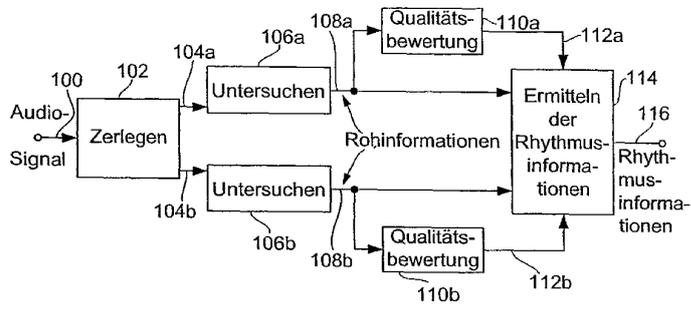


Fig. 1

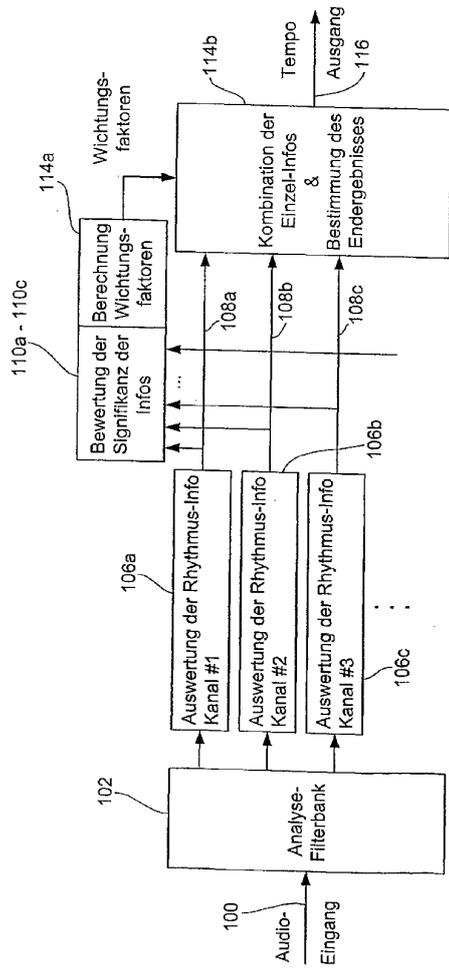


Fig. 2

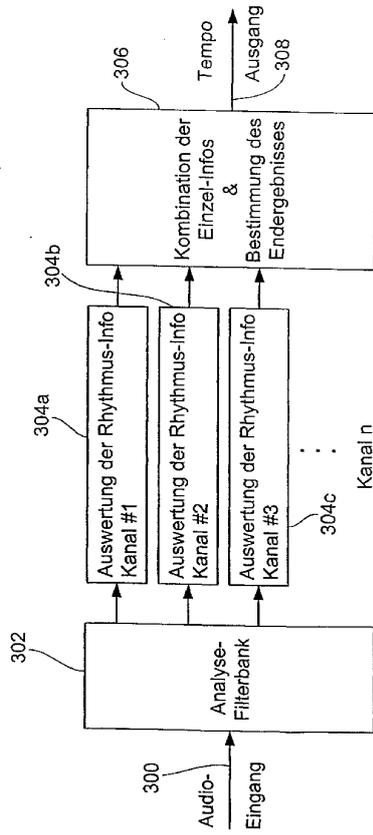


Fig. 3 (Stand der Technik)

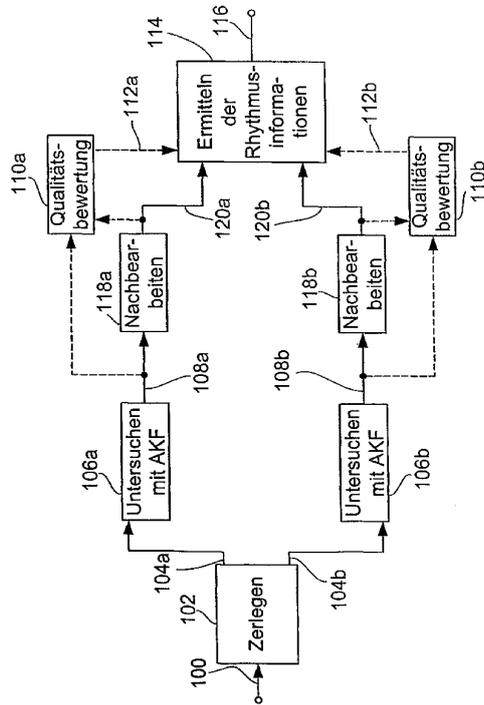


Fig. 4

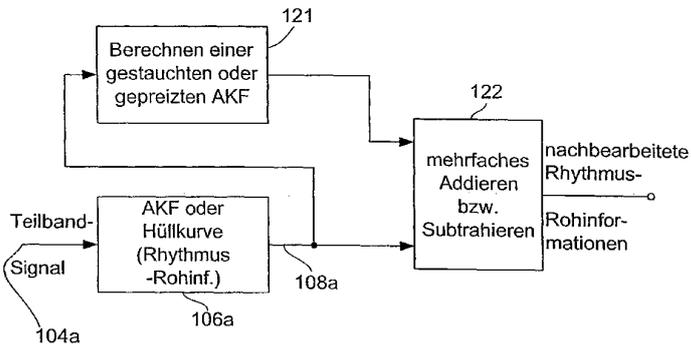


Fig. 5

## 【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International Application No. PCT/EP 02/04618
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 G10L19/02		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 G10H G10L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) INSPEC, EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCHEIRER E D: "Tempo and beat analysis of acoustic musical signals" JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, JAN. 1998, ACOUST. SOC. AMERICA THROUGH AIP, USA, vol. 103, no. 1, pages 588-601, XP002210418 ISSN: 0001-4966 cited in the application the whole document --- -/-	1-11
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents:		
*A* document defining the general state of the art which is not considered to be of particular relevance		*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*E* earlier document but published on or after the international filing date		*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (see specific)		*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
*O* document referring to an oral disclosure, use, exhibition or other means		*Z* document member of the same patent family
*P* document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 21 August 2002		Date of mailing of the international search report 19/09/2002
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patenlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2240, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Quélayoive, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		International Application No. PCT/EP 02/04618
C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GOTO M ET AL: "Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 27, no. 3-4, April 1999 (1999-04), pages 311-335, XP004163257 ISSN: 0167-6393 abstract	1,11
A	SCHEIRER E D: "Pulse tracking with a pitch tracker" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1997. 1997 IEEE ASSP WORKSHOP ON NEW PALTZ, NY, USA 19-22 OCT. 1997, NEW YORK, NY, USA, IEEE, US, 19 October 1997 (1997-10-19), page 4pp XP010248228 ISBN: 0-7803-3908-8 cited in the application the whole document	1-11
A	TOLONEN T ET AL: "A computationally efficient multipitch analysis model" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, NOV. 2000, IEEE, USA, vol. 8, no. 6, pages 708-716, XP002210419 ISSN: 1063-6676 cited in the application abstract	1,11

INTERNATIONALER RECHERCHENBERICHT		Internationales Aktenzeichen PCT/EP 02/04618
A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES IPK 7 610L19/02		
Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK		
B. RECHERCHIERTE GEBIETE Recherchierte Mindestprüfstoffe (Klassifikationssystem und Klassifikations Symbole) IPK 7 610H 610L		
Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe) INSPEC, EPO-Internal, WPI Data		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	SCHEIRER E D: "Tempo and beat analysis of acoustic musical signals" JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, JAN. 1998, ACOUST. SOC. AMERICA THROUGH AIP, USA, Bd. 103, Nr. 1, Seiten 588-601, XP002210418 ISSN: 0001-4966 in der Anmeldung erwähnt das ganze Dokument --- -/--	1-11
<input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen <input type="checkbox"/> Siehe Anhang Patentfamilie		
* Besondere Kategorien von angegebenen Veröffentlichungen : *A* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist *E* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist *L* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausglüht) *O* Veröffentlichung, die sich auf eine mündliche Offenbarung, ohne Benutzung, ohne Ausstellung oder andere Maßnahmen bezieht *P* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist *T* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist *X* Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderscher Tätigkeit beruhend betrachtet werden *Y* Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann nicht als auf erfinderscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist *Z* Veröffentlichung, die Mitglied derselben Patentfamilie ist		
Datum des Abschlusses der internationalen Recherche		Absenddatum des internationalen Recherchenberichts
21. August 2002		19/09/2002
Name und Postanschrift der internationalen Recherchenbehörde Europäisches Patentamt, P.B. 6816 Palatinlaan 2 NL - 2200 HV Rijswijk Tel. (+31-70) 340-2340, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Bevollmächtigter Bediensteter Quévaioine, R

INTERNATIONALER RECHERCHENBERICHT		nationales Aktenzeichen PCT/EP 02/04618
C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Beitr. Anspruch Nr.
A	GOTO M ET AL: "Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions" SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, Bd. 27, Nr. 3-4, April 1999 (1999-04), Seiten 311-335, XP004163257 ISSN: 0167-6393 Zusammenfassung ----	1,11
A	SCHEIRER E D: "Pulse tracking with a pitch tracker" APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS, 1997. 1997 IEEE ASSP WORKSHOP ON NEW PALTZ, NY, USA 19-22 OCT. 1997, NEW YORK, NY, USA, IEEE, US, 19. Oktober 1997 (1997-10-19), Seite 4pp XP010248228 ISBN: 0-7803-3908-8 in der Anmeldung erwähnt das ganze Dokument ----	1-11
A	TOLONEN T ET AL: "A computationally efficient multipitch analysis model" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, NOV. 2000, IEEE, USA, Bd. 8, Nr. 6, Seiten 708-716, XP002210419 ISSN: 1063-6676 in der Anmeldung erwähnt Zusammenfassung -----	1,11

---

フロントページの続き

(72)発明者 ウーレ、クリスティアーン

ドイツ国、D - 9 8 6 9 3 イルメナオ、シュヴァニッツェストラーセ 3

(72)発明者 クレーマー、マルクス

ドイツ国、D - 9 8 6 9 3 イルメナオ、ヘルダーストラーセ 4

Fターム(参考) 5D082 BB01

5D378 HB00 QQ27