

(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2004-515808
(P2004-515808A)

(43) 公表日 平成16年5月27日(2004.5.27)

(51) Int. Cl. ⁷	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 1/00	G 1 0 H 1/00	5 D 3 7 8

審査請求 有 予備審査請求 有 (全 124 頁)

(21) 出願番号	特願2002-548707 (P2002-548707)	(71) 出願人	501348243
(86) (22) 出願日	平成13年12月3日 (2001.12.3)		アミューズテック カンパニー リミテッ ド
(85) 翻訳文提出日	平成15年5月26日 (2003.5.26)		大韓民国, ソウル 156-010, トン ジャーク, シンデバンドン 395- 73, ウーソン キャラクター 312
(86) 国際出願番号	PCT/KR2001/002081	(74) 代理人	100080034
(87) 国際公開番号	W02002/047064		弁理士 原 謙三
(87) 国際公開日	平成14年6月13日 (2002.6.13)	(74) 代理人	100113701
(31) 優先権主張番号	2000-0073452		弁理士 木島 隆一
(32) 優先日	平成12年12月5日 (2000.12.5)	(74) 代理人	100116241
(33) 優先権主張国	韓国 (KR)		弁理士 金子 一郎

最終頁に続く

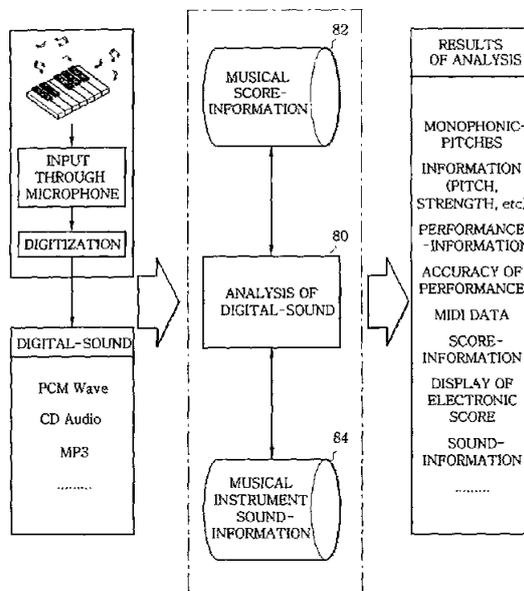
(54) 【発明の名称】 演奏楽器の音情報を用いた音楽分析方法

(57) 【要約】

本発明は、デジタル音響を分析するために楽器の音情報または音情報および楽譜情報を用いる技術に関する。特に、本発明は、入力されるデジタル音響を生成するために使用された、または、使用中の楽器の音情報、または、該音情報と入力されるデジタル音響を生成するために使用された、または、使用中の楽譜情報を用いるものに関する。

本発明によれば、入力されるデジタル音響を生成するために使用された演奏楽器の音の高さ別、強さ別の情報などを予め格納しておいているため、該楽器で演奏する単音や複音を容易に分析することができる。

また、本発明によれば、演奏楽器の音情報と楽譜情報とを共に用いることによって、入力されるデジタル音響を正確に分析可能で、定量的なデータとして抽出することができる。



【特許請求の範囲】**【請求項 1】**

デジタル音響を分析するため、演奏楽器の音情報を利用する方法であって、

(a) 演奏楽器別の音情報を生成し格納するステップと、

(b) 前記格納された演奏楽器別の音情報のうち実際に演奏される楽器の音情報を選択するステップと、

(c) デジタル音響信号を入力されるステップと、

(d) 前記入力されたデジタル音響信号を単位フレーム別の周波数成分に分解するステップと、

(e) 前記入力されたデジタル音響信号の周波数成分と前記選択された演奏楽器の音情報の周波数成分とを比較・分析し、前記入力されたデジタル音響信号に含まれた単音情報を導出するステップと、

(f) 前記導出された単音情報を出力して送るステップと、

を含むことを特徴とするデジタル音響分析方法。

【請求項 2】

請求項 1 において、前記ステップ (e) は、

前記区分された単位フレーム別の時間情報を導出した後、該単位フレーム別の周波数成分と演奏楽器の音情報の周波数成分とを比較・分析し、各単位フレーム別に含まれた単音の音高さ別および強さ別の情報を時間情報と共に導出することを特徴とするデジタル音響分析方法。

【請求項 3】

請求項 2 において、前記ステップ (e) は、

前記導出された単音情報が、前フレームに含まれていない新しい単音である場合、これに相当するフレームを前フレームより大きさが小さなサブフレームに分け、新しい音の含まれたサブフレームを探して当該サブフレームの時間情報を前記導出された単音の音高さ別および強さ別の情報と共に導出することを特徴とするデジタル音響分析方法。

【請求項 4】

請求項 1 において、前記ステップ (a) は、

前記演奏楽器の音情報を周期的に更新することを特徴とするデジタル音響分析方法。

【請求項 5】

請求項 1 において、前記ステップ (a) は、

前記演奏楽器の音情報を 1 つ以上の強さ別の音のサンプルとして格納する時、音情報として表現可能な各単音をウェーブ形式のデータとして格納し、該ウェーブ形式のデータから演奏楽器別の音情報の周波数成分を導出するようにすることを特徴とするデジタル音響分析方法。

【請求項 6】

第 1 項において、前記ステップ (a) は、

前記演奏楽器の音情報を 1 つ以上の強さ別の音のサンプルとして格納する時、音情報として表現可能な各単音を周波数成分別の強さに直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

【請求項 7】

請求項 6 において、前記ステップ (a) は、

前記演奏楽器の音情報をフーリエ変換を経て直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

【請求項 8】

請求項 6 において、前記ステップ (a) は、

前記演奏楽器の音情報をウェーブレット変換を経て直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

【請求項 9】

請求項 5 または請求項 6 において、前記ステップ (a) は、

10

20

30

40

50

鍵盤楽器の音情報を格納する時、ペダルの使用有無について分類格納することを特徴とするデジタル音響分析方法。

【請求項 10】

請求項 5 または請求項 6 において、前記ステップ (a) は、弦楽器の音情報を格納する時、該音情報を各弦別に分類格納することを特徴とするデジタル音響分析方法。

【請求項 11】

請求項 1 または請求項 2 において、前記ステップ (e) は、前記入力されたデジタル音響信号のフレーム別に当該フレームに含まれた最も低いピーク周波数を選択するステップ (e 1) と、前記選択された演奏楽器の音情報のうちから前記選択されたピーク周波数を含む音情報を導出するステップ (e 2) と、前記導出された音情報のうち当該フレームのピーク周波数成分と最も近接したピーク情報を有する音情報を単音情報として導出するステップ (e 3) と、前記導出された単音情報に相当する音情報の周波数成分を当該フレームから除去するステップ (e 4) と、当該フレームのピーク周波数成分が残っている場合、前記ステップ (e 1 ~ e 4) を繰り返して行うステップ (e 5) と、を含むことを特徴とするデジタル音響分析方法。

10

【請求項 12】

デジタル音響を分析するため、演奏楽器の音情報および楽譜情報を利用する方法であって、

20

(a) 演奏楽器別の音情報を生成し格納するステップと、
 (b) 演奏される楽譜の楽譜情報を生成し格納するステップと、
 (c) 前記格納された演奏楽器別の音情報および楽譜情報のうち実際に演奏される楽器の音情報および楽譜情報を選択するステップと、
 (d) デジタル音響を入力されるステップと、
 (e) 前記入力されたデジタル音響信号を単位フレーム別の周波数成分に分解するステップと、
 (f) 前記入力されたデジタル音響信号の周波数成分と前記選択された演奏楽器の音情報の周波数成分および楽譜情報を比較・分析し、前記入力されたデジタル音響信号に含まれた演奏誤り情報および単音情報を導出するステップと、
 (g) 前記導出された単音の情報を出力して送るステップと、
 を含むことを特徴とするデジタル音響分析方法。

30

【請求項 13】

請求項 12 において、前記ステップ (f) は、前記単位フレーム別の時間情報を導出した後、当該単位フレーム別の周波数成分と演奏楽器の音情報の周波数成分および楽譜情報を比較・分析し、各単位フレーム別に含まれた単音の音高さ別および強さ別の情報を時間情報と共に導出することを特徴とするデジタル音響分析方法。

40

【請求項 14】

請求項 12 または請求項 13 において、前記ステップ (f) は、前記導出された単音情報が前フレームに含まれていない新しい単音である場合、これに相当するフレームを前フレームより大きさが小さなサブフレームに分け、新しい音の含まれたサブフレームを探して当該サブフレームの時間情報を、前記導出された単音の音高さ別および強さ別の情報と共に導出するステップをさらに含むことを特徴とするデジタル音響分析方法。

【請求項 15】

請求項 12 において、前記ステップ (a) は、前記演奏楽器の音情報を周期的に更新することを特徴とするデジタル音響分析方法。

50

【請求項 16】

請求項 12 において、前記ステップ (a) は、
前記演奏楽器の音情報を 1 つ以上の強さ別の音のサンプルとして格納する時、音情報として表現可能な各単音をウェーブ形式のデータとして格納することを特徴とするデジタル音響分析方法。

【請求項 17】

請求項 12 において、前記ステップ (a) は、
前記演奏楽器の音情報を 1 つ以上の強さ別の音のサンプルとして格納する時、音情報として表現可能な各単音を周波数成分別の強さに直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

10

【請求項 18】

請求項 17 において、前記ステップ (a) は、
前記演奏楽器の音情報をフーリエ変換を経て直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

【請求項 19】

請求項 17 において、前記ステップ (a) は、
前記演奏楽器の音情報をウェーブレット変換を経て直接表現可能な形態で格納することを特徴とするデジタル音響分析方法。

【請求項 20】

請求項 16 または請求項 17 において、前記ステップ (a) は、
鍵盤楽器の音情報を格納する時、ペダルの使用有無について分類格納することを特徴とするデジタル音響分析方法。

20

【請求項 21】

請求項 16 または請求項 17 において、前記ステップ (a) は、
弦楽器の音情報を格納する時、その音情報を各弦別に分類格納することを特徴とするデジタル音響分析方法。

【請求項 22】

請求項 12 または請求項 13 において、前記ステップ (f) は、
前記楽譜情報を参照して当該演奏楽器の演奏進行に従うデジタル音響信号のフレーム別の演奏期待値をリアルタイムで生成し、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在するか否かをフレーム単位で確認するステップ (f1) と、

30

前記ステップ (f1) の結果、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在しない場合、当該フレームのデジタル音響信号に含まれた周波数成分が演奏誤り情報であるか否かを確認し、演奏誤り情報および単音情報を導出した後、該演奏誤り情報および単音情報として導出された音情報の周波数成分を当該フレームのデジタル音響信号から除去するステップ (f2) と、

前記ステップ (f1) の結果、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在する場合、前記デジタル音響信号と前記演奏期待値とを比較・分析し、当該フレームのデジタル音響信号に含まれた単音情報を導出し、前記単音情報として導出された音情報の周波数成分を該単音情報が含まれたフレームのデジタル音響信号から除去するステップ (f3) と、

40

フレーム単位のデジタル音響信号にピーク周波数成分が残っている場合、前記ステップ (f1 ~ f3) を繰り返して行うステップ (f4) と、
を含むことを特徴とするデジタル音響分析方法。

【請求項 23】

請求項 22 において、前記ステップ (f2) は、
前記入力されたデジタル音響信号のフレーム別に当該フレームに含まれた最も低いピーク周波数を選択するステップ (f2_1) と、
当該演奏楽器の音情報のうちから前記選択されたピーク周波数を含む音情報を導出するス

50

テップ (f 2 _ 2) と、
前記導出された音情報のうち当該フレームのピーク周波数成分と最も近接したピーク情報を有する音情報を演奏誤り情報として導出するステップ (f 2 _ 3) と、
前記演奏誤り情報が楽譜情報において次に演奏される音符に含まれる場合、前記演奏誤り情報に含まされた音を演奏期待値に追加した後、該演奏誤り情報を単音情報として導出するステップ (f 2 _ 4) と、
前記演奏誤り情報および単音情報として導出された音情報の周波数成分を当該フレームから除去するステップ (f 2 _ 5) と、
を含むことを特徴とするデジタル音響分析方法。

【請求項 2 4】

10

請求項 2 3 において、前記ステップ (f 2 _ 3) は、
当該演奏音の高さおよび強さを演奏誤り情報として導出することを特徴とするデジタル音響分析方法。

【請求項 2 5】

請求項 2 2 において、前記ステップ (f 3) は、
前記フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値に含まれた音情報のうち最も低い音の情報を選択するステップ (f 3 _ 1) と、
前記選択された音情報の周波数成分が当該フレームのデジタル音響信号に含まれた周波数成分に含まれる場合、該音情報を単音情報として導出した後、該音情報の周波数成分を当該フレームのデジタル音響信号から除去するステップ (f 3 _ 2) と、
前記選択された音情報の周波数成分が当該フレームのデジタル音響信号に含まれていない場合、前記演奏期待値を訂正するステップ (f 3 _ 3) と、
を含むことを特徴とするデジタル音響分析方法。

20

【請求項 2 6】

請求項 2 5 において、前記ステップ (f 3 _ 3) は、
前記選択された音情報の周波数成分がある時点で前記デジタル音響信号に含まれた履歴があるが、当該フレームの前フレームに所定の回数以上連続して含まれない場合、該音情報を演奏期待値から除去することを特徴とするデジタル音響分析方法。

【請求項 2 7】

30

請求項 1 2 において、
前記ステップ (f) で導出された前記演奏誤り情報に基づいて演奏正確度を判別するステップ (h) をさらに含むことを特徴とするデジタル音響分析方法。

【請求項 2 8】

請求項 1 2 において、
前記ステップ (f) で導出された前記演奏誤り情報に基づいて前記演奏誤り情報に相当するデジタル音響信号の単音情報を前記ステップ (b) で生成された楽譜情報に追加するステップ (i) をさらに含むことを特徴とするデジタル音響分析方法。

【請求項 2 9】

40

請求項 1 2 において、前記ステップ (b) は、
演奏される楽譜に基づいて、
時間の流れに従う音の高さ、音の長さ情報、速さ情報、拍子情報、音の強さ情報と、
スタッカート、スタッカティッシモ、プラルトリラーなどを含む細部演奏情報と、
両手演奏および多数楽器の演奏に係る演奏区分情報のうち 1 つ以上の情報と、を含む楽譜情報を生成し格納することを特徴とするデジタル音響分析方法。

【発明の詳細な説明】

[技術分野]

本発明は、デジタル音響信号を分析する方法に関し、特に、入力されるデジタル音響信号の周波数成分と演奏楽器の音の周波数成分とを比較することによって、デジタル音響信号を分析する方法に関する。

50

【0001】

[背景技術]

1980年代にパソコンの補給が始まって以来、コンピュータの技術、性能および環境が急速に発展し、1990年代に入ってはインターネットが会社の各部門および個人性格領域にまでも急に広がりつつある。これによって、21世紀には全世界においてコンピュータの利用が全分野で非常に重要視され、音楽分野においてこれを応用したものの1つとしてミディ(MIDI: Music Instrument Digital Interface)がある。「ミディ」とは、人の音声や楽器の演奏音を合成(synthesis)し、格納することのできるコンピュータ音楽技術であって、音楽家が利用している代表的なものであると言え、現在、大衆音楽の作曲家や演奏者が主に利用しているものである。

10

【0002】

例えば、作曲家は、ミディ機能を有する電子楽器をコンピュータと連結して手軽に作曲でき、このように作曲した演奏曲は、コンピュータまたはシンセサイザーの音響合成を用いて手軽に再生できるようになっている。また、録音過程においてミディ装備で製作された音響が、歌手の歌とミキシングされて人々に好まれる音楽に作られることもあった。

【0003】

ミディ技術は、このように大衆音楽と関連しながら持続的に発達すると共に、音楽教育分野へも進出している。すなわち、ミディは、音楽演奏の実際音とは無関係に、単に楽器の種類、音の高さおよび強さ、音の始まりと終わり情報のみを用いているため、ミディ楽器とコンピュータを用いてミディ情報をやり取りすることができる。従って、ミディ機能を有する電子ピアノとコンピュータをミディケーブルで連結した後、電子ピアノの演奏時に生成されるミディ情報を用いることで、音楽教育に活用し得る。これによって、日本のヤマハを始めとした多くの会社でミディを用いた音楽教育ソフトウェアを開発して適用してきている。

20

【0004】

しかし、このようなミディ技術は、アコースティック楽器の音とアコースティック楽器の演奏から感じられる感じを大事にする多くのクラシック音楽家の欲求を充足させていない。即ち、多くのクラシック音楽家は、電子楽器の音や感じをあまり好まないため、伝統的な方法で音楽を勉強し、アコースティック楽器の演奏を学ぶのが実状である。従って、音楽教師がクラシック音楽を教えるか学生が音楽を学ぶには、音楽学院や音楽学校などを用い、学生は音楽教師の指導に全的に依存している現状である。このようなクラシック音楽教育環境に鑑みても、クラシック音楽分野にコンピュータ技術とデジタル信号処理技術を導入して、アコースティック楽器で演奏された音楽を分析し、その分析結果を定量的な演奏情報として示すことができれば好ましいだろう。

30

【0005】

このため、アコースティック楽器の演奏音をデジタル音響に変換した後、該デジタル音響をコンピュータで分析する技術の工夫が種々試みられている。

【0006】

このような例の1つとして、「Eric D. Scheirer」の修士論文である『Extracting Expressive Performance Information from Recorded Music』においては、録音されたデジタル音響からミディ音楽を抽出するため楽譜情報を利用している。前記論文に記載のように、彼は、各音の強さ、始まりタイミング、終わりタイミングを抽出してミディ音楽に変化させる方法について研究した。しかし、前記論文に記載の実験結果を見ると、楽譜情報を用いた場合、録音されたデジタル音響から始まるタイミングは正確に抽出されるが、終わるタイミングと音の強さ情報の抽出は、やはり正確でないという問題点がある。

40

【0007】

なお、全世界的に少数の会社が音楽認識技術を用いて簡単なデジタル音響を分析できる初期の製品を出しており、インターネット上のalt.music.midi.FAQによ

50

れば、ウェーブ (wave) 形式のデジタル音響を分析してミディ形式に変換または楽譜化する製品が多数開発されている。これらのものとしては、A k o f f M u s i c C o m p o s e r、S o u n d 2 M I D I、G a m a、W I D I、D i g i t a l E a r、W A V 2 M I D、P o l y a x e D r i v e r、W A V 2 M I D I、I n t e l l i S c o r e、P F S - S y s t e m、H a n a u t a M u s i c i a n、A u d i o t o M I D I、A m a z i n g M I D I、C a p e l l a - A u d i o、A u t o S c o r e およびごく最近に発表された Wave Goodbye などが挙げられる。

【0008】

これらの幾つかは、複音の分析も可能であると広報しているが、実際に試験を行った結果、複音の分析は不可能であった。そのため、前記FAQ文書では、コンピュータでウェーブ形式の音楽をミディ形式に変更した後、再生時には元のウェーブ音楽と同様に聞こえるか否かに対して、不可であると説明している。さらに、現在としては、ウェーブ音楽をミディ音楽に変更するソフトウェアに対して全て価値がないと明示されている。

10

【0009】

以下、これらの製品の中で Araki Software 社の Amazing MIDI を使用してウェーブ音楽がどのように分析されるかに関する実験結果について記述する。

【0010】

図1は、実験に使用した曲の楽譜の、ベートーベンのピアノソナタ第8番第2楽章の初の2小節に相当し、図2は、図1に示された楽譜を分析の便宜のため、それぞれの単音に区分して各音符に音程を表示する記号を付したものである。図3は、Amazing MIDI でウェーブ音楽をミディ音楽に変更するために使用者が設定するパラメータ入力ウィンドウであり、図4は、各種パラメータの値を最右側に設定した時に変換されたミディウィンドウであり、図5は、図2の楽譜情報を用いて元のウェーブ音楽に相当する部分を、図4に黒棒で表示したものである。図6は、各種パラメータ値を最左側に設定した時に変換されたミディウィンドウであり、図7は、図5と同様に、図6にウェーブ音楽に相当する部分を黒棒で表示したものである。

20

【0011】

まず、図1および図2を参照する。最初にC4、A3、A2の3つの音が示され、C4およびA2の鍵盤が押された状態で、A3の代わりにE3の鍵盤が押され、再度A3、E3の順に繰り返される。次いで、C4の音はB3の音に変わり、A2の音はD3の音に変わり、E3の音がG3の音に変わり、再度B3およびD3の鍵盤が押された状態でE3、G3、E3の順に繰り返される。従って、これらの楽譜をミディに変換する場合、図5に黒棒で示されたようなミディ情報が構成される必要があるが、実際には図4に示されたようなミディ情報が構成されるという問題点がある。

30

【0012】

図3に示したように、Amazing MIDIでは、使用者が各種パラメータを設定してウェーブ音楽をミディ音楽に変更しているが、これらの設定値によって構成されるミディ情報が大きな差を示すことがわかる。図3の設定値のうち、Minimum Analysisの値(最小分析値)、Minimum Relativeの値(最小相関値)、Minimum Noteの値(最小音符値)を全て最右側に設定してミディ音楽に変換した結果が、図4に、各値を全て最左側に設定してミディ音楽に変換した結果が、図6に示されている。図4と図6とを比較すると、大きな違いがあることがわかる。すなわち、図4では、周波数領域において強さの大きな周波数部分のみを認識してミディとして示しており、図6では、強さの小さな部分も認識してミディに示していることがわかる。従って、図4のミディ情報項目は基本的に図6に含まれて示されるようになる。

40

【0013】

図4と図5とを比較すると、図4の場合、実際に演奏された部分であるA2、E3、G3、D3が全く認識されておらず、C4、A3、B3の場合も認識された部分が実際演奏される部分と大きく相違していることがわかる。すなわち、C4の場合、認識された部分が全体音長さのうち初期25%に止めており、B3の場合、20%よりも小さ

50

く認識されていることがわかる。また、A 3 の場合、35%程度に止めている。これに対して、非演奏の部分が非常に多く認識されていることがわかる。E 4 の場合、非常に強い大きさに認識されており、A 4、G 4、B 4、D 5、F 5などの音も誤って認識されている。

【0014】

図6と図7とを比較すると、図6の場合、実際に演奏された部分であるA 2、E 3、G 3、D 3、C 4、A 3、B 3が全て認識されているが、認識された部分と実際演奏された部分とが非常に相違していることがわかる。すなわち、C 4、A 2の場合、鍵盤を続けて押していたので、実際に続けて音が出ているにもかかわらず、少なくとも一度音の切れた状態に認識されていることがわかる。さらに、A 3、E 3などの場合、実際に演奏された時点と音の長さとは相当に違って認識された。図6および図7に示されたように、黒棒で示した部分以外に多くの部分がグレー色の棒で示されていることがわかる。これらの部分は、実際に演奏されなかったにもかかわらず、誤って認識されたもので、正しく認識されたものより遥かに多く占めている。本明細書では、Amazing MIDIプログラム以外の他のプログラムを実験した結果は記載しないが、現在まで公開された全てのプログラムの音楽認識結果は、前述のようなAmazing MIDIプログラムを用いて実験を行った結果と同様であり、あまり満足でないことが確認された。

10

【0015】

すなわち、コンピュータ技術とデジタル信号処理技術を導入して、アコースティック楽器で演奏された音楽を分析するための技術の工夫が種々試みられているが、未だ満足な結果

20

【0016】

[発明の開示]

従って、本発明では、演奏に使用された楽器に対して予め格納されている情報を用いることによって、より正確な演奏分析結果を導出することができ、その結果を定量的なデータとして抽出することができる、音楽分析方法を提供することを目的とする。

【0017】

すなわち、本発明の目的は、演奏楽器の音情報を用いてデジタル音響に含まれた構成信号と当該音情報の構成信号とを比較・分析することによって、単音は勿論、複音の分析も正確に行える音楽分析方法を提供することにある。

30

【0018】

また、本発明の他の目的は、演奏楽器の音情報および演奏曲の楽譜情報を用いて、上記のように正確な分析結果が得られると共に、分析にかかる時間を減らすことのできる音楽分析方法を提供することにある。

【0019】

上記の目的を達成するための本発明が提供する演奏楽器の音情報を用いた音楽分析方法は、演奏楽器別の音情報を生成し格納するステップ(a)と、前記格納された演奏楽器別の音情報のうち実際に演奏される楽器の音情報を選択するステップ(b)と、デジタル音響を入力で受けるステップ(c)と、前記入力されたデジタル音響信号を単位フレーム別の周波数成分に分解するステップ(d)と、前記入力されたデジタル音響信号の周波数成分と前記選択された演奏楽器の音情報の周波数成分とを比較・分析し、前記入力されたデジタル音響信号に含まれた単音情報を導出するステップ(e)と、前記導出された単音情報を出力で送るステップ(f)とを含むことを特徴とする。

40

【0020】

また、上記の他の目的を達成するため、本発明が提供する演奏楽器の音情報および楽譜情報を用いた音楽分析方法は、演奏楽器別の音情報を生成し格納するステップ(a)と、演奏される楽譜の楽譜情報を生成し格納するステップ(b)と、前記格納された演奏楽器別の音情報および楽譜情報のうち実際に演奏される楽器の音情報および楽譜情報を選択するステップ(c)と、デジタル音響を入力で受けるステップ(d)と、前記入力されたデジタル音響信号を単位フレーム別の周波数成分に分解するステップ(e)と、前記入力され

50

たデジタル音響信号の周波数成分と前記選択された演奏楽器の音情報の周波数成分および楽譜情報を比較・分析し、前記入力されたデジタル音響信号に含まれた演奏誤り情報および単音情報を導出するステップ (f) と、前記導出された単音の情報を出力で送るステップ (g) とを含むことを特徴とする。

【 0 0 2 1 】

[発明を実施するための最良の形態]

以下、添付の図面を参照して本発明の音楽分析方法の詳細を説明する。

【 0 0 2 2 】

図 8 は、デジタル音響を分析する方法に関する概念図である。同図に示されたように、デジタル音響を入力され演奏楽器の音情報 (8 4) および楽譜情報 (8 2) を用いて入力されたデジタル音響の信号を分析 (8 0) し、その結果として演奏情報および正確度、ミディ音楽などを導出して電子楽譜を描く概念が示されている。

10

【 0 0 2 3 】

ここで、デジタル音響は、PCMウェーブ、CDオーディオ、MP3ファイルなど、入力される音をデジタル変換してコンピュータ読取り可能な形態で格納された全てのものを意味し、実際に演奏されている音楽の場合は、コンピュータに連結されたマイクを介して入力された後、デジタル変換して格納しながら分析を行うこともできる。

【 0 0 2 4 】

入力音楽の楽譜情報 (8 2) は、例えば、音の高さ、音の長さ、速さ情報 (例えば、四分音符 = 6 4、フェルマータなど)、拍子情報、音の強さ情報 (例えば、*f o r t e*、*p i a n o* > - *a c c e n t*、< *だんだん強く (c r e s c e n d o)* など)、細部演奏情報 (例えば、スタッカート、スタッカティッシモ、プラルトリラーなど) およびピアノのように両手を用いて演奏する場合、左手の演奏および右手の演奏に相当する部分を区分するための情報などを含む。また、2つ以上の楽器で演奏された場合は、各楽器に相当する楽譜の部分情報などを含む。すなわち、楽器の演奏時に、目で見ても適用する楽譜上の情報は全て楽譜情報として活用できるが、作曲家や時代によって表記法が相違しているため、細部の表記法については本明細書で言及しない。

20

【 0 0 2 5 】

なお、演奏楽器の音情報 (8 4) は、図 9 A ~ 図 9 E に示されたように、演奏される特定の楽器別に予め構築されているものであって、音の高さ、音の強さ、ペダル表などがあるが、これについては、図 9 A ~ 図 9 E を参照して説明する部分において再度説明する。

30

【 0 0 2 6 】

図 8 に示されたように、本発明では、入力されるデジタル音響を分析する時、音情報または音情報および楽譜情報を活用し、ピアノ演奏曲のように同時に複数の音が演奏される場合、各構成音の高さおよび強さを正確に分析することができ、各時間別に分けて分析した構成音に関する情報から、どの音がどの強さで演奏されたかに関する演奏情報を分析結果として導出することができる。

【 0 0 2 7 】

この時、音楽分析のために演奏楽器の音情報を用いる理由は、一般に、音楽において使用する各音は、その高さに応じて固有なピッチ (*P i t c h*) 周波数およびハーモニック (*H a r m o n i c*) 周波数を有するが、アコースティック楽器の演奏音および人音声の分析において基本となるものが、ピッチ周波数とハーモニック周波数であるからである。

40

【 0 0 2 8 】

なお、このようなピーク周波数 - ピッチ周波数およびハーモニック周波数 - 成分は、一般に、各楽器の種類別に違って現われる。従って、このような楽器の種類別のピーク周波数成分を予め導出し、当該楽器の音情報として格納した後、入力されるデジタル音響に含まれたピーク周波数成分と、既に格納されている演奏楽器の音情報とを比較することによって、前記デジタル音響を分析することが可能となる。

【 0 0 2 9 】

例えば、ピアノの場合、88個の鍵盤に関する音情報を予め知っていれば、該ピアノで演

50

奏される演奏音の分析結果、同時に複数の音が演奏されたとしても、該演奏音を予め格納された 88 個の音情報の組合わせと比較することができるため、それぞれの個別音を正確に分析することができる。

【0030】

図 9 A ~ 図 9 E は、デジタル音響を分析するために使用するピアノの音情報を示す例示図である。すなわち、韓国ของบริษัทである YOUNG CHANG ピアノ社製のピアノの 88 個の鍵盤に関する音情報の例を示している。この時、図 9 A ~ 図 9 C は、当該ピアノの音情報を導出するための条件を示すもので、図 9 A は、88 個の鍵盤のそれぞれに関する音の高さ (A0、(、C8) を区分して示したものであり、図 9 B は、音の強さ別の識別情報を示したものであり、図 9 C は、ペダルの使用有無に関する識別情報を示したものである。図 9 B に示されたように、各音の強さは、「 - 」 ~ 「 0 」を所定の段階に分類して格納することができる。図 9 C に示されたように、ペダルを使用した場合を「 1」、ペダルを使用しない場合を「 0」とし、3つのペダルを有するピアノにおいてペダルの使用形態に関する全ての場合の数について示した。

10

【0031】

また、図 9 D および図 9 E は、前記ピアノに関する音情報の実際的な格納形態を示したものであって、図 9 A ~ 図 9 C に示された各音情報の条件から、音の高さが「 C4」、音の強さが「 - 7 dB」で、ペダルを全く使用しない場合の音情報を示している。特に、図 9 D は、当該音情報がウェーブで格納された形態を示し、図 9 E は、当該音情報がスペクトログラムで格納された形態を示している。ここで、スペクトログラムとは、時間の流れに従う各周波数成分の強さを示すもので、スペクトログラムの横軸は時間情報を示し、縦軸は周波数情報を示すことで、図 9 E のようなスペクトログラムを参照すると、各時間別の周波数成分に関する強さの情報がわかるようになる。

20

【0032】

すなわち、当該演奏楽器の音情報を 1 つ以上の強さ別の音のサンプルとして格納する時、音情報として表現可能な各単音を、図 9 D のようにウェーブ形態で格納した後、デジタル音響の分析時に該ウェーブから周波数成分を導出可能にすることができ、図 9 E のように周波数成分別の強さをもって直接格納することもできる。

【0033】

このように演奏楽器の音情報を周波数成分別の強さで直接表現するためには、フーリエ変換およびウェーブレット変換などのような周波数分析方法を使用することができる。

30

【0034】

一方、前記演奏楽器が、ピアノでなくバイオリンのような弦楽器である場合は、該弦楽器の各弦別の音情報を分類して格納する。

【0035】

なお、このような各演奏楽器別の音情報は、使用者の選択によって周期的に更新して格納するが、これは、当該楽器の音情報が時間が経つにつれて、または、温度などの周辺環境によって変化するためである。

【0036】

図 10 ~ 図 10 B は、本発明の一実施例に係るデジタル音響分析過程を示す図であって、以下、これらを参照して本発明の実施例をより詳しく説明する。

40

【0037】

まず、図 10 は、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響を分析する過程に関する処理フローを示す図である。同図を参照して、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響を分析する過程を説明すると、次の通りである。

【0038】

本発明の一実施例を行うためには、事前に楽器別の音情報を生成し格納する過程を経た後 (図示せず)、格納された楽器別の音情報のうちから実際に演奏される楽器の音情報を選択する (s100)。この時、楽器別の音情報の格納形態例は、図 9 A ~ 図 9 E の通りで

50

ある。

【0039】

また、外部からデジタル音響が入力されると（s200）、該デジタル音響信号を単位フレーム別の周波数成分に分解した後（s400）、前記デジタル音響信号の周波数成分と前記選択された演奏楽器の音情報の周波数成分とを比較・分析することによって、前記デジタル音響に含まれた単音情報をフレーム別に導出する（s500）。

【0040】

このように、外部から入力されたデジタル音響の単音情報が導出されると、該単音情報を出力する（s600）。

【0041】

このような一連の過程（s200乃至s600）を、デジタル音響入力中断されるかまたは終了命令が入力される（s300）まで繰り返して行う。

【0042】

図10Aは、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響のフレーム別の単音情報を導出する過程（s500）を示す処理フローチャートである。同図では、1つのフレームに対する情報導出過程を説明している。同図に示されたように、デジタル音響のフレーム別の単音情報を導出するため、まず、当該フレームの時間情報を導出し（s510）、該単位フレーム別の周波数成分と演奏楽器の音情報の周波数成分とを比較・分析し（s520）、各単位フレーム別に含まれた単音の音高さ別、強さ別の情報を時間情報と共に導出する。また、その導出結果を単音情報として導出する（s530）。

【0043】

また、前記過程（s530）によって導出された単音が前フレームに含まれていない新しい単音である場合（s540）、現フレームをサブフレームに区分した後（s550）、該サブフレームのうち新しい単音が含まれたサブフレームを導出し（s560）、前記サブフレームの時間情報を導出し（s570）、このとき、新しく導出された時間情報を現在導出された単音情報の時間情報に変更する（s580）。但し、このような一連の過程（s540乃至s580）は、導出された単音が低音の帯域に含まれている場合、または、時間情報の正確性が要求されない場合は、省略可能である。

【0044】

図10Bは、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響および当該演奏楽器の音情報のフレーム別の周波数情報を比較・分析する過程（s520）を示す処理フローチャートである。

【0045】

同図に示されたように、まず、前記入力されたデジタル音響信号のフレーム別に当該フレームに含まれた最も低いピーク周波数を選択する（s521）。また、当該演奏楽器の音情報のうちから前記選択されたピーク周波数を含む音情報を導出し（s522）、該音情報のうち前記過程（s521）で選択されたピーク周波数の周波数成分と最も近接したピーク情報を有する音情報を単音情報として導出する（s523）。

【0046】

このように、最も低いピーク周波数に相当する単音情報を導出し、次いで当該フレームに含まれた周波数成分のうち前記導出された単音情報に含まれる周波数成分を除去した後（s524）、当該フレームに残余のピーク周波数が存在する場合、前記過程（s521乃至s524）を繰り返して行う。

【0047】

例えば、外部から入力されたデジタル音響信号の当該フレームに、「C4、E4、G4」の3つの音が含まれた場合、前記過程（s521）によって、「C4」の基本周波数成分を、現フレームに含まれたピーク周波数成分のうち最も低いピーク周波数成分として選択する。

【0048】

10

20

30

40

50

そして、前記過程 (s 5 2 2) によって予め設定されている当該演奏楽器の音情報のうち「 C 4 」の基本周波数成分を含む音情報を導出する。この場合、一般に、「 C 4 」、「 C 3 」、「 G 2 」(など、多数の音情報が導出されるようになる。

【 0 0 4 9 】

次いで、前記過程 (s 5 2 3) によって導出された多数個の音情報のうち前記過程 (s 5 2 1) で選択されたピーク周波数成分と最も近似した「 C 4 」の音情報を当該周波数成分に関する単音情報として導出する。

【 0 0 5 0 】

また、該単音情報 (「 C 4 」) を、前記デジタル音響信号の当該フレームに含まれた周波数成分 (「 C 4 、 E 4 、 G 4 」) から除去する (s 5 2 4) 。そうすると、当該フレームには「 E 4 、 G 4 」に相当する周波数成分のみが残るようになる。このようにして当該フレームにおいて周波数成分が完全に消滅するまでに前記過程 (s 5 2 1 乃至 s 5 2 4) を繰り返して行うことで、当該フレームに含まれた全ての音に関する単音情報を導出することができる。

10

【 0 0 5 1 】

上記の例において、前記過程 (s 5 2 1 乃至 s 5 2 4) を、3回繰り返して行い、「 C 4 、 E 4 、 G 4 」の音に関する単音情報を全て導出している。

【 0 0 5 2 】

以下、〔疑似コード (p s e u d o - c o d e) 1 〕に基づいて、このような音情報を用いたデジタル音響分析方法に関する処理過程を説明する。なお、疑似コードのうち本明細書において説明しない部分は、従来のデジタル音響分析方法を参照したものである。

20

〔疑似コード 1 〕

```

line 1   デジタル音響信号(das)入力;
line 2   // FFT ウィンドウのサイズ別、ウィンドウ間の間隔別(重畳許容)に
         // dasを フレーム (frame) 単位で区分
line 3   frame = dasをフレーム別に区分(das, fft-size, overlap-size);
line 4   for 全てのframe
line 5     x = fft(frame); // フーリエ変換
line 6     peak = 最も低いピーク周波数(x);
line 7     timing = frameの時間情報;
line 8     while (peakが存在する間)
line 9       candidates = ピークを含む音情報(peak);
line 10      sound = 最も近似した音情報(candidates, x);
line 11      if soundが新しい音の始まりであれば
line 12        subframe = フレームをサブフレーム (subframe) 別に区分
                (frame, sub-size, overlap-size);
line 13        for 全てのsubframe
line 14          subx = fft(subframe);
line 15          if subxに peakがあれば
line 16            timing = subframeの時間情報;
line 17            exit-for;
line 18          end-if
line 19        end-for
line 20      end-if
line 21      result = 分析結果格納(result, timing, sound);
line 22      x = x - sound;
line 23      peak = 最も低いピーク周波数(x);
line 24    end-while
line 25  end-for
line 26  performance = 楽器別の演奏補正(result);

```

30

40

50

〔疑似コード1〕に示されたように、まず、デジタル音響信号を入力され（line 1）、フレーム別に区分（line 3）した後、各フレーム別にループ（for loop）を回りながら分析を行う（line 4～line 25）。フーリエ変換を経て周波数成分を導出し（line 5）、最も低いピーク周波数成分を探す（line 6）。次いで（line 7）では、現フレームに相当する時間情報を（line 21）で格納するために導出する。次いで、現フレームに対してピーク周波数成分が存在する間、ループを（while loop）を回りながら分析を行う（line 8～line 24）。（line 9）で現フレームのピーク周波数成分を有する音情報を導出し、（line 10）でこれらの中で現フレーム上のピーク周波数成分およびハーモニック周波数成分を比較して最も近似した音情報を導出するようになる。このとき、音情報は、現フレーム上のピーク周波数の強さに準じる強さに調節される。導出された音情報が新しい音の始まりであることを意味すると（line 11）、FFTウィンドウのサイズを小さくして正確な時間情報を抽出するようになる。

10

【0053】

この過程において、現フレームを複数個のサブフレームに分け（line 12）、各サブフレームに対してループ（for loop）を回りながら分析を行う（line 13～line 19）。フーリエ変換を経て周波数成分を導出し（line 14）、（line 6）で導出したピーク周波数が含まれているサブフレームが見つかり（line 15）、（line 16）では、サブフレームに相当する時間情報を（line 21）で格納するために導出する。（line 7）の時間情報は、FFTウィンドウのサイズが大きなフレーム単位の時間情報であるため、時間誤差の多い情報であるが、（line 16）の時間情報は、FFTウィンドウのサイズが小さなサブフレーム単位の情報であるため、時間誤差がほとんどない情報である。（line 17）で（line 13～line 19）のループ（for loop）を抜け出るため、（line 21）で格納される時間情報は、（line 7）で導出した時間情報の代わりに（line 16）で導出した正確な時間情報となる。

20

【0054】

このように（line 11）から（line 20）に至るまで新しい音が始まった場合は、単位フレームの大きさを小さくして正確な時間情報を導出するようになる。（line 21）では、導出された単音の音高さ情報および強さ情報を時間情報と共に格納し、（line 22）では（line 10）で導出された音情報を、現フレームから減じた後、再度ピーク周波数を探すようになる（line 23）。上記の過程を繰り返しながら全てのデジタル音響の分析結果は、（line 21）において結果変数（result）に格納される。

30

【0055】

なお、該結果変数（result）に格納された分析結果は、実際演奏情報として活用するには十分でない。ピアノの場合、鍵盤を押して音が発生した最初には、正確な周波数帯域に表現されないため、1つ以上のフレームを経た上で、次フレームにおいて正確に分析される可能性が多い。従って、このような場合、ピアノの音はごく短い時間（例えば、3～4つのフレームに相当する時間）の間は音が変更されないという特性を用いれば、より正確な演奏情報を導出することができる。そのため、（line 26）の過程においてこのような楽器別特性を用いて分析された結果変数（result）に格納された値を、より正確な演奏情報であるパフォーマンス（performance）に補正するようになる。

40

【0056】

図11乃至図11Dは、本発明の他の実施例によるデジタル音響分析過程を示す図であって、これらの図を参照して本発明の実施例をより詳しく説明すると、次の通りである。

【0057】

本発明の他の実施例は、楽器別の音情報と演奏しようとする楽譜情報とを同時に用いる方法に関するものであって、音情報の構築時に各単音別に含まれた周波数成分の変化に応じ

50

てできる限り全ての情報を構築することができれば、ほぼ正確に入力されたデジタル音響信号を分析することができるが、実際状況ではこのような音情報の構築は容易でないため、これを補完するための方法に関するものである。すなわち、本発明の他の実施例は、演奏しようとする演奏曲に関する楽譜情報を導出した後、既に抽出された楽器別の音情報と前記楽譜情報によって入力される音を予想し、予想された音の情報をを用いてデジタル音響を分析する方法に関するものである。

【0058】

まず、図11は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響を分析する過程を示す処理フローチャートである。同図に示したように、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響を分析する過程を説明すると、次の通りである。

10

【0059】

本発明の他の実施例を行うため、演奏楽器別の音情報を生成し格納する過程と、演奏される楽譜の楽譜情報を生成し格納する過程を予め行った後（図示せず）、その格納された演奏楽器別の音情報および楽譜情報のうちから実際に演奏される楽器の音情報および楽譜情報を選択する（t100、t200）。このとき、楽器別の音情報の格納形態の例は、「図9A～図9E」の通りである。なお、演奏される楽譜から楽譜情報を生成する方法は、本発明の範囲に属せず、現在、紙の楽譜をスキャンすると、直にミディ音楽の演奏情報に変換して格納する技術などが多数存在しているため、楽譜情報を生成し格納する方法に関する説明は省略する。

20

【0060】

但し、楽譜情報に含まれる情報の例としては、時間の経過に従う音の高さ、音の長さ情報、速さ情報、拍子情報、音の強さ情報、細部演奏情報（例えば、スタッカート、スタッカティッシモ、プラルトリラーなど）や両手演奏または多数の楽器演奏に対する演奏区分情報などが挙げられる。

【0061】

上記のように、当該演奏楽器の音情報および楽譜情報を選択（t100、t200）した後、外部からデジタル音響が入力されると（t300）、該デジタル音響信号を単位フレーム別の周波数成分に分解し（t500）、前記デジタル音響信号の周波数成分と前記選択された演奏楽器別の音情報の周波数成分および楽譜情報を比較・分析し、前記デジタル音響信号に含まれた演奏誤り情報および単音情報を導出する（t600）。

30

【0062】

このように、外部から入力されたデジタル音響の単音情報および演奏誤り情報が導出されると、該単音情報を出力する（t700）。

【0063】

なお、前記演奏誤り情報に基づいて演奏の正確度を判別したり（t800）、該演奏誤り情報が演奏者の意図によって演奏された音（例えば、変奏音）である場合、これを既存の楽譜情報に追加する過程（t900）を選択的に行うこともできる。

【0064】

図11Aは、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響のフレーム別の単音情報および演奏誤り情報を導出する過程（t600）に関する処理フローチャートである。同図では、1つのフレームに関する単音情報および演奏誤り情報の導出過程を説明している。同図に示されたように、デジタル音響のフレーム別の単音情報および演奏誤り情報を導出するため、先ず当該フレームの時間情報を導出し（t610）、該単位フレーム別の周波数成分と演奏楽器の音情報の周波数成分および楽譜情報を比較・分析し（t620）、各単位フレーム別に含まれた単音の音高さ別、強さ別の情報を時間情報と共に導出する。また、その分析結果として導出された単音情報および演奏誤り情報をフレーム別に導出する（t640）。

40

【0065】

なお、前記過程（t640）によって導出された単音が前フレームに含まれていない新し

50

い単音である場合 (t 6 5 0)、現フレームをサブフレームに区分した後 (t 6 6 0)、該サブフレームのうち新しい単音が含まれたサブフレームを導出し (t 6 7 0)、前記導出されたサブフレームの時間情報を導出する (t 6 8 0)。また、この時に新しく導出された時間情報によって現在導出されている単音情報の時間情報を変更する (t 6 9 0)。但し、このような一連の過程 (t 6 5 0 乃至 t 6 9 0) は、前述の本発明の一実施例のように、単音が低音の帯域に含まれている場合や時間情報の正確性が要求されない場合は省略可能である。

【 0 0 6 6 】

図 1 1 B および図 1 1 C は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響および当該楽器の音情報および楽譜情報のフレーム別の周波数成分を比較・分析する過程 (t 6 2 0) を示す処理フローチャートである。

10

【 0 0 6 7 】

同図に示されたように、先ず、楽譜情報を参照して当該演奏楽器の演奏進行に従うデジタル音響信号のフレーム別の演奏期待値をリアルタイムで生成し、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在するか否かをフレーム単位で確認する (t 6 2 1)。

【 0 0 6 8 】

また、前記確認 (t 6 2 1) の結果、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在しない場合、当該フレームのデジタル音響信号に含まれた周波数成分が演奏誤り情報であるか否かを確認し、演奏誤り情報および単音情報を導出した後、該演奏誤り情報および単音情報として導出された音情報の周波数成分を当該フレームのデジタル音響信号から除去する過程 (t 6 2 2 乃至 t 6 2 8) を行う。

20

【 0 0 6 9 】

すなわち、前記入力されたデジタル音響信号のフレーム別に当該フレームに含まれた最も低いピーク周波数を選択した後 (t 6 2 2)、当該演奏楽器の音情報のうちから前記選択されたピーク周波数を含む音情報を導出し (t 6 2 3)、該音情報のうち前記過程 (t 6 2 2) で選択されたピーク周波数成分と最も近接したピーク情報を有する音情報を演奏誤り情報として導出する (t 6 2 4)。また、前記演奏誤り情報が、連続する音の楽譜情報において次フレームで演奏される音符に含まれる場合 (t 6 2 5)、該演奏誤り情報に相応する音を演奏期待値に追加した後 (t 6 2 6)、該演奏誤り情報を単音情報として導出する (t 6 2 7)。なお、このように前記過程 (t 6 2 4、t 6 2 7) で演奏誤り情報または単音情報として導出された音情報の周波数成分をデジタル音響信号の当該フレームから除去する (t 6 2 8)。

30

【 0 0 7 0 】

また、前記確認 (t 6 2 1) の結果、フレーム別の演奏期待値のうち当該フレームのデジタル音響信号とマッチングされない演奏期待値が存在する場合、前記デジタル音響信号と前記演奏期待値とを比較・分析して当該フレームのデジタル音響信号に含まれた単音情報を導出し、前記単音情報として導出された音情報の周波数成分を該単音情報が含まれたフレームのデジタル音響信号から除去する過程 (t 6 3 0 乃至 t 6 3 4) を行う。

40

【 0 0 7 1 】

すなわち、前記演奏期待値に含まれた音情報のうち前記デジタル音響信号の当該フレームに含まれた周波数成分とマッチングされない最も低い音の音情報を選択し (t 6 3 0)、前記選択された音情報の周波数成分が前記デジタル音響信号の当該フレームに含まれた周波数成分に含まれる場合 (t 6 3 1)、該音情報を単音情報として導出した後 (t 6 3 2)、該音情報の周波数成分をデジタル音響信号の当該フレームから除去する (t 6 3 3)。このとき、前記過程 (t 6 3 0) で選択された音情報の周波数成分が前記デジタル音響信号の当該フレームに含まれた周波数成分に含まれていない場合、既に設定された演奏期待値を訂正する過程 (t 6 3 5) を行う。このような一連の処理過程 (t 6 3 0 乃至 t 6

50

33およびt635)を、演奏期待値のうちマッチングされない音が消滅するまで(t634)繰り返して行う。

【0072】

また、図11Bおよび図11Cに示された全ての処理過程(t621乃至t628およびt630乃至t635)は、フレーム単位のデジタル音響信号に含まれたピーク周波数成分が全て消滅するまで(t629)繰り返して行う。

【0073】

図11Dは、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて生成された演奏期待値を訂正する過程(t635)に関する処理フローチャートである。同図に示されたように、前記選択された音情報の周波数成分が当該フレームの前フレームに所定の回数(N)以上連続して含まれておらず(t636)、該音情報の周波数成分がある時点で前記デジタル音響信号に一度以上含まれた履歴がある(t637)音情報である場合、演奏期待値から除去する(t639)。なお、前記選択された音情報の周波数成分のうちデジタル音響信号の当該フレームに所定の回数(N)以上連続して含まれておらず(t636)、当該フレームに一度も含まれていない(t637)音情報である場合は、演奏誤り情報として導出した後(t638)、演奏期待値から除去する(t639)。

10

【0074】

以下、〔疑似コード(pseudo-code)2〕に基づいて、音情報および楽譜情報を用いたデジタル音響分析方法に関する処理過程を説明する。

〔疑似コード2〕

20

```

line 1 楽譜情報(score)入力;
line 2 デジタル音響信号(das)入力;
line 3 frame = フレーム別に区分(das, fft-size, overlap-size);
line 4 現演奏値(current) = 前演奏値(prev) = NULL;
line 5 次演奏値(next) = 最初に演奏される単音;
line 6 for 全てのframe
line 7   x = fft(frame);
line 8   timing = frameの時間情報;
line 9   for 全ての単音(sound) in next & not in (current, prev)
line 10     if soundが frameに含まれていれば
line 11       prev = prev + current;
line 12       current = next;
line 13       next = 次に演奏される単音;
line 14       exit-for;
line 15     end-if
line 16   end-for
line 17   for 全ての単音(sound) in prev
line 18     if soundが frameに含まれていなければ
line 19       prev = prev - sound;
line 20     end-if
line 21   end-for
line 22   for 全ての単音(sound) in (current, prev)
line 23     if soundが frameに含まれていなければ
line 24       result = 演奏誤り格納(result, timing, sound);
line 25     else // soundが frameに含まれていれば
line 26       sound = 強さの調節(sound, x);
line 27       result = 分析結果格納(result, timing, sound);
line 28       x = x - sound;
line 29     end-if

```

30

40

```

line 30   end-for
line 31   peak = 最も低いピーク周波数(x);
line 32   while (peakが存在する間)
line 33     candidates = ピークを含む音情報(peak);
line 34     sound = 最も近似した音情報(candidates, x);
line 35     result = 演奏誤り情報 (result, timing, sound);
line 36     x = x - sound;
line 37     peak = 最も低いピーク周波数(x);
line 38   end-while
line 39   end-for
line 40   performance = 楽器別の演奏情報(result);

```

10

〔疑似コード2〕に示されたように、楽譜情報と音情報とを同時に使用するため、先ず、(line 1)で楽譜情報を入力される。該疑似コードは、楽譜情報(score)のうちから音符情報のみを用いて演奏される各単音の情報をデジタル音響と比較・分析する例であって、最も基本的なものと理解すれば良い。(line 1)で入力された楽譜情報は、(line 5)および(line 13)で次演奏値(next)を導出するために使用される。すなわち、フレーム別の演奏期待値の導出を行うために使用される。次いで、次のステップは、前述の音情報を用いる疑似コードにおける説明と同様に、デジタル音響信号を入力され(line 2)、フレーム別に区分する(line 3)。(line 4)で現演奏値(current)および前演奏値(prev)をナル(null)として指定するが、楽譜上でデジタル音響信号の現フレームに含まれた音に対応される音の情報が現演奏値(current)であり、デジタル音響信号の前フレームに含まれた音に対応される音の情報が前演奏値(prev)であり、デジタル音響信号の次フレームに含まれるものと予想される音の情報が次演奏値(next)である。

20

【0075】

次いで、全てのフレームに対してfor loopを回りながら分析を行う(line 6)。各フレーム別にフーリエ変換を経て周波数成分を導出し(line 7)、(line 9)において楽譜上の次の演奏部分に進行されたか否かを判断する。すなわち、現演奏値(current)および前演奏値(prev)に含まれておらず、次演奏値(next)にのみ含まれている新しい音がデジタル音響信号の現フレームに含まれていれば、楽譜上の次の演奏部分に進行されたものと判断する。このとき、前演奏値(prev)、現演奏値(current)、また、次演奏値(next)の値を適当に変更させる。(line 17~line 21)では、前演奏値(prev)に含まれた音のうちデジタル音響信号の現フレームに含まれていない音(消滅された音)を探して前演奏値(prev)から除去する。この過程によって、楽譜上では既に過ぎたが、実際の演奏では音が残っている部分に対して消滅処理を行うようになる。(line 22~line 30)では、現演奏値(current)および前演奏値(prev)に含まれた全ての音情報(sound)に対して、各音情報がデジタル音響信号の現フレームにあるかを判断し、各音情報がデジタル音響信号の現フレームに含まれていなければ、楽譜と違って演奏されたことを結果として格納し、各音情報がデジタル音響信号の現フレームに含まれていれば、現フレームに含まれた音の強さに合わせて音情報(sound)を導出した後、音の高さ情報および強さ情報を時間情報と一緒に格納する。このように、(line 9)~(line 30)は、楽譜情報からデジタル音響信号の現フレームに含まれた音に対応する単音を現演奏値(current)と、デジタル音響信号の前フレームに含まれた音に対応される単音を前演奏値(prev)と、デジタル音響信号の次フレームに含まれるものと予想される単音を次演奏値(next)と設定し、前演奏値と現演奏値を演奏期待値にしておき、ここに含まれた音を基準にしてデジタル音響信号を分析しているため、一層正確かつ速やかに分析することが可能となる。

30

40

【0076】

さらに、楽譜情報と違って演奏される場合に備えて(line 31)のステップを追加

50

した。楽譜情報に含まれた音を分析した後、未だピーク周波数が残っていれば、違って演奏された部分であるため、前述の音情報を用いる疑似コード1に記載のアルゴリズムを用いて各ピーク周波数に相当する単音を導出し、これを疑似コード2 (line 23) のように楽譜と違って演奏されたことを格納する。疑似コード2では、楽譜情報を用いる方法について重点を置いて説明しているため、細部的な内容は省略しているが、音情報および楽譜情報を用いる方法は、前述の音情報のみを用いる方法と同様に、正確な時間情報を導出するために単位フレームの大きさを減らしながら分析する方法で、疑似コード1の (line 11) ~ 疑似コード1の (line 20) に相当するステップを含むことができる。

【0077】

疑似コード2においても、結果変数 (result) に格納された分析結果および演奏誤り結果は、実際演奏の情報として活用するには十分でない。疑似コード1の場合と同じ理由から、また、楽譜情報では同じ時点で複数個の音が始まるに対して実際演奏では各音の間にごく短い時間差があり得るという点を勘案し、(line 40) で楽器別の特性および演奏者の特性を勘案して分析された結果変数 (result) をパフォーマンス (performance) に補正するようになる。

【0078】

以下、上記のような本発明を立証するため、デジタル音響および演奏楽器の音情報の周波数特性についてより詳しく説明する。

【0079】

図12は、図1および図2に示された楽譜の第1小節をピアノで演奏し、該音を周波数別に分析した図である。即ち、図1および図2に示されたベートーベンのピアノソナタ第8番第2楽章の第1小節のピアノ演奏曲をスペクトログラム (Spectrogram) で示したものである。ここで、使用されたピアノはYOUNG CHANGピアノ社製のグランドピアノであり、ソニ社のノート型パソコンにマイクを連結してウィンドウ補助プログラムに備えられた録音機で録音した。スペクトログラムを分析して示すプログラムは、R. S. Horneが開発して公開したフリーウェア「Spectrogram 5.1.6」バージョンを用いた。設定項目としては、Scaleを90dBとし、Time Scaleは、5msec、FFT (Fast Fourier Transformation) sizeは、8192とし、他のものには基本値を適用した。Scaleは、-90dBより小さな音は無視して表示しないという意味であり、Time Scaleは、FFTウィンドウを5msec毎に重畳させながらフーリエ変換 (Fourier) を経て図示化したことを示すものである。

【0080】

図12の上部に示された線 (100) は、入力されたデジタル音響信号の強さを入力された、そのまま示したものであり、その下に各周波数帯域別にデジタル音響に含まれた周波数成分が表示されている。黒色に近いほど強さの大きな周波数成分を包含している。スペクトログラムの縦線の1つは、当該地点を基準にして8192個のサンプルをフーリエ変換して各周波数の成分別に強さに従って色を示すものであり、横軸は、時間の進行を示すものであるため、個別周波数成分に対して時間の流れによる強さの変化が一目で把握できる。図12と図2とを共に参照して、図2の楽譜上に示される各音の高さに応じて相当する基本周波数とハーモニック周波数とが図12に示されていることがわかる。

【0081】

図13A乃至図13Gは、前記楽譜の第1小節に含まれた各音をピアノで演奏し、該演奏音を周波数別に分析した図であって、図2の第1小節に含まれた各音を同じ環境下で別に録音した後、その結果をスペクトログラムで示したものである。即ち、図13Aは、C4、図13Bは、A2、図13Cは、A3、図13Dは、E3、図13Eは、B3、図13Fは、D3、図13Gは、G3に相当するピアノ音をスペクトログラムで示したものである。各図は、すべて4秒間の周波数別の強さ情報を有しており、分析時の設定項目は、図12の場合と同様にした。C4音の場合、262Hzを基本周波数として有し

10

20

30

40

50

、その n 倍数である 523 Hz 、 785 Hz 、 1047 Hz などをハーモニック周波数として有するが、図 13 A でこれを確認することができる。即ち、 262 Hz と 523 Hz に示される周波数成分は強く（黒色に近く）示し、 785 Hz の場合より高い倍数の周波数帯域にいくほど概ね強さが弱くなるのがわかる。図面では、 $C4$ の基本周波数およびハーモニック周波数は、いずれも「 $C4$ 」と表示した。

【0082】

なお、 $A2$ 音は、基本周波数として 104 Hz を有するが、図 13 B に示されたように、 $A2$ 音の基本周波数は弱く示され、 $A2$ 音のハーモニック周波数が基本周波数より遥かに強く示されることがわかる。図 13 のみを参照して、特に、 $A2$ 音の 3 倍数ハーモニック周波数（ 311 Hz ）帯域が最も強く示されるため、単に周波数の強さで音の高さを決定するとすれば、 $A2$ 音を 311 Hz を基本周波数として有する $E4$ 音と誤って認識することがあり得る。

10

【0083】

また、図 13 C 乃至図 13 G においても、音の強さのみをもって音の高さを決定する場合、このような誤りを犯すことがある。

【0084】

図 14 A 乃至図 14 G は、前記楽譜の第 1 小節に含まれた各音に対する周波数成分が図 12 に示された状態を示すものである。

【0085】

図 14 A は、 $C4$ 音に相当する図 13 A の周波数成分が図 12 に示された状態を表示したものである。図 13 A に示された音の強さは、図 12 に使用された音の強さより大きいため、図 12 の上端部にある $C4$ 音のハーモニック周波数は薄く示されるか、見分け難いほど小さな強さに示されている。しかし、図 13 A の周波数強さを図 12 に示された $C4$ 基本周波数の強さに合うように低くした上で比較すると、図 12 において $C4$ 音の周波数成分が、図 14 A でのように含まれていることがわかる。

20

【0086】

図 14 B は、 $A2$ 音に相当する図 13 B の周波数成分が図 12 に示される状態を表示したものである。図 13 B に示された音の強さは、図 12 で演奏された $A2$ 音の強さより遥かに強いため、基本周波数の成分と上端部のハーモニック周波数成分が明確に示されるが、図 12 では薄く示されており、特に上端部ではハーモニック周波数がほとんど表示されていない。これもやはり図 13 B の周波数の強さを図 12 に示された $A2$ の基本周波数の強さに合うように低くした上で比較すると、図 12 において $A2$ 音の周波数成分が図 14 B でのように含まれていることがわかる。図 14 B において 5 倍数のハーモニック周波数成分が強く示されるのは、図 14 A において $C4$ 音の 2 倍数ハーモニック周波数成分と重畳されるためであり、 $A2$ 音の 5 倍ハーモニック周波数は 519 Hz であり、 $C4$ 音の 523 Hz であるため、同一の帯域に重畳されて図に表われている。続けて図 14 B を分析すると、 $A2$ 音の 5 倍数、10 倍数、15 倍数などに相当するハーモニック周波数帯域が $C4$ 音の 2 倍数、4 倍数、6 倍数ハーモニック周波数などと重畳され、図 13 B に示される基本周波数とハーモニック周波数との間の相対的な大きさより大きく示されていることがわかる（参考として、弱い強さの音は、スペクトログラムで薄く示されるため、目で各周波数成分を明確に見分け得るようにするため、図 13 A ~ 図 13 G の単音を録音する時、図 12 に含まれた実際音の強さより大きな音で録音した）。

30

40

【0087】

図 14 C は、 $A3$ 音に相当する図 13 C の周波数成分が図 12 に示される状態を表示した図である。図 13 C に示された音の強さは、図 12 で演奏された $A3$ 音の強さより強いため、図 13 C に表示された周波数成分が、図 14 C より強く示されている。なお、上記の単音に比べて図 14 C では $A3$ 音の成分のみを探すのはあまり容易でない。その理由は、他の音の基本周波数およびハーモニック周波数帯域が重畳される部分が多く、他の音の演奏時に $A3$ 音が弱く立て、その内消えるためである。先ず、 $A3$ 音の全ての周波数成分は、 $A2$ 音において 2 倍数に相当するハーモニック周波数の成分がそのまま重

50

畳されている。さらに、A₃ 音の5倍音ハーモニック周波数は、C₄ 音の4倍音ハーモニック周波数と重畳されるため、この部分では、音の二度演奏の間、切れる部分を見つけ難い。他の周波数成分は、共通的に中間部分において強さが弱くなるため、この部分では、A₂ 音のハーモニック周波数成分が示されていることがわかり、この部分においてA₃ 音は、中断されてから再度演奏されたことがわかる。

【0088】

図14Dは、E₃ 音に相当する図13Dの周波数成分が図12に示される状態を表示した図である。図13Dに示された音の強さは、図12において演奏されたE₃ 音の強さより強いため、図13Dに表示された周波数成分が図14Dのそれより強く示されている。E₃ 音は、4回演奏されたが、初の二回演奏の間は、E₃ 音の2倍音および4倍音のハーモニック周波数成分が、A₂ 音の3倍音および6倍音のハーモニック成分と重畳するため、この部分では、二回演奏の間、続けて他の音のハーモニック周波数成分が示されていることが確認される。また、E₃ 音の5倍音ハーモニック周波数成分は、C₄ 音の3倍音ハーモニック周波数成分と重畳されたように示されるため、一番目の音と二番目の音との間で当該周波数成分が続けて示されていることがわかる。三番目の音と四番目の音においては、E₃ 音の3倍音ハーモニック周波数成分がB₃ 音の2倍音ハーモニック周波数成分と重畳するため、E₃ 音の非演奏の間も続けて当該周波数成分が示されていることがわかる。また、E₃ 音の5倍音ハーモニック周波数成分が、G₃ 音の4倍音ハーモニック周波数成分と重畳するため、G₃ 音とE₃ 音とが交代に演奏されたにもかかわらず、G₃ 音の4倍音ハーモニック周波数成分とE₃ 音の5倍音ハーモニック周波数成分とはずっと連結されて示されている。

10

20

【0089】

図14Eは、B₃ 音に相当する図13Eの周波数成分が図12に示されている状態を表示したものである。図13Eに示された音の強さは、図12で演奏されたB₃ 音の強さより少し強いため、図13Eに表示された周波数成分が強く示されている。しかし、図14Eでは、図13Eに表示された周波数成分がほとんどそのままマッチングされていることが確認される。図13Eに示されたように、音が弱くなりながらB₃ 音の上端部のハーモニック周波数は良く表示されない程度に弱くなることがわかり、図14Eにおいても同様に右側に進行しながら上端部のハーモニック周波数の強さが弱くなることが確認できた。

30

【0090】

図14Fは、D₃ 音に相当する図13Fの周波数成分が図12に示される状態を表示した図である。図13Fに示された音の強さは、図12で演奏されたD₃ 音の強さより強いため、図13Fに表示された周波数成分が図14Dのそれより強く示されている。しかし、図14Fにおいても図13Fに表示された周波数成分がそのままマッチングされていることが確認される。特に、図13FにおいてD₃ 音の9倍音ハーモニック周波数の強さは、10倍音ハーモニック周波数の強さより小さく示されているが、図14FにおいてもD₃ 音の9倍音ハーモニック周波数の強さは、非常に弱く示され、10倍音ハーモニック周波数の強さより小さく示されていることが確認される。なお、図14FにおいてD₃ 音の5倍音および10倍音ハーモニック周波数は、図14Eにおいて表示されたB₃ 音の3倍音および6倍音ハーモニック周波数と重畳するため、他のハーモニック周波数より強く示されている。D₃ 音の5倍音ハーモニック周波数は、693 Hzで、B₃ 音の3倍音ハーモニック周波数は、699 Hzで、隣接しているため、スペクトログラムで重畳して現われるようになる。

40

【0091】

図14Gは、G₃ 音に相当する図13Gの周波数成分が図12に示される状態を表示した図である。図13Gに示された音の強さは、図12で演奏されたG₃ 音の強さより少し強いため、図13Gに表示された周波数成分が図14Gのそれより強く示されている。図14Gでは、G₃ 音の強さが図14CでのA₃ 音の強さより強く演奏されたので、明確に各周波数成分を探し出すことができる。さらに、図14Cや図14Fでのように他の音の

50

周波数成分と重畳する部分がほとんど示されていないため、個別周波数成分が目視で容易に確認可能である。但し、G 3 音の 4 倍数ハーモニック周波数帯域と図 1 4 D に表示された E 3 音の 5 倍数ハーモニック周波数帯域とがそれぞれ 7 8 4 H z、7 7 8 H z で近似しているが、E 3 音と G 3 音の演奏時間が互いに異なっているため、図 1 4 G において G 3 音の 4 倍数ハーモニック周波数成分の 2 つ部分の間で僅かに下方に E 3 音の 5 倍数ハーモニック周波数成分が示されていることがわかる。

【 0 0 9 2 】

図 1 5 は、図 1 2 に表示された各周波数を、図 2 の楽譜に含まれた音の周波数のそれぞれと比較して示した図である。すなわち、今まで図 1 2 の周波数成分を分析した内容が一目で確認できるように、相当する各周波数成分に表示をした。前述の本発明の方法では、図 1 2 に示された周波数成分を分析するため、図 1 3 A ~ 図 1 3 G に表現された各単音の周波数成分を用いた。その結果として、図 1 5 の図面が得られ、これによって入力されたデジタル音響を分析するために演奏楽器の音情報を用いる方法の概要を確認することができる。すなわち、前述の本発明の方法では、演奏楽器の音情報として各単音の実際音を入力され、各音に含まれた周波数成分を用いた。

10

【 0 0 9 3 】

本明細書では、フーリエ変換の F F T を用いて周波数成分を分析したが、周波数成分の分析技術としては、その他にも、ウェーブレット (W a v e l e t) や他のデジタル信号処理アルゴリズムで開発した多くの技術が使用できるのは勿論である。即ち、本明細書では、説明を容易にするため、最も代表的なフーリエ変換技術を利用しているが、本発明を限定するためのものではない。

20

【 0 0 9 4 】

なお、図 1 4 A 乃至図 1 4 G および図 1 5 に至るまで、各音の周波数成分が表現される時間情報は、実際に演奏された時間情報と違って示されていることがわかる。特に、図 1 5 において参照符号 (1 5 0 0、1 5 0 1、1 5 0 2、1 5 0 3、1 5 0 4、1 5 0 5、1 5 0 6、1 5 0 7) で表示された時間が、実際演奏が始まった部分であるにもかかわらず、周波数成分は、以前の時間から表現されていることがわかる。また、各周波数成分が、音の終わり時間を超えて示されることもある。これは、時間の流れに従う周波数の分析を正確にするため、F F T ウィンドウのサイズを 8 1 9 2 にしたので発生した時間上の誤差である。この誤差の範囲は、F F T ウィンドウのサイズに応じて決定され、本例で使用されたサンプリング比率は、2 2 , 0 5 0 H z であり、F F T ウィンドウは、8 1 9 2 サンプルであるため、8 , 1 9 2 ÷ 2 2 , 0 5 0 = 0 . 3 7 秒の誤差を含んでいる。即ち、F F T ウィンドウのサイズを大きくすると、単位フレームの大きさが大きくなって区分可能な周波数の間隔が狭くなるため、各音の高さによる周波数成分を正確に分析できるが、時間上の誤差を有するようになり、F F T ウィンドウのサイズを小さくすると、区分可能な周波数の間隔が広がるため、低音域において隣接した音が区分できないが、時間上の誤差は少なくなる。選択的に、サンプリング比率を高くしても前述のように時間上の誤差範囲が狭くなる。

30

【 0 0 9 5 】

図 1 6 A 乃至図 1 6 D は、このように F F T ウィンドウのサイズ変化による誤差範囲の変化を説明するため、図 1 および図 2 に示された楽譜の第 1 小節に関する演奏音を互いに異なった F F T ウィンドウのサイズで分析した結果を示す図である。

40

【 0 0 9 6 】

図 1 6 A は、F F T ウィンドウのサイズを 4 0 9 6 に設定した後、F F T 変換を行った場合に関する分析結果を示し、図 1 6 B は、F F T ウィンドウのサイズを 2 0 4 8 に設定した後、F F T 変換を行った場合に関する分析結果を示し、図 1 6 C は、F F T ウィンドウのサイズを 1 0 2 4 に設定した後、F F T 変換を行った場合に関する分析結果を示し、図 1 6 D は、F F T ウィンドウのサイズを 5 1 2 に設定した後、F F T 変換を行った場合に関する分析結果を示すものである。

【 0 0 9 7 】

50

一方、図15の場合、FFTウィンドウのサイズを8192に設定した後、FFT変換を行った場合に関する分析結果を示しているため、図15乃至16A~16Dに至るまで分析結果を比較すると、FFTウィンドウのサイズが大きくなるほど周波数帯域を細かく分けて細密に分析できるが、時間上の誤差が大きくなり、FFTウィンドウのサイズが小さくなるほど区分可能な周波数帯域が広がるため、細密な分析が難しい反面、時間上の誤差はそれ程小さくなることわかる。

【0098】

従って、時間上の正確度および周波数帯域区分の正確度の要求に応じてFFTウィンドウのサイズを可変的に変更しながら分析し、または、時間上の情報と周波数上の情報とを互いに異なったFFTウィンドウのサイズを用いて分析する方法などを適用すれば良い。

10

【0099】

図17Aおよび図17Bは、FFTウィンドウのサイズに応じてデジタル音響分析過程の時間誤差が変化する模様を示すための図である。このとき、図面上に白く表示された領域が当該音が発見されたウィンドウを示す。従って、図17Aの場合、FFTウィンドウのサイズが大きいため(8192)、当該音が発見されたウィンドウを示す領域が広く示され、図17Bの場合、FFTウィンドウのサイズが相対的に小さいため(1024)、当該音が発見されたウィンドウを示す領域が狭く示される。

【0100】

図17Aは、FFTウィンドウのサイズを「8192」に設定した場合に関するデジタル音響の分析結果を示す図であって、同図に示されたように、実際音が始まった位置は、「9780」であるが、FFT変換の結果、音の始まり位置と分析された位置は、当該音が発見されたウィンドウの中間地点である「12288」($= 8192 + 16384$) / 2)となる。この場合は、12288番目のサンプルと9780番目のサンプルとの間の差である2508個のサンプルに相当する時間分だけの誤差が発生する。即ち、22.5kHzサンプリングの場合、 $2508 * (1 / 22500) =$ およそ0.11秒の誤差が発生する。

20

【0101】

また、図17Bは、FFTウィンドウのサイズを「1024」に設定した場合に関するデジタル音響の分析結果を示す図であって、同図に示されたように、実際音が始まった位置は、図17Aと同様に「9780」であるが、FFT変換の結果、音の始まり位置と分析された位置は、「9728」($= 9216 + 10240$) / 2)となる。この場合は、9216番目のサンプルと10239番目のサンプルとの中間値である9728番目のサンプルの時間として分析するようになるが、その誤差は、単に52個のサンプルに過ぎず、上記のように22.5kHzサンプリングの場合、上記のような計算方法によって時間としてはおよそ0.002秒の誤差が発生する。従って、図17BのようにFFTウィンドウのサイズを小さく設定した場合により正確な分析結果が得られることわかる。

30

【0102】

図18は、本発明の実施例によって音情報および楽譜情報を用いて導出した単音情報を再度合成した音を周波数別に分析した結果を示す図である。すなわち、図1の楽譜から楽譜情報を導出し、図13A乃至図13Gに関する部分で説明した音情報を適用した。

40

【0103】

すなわち、図1から導出された楽譜情報から初の0.5秒間、C4、A3、そしてA2音が演奏されることを導出し、図13A乃至図13Cの情報から、当該音の音情報を導出した後、該情報を用いて外部から入力されたデジタル音響を分析した後、その分析結果を導出した図である。このとき、図12の初の0.5秒間に相当する部分と図14Dの前部とを比較すると、ほとんど一致していることが確認された。従って、図18の初の0.5秒間に相当する部分は、前述の〔疑似コード2〕のうちで結果変数(result)またはパフォーマンス(performance)に相当する部分であって、図12の初の0.5秒部分と一致する。

【0104】

50

以上、本発明について好適な実施例を中心に説明したが、本発明の属する技術分野で通常の知識を有する者であれば、本発明の本質的な特性から逸脱することなく変更された形態に本発明を実施することができる。したがって、本明細書に開示の実施例は、本発明を限定するものではなく、本発明を説明するためのものと解釈すべきである。すなわち、本発明の範囲は、添付の請求の範囲によってのみ制限される。

【0105】

[産業上の利用可能性]

本発明によれば、デジタル音響分析において音情報または音情報および楽譜情報を用いることによって、入力されるデジタル音響を速やかに分析することができ、その正確性も向上する。また、従来の方法では、ピアノ演奏曲など複音で構成された音楽がほとんど分析できなかったが、本発明では、音情報または音情報および楽譜情報を用いることによって、デジタル音響に含まれた単音演奏部分は勿論、複音演奏部分までも速やかにかつ正確に分析することができる効果を奏する。

10

【0106】

よって、このようなデジタル音響の分析結果を電子楽譜に直接適用することができ、また、正確な分析結果を用いて演奏情報を定量的に導出することができる。従って、このような分析結果は、その利用範囲が広く、子供向けの音楽教育から専門家の演奏レッスンに至るまで、幅広く適用することができる。

【0107】

即ち、外部から入力されるデジタル音響をリアルタイムで分析することのできる本発明の技術を応用し、現在演奏されている演奏音に対する電子楽譜上の位置をリアルタイムで把握し、次の演奏位置を電子楽譜上に自動に表示できるようにすることで、演奏者が手で楽譜をめくる必要がないため、演奏に専念することができる効果が得られる。

20

【0108】

また、分析結果である演奏情報を、予め格納された楽譜情報と比較して演奏正確度を導出することによって、演奏者に、違った演奏部分を指摘して知らせるか、前記演奏正確度をもって当該演奏者の演奏を評価することのできる資料として活用することもできる。

【図面の簡単な説明】

【図1】

図1は、ベートーベンのピアノソナタ第8番第2楽章の初めの2小節に相当する楽譜を示す図である。

30

【図2】

図2は、図1に示された楽譜の複音を単音に区分して示した楽譜図である。

【図3】

図3は、Amazing MIDIプログラムのパラメータ設定ウィンドウを示す図である。

【図4】

図4は、Amazing MIDIプログラムにおいて図1に示された楽譜の実際演奏音をミディデータに変換した結果を示す図である。

【図5】

図5は、図4の中、実際演奏音に相当する部分を黒棒で表示して組み合わせた図である。

40

【図6】

図6は、Amazing MIDIプログラムにおいて図1に示された楽譜の実際演奏音をミディデータに変換した、他の結果を示す図である。

【図7】

図7は、図6の中、実際演奏音に相当する部分を黒棒で表示して組み合わせた図である。

【図8】

図8は、デジタル音響を分析する方法に関する概念図である。

【図9A】

図9Aは、デジタル音響を分析するために使用するピアノの音情報に関する例示図である

50

。

【図 9 B】

図 9 B は、デジタル音響を分析するために使用するピアノの音情報に関する例示図である

。

【図 9 C】

図 9 C は、デジタル音響を分析するために使用するピアノの音情報に関する例示図である

。

【図 9 D】

図 9 D は、デジタル音響を分析するために使用するピアノの音情報に関する例示図である

。

【図 9 E】

図 9 E は、デジタル音響を分析するために使用するピアノの音情報に関する例示図である

。

【図 10】

図 10 は、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響を分析する過程を示す処理フローチャートである。

【図 10 A】

図 10 A は、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響のフレーム別の単音情報を導出する過程を示す処理フローチャートである。

【図 10 B】

図 10 B は、本発明の一実施例によって楽器別の音情報に基づいて外部から入力されるデジタル音響および当該楽器の音情報のフレーム別の周波数成分を比較・分析する過程を示す処理フローチャートである。

【図 11】

図 11 は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響を分析する過程を示す処理フローチャートである。

【図 11 A】

図 11 A は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響のフレーム別の単音情報および演奏誤り情報を導出する過程を示す処理フローチャートである。

【図 11 B】

図 11 B は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響および当該楽器の音情報および楽譜情報のフレーム別の周波数成分を比較・分析する過程を示す処理フローチャートである。

【図 11 C】

図 11 C は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて外部から入力されるデジタル音響および当該楽器の音情報および楽譜情報のフレーム別の周波数成分を比較・分析する過程を示す処理フローチャートである。

【図 11 D】

図 11 D は、本発明の他の実施例によって楽器別の音情報および楽譜情報に基づいて生成された演奏期待値を訂正する過程を示す処理フローチャートである。

【図 12】

図 12 は、図 1 および図 2 に示された楽譜の第 1 小節をピアノで演奏し、その音を周波数別に分析した図である。

【図 13 A】

図 13 A は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

【図 13 B】

図 13 B は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

10

20

30

40

50

【図 1 3 C】

図 1 3 C は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

【図 1 3 D】

図 1 3 D は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

【図 1 3 E】

図 1 3 E は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

【図 1 3 F】

図 1 3 F は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

10

【図 1 3 G】

図 1 3 G は、前記楽譜の第 1 小節に含まれた各音をピアノで演奏し、その音を周波数別に分析した図である。

【図 1 4 A】

図 1 4 A は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

【図 1 4 B】

図 1 4 B は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

20

【図 1 4 C】

図 1 4 C は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

【図 1 4 D】

図 1 4 D は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

【図 1 4 E】

図 1 4 E は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

30

【図 1 4 F】

図 1 4 F は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

【図 1 4 G】

図 1 4 G は、前記楽譜の第 1 小節に含まれた各音に関する周波数成分が図 1 2 に示された状態を示す図である。

【図 1 5】

図 1 5 は、図 1 2 において表示される各周波数を、図 2 の楽譜に含まれた音の周波数のそれぞれと比較して示す図である。

【図 1 6 A】

図 1 6 A は、図 1 および図 2 に示された楽譜の第 1 小節の演奏音を F F T 変換する時、設定されるウィンドウのサイズ別の周波数分析結果を示す図である。

40

【図 1 6 B】

図 1 6 B は、図 1 および図 2 に示された楽譜の第 1 小節の演奏音を F F T 変換する時、設定されるウィンドウのサイズ別の周波数分析結果を示す図である。

【図 1 6 C】

図 1 6 C は、図 1 および図 2 に示された楽譜の第 1 小節の演奏音を F F T 変換する時、設定されるウィンドウのサイズ別の周波数分析結果を示す図である。

【図 1 6 D】

図 1 6 D は、図 1 および図 2 に示された楽譜の第 1 小節の演奏音を F F T 変換する時、設

50

定されるウィンドウのサイズ別の周波数分析結果を示す図である。

【図17A】

図17Aは、FFTウィンドウのサイズに応じてデジタル音響分析過程の時間誤差が相違していることを示す図である。

【図17B】

図17Bは、FFTウィンドウのサイズに応じてデジタル音響分析過程の時間誤差が相違していることを示す図である。

【図18】

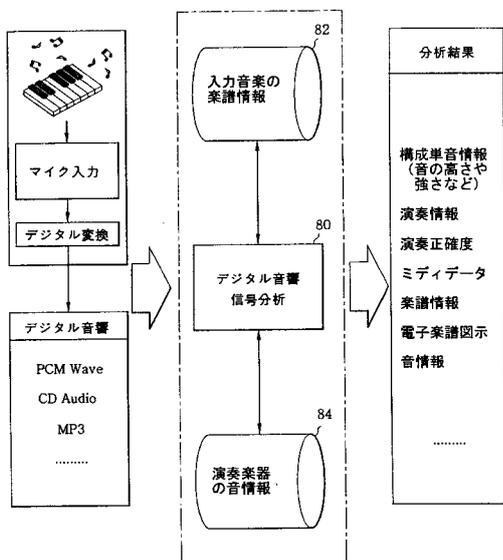
図18は、本発明の実施例によって音情報および楽譜情報に基づいて導出した単音情報を再度合成した音を周波数別に分析した結果を示す図である。

【図1】



ベートーヴェンのピアノソナタ第8番第2楽章の初めの2小節

【図8】



【図9A】

音高さ	A0	B0b	B0	C1	D1b	...	C4	...	A7	B7b	B7	C8
-----	----	-----	----	----	-----	-----	----	-----	----	-----	----	----

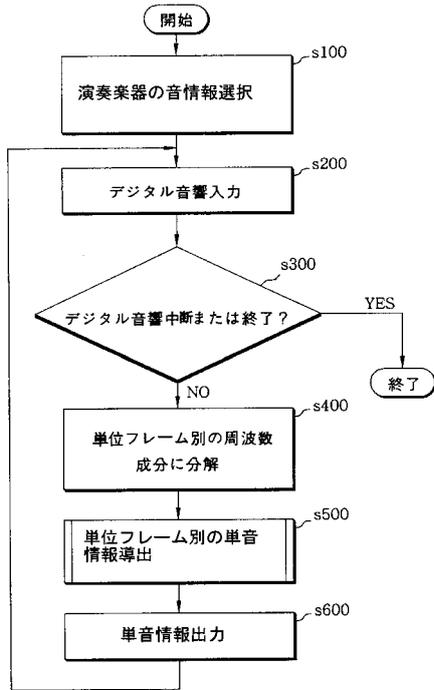
【図9B】

音強さ	-∞	...	-7dB	...	0
-----	----	-----	------	-----	---

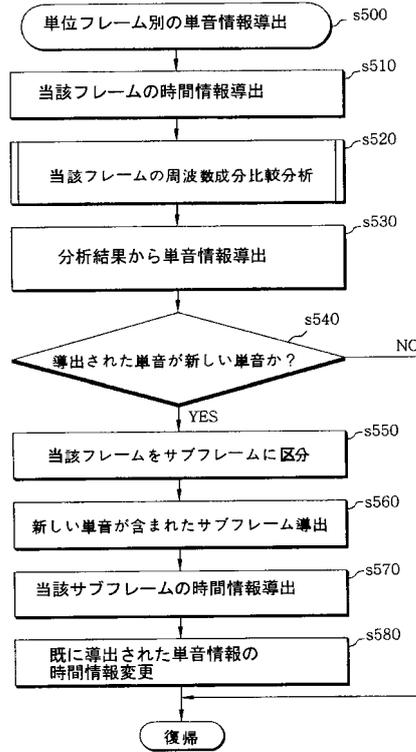
【図9C】

ペダル表	000	001	010	100	101	110	011	111
------	-----	-----	-----	-----	-----	-----	-----	-----

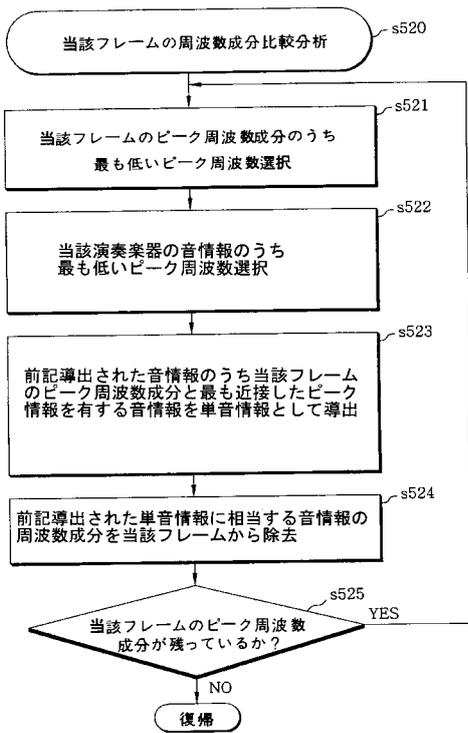
【図10】



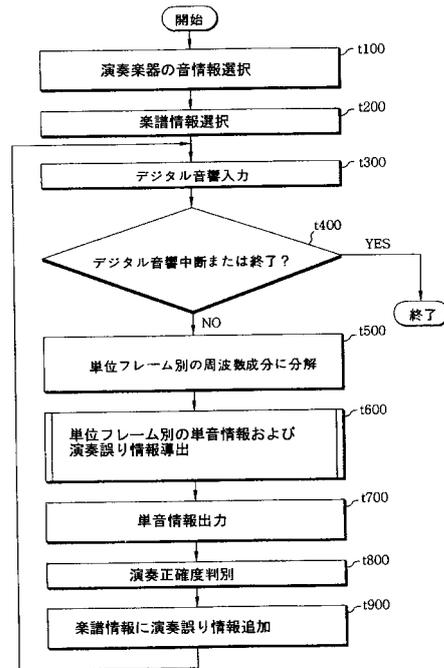
【図10A】



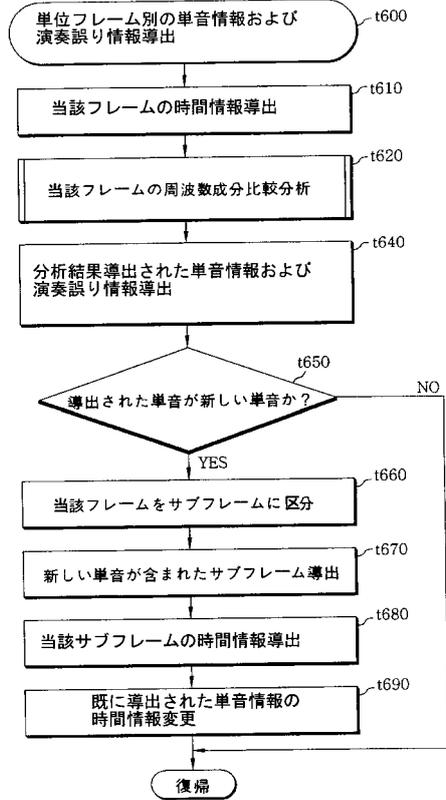
【図10B】



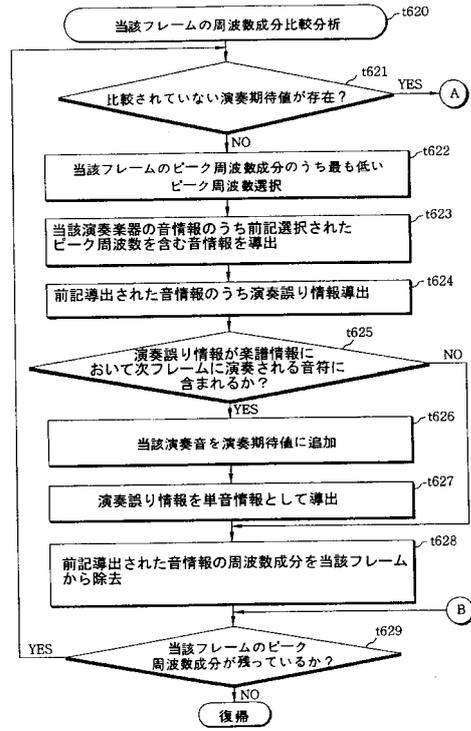
【図11】



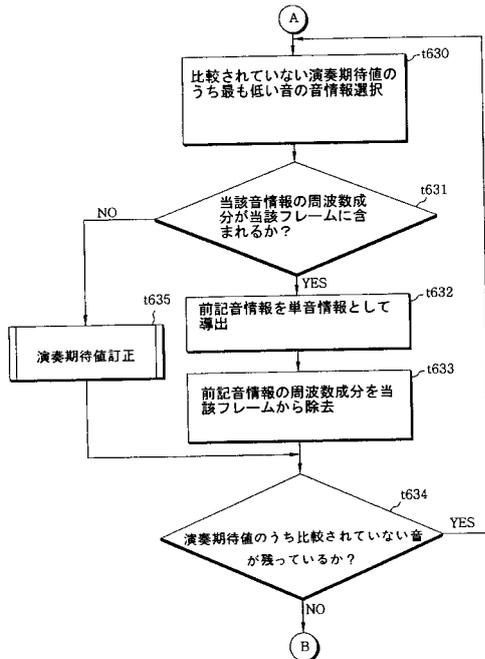
【図11A】



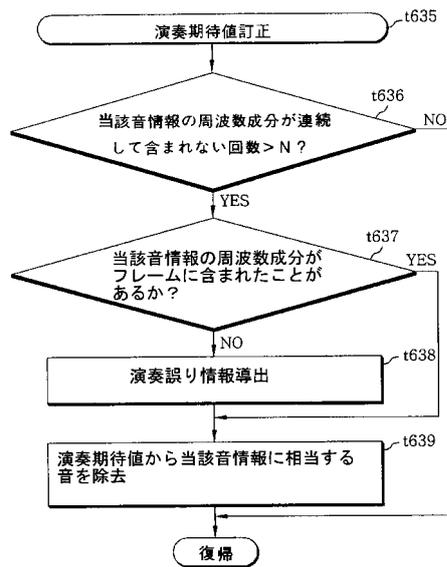
【図11B】



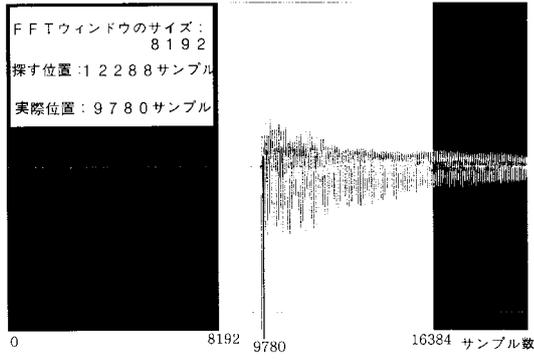
【図11C】



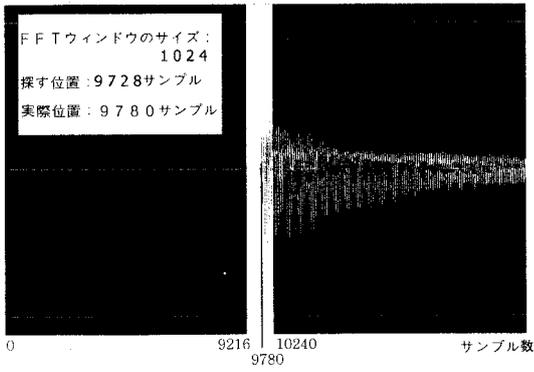
【図11D】



【図 17 A】



【図 17 B】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/47064 A1

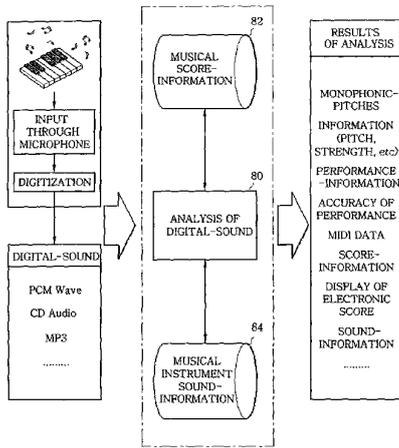
- (51) International Patent Classification⁷: G10H 1/00 [KR/KR]; 3103 Boramae-Nasan-Suite, Shindae-bang-2-Dong, Dongjak-Ku, Seoul 156-710 (KR).
- (21) International Application Number: PCT/KR01/02081
- (22) International Filing Date: 3 December 2001 (03.12.2001)
- (25) Filing Language: Korean
- (26) Publication Language: English
- (30) Priority Data: 2000-0073452 5 December 2000 (05.12.2000) KR
- (71) Applicant (for all designated States except US): AMUSETEC CO. LTD. [KR/KR]; 312 Woosung-Character, 395-73 Shindae-bang-Dong, Dongjak-Ku, Seoul 156-010 (KR).
- (72) Inventor: and
- (75) Inventor/Applicant (for US only): JUNG, Doill
- (74) Agent: KIM, Ham, Kon; Hyundai Life Insurance Bldg., 5f, 649-14 Yoksam-dong, Kangnam-ku, Seoul 135-080 (KR).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TT, TM, TR, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM).

[Continued on next page]

(54) Title: METHOD FOR ANALYZING MUSIC USING SOUNDS OF INSTRUMENTS



WO 02/47064 A1



(57) Abstract: A method for analyzing digital-sounds using sound-information of instruments and/or score-information is provided. Particularly, sound-information of instruments which were used or which are being used to generate input digital-sounds is used. Alternatively, in addition to the sound-information, score-information which were used or which are being used to generate the input digital-sounds is also used. According to the method, sound-information including pitches and strengths of notes performed on instruments used to generate the input digital-sounds is stored in advance so that monophonic or polyphonic pitches performed on the instruments can be easily analyzed. Since the sound-information of instruments and the score-information are used together, the input digital-sounds can be accurately analyzed and output as quantitative data.

WO 02/47064 A1



European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

- with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**METHOD FOR ANALYZING MUSIC USING SOUNDS OF
INSTRUMENTS**

Technical Field

5 The present invention relates to a method for analyzing
digital-sound-signals, and more particularly to a method for analyzing
digital-sound-signals by comparing frequency-components of input
digital-sound-signals with frequency-components of
performing-instruments'-sounds.

10

Background Art

 Since personal computers started to be spread in 1980's,
technology, performance and environment of computers have been
rapidly developed. In 1990's, the Internet was rapidly applied to various
15 fields of companies and personal lives. Therefore, computers are going
to be very important in every field throughout the world in the 21st
century. One of the computer music applications is musical instrument
digital interface (MIDI). MIDI is a representative computer music
technique used by musicians to synthesize and/or store musical sounds
20 of instruments or voices. At present, MIDI is a technique mainly used by
popular music composers or players.

 For example, composers can easily compose music using
computers connected to electronic MIDI instruments, and computers or
synthesizers can easily reproduce the composed MIDI music. In addition,
25 sounds produced using MIDI equipments can be mixed with vocals in
studios to be recreated as a popular song having support of the public.

 The MIDI technique has been developed in combination with
popular music and has been entered to musical education field. In other
words, MIDI uses only simple musical-information like instrument-types,
30 notes, notes'-strength, onset and offset of notes regardless of the actual

sounds of musical performance so that MIDI data can be easily exchanged between MIDI instruments and computers. Accordingly, the MIDI data generated by electronic-MIDI-pianos can be utilized in musical education using computers, which are connected to those electronic-MIDI-pianos. Therefore, many companies including Yamaha in Japan develop musical education software using MIDI.

However, the MIDI technique does not satisfy the desires of most classical musicians treasuring sounds of acoustic instruments and feelings arising when playing acoustic instruments. Because most of the classical musicians do not like the sounds and feelings of electronic instruments, they study music through traditional methods and learn how to play acoustic instruments. Accordingly, music teachers and students teach and learn classical music in academies of music or schools of music, and there is no other way for students but to fully depend on music teachers. In this situation, it is desired to apply computer technology and digital signal processing technology to the field of classical music education so that the music performed on acoustic instruments can be analyzed and the result of analysis can be expressed by quantitative performance information.

For this, digital sound analysis technology, which digital sounds are converted from the performing sounds on acoustic instruments, has been developed using computers in various viewpoints.

For example, the method of using score information to extract MIDI data from recorded digital sounds is disclosed in a master's thesis entitled "Extracting Expressive Performance Information from Recorded Music," written by Eric D. Scheirer. This thesis relates to extracting of the notes'-strength, onset timing, offset timing of each note and converting the extracted information into MIDI data. However, referring to the results of experiments described in the thesis, onset timings were accurately extracted from recorded digital sounds to some extent, but extraction of

offset timings and notes'-strength of notes were inaccurate.

Meanwhile, several small companies in the world have put initial products that can analyze simple digital sounds using a music recognition technique on the market. According to the official
5 alt.music.midi newsgroup FAQ (frequently asked questions), which is on the Internet page <http://home.sc.rr.com/cosmogony/ammfag.html>, there are some products to convert wave files into MIDI data or score data by analyzing the digital sounds in wave files. The products include Akoff Music Composer, Sound2MIDI, Gama, WIDI, Digital Ear, WAV2MID,
10 Polyaxe Driver, WAV2MIDI, IntelliScore, PFS-System, Hanauta Musician, Audio to MIDI, AmazingMIDI, Capella-Audio, AutoScore, and most recently published WaveGoodbye.

Some of these products are advertised as being able to analyze polyphonic-sounds. However, it was found that they could not analyze
15 polyphonic-sounds as a result of experiments. For this reason, the FAQ document describes that the reproduced MIDI sounds cannot be heard just like the original sounds after the sounds have been converted into MIDI format. Moreover, the FAQ document plainly states that all software published at present for converting wave files into MIDI files are of no
20 worth.

The following description concerns the result of the experiment on AmazingMIDI by Araki Software to find how it analyzes polyphonic-sounds in a wave file.

FIG. 1 is a piece of musical score used in the experiment and
25 shows first two measures of the second movement in Beethoven's Piano Sonata No. 8. In FIG. 2, the score is divided in units of monophonic notes for convenience of analysis, and the note names are assigned to the individual notes. FIG. 3 shows a parameter setting window on which a user sets parameters for converting a wave file into a MIDI file in
30 AmazingMIDI. FIG. 4 is a window showing the converted MIDI data

obtained when all parameter control bars are fixed at the right-most ends of control sections. FIG. 5 shows the expected original notes based on the score of FIG. 2 using black bars on the MIDI window of FIG. 4. FIG. 6 is another MIDI window showing the converted MIDI data obtained when
5 all the parameter control bars are fixed at the left-most ends of the control sections. FIG. 7 shows the expected original notes using black bars on the MIDI window of FIG. 6, like FIG. 5.

Referring to FIGS. 1 and 2, three notes C4, A3 \flat , and A2 \flat initially start. Then, in a state where piano keys corresponding to the notes C4 and A2 \flat are pressed, keys corresponding to notes E3 \flat , A3 \flat , and E3 \flat
10 are sequentially pressed. Next, a note B3 \flat follows the note C4, and simultaneously, notes D3 \flat and G3 follows the notes A2 \flat and E3 \flat , respectively. Then, in a state where keys corresponding to the notes B3 \flat and D3 \flat are pressed, keys corresponding to notes E3 \flat , G3, and E3 \flat
15 are sequentially pressed. Accordingly, when this wave file based on the score is converted to MIDI data, MIDI data must be configured as expressed by black bars shown in FIG. 5. However, in the real experiment, MIDI data was configured as shown in FIG. 4.

Referring to FIG. 3, AmazingMIDI allows a user to set various
20 parameters for converting wave files into MIDI files. Configuration of the MIDI data varied with the set values of these parameters very much. When the values of Minimum Analysis, Minimum Relative, and Minimum Note were set to the right-most values on the parameter input window of FIG. 3, MIDI data resulting from conversion was obtained as shown in
25 FIG. 4. When these values were set to the left-most values, MIDI data resulting from conversion was obtained as shown in FIG. 6. When FIG. 4 is compared with FIG. 6, it can be seen that there is a lot of difference between them. In other words, only frequencies having large magnitudes in a frequency domain were recognized and expressed in the form of
30 MIDI in FIG. 4, but frequencies having small magnitudes were

recognized and expressed in the form of MIDI in FIG. 6. Accordingly, MIDI data shown in FIG. 6 basically contains MIDI data of FIG. 4.

When compared with FIG. 5, FIG. 4 shows that the notes A2 \flat , E3 \flat , G3, and D3 \flat were not recognized at all, and recognition of the notes C4, A3 \flat , and B3 \flat was very different from actual performance based on the score of FIG. 2. In detail, in the case of the note C4, recognized length is only initial 25% of original length. In the case of the note B3 \flat , recognized length is less than 20% of original length. In the case of the note A3 \flat , recognized length is only 35% of original length. Moreover, many notes that were not performed were recognized. A note E4 \flat was recognized with loud notes'-strength, and unperformed notes A4 \flat , G4, B4 \flat , D5, and F5 were wrongly recognized.

When compared with FIG. 7, FIG. 6 shows that although the notes A2 \flat , E3 \flat , G3, D3 \flat , C4, A3 \flat , and B3 \flat that were actually performed were all recognized, recognized notes were very different from the performed notes. In other words, the actual sounds of the notes C4 and A2 \flat were continued since the keys were maintained pressed, but the notes C4 and A2 \flat were recognized as being stopped at least one time. In the case of the notes A3 \flat and E3 \flat , recognized onset timings and note lengths were very different from actually performed ones. In FIGS. 6 and 7, many gray bars show in addition to black bars. The gray bars indicate notes that were wrongly recognized although they were not actually performed. These wrongly recognized gray bars are more than correctly recognized bars. Although the results of experiments on programs other than AmazingMIDI program will not be described in this specification, it was proved that the results of experiments on all published programs for recognizing music were similar to the result of the experiment on AmazingMIDI program and were not satisfactory.

Although techniques of analyzing music performed on acoustic instruments using computer technology and digital signal processing

technology have been developed in various viewpoints, satisfactory results have never been obtained.

Disclosure of the Invention

5 Accordingly, the present invention aims at providing a method for analyzing music using sound-information previously stored with respect to the instruments used in performance so that the more accurate result of analyzing the performance can be obtained and the result can be extracted in the form of quantitative data.

10 In other words, it is a first object of the present invention to provide a method for analyzing music by comparing components contained in digital-sounds with components contained sound-information of musical instruments and analyzing the components so that polyphonic pitches as well as monophonic pitches can be accurately analyzed.

15 It is a second object of the present invention to provide a method for analyzing music using sound-information of musical instruments and score-information of the music so that the accurate result of analysis can be obtained and time for analyzing music can be reduced.

20 To achieve the first object of the present invention, there is provided a method for analyzing music using sound-information of musical instruments. The method includes the steps of (a) generating and storing sound-information of different musical instruments; (b) selecting the sound-information of a particular instrument to be actually played from among the stored sound-information of different musical instruments; (c) receiving digital-sound-signals; (d) decomposing the digital-sound-signals into frequency-components in units of frames; (e) comparing the frequency-components of the digital-sound-signals with the frequency-components of the selected sound-information, and
25
30 analyzing the frequency-components of the digital-sound-signals to

detect monophonic-pitches-information from the digital-sound-signals; and (f) outputting the detected monophonic-pitches-information.

To achieve the second object of the present invention, there is provided a method for analyzing music using sound-information of musical instruments and score-information. The method includes the steps of (a) generating and storing sound-information of different musical instruments; (b) generating and storing score-information of a score to be performed; (c) selecting the sound-information of a particular instrument to be actually played and score-information of a score to be actually performed from among the stored sound-information of different musical instruments and the stored score-information; (d) receiving digital-sound-signals; (e) decomposing the digital-sound-signals into frequency-components in units of frames; (f) comparing the frequency-components of the digital-sound-signals with the frequency-components of the selected sound-information and the selected score-information, and analyzing the frequency-components of the digital-sound-signals to detect performance-error-information and monophonic-pitches-information from the digital-sound-signals; and (g) outputting the detected monophonic-pitches-information and/or the detected performance-error-information.

Brief Description of the Drawings

FIG. 1 is a diagram of a score corresponding to the first two measures of the second movement in Beethoven's Piano Sonata No. 8.

FIG. 2 is a diagram of a score in which polyphonic-notes in the score shown in FIG. 1 are divided into monophonic-notes.

FIG. 3 is a diagram of a parameter-setting-window of AmazingMIDI program.

FIG. 4 is a diagram of one result of converting actual performed notes of the score shown in FIG. 1 into MIDI data using AmazingMIDI

program.

FIG. 5 is a diagram in which the actual performed notes are expressed as black bars on FIG. 4.

FIG. 6 is a diagram of another result of converting actual performed notes of the score shown in FIG. 1 into MIDI data using AmazingMIDI program.

FIG. 7 is a diagram in which the actual performed notes are expressed as black bars on FIG. 6.

FIG. 8 is a conceptual diagram of a method for analyzing digital-sounds.

FIGS. 9A through 9E are diagrams of examples of piano sound-information used to analyze digital sounds.

FIG. 10 is a flowchart of a process for analyzing input digital-sounds based on sound-information of different kinds of instruments according to a first embodiment of the present invention.

FIG. 10A is a flowchart of a step of detecting monophonic-pitches-information from the input digital-sounds in units of sound frames based on the sound-information of different kinds of instruments according to the first embodiment of the present invention.

FIG. 10B is a flowchart of a step of comparing frequency-components of the input digital-sounds with frequency-components of sound-information of a performed instrument in frame units and analyzing the frequency-components of the digital-sounds based on the sound-information of different kinds of instruments according to the first embodiment of the present invention.

FIG. 11 is a flowchart of a process for analyzing input digital-sounds based on sound-information of different kinds of instruments and score-information according to a second embodiment of the present invention.

FIG. 11A is a flowchart of a step of detecting

WO 02/47064

PCT/KR01/02081

monophonic-pitches-information and performance-error-information from the input digital-sounds in units of frames based on the sound-information of different kinds of instruments and the score-information according to the second embodiment of the present invention.

FIGS. 11B and 11C are flowcharts of a step of comparing frequency-components of the input digital-sounds with frequency-components of the sound-information of a performed instrument in frame units and analyzing the frequency-components of the digital-sounds based on the sound-information and the score-information according to the second embodiment of the present invention.

FIG. 11D is a flowchart of a step of adjusting the expected-performance-value based on the sound-information of different kinds of instruments and the score-information according to the second embodiment of the present invention.

FIG. 12 is a diagram of the result of analyzing the frequency-components of the sound of a piano played according to the first measure of the score shown in FIGS. 1 and 2.

FIGS. 13A through 13G are diagrams of the results of analyzing the frequency-components of the sounds of individual notes performed on a piano, which are contained in the first measure of the score.

FIGS. 14A through 14G are diagrams of the results of indicating the frequency-components of each of the notes contained in the first measure of the score on FIG. 12.

FIG. 15 is a diagram in which the frequency-components shown in FIG. 12 are compared with the frequency-components of the notes contained in the score of FIG. 2.

FIGS. 16A through 16D are diagrams of the results of analyzing the frequency-components of the notes, which are performed according to the first measure of the score shown in FIGS. 1 and 2, by performing

staves for right hand in the case where both hands are used for performing music on, for example, piano. In addition, in the case where at least two instruments are used, information about the staves for each instrument is included. In other words, all information on a score which
5 people applies to perform music on musical-instruments can be used as score-information. Since notation is different among composers and ages, detailed notation will not be described in this specification.

The musical-instrument sound-information 84 is previously constructed for each of the instruments used for performance, as shown
10 in FIGS. 9A through 9E, and includes information such as pitch, note strength, and pedal table. This will be further described later with reference to FIGS. 9A through 9E.

As shown in FIG. 8, in the present invention, sound-information or both sound-information and score-information are utilized to analyze
15 input digital-sounds. The present invention can accurately analyze the pitch and strength of each note even if many notes are simultaneously performed as in piano music and can detect performance-information including which notes are performed at what strength from the analyzed information in each time slot.

To analyze input digital-sounds, sound-information of
20 musical-instruments is used because each musical-note has an inherent pitch-frequency and inherent harmonic-frequencies, and pitch-frequencies and harmonic-frequencies are basically used to analyze performance sounds of acoustic-instruments and human-voices.

Different types of instruments usually have different
25 peak-frequency-components (pitch-frequencies and harmonic-frequencies). Accordingly, it is possible to analyze digital-sounds by comparing the peak-frequency-components of the digital-sounds with the peak-frequency-components of different types of
30 instruments that are previously detected and stored as sound-information

by the types of instruments.

For example, if sound-information of 88 keys of a piano is previously detected and stored, even if different notes are simultaneously performed on the piano, the sounds of simultaneously performed notes can be compared with combinations of 88 sounds previously stored as sound information. Therefore, each of the simultaneously performed notes can be accurately analyzed.

FIGS. 9A through 9E are diagrams of examples of piano sound-information used to analyze digital-sounds. FIGS. 9A through 9E show examples of sound-information of 88 keys of a piano made by Young-chang. FIGS. 9A through 9C show the conditions used for detecting sound-information of the piano. FIG. 9A shows the pitches A0 through C8 of the respective 88 keys. FIG. 9B shows note strength identification information. FIG. 9C shows identification information indicating which pedals are used. Referring to FIG. 9B, the note strengths can be classified into predetermined levels from " $-\infty$ " to "0". Referring to FIG. 9C, the case where a pedal is used is expressed by "1", and the case where a pedal is not used is expressed by "0". FIG. 9C shows all cases of use of three pedals of the piano.

FIGS. 9D and 9E show examples of the actual formats in which the sound-information of the piano is stored. FIGS. 9D and 9E show sound-information with respect to the case where the note is C4, the note strength is -7dB, and no pedals are used under the conditions of sound-information shown in FIGS. 9A through 9C. Specifically, FIG. 9D shows the sound-information stored in wave format, and FIG. 9E shows the sound-information stored in frequency format, spectrogram. Here, a spectrogram shows the magnitudes of individual frequencies in a temporal domain. The horizontal axis of the spectrogram indicates time information, and the vertical axis thereof indicates frequency information. Referring to a spectrogram as shown in FIG. 9E, frequency-components'

magnitudes can be obtained at each time.

In other words, when the sound-information of each musical-instrument is stored in the form of samples of sounds having at least one strength, sounds of each note can be stored as the sound information in wave forms, as shown in FIG. 9D, so that
5 frequency-components can be detected from the waves during analysis of digital-sounds, or the magnitudes of individual frequency-components can be directly stored as the sound-information, as shown in FIG. 9E.

In order to directly express the sound-information of each
10 musical-instrument as the magnitudes of individual frequency-components, frequency analysis methods such as Fourier transform or wavelet transform can be used.

If a string-instrument, for example a violin, is used as a musical-instrument, sound-information can be classified by different
15 strings for the same notes and stored.

Such sound-information of each musical-instrument can be periodically updated according to a user's selection, considering the fact that sound-information of the musical-instrument can vary with the lapse of time or with circumstances such as temperature.

20 FIGS. 10 through 10B are flowcharts of a method of analyzing digital-sounds according to a first embodiment of the present invention. The first embodiment of the present invention will be described in detail with reference to the attached drawings.

FIG. 10 is a flowchart of a process for analyzing input
25 digital-sounds based on sound-information of different kinds of instruments according to the first embodiment of the present invention. The process for analyzing input digital-sounds based on sound-information of different kinds of instruments according to the first embodiment of the present invention will be described with reference to
30 FIG. 10.

WO 02/47064

PCT/KR01/02081

After sound-information of different kinds of instruments is generated and stored (not shown), sound-information of the instrument for actual performance is selected in step s100. Here, the sound-information of different kinds of instruments is stored in formats as shown in FIGS. 9A through 9E.

Next, if digital-sound-signals are input in step s200, the digital-sound-signals are decomposed into frequency-components in units of frames in step s400. The frequency-components of the digital-sound-signals are compared with the frequency-components of the selected sound-information and analyzed to detect monophonic-pitches-information from the digital-sound-signals in units of frames in step s500. The detected monophonic-pitches-information is output in step s600.

The steps s200 and s400 through s600 are repeated until the input digital-sound-signals are stopped or an end command is input in step s300.

FIG. 10A is a flowchart of the step s500 of detecting monophonic-pitches-information from the input digital-sounds in units of sound frames based on the sound-information of different kinds of instruments according to the first embodiment of the present invention.

FIG. 10A shows a procedure for detecting monophonic-pitches-information with respect to a single current-frame. Referring to FIG. 10A, time-information of a current-frame is detected in step s510. The frequency-components of the current-frame are compared with the frequency-components of the selected sound-information and analyzed to detect current pitch and strength information of each of monophonic-notes in the current-frame in step s520. In step s530, monophonic-pitches-information is detected from the current pitch-information, note-strength-information and time-information.

If it is determined that current pitch in the detected

monophonic-pitches-information is a new-pitch that is not included in the previous frame in step s540, the current-frame is divided into a plurality of subframes in step s550. A subframe including the new-pitch is detected from among the plurality of subframes in step s560.

5 Time-information of the detected subframe is detected s570. The time-information of the new-pitch is updated with the time-information of the subframe in step s580. The steps s540 through s580 can be omitted when the new-pitch is in a low frequency range, or when the accuracy of time-information is not required.

10 FIG. 10B is a flowchart of the step s520 of comparing the frequency components of the input digital-sounds with the frequency-components of the sound-information of the performed instrument in frame units and analyzing the frequency-components of the digital-sounds based on the sound-information of different kinds of
15 instruments according to the first embodiment of the present invention.

Referring to FIG. 10B, the lowest peak frequency-components contained in the current frame is selected in step s521. Next, the sound-information (S_CANDIDATES) containing the selected peak frequency-components is detected from the sound-information of the performed instrument in step s522. In step s523, the sound-information (S_DETECTED) having most similar peak-frequency-components to the selected peak-frequency-components is detected as
20 monophonic-pitches-information from the sound-information (S_CANDIDATES) detected in step s522.

25 If the monophonic-pitches-information corresponding to the lowest peak frequency-components is detected, the lowest peak frequency-components are removed from the frequency-components contained in the current-frame in step s524. Thereafter, it is determined whether there are any peak frequency-components in the current-frame
30 in step s525. If it is determined that there is any, the steps s521 through

WO 02/47064

PCT/KR01/02081

s524 are repeated.

For example, in the case where three notes C4, E4, and G4 are contained in the current-frame of the input digital-sound-signals, the reference frequency-components of the note C4 is selected as the lowest
5 peak frequency-components from among peak frequency-components contained in the current-frame in step s521.

Next, the sound-information (S_CANDIDATES) containing the reference frequency-component of the note C4 is detected from the sound-information of the performed instrument in step s522. Here,
10 generally, sound-information of the note C4, sound-information of a note C3, sound-information of a note G2, and so on can be detected.

Then, in step s523, among the several sound-information (S_CANDIDATES) detected in step of s522, the sound-information (S_DETECTED) of C4 is selected as monophonic-pitches-information
15 because of the high resemblance of the selected peak frequency-components.

Thereafter, the frequency-components of the detected sound-information (S_DETECTED) (i.e., the note C4) are removed from frequency-components (i.e., the notes C4, E4, and G4) contained in the
20 current-frame of the digital-sound-signals in step s524. Then, the frequency-components corresponding to the notes E4 and G4 remain in the current-frame. The steps s521 through s524 are repeated until there are no frequency-components in the current-frame. Through the above steps, monophonic-pitches-information with respect to all of the notes
25 contained in the current-frame can be detected. In the above case, monophonic-pitches-information with respect to all of the notes C4, E4, and G4 can be detected by repeating the steps s521 through s524 three times.

Hereinafter, a method for analyzing digital-sounds using
30 sound-information according to the present invention will be described

WO 02/47064

PCT/KR01/02081

based on the following pseudo-code 1. Refer to conventional methods for analyzing digital-sounds for a part of [Pseudo-code 1] which is not described.

```
5 [Pseudo-code 1]
line 1 input of digital-sound-signals (das)
line 2 // division of the das into frames considering the size of an FFT
      // window and a space between FFT windows (overlap is
      // permitted)
10 line 3 frame = division of das into frames (das, fft-size, overlap-size)
line 4 for all frames
line 5   x = fft (frame) // Fourier transform
line 6   peak = lowest peak frequency components (x)
line 7   timing = time information of a frame
15 line 8   while (peak exist)
line 9     candidates = sound information contains (peak)
line 10    sound = most similar sound information (candidates, x)
line 11    if sound is new pitch
line 12      subframe = division of the frame into subframes
20      (frame, sub-size, overlap size)
line 13      for all subframes
line 14        subx = fft (subframe)
line 15        if subx includes the peak
line 16          timing = time information of a subframe
25 line 17          exit-for
line 18        end-if
line 19      end-for
line 20    end-if
line 21    result = new result of analysis (result, timing, sound)
30 line 22    x = x - sound
```

WO 02/47064

PCT/KR01/02081

line 23 peak = lowest peak frequency components (x)
line 24 end-while
line 25 end-for
line 26 performance = correction by instrument types (result)

5

Referring to [Pseudo-code 1], digital-sound-signals are input in line 1 and are divided into frames in line 3. Each of the frames is analyzed by repeating a for-loop in lines 4 through 25. Frequency-components are calculated through Fourier transform in line 5, and the lowest peak frequency-components are selected in line 6. Subsequently, in line 7, time-information of a current-frame to be stored in line 21 is detected. The current-frame is analyzed by repeating a while-loop while peak frequency-components exist in lines 8 through 24. Sound-information (candidates) containing the peak frequency-components of the current-frame is detected in line 9. Peak frequency-components contained in the current-frame are compared with those contained in the detected sound-information (candidates) to detect sound-information (sound) containing most similar peak frequency-components to those contained in the current-frame in line 10. Here, the detected sound-information is adjusted to a strength the same as the strength of the peak-frequency of the current-frame. If it is determined that a pitch corresponding to the sound-information detected in line 10 is new one which is not contained in the previous frame in line 11, the size of an FFT window is reduced to extract accurate time information.

To extract the accurate time-information, the current-frame is divided into a plurality of subframes in line 12, and each of the subframes is analyzed by repeating a for-loop in lines 13 through 19. Frequency-components of a subframe are calculated through Fourier transform in line 14. If it is determined that the subframe contains the

lowest peak frequency-components selected in line 6 in line 15, time-information corresponding to the subframe is detected in line 16 to be stored in line 21. The time-information detected in line 7 has a large time error in the time-information since a large-size FFT window is applied. However, the time-information detected in line 16 has a small time error in the time-information since a small-size FFT window is applied. Because the for-loop from line 13 to line 19 exits in line 17, not the time-information detected in line 7 but the more accurate time-information detected in line 16 is stored in line 21.

10 As described above, when it is determined that a pitch is new, the size of a unit frame is reduced to detect accurate time-information in lines 11 through 20. As well as the time-information, the pitch-information and the strength-information of the detected pitch are stored in line 21. The frequency-components of the sound-information detected in line 10
15 is subtracted from the current-frame in line 22, and the next lowest peak frequency-components are searched in line 23 again. The above procedure from line 9 to line 20 is repeated, and the result of analyzing the digital-sound-signals is stored as a result-variable (result) in line 21.

20 However, the stored result (result) is insufficient to be used as information of actually performed music. In the case of a piano, when a pitch is performed by pressing a key, the pitch is not represented by an accurate frequency-components during an initial stage, onset. Accordingly, the pitch can be usually analyzed accurately only after at least one frame is processed. In this case, if it is considered that a pitch
25 performed on a piano does not change within a very short time (for example, a time corresponding to three or four frames), more accurate performance-information can be detected. Therefore, the result variable (result) is analyzed considering the characteristics of a corresponding instrument and the result of analysis is stored as more accurate
30 performance-information (performance) in line 26.

FIGS. 11 through 11D are flowcharts of a method of analyzing digital sounds according to a second embodiment of the present invention. The second embodiment of the present invention will be described in detail with reference to the attached drawings.

5 In the second embodiment, both sound-information of different kinds of instruments and score-information of music to be performed are used. If all available kinds of information according to changes in frequency-components of each pitch can be constructed as sound-information, input digital-sound-signals can be analyzed very
10 accurately. However, it is difficult to construct such sound-information in an actual state. The second embodiment is provided considering the above difficulty. In other words, in the second embodiment, score-information of music to be performed is selected so that next input notes can be predicted based on the score-information. Therefore, input
15 digital-sounds are analyzed using the sound-information corresponding to the predicted notes.

FIG. 11 is a flowchart of a process for analyzing input digital-sounds based on sound-information of different kinds of instruments and score-information according to the second embodiment
20 of the present invention. The process for analyzing input digital sounds based on sound-information of different kinds of instruments and score-information according to the second embodiment of the present invention will be described with reference to FIG. 11.

After sound-information of different kinds of instruments and
25 score-information of music to be performed are generated and stored (not shown), sound-information of the instrument for actual performance and score-information of music to be actually performed are selected among stored sound-information and score-information in steps t100 and t200. Here, the sound-information of different kinds of instruments is
30 stored in formats as shown in FIGS. 9A through 9E. Meanwhile, a

method of generating score-information of music to be performed is beyond the scope of the present invention. At present, there are many types of techniques of scanning printed scores, converting the scanned scores into MIDI data, and storing the performance-information. Thus, a
5 detailed description of generating and storing score-information will be omitted.

The score-information includes pitch-information, note length-information, speed-information, tempo-information, note strength-information, detailed performance-information (e.g., staccato,
10 staccatissimo, and pralltriller), and discrimination-information for performance using two hands or a plurality of instruments.

After the sound-information and score-information are selected in steps t100 and t200, if digital-sound-signals are input in step t300, the digital-sound-signals are decomposed into frequency-components in
15 units of frames in step t500. The frequency-components of the digital-sound-signals are compared with the selected score-information and the frequency-components of the selected sound-information of the performed instrument and analyzed to detect performance-error-information and monophonic-pitches-information from
20 the digital-sound-signals in step t600. Thereafter, the detected monophonic-pitches-information is output in step t700.

Performance accuracy can be estimated based on the performance-error-information in step t800. If the performance-error-information corresponds to a pitch (for example, a
25 variation) intentionally performed by a player, the performance-error-information is added to the existing score-information in step t900. The steps t800 and t900 can be selectively performed.

FIG. 11A is a flowchart of the step t600 of detecting monophonic-pitches-information and performance-error-information from
30 the input digital-sounds in units of frames based on the

sound-information of different kinds of instruments and the score-information according to the second embodiment of the present invention. FIG. 11A shows a procedure for detecting monophonic-pitches-information and performance-error-information with respect to a single current-frame. Referring to FIG. 11A, time-information of the current-frame is detected in step t610. The frequency-components of the current-frame are compared with the frequency-components of the selected sound-information of the performed instrument and with the score-information and analyzed to detect current pitch and strength information of each of pitches in the current-frame in step t620. In step t640, monophonic-pitches-information and performance-error-information are detected from the detected pitch-information, note strength-information and time-information.

If it is determined that current pitch in the detected monophonic-pitches-information is a new one that is not included in the previous frame in step t650, the current-frame is divided into a plurality of subframes in step t660. A subframe including the new pitch is detected from among the plurality of subframes in step t670. Time-information of the detected subframe is detected t680. The time-information of the new pitch is updated with the time-information of the subframe in step t690. Similar to the first embodiment, the steps t650 through t690 can be omitted when the new pitch is in a low frequency range, or when the accuracy of time-information is not required.

FIGS. 11B and 11C are flowcharts of the step t620 of comparing frequency-components of the input digital-sounds with frequency-components of the sound-information of a performed instrument in frame units based on the score-information, and analyzing the frequency-components of the digital-sounds based on the sound-information and the score-information according to the second embodiment of the present invention.

Referring to FIGS. 11B and 11C, in step t621, an expected-performance-value of the current-frame is generated referring to the score-information in real time, and it is determined whether there is any note in the expected-performance-value that is not compared with the digital-sound-signals in the current-frame.

If it is determined that there is no note in the expected-performance-value which is not compared with the digital-sound-signals in the current-frame in step t621, it is determined whether frequency-components of the digital-sound-signals in the current-frame correspond to performance-error-information, and performance-error-information and monophonic-pitches-information are detected, and the frequency-components of sound-information corresponding to the performance-error-information and the monophonic-pitches-information are removed from the digital-sound-signals in the current-frame, in steps t622 through t628.

More specifically, the lowest peak frequency-components of the input digital-sound-signals in the current-frame are selected in step t622. Sound-information containing the selected peak frequency-components is detected from the sound-information of the performed instrument in step t623. Sound-information containing most similar peak frequency-components to the frequency-components of the selected peak frequency-components is detected from the sound-information detected in step t623 as performance-error-information in step t624. If it is determined that the current pitches of the performance-error-information are contained in next notes in the score-information in step t625, the current pitches of the performance-error-information are added to the expected-performance-value in step t626. Next, the current pitches of the performance-error-information are moved into the monophonic-pitches-information in step t627. The

WO 02/47064

PCT/KR01/02081

frequency-components of the sound-information detected as the performance-error-information or the monophonic-pitches-information in step t624 or t627 are removed from the current-frame of the digital-sound-signals in step t628.

5 If it is determined that there is any note in the expected-performance-value which is not compared with the digital-sound-signals in the current-frame in step t621, the digital-sound-signals are compared with the expected-performance-value and analyzed to detect monophonic-pitches-information from the
10 digital-sound-signals in the current-frame, and the frequency-components of the sound-information detected as the monophonic-pitches-information are removed from the digital-sound-signals, in steps t630 through t634.

More specifically, sound-information of the lowest pitch which is
15 not compared with frequency-components contained in the current-frame of the digital-sound-signals is selected from the sound-information corresponding to the expected-performance-value which has not undergone comparison in step t630. If it is determined that the frequency-components of the selected sound-information are included in
20 frequency-components contained in the current-frame of the digital-sound-signals in step t631, the selected sound-information is detected as monophonic-pitches-information in step t632. Then, the frequency-components of the selected sound-information are removed from the current-frame of the digital-sound-signals in step t633. If it is
25 determined that the frequency-components of the selected sound-information are not included in the frequency-components contained in the current-frame of the digital-sound-signals in step t631, the expected-performance-value is adjusted in step t635. The steps t630 through t633 are repeated until it is determined that every pitch in the
30 expected-performance-value has undergone comparison in step t634.

The steps t621 through t628 and t630 through t635 shown in FIGS. 11B and 11C are repeated until it is determined that no peak frequency-components are left in the digital-sound-signals in the current-frame in step t629.

5 FIG. 11D is a flowchart of the step t635 of adjusting the expected performance value according to the second embodiment of the present invention. Referring to FIG. 11D, if it is determined that the frequency-components of the selected sound-information are not included in at least a predetermined-number (N) of consecutive previous frames in step t636, and if it is determined that the frequency-components of the selected sound-information are included in the digital-sound-signals at one or more time points in step t637, the notes corresponding to the selected sound-information are removed from the expected-performance-value in step t639. Alternatively, if it is determined that the frequency-components of the selected sound-information are not included in at least a predetermined number (N) of consecutive previous frames in step t636, and if it is determined that the frequency-components of the selected sound-information are never included in the digital-sound-signals in step t637, the selected sound-information is detected as the performance-error-information in step t638, and the notes corresponding to the selected sound-information are removed from the expected-performance-value in step t639.

15 Hereinafter, a method for analyzing digital-sounds using sound-information and score-information according to the present invention will be described based on the following pseudo-code 2.

[Pseudo-code 2]

line 1 input of score information (score)
30 line 2 input of digital sound signals (das)

WO 02/47064

PCT/KR01/02081

```
line 3  frame = division of das into frames (das, fft-size, overlap-size)
line 4  current performance value (current) =
        previous performance value (prev) = NULL
line 5  next performance value (next) = pitches to be initially performed
5 line 6  for all frames
line 7    x = fft (frame)
line 8    timing = time information of a frame
line 9    for all pitches (sound) in next & not in (current, prev)
line 10      if sound is contained in the frame
10 line 11        prev = prev + current
line 12        current = next
line 13        next = pitches to be performed next
line 14      exit-for
line 15    end-if
15 line 16  end-for
line 17  for all pitches (sound) in prev
line 18    if sound is not contained in the frame
line 19      prev = prev - sound
line 20    end-if
20 line 21  end-for
line 22  for all pitches (sound) in (current, prev)
line 23    if sound is not contained in the frame
line 24      result = performance error (result, timing, sound)
line 25    else // if sound is contained in the frame
25 line 26      sound = adjustment of strength (sound, x)
line 27      result = new result of analysis (result, timing, sound)
line 28      x = x - sound
line 29    end-if
line 30  end-for
30 line 31  peak = lowest peak frequency (x)
```

WO 02/47064

PCT/KR01/02081

```
line 32 while (peak exist)
line 33   candidates = sound information contains (peak)
line 34   sound = most similar sound information (candidates, x)
line 35   result = performance error (result, timing, sound)
5 line 36   x = x - sound
line 37   peak = lowest peak frequency components (x)
line 38   end-while
line 39 end-for
line 40 performance = correction by instrument types (result)
```

10

Referring to [Pseudo-code 2], in order to use both score-information and sound-information, first, score-information is received in line 1. This pseudo-code is a most basic example of analyzing digital-sounds by comparing information of each of performed

15 pitches with the digital-sounds using only note-information in the score-information. Score-information input in line 1 is used to detect a next-performance-value (next) in lines 5 and 13. That is, the score-information is used to detect expected-performance-value for each frame. Subsequently, like Pseudo-code 1 using sound-information,

20 digital-sound-signals are input in line 2 and are divided in to a plurality of frames in line 3. The current-performance-value (current) and the previous-performance-value (prev) are set as NULL in line 4. The current-performance-value (current) corresponds to information of notes on the score corresponding to pitches contained in the current-frame of

25 the digital-sound-signals, the previous-performance-value (prev) corresponds to information of notes on the score corresponding to pitches included in the previous frame of the digital-sound-signals, and the next-performance-value (next) corresponds to information of notes on the score corresponding to pitches predicted to be included in the

30 next frame of the digital-sound-signals.

Thereafter, analysis is performed on all of the frames by repeating a for-loop in line 6 through line 39. Fourier transform is performed on a current-frame to detect frequency-components in line 7. It is determined whether performance proceeds to the next according to the score in lines 5 9 through 16. In other words, if a new pitch which is not contained in the current-performance-value (current) and the previous-performance-value (prev) but is contained only in the next-performance-value (next) is contained in the current-frame of the digital-sound-signals, it is determined that performance has proceeded to the next position in the 10 score-information. Here, the previous-performance-value (prev), the current-performance-value (current), and the next-performance-value (next) are appropriately changed. Among notes included in the previous-performance-value (prev), notes which are not included in the current frame of the digital-sound-signals are found and removed from 15 the previous-performance-value (prev) in lines 17 through 21, thereby nullifying pitches which are continued in the real performance but have passed away in the score. It is determined whether each of the pieces of sound-information (sound) contained in the current-performance-value (current) and the previous-performance-value (prev) is contained in the 20 current frame of the digital sound signals in lines 22 through 30. If it is determined that the corresponding sound-information (sound) is not contained in the current frame of the digital sound signals, the fact that the performance is different from the score is stored as the result. If it is determined that the sound-information (sound) is contained in the current 25 frame of the digital sound signals, sound-information (sound) is detected according to the strength of the sound contained in the current frame and pitch information, strength information, and time information are stored. As described above, in lines 9 through 30, score information corresponding to the pitches included in the current frame of the digital 30 sound signals is set as the current-performance-value (current),

score-information corresponding to pitches included in the previous frame of the digital-sound-signals is set as the previous-performance-value (prev), score-information corresponding to pitches predicted to be included in the next frame of the digital-sound-signals is set as the next-performance-value (next), the previous-performance-value (prev) and the current-performance-value (current) are set as expected-performance-value, and the digital-sound-signals is analyzed based on notes corresponding to the expected-performance-value, so analysis of the digital-sound-signals can be performed very accurately and quickly.

Moreover, considering the case where music is differently performed from the score-information, line 31 is added. When peak frequency-components are left after analysis of pitches contained in the score-information was completed, the remained peak frequency-components correspond to notes differently performed from the score-information. Accordingly, the notes corresponding to the remained peak frequency-components are detected using the algorithm of Pseudo-code 1 using sound-information, and the fact that the music is differently performed from the score is stored as in line 23 of Pseudo-code 2. For Pseudo-code 2, a method of using score-information has been mainly described, and other detailed descriptions are omitted. Like a method using only sound-information, the method using sound-information and score-information can include lines 11 through 20 of Pseudo-code 1 in which the size of a unit frame for analysis is reduced in order to detect accurate time-information.

However, the result of analysis and the performance error as the result-variable (result) are insufficient to be used as information of actually performed music. For the same reason as described in Pseudo-code 1, and considering that although different pitches start at the same time according to the score-information, a very slight time

difference among the pitches can occur in actual performance, the result-variable (result) is analyzed considering the characteristics of a corresponding instrument and the characteristics of a player, and the result of analysis is revised with (performance) in line 40.

5 Hereinafter, the frequency characteristics of digital-sounds and musical-instrument sound-information will be described in detail.

FIG. 12 is a diagram of the result of analyzing the frequency-components of the acoustic-piano-sounds according to the first measure of the score shown in FIGS. 1 and 2. In other words, FIG.
10 12 is a spectrogram of piano sounds performed according to the first measure of the second movement in Beethoven's Piano Sonata No. 8. Here, a grand piano made by the Young-chang piano company was used. A microphone was connected to a notebook computer made by Sony, and the sound was recorded using a recorder in a Windows
15 auxiliary program. Freeware, a Spectrogram 5.1.6 version, developed and published by R. S. Horne was used as a program for analyzing and displaying the spectrogram. A scale was set to 90 dB, a time scale was set to 5 msec, a fast Fourier transform (FFT) size was set to 8192, and default values are used for the others. Here, the scale set to 90 dB
20 indicates that sound of less than -90 dB is ignored and not displayed. The time scale set to 5 msec indicates that Fourier transform is performed with FFT windows overlapping every 5 msec to display an image.

A line 100 shown at the top of FIG. 12 indicates the strength of
25 input digital sound signals. Below the line 100, frequency-components contained in the digital sound signals are displayed by frequencies. A darker portion shows the magnitude of the frequency-component is larger than the bright ones. Accordingly, changes in the magnitude of the individual frequency-components in the flow of time can be caught at a
30 glance. Referring to FIGS. 12 and 2, it can be seen that

pitch-frequencies and harmonic-frequencies corresponding to the individual notes shown in the score of FIG. 2 are shown in FIG. 12.

FIGS. 13A through 13G are diagrams of the results of analyzing the frequency-components of the sounds of individual notes performed on the piano, which are contained in the first measure of the score of FIG. 2. Each of the notes contained in the first measure of FIG. 2 was independently performed and recorded in the same environment, and the result of analyzing each recorded note was displayed as a spectrogram. In other words, FIGS. 13A through 13G are spectrograms of the piano sounds corresponding to the notes C4, A2 \flat , A3 \flat , E3 \flat , B3 \flat , D3 \flat , and G3, respectively. FIGS. 13A through 13G show the magnitudes of each of frequency-components for 4 seconds. The conditions of analysis were set to be the same as those in the case of FIG. 12. The note C4 has a pitch-frequency of 262 Hz and harmonic-frequencies of n multiples of the pitch-frequency, for example, 523 Hz, 785 Hz, and 1047 Hz. This can be confirmed in FIG. 13A. In other words, it shows that frequency-components of 262 Hz and 523 Hz are strong in near black portions, and the magnitude roughly decreases from a frequency of 785 Hz toward a higher multiple harmonic-frequencies. The pitch-frequency and harmonic-frequencies of the note C4 are denoted by C4.

The note A2 \flat has a pitch frequency of 104 Hz. Referring to FIG. 13B, the harmonic-frequencies of the note A2 \flat is much stronger than its pitch frequency. Referring to FIG. 13B only, because that the note A2 \flat 's 3rd harmonic-frequency 311 Hz is strongest among the frequency-components displayed, this note A2 \flat may be erroneously recognized as the note E4 \flat having pitch-frequency 311 Hz if the note is determined by order of the magnitude of frequency-components.

In addition, if the notes are determined by their magnitudes of the frequency-components in FIGS. 13C through 13G, the same error can occur.

FIGS. 14A through 14G are diagrams of the results of indicating the frequency-components of each of the notes contained in the first measure of the score of FIG. 2 on FIG. 12.

FIG. 14A shows the frequency-components of the note C4 shown in FIG. 13A indicated on FIG. 12. Since the strength of the note C4 shown in FIG. 13A is greater than that shown in FIG. 12, the harmonic-frequencies of the note C4 shown in the upper portion of FIG. 12 are vague or too weak to be identified. However, if the frequency-magnitudes of FIG. 13A are lowered to match the magnitude of the pitch-frequency of the note C4 shown in FIG. 12 and compared with those of FIG. 12, it can be seen that the frequency-components of the note C4 are included in FIG. 12, as shown in FIG. 14A.

FIG. 14B shows the frequency-components of the note A2 \flat shown in FIG. 13B indicated on FIG. 12. Since the strength of the note A2 \flat shown in FIG. 13B is greater than that shown in FIG. 12, the pitch-frequency and harmonic-frequencies of the note A2 \flat are clearly shown in FIG. 13B but vaguely shown in FIG. 12, and particularly, higher harmonic-frequencies are barely shown in the upper portion of FIG. 12. If the frequency-magnitudes of FIG. 13B are lowered to match the magnitude of the pitch-frequency of the note A2 \flat shown in FIG. 12 and compared with those of FIG. 12, it can be seen that the frequency-components of the note A2 \flat are included in FIG. 12, as shown in FIG. 14B. In FIG. 14B, the 5th harmonic-frequency-component of the note A2 \flat is strong because it overlaps with the 2nd harmonic-frequency-component of the note C4. That is, because the 5th harmonic-frequency of the note A2 \flat is 519 Hz and the 2nd harmonic-frequency of the note C4 is 523 Hz, they overlap in the same frequency range in FIG. 14B. In addition, referring to FIG. 14, the ranges of 5th, 10th, and 15th harmonic-frequencies of the note A2 \flat respectively overlap with the ranges of the 2nd, 4th, and 6th harmonic-frequencies of

the note C4, so the corresponding harmonic-frequencies show stronger than in FIG. 13B. (Here, considering the fact that weak sound is vaguely illustrated on a spectrogram, the sounds of individual notes were recorded at greater strengths than the actual performance as shown in FIG. 12 to obtain FIGS. 13A through 13G so that frequency-components could be clearly distinguished from one another visually.)

FIG. 14C shows the frequency-components of the note A3 \flat shown in FIG. 13C indicated on FIG. 12. Since the strength of the note A3 \flat shown in FIG. 13C is greater than that shown in FIG. 12, the frequency-components shown in FIG. 13C are expressed as stronger than in FIG. 14C. Unlike the above-described notes, it is not easy to find only the components of the note A3 \flat in FIG. 14C because a lot of portions of the frequency-components of the note A3 \flat overlap with the pitch and harmonic-frequency-components of other notes and the note A3 \flat was weakly performed for a while and disappeared while other notes were continuously performed. All of the frequency-components of the note A3 \flat overlap with harmonic-frequencies of the note A2 \flat of multiples of 2. In addition, the 5th harmonic-frequency of the note A3 \flat overlaps with the 4th harmonic-frequency of the note C4, so it is difficult to identify a discontinued portion between two portions of the note A3 \flat separately performed two times while the note C4 was continuously performed. Nevertheless, other frequency-components become weaker in the middle, so the harmonic-frequency-components of the note A2 \flat and the discontinued portion of the note A3 \flat can be identified.

FIG. 14D shows the frequency-components of the note E3 \flat shown in FIG. 13D indicated on FIG. 12. Since the strength of the note E3 \flat shown in FIG. 13D is greater than that shown in FIG. 12, the frequency-components shown in FIG. 13D are expressed as stronger than in FIG. 14D. The note E3 \flat was separately performed four times. For the time during which the note E3 \flat was performed first two times,

the 2nd and 4th harmonic-frequency-components of the note E3 \flat overlap with the 3rd and 6th harmonic-frequency-components of the note A2 \flat , so harmonic-frequency-components of the note A2 \flat show in the discontinued portion between the separate two portions of the note E3 \flat performed separately. In addition, the 5th harmonic-frequency-component of the note E3 \flat overlaps with the 3rd harmonic-frequency-component of the note C4, so the frequency-components of the note E3 \flat are continued in the discontinued portion in the actual performance. For the time during which the note E3 \flat was performed next two times, the 3rd harmonic-frequency-component of the note E3 \flat overlaps with the 2nd harmonic-frequency-component of the note B3 \flat , so the frequency-component of the note E3 \flat shows even while the note E3 \flat is not actually performed. In addition, the 5th harmonic-frequency-component of the note E3 \flat overlaps with the 4th harmonic-frequency-component of the note G3, so the 4th harmonic-frequency-component of the notes G3 and the 5th harmonic-frequency-component of the note E3 \flat are continued even if the notes G3 and E3 \flat were alternately performed.

FIG. 14E shows the frequency-components of the note B3 \flat shown in FIG. 13E indicated on FIG. 12. Since the strength of the note B3 \flat shown in FIG. 13D is a little greater than that shown in FIG. 12, the frequency-components shown in FIG. 13E are expressed as stronger than in FIG. 14E. However, the frequency-components of the note B3 \flat shown in FIG. 13E almost match those in FIG. 14E. As shown in FIG. 13E, harmonic-frequencies of the note B3 \flat shown in the upper portion of FIG. 13E become very weak showing vaguely, as the sound of the note B3 \flat becomes weaker. Similarly, in FIG. 14E, harmonic-frequencies shown in the upper portion become weaker toward the right end.

FIG. 14F shows the frequency-components of the note D3 \flat

shown in FIG. 13F indicated on FIG. 12. Since the strength of the note D3 \flat shown in FIG. 13F is greater than that shown in FIG. 12, the frequency-components shown in FIG. 13F are expressed as stronger than in FIG. 14F. However, the frequency-components of the note D3 \flat shown in FIG. 13F almost match those in FIG. 14F. Particularly, like FIG. 13F in which the 9th harmonic-frequency of the note D3 \flat is weaker than the 10th harmonic-frequency of the note D3 \flat , the 9th harmonic-frequency of the note D3 \flat is very weak and weaker than the 10th harmonic-frequency of the note D3 \flat in FIG. 14F. However, since the 5th and 10th harmonic-frequencies of the note D3 \flat shown in FIG. 14F overlap with the 3rd and 6th harmonic-frequencies of the note B3 \flat shown in FIG. 14E, the 5th and 10th harmonic-frequencies of the note D3 \flat show stronger than the other harmonic-frequencies of the note D3 \flat . Since the 5th harmonic-frequency of the note D3 \flat is 693 Hz, and the 3rd harmonic-frequency of the note B3 \flat is very close to 699 Hz, they overlap in a spectrogram.

FIG. 14G shows the frequency-components of the note G3 shown in FIG. 13G indicated on FIG. 12. Since the strength of the note G3 shown in FIG. 13G is a little greater than that shown in FIG. 12, the frequency-components shown in FIG. 13G are expressed as stronger than in FIG. 14G. Since the note G3 shown in FIG. 14G was performed stronger than the note A3 \flat shown in FIG. 14C, each of the frequency-components of the note G3 could be found clearly. In addition, unlike FIGS. 14C and 14F, the frequency-components of the note G3 rarely overlap with frequency-components of the other notes, so each of the frequency-components of the note G3 can be visually identified easily. However, although the 4th harmonic-frequency of the note G3 and the 5th harmonic-frequency of the note E3 \flat shown in FIG. 14D are similar at 784 Hz and 778 Hz, respectively, since the notes E3 \flat and G3 are performed at different time points, the 5th

harmonic-frequency-component of the note E3_b shows a little below a portion between two separate portions of the 4th harmonic-frequency-component of the note G3.

FIG. 15 is a diagram in which the frequencies shown in FIG. 12 are compared with the frequency-components of the individual notes contained in the score of FIG. 2. In other words, the results of analyzing the frequency-components shown in FIG. 12 are displayed in FIG. 15 so that the results can be understood at one sight. In the above-described method for analyzing music according to the present invention, the frequency-components of the individual notes shown in FIGS. 13A through 13G are used to analyze the frequency-components shown in FIG. 12. As a result, FIG. 15 can be obtained. A method of analyzing input digital-sounds using sound-information of musical-instrument according to the present invention can be summarized through FIG. 15. In other words, in the above-described method of the present invention, the sounds of individual notes actually performed are received, and the frequency-components of the received sounds are used as sound-information of musical-instrument.

It has been described that frequency-components are analyzed using FFT. However, it is apparent that wavelet or other techniques developed from digital signal processing algorithms instead of FFT can be used to analyze frequency-components. In other words, a most representative Fourier transform technique is used in descriptive sense only, and the present invention is not restricted thereto.

Meanwhile, in FIGS. 14A through 15, time-information of frequency-components of the notes is different from that of actual performance. Particularly, in FIG. 15, the notes start at 1500, 1501, 1502, 1503, 1504, 1505, 1506, and 1507 in the actual performance, but their frequency-components show before the start-points. Moreover, the frequency-components show after end-points of the actually performed

notes. These timing-errors occur because the size of an FFT window is set to 8192 in order to accurately analyze frequency-components according to the flow of time. The range of timing-errors depends on the size of an FFT window. In the above embodiment, the sampling rate is 22050 Hz, and the FFT window is 8192 samples, so an error is $8192 \div 22050 \approx 0.37$ seconds. In other words, when the size of the FFT window increases, the size of a unit frame also increases, thereby decreasing a gap between identifiable frequencies. As a result, frequency-components can be accurately analyzed according to the pitches, but timing-errors increase. When the size of the FFT window decreases, a gap between identifiable frequencies increases. As a result, notes close to each other in a low frequency range cannot be distinguished from one another, but timing errors decrease. Alternatively, increasing the sampling rate can decrease the range of timing-errors.

FIGS. 16A through 16D are diagrams of the results of analyzing notes performed according to the first measure of the score shown in FIGS. 1 and 2 using FFT windows of different sizes in order to explain changes in timing-errors according to changes in the size of an FFT window.

FIG. 16A shows the result of analysis in the case where the size of an FFT window is set to 4096 for FFT. FIG. 16B shows the result of analysis in the case where the size of an FFT window is set to 2048 for FFT. FIG. 16C shows the result of analysis in the case where the size of an FFT window is set to 1024 for FFT. FIG. 16D shows the result of analysis in the case where the size of an FFT window is set to 512 for FFT.

Meanwhile, FIG. 15 shows the result of analysis in the case where the size of an FFT window is set to 8192 for FFT. Accordingly, by comparing the results shown in FIGS. 15 through 16D, it can be inferred that a gap between identifiable frequencies becomes narrower to thus

allow fine analysis but a timing-error increases when the size of an FFT window increases, whereas a gap between identifiable frequencies becomes wider to thus make it difficult to perform fine analysis but a timing-error decreases when the size of an FFT window decreases.

5 Therefore, when analysis is performed, the size of an FFT window can be changed according to required time accuracy and required frequency accuracy. Alternatively, time-information and frequency-information can be analyzed using FFT windows of different sizes.

10 FIGS. 17A and 17B show timing errors occurring during analysis of digital-sounds, which vary with the size of an FFT window. Here, a white area corresponds to an FFT window in which a particular note is found. In FIG. 17A, the size of an FFT window is large at 8192, so a white area corresponding to a window in which the particular note is
15 found is wide. In FIG. 17B, the size of an FFT window is small at 1024, so a white area corresponding to a window in which the particular note is found is narrow.

FIG. 17A is a diagram of the result of analyzing digital-sounds when the size of an FFT window is set to 8192. Referring to FIG. 17A,
20 the note actually starts at a point 9780, but the note starts at a point 12288 $(=(8192+16384)/2)$ in the middle of the window in which the particular note is found according to the result of FFT. Here, there occurs an error of a time corresponding to 2508 samples, i.e., a difference between a 12288th sample and a 9780th sample. In other words, in the
25 case of sampling rate 22.5 KHz, an error of about $2508*(1/22500)=0.11$ seconds occurs.

FIG. 17B is a diagram of the result of analyzing digital-sounds when the size of an FFT window is set to 1024. Referring to FIG. 17B,
like FIG. 17A, the note actually starts at a point 9780, but the note starts
30 at a point 9728 $(=(9216+10240)/2)$ according to the result of FFT. Here, it

is determined that the note starts at a time point corresponding to a 9728th sample in the middle of the range between a 9216th sample and a 10239th sample. An error is only a time corresponding to 52 samples. In the case of sampling rate 22.5 KHz, the error of about 0.002 seconds occurs according to the above-described calculation method. Therefore, it can be inferred that the more accurate result of analysis can be obtained as the size of an FFT window decreases.

FIG. 18 is a diagram of the result of analyzing the frequency-components of the sounds obtained by putting together a plurality of pieces of individual pitches detected using the sound-information and the score-information according to the second embodiment of the present invention. In other words, the score-information is detected from the score shown in FIG. 1, and the sound-information described with reference to FIGS. 13A through 13G are used.

More specifically, it is detected from the score-information detected from the score of FIG. 1 that the notes C4, A3 \flat , and A2 \flat are initially performed for 0.5 seconds. Sound information of the notes C4, A3 \flat , and A2 \flat is detected from the information shown in FIGS. 13A through 13C. Input digital-sounds are analyzed using the selected score-information and the selected sound-information. The result of analysis is shown in FIG. 18. Here, it can be found that a portion of FIG. 12 corresponding to the initial 0.5 seconds is almost the same as the corresponding portion of FIG. 14D. Accordingly, the portion of FIG. 18 corresponding to the initial 0.5 seconds, which corresponds to (result) or (performance) in Pseudo-code 2, is the same as the portion of FIG. 12 corresponding to the initial 0.5 seconds.

While this invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that various changes may be made within the

scope which does not beyond the essential characteristics of this invention. The above embodiments have been used in a descriptive sense only and not for purpose of limitation. Therefore, it will be understood that the scope of the invention will be defined by the
5 appended claims.

Industrial Applicability

According to the present invention, input digital-sounds can be quickly analyzed using sound-information or both sound-information and
10 score-information. In conventional methods for analyzing digital-sounds, music composed of polyphonic-pitches, for example, piano music, cannot be analyzed. However, according to the present invention, as well as monophonic-pitches, polyphonic-pitches contained in digital-sounds can be quickly and accurately analyzed using sound-information or both
15 sound-information and score-information.

Accordingly, the result of analyzing digital-sounds according to the present invention can be directly applied to an electronic-score, and performance-information can be quantitatively detected using the result of analysis. This result of analysis can be widely used in from musical
20 education for children to professional players' practice.

That is, by using a technique of the present invention allowing input digital-sounds to be analyzed in real time, positions of currently performed notes on an electronic-score are recognized in real time and positions of notes to be performed next are automatically indicated on
25 the electronic-score, so that players can concentrate on performance without caring about turning over the leaves of a paper-score.

In addition, the present invention compares performance-information obtained as the result of analysis with previously stored score-information to detect performance accuracy so
30 that players can be informed about wrong-performance. The detected

WO 02/47064

PCT/KR01/02081

performance accuracy can be used as data by which a player's performance is evaluated.

What is claimed is:

1. A method for analyzing digital-sounds using sound-information of musical-instruments, the method comprising the steps of:
 - 5 (a) generating and storing sound-information of different musical instruments;
 - (b) selecting the sound-information of the particular instrument to be actually played from among the stored sound-information of different musical-instruments;
 - 10 (c) receiving digital-sound-signals;
 - (d) decomposing the digital-sound-signals into frequency-components in units of frames;
 - (e) comparing the frequency-components of the digital-sound-signals with frequency-components of the selected
15 sound-information of the particular instrument and analyzing the frequency-components of the digital-sound-signals to detect monophonic-pitches-information from the digital-sound-signals; and
 - (f) outputting the detected monophonic-pitches-information.
- 20 2. The method of claim 1, wherein the step (e) comprises detecting time-information of each frame, comparing the frequency-components of the digital-sound-signals with the frequency-components of the selected sound-information of the particular instrument and analyzing the frequency-components of the
25 digital-sound-signals in units of frames, and detecting pitch-information, strength-information, and time-information of each of individual pitches contained in each of the frames.
3. The method of claim 2, wherein the step (e) further
30 comprises determining whether the detected

monophonic-pitches-information contains any new-pitch which is not included in a previous-frame, dividing a current-frame including the new-pitch into subframes if it is determined that the detected monophonic-pitches-information contains the new-pitch, finding a
5 subframe including the new-pitch, and detecting pitch-information and strength-information of the new-pitch and time-information of the found subframe.

4. The method of claim 1, wherein the step (a) comprises
10 periodically updating the sound-information of different musical instruments.

5. The method of claim 1, wherein the step (a) comprises storing each individual pitch which can be expressed by the
15 sound-information in the form of wave data when storing the sound-information of different musical instruments in the form of samples of sounds having at least one strength, and extracting the frequency-components of the sound-information of different musical instruments from the wave data stored.

20 6. The method of claim 1, wherein the step (a) comprises storing each individual pitch which can be expressed by the sound-information in a form which can directly expressing the magnitudes of each frequency-components of the pitch when storing the
25 sound-information of different musical instruments in the form of samples of sounds having at least one strength.

7. The method of claim 6, wherein the step (a) comprises performing Fourier transform on the sound-information of different
30 musical instruments and storing the sound-information in a form in which

the sound-information can be directly displayed.

8. The method of claim 6, wherein the step (a) comprises performing wavelet transform on the sound-information of different musical instruments and storing the sound-information in a form in which the sound-information can be directly displayed.

9. The method of claim 5 or 6, wherein the step (a) comprises separately storing sound-information of keyboard-instruments according to use/nonuse of pedals.

10. The method of claim 5 or 6, wherein the step (a) comprises separately storing sound-information of string-instruments by each string.

11. The method of claim 1 or 2, wherein the step (e) comprises the steps of:

(e1) selecting the lowest peak frequency-components contained in a current frame of the digital-sound-signals;

(e2) detecting the sound-information containing the lowest peak frequency-components from the selected sound-information of the particular instrument;

(e3) detecting, as monophonic-pitches-information, the sound-information containing most similar peak frequency-components to those of the current-frame from among the detected sound-information in step (e2);

(e4) removing the frequency-components of the sound-information detected as the monophonic-pitches-information in step (e3) from the current-frame; and

(e5) repeating steps (e1) through (e4) when there are any peak frequency-components left in the current-frame.

12. A method for analyzing digital-sounds using sound-information of musical-instruments and score-information, the method comprising the steps of:
- 5 (a) generating and storing sound-information of different musical instruments;
 - (b) generating and storing score-information of a score to be performed;
 - (c) selecting the sound-information of the particular instrument to
10 be actually played and the score-information of the score to be actually performed from among the stored sound-information of different musical instruments and the stored score-information;
 - (d) receiving digital-sound-signals;
 - (e) decomposing the digital-sound-signals into
15 frequency-components in units of frames;
 - (f) comparing the frequency-components of the digital-sound-signals with frequency-components of the selected sound-information of the particular instrument and the selected score-information, and analyzing the frequency-components of the
20 digital-sound-signals to detect performance-error-information and monophonic-pitches-information from the digital-sound-signals; and
 - (g) outputting the detected monophonic-pitches-information.

13. The method of claim 12, wherein the step (f) comprises
25 detecting time-information of each-frame, comparing the frequency-components of the digital-sound-signals with the frequency-components of the selected sound-information of the particular instrument and the selected score-information, analyzing the frequency-components of the digital-sound-signals in units of frames,
30 and detecting pitch-information, strength-information, and

time-information of each of individual pitches contained in each of the frames.

14. The method of claim 12 or 13, wherein the step (f) further
5 comprises determining whether the detected monophonic-pitches-information contains any new-pitch which is not included in a previous frame, dividing a current frame including a new-pitch into subframes if it is determined that the detected monophonic-pitches-information contains the new-pitch, finding a
10 subframe including the new-pitch, and detecting pitch-information and strength-information of the new-pitch and time-information of the found subframe.

15. The method of claim 12, wherein the step (a) comprises
15 periodically updating the sound-information of different musical instruments.

16. The method of claim 12, wherein the step (a) comprises
storing each individual pitch which can be expressed by the
20 sound-information in the form of wave data when storing the sound-information of different musical instruments in the form of samples of sounds having at least one strength.

17. The method of claim 12, wherein the step (a) comprises
25 storing each individual pitch which can be expressed by the sound-information in a form which can directly expressing the magnitudes of each frequency-components of the pitch when storing the sound-information of different musical instruments in the form of samples of sounds having at least one strength.

30

18. The method of claim 17, wherein the step (a) comprises performing Fourier transform on the sound-information of different musical instruments and storing the sound-information in a form in which the sound-information can be directly displayed.

5

19. The method of claim 17, wherein the step (a) comprises performing wavelet transform on the sound-information of different musical instruments and storing the sound-information in a form in which the sound-information can be directly displayed.

10

20. The method of claim 16 or 17, wherein the step (a) comprises separately storing sound-information of keyboard-instruments according to use/nonuse of pedals.

15

21. The method of claim 16 or 17, wherein the step (a) comprises separately storing sound-information of string-instruments by each string.

22. The method of claim 12 or 13, wherein the step (f) comprises the steps of:

20

(f1) generating expected-performance-values of the current-frame referring to the score-information in real time; and determining whether there is any note in the expected-performance-values which is not compared with the digital-sound-signals in the current-frame;

25

(f2) if it is determined that there is no note in the expected-performance-value which is not compared with the digital-sound-signals in the current-frame in step (f1), determining whether frequency-components of the digital-sound-signals in the current-frame correspond to performance-error-information, detecting performance-error-information and monophonic-pitches-information, and

30

removing the frequency-components of the sound-information corresponding to the performance-error-information and the monophonic-pitches-information from the digital-sound-signals in the current-frame;

5 (f3) If it is determined that there is any note in the expected-performance-value which is not compared with the digital-sound-signals in the current-frame in step (f1), comparing the digital-sound-signals in the current-frame with the expected-performance-values and analyzing to detect
10 monophonic-pitches-information from the digital-sound-signals in the current-frame, and removing the frequency-components of the sound-information detected as the monophonic-pitches-information from the digital-sound-signals in the current-frame; and

(f4) repeating steps (f1) through (f4) when there are any peak
15 frequency-components left in the current-frame of the digital-sound-signals.

23. The method of claim 22, wherein the step (f2) comprises the steps of:

20 (f2_1) selecting the lowest peak frequency-components contained in the current-frame of the digital-sound-signals;

(f2_2) detecting the sound-information containing the lowest peak frequency-components from the selected sound-information of the particular instrument;

25 (f2_3) detecting, as performance-error-information, the sound-information containing most similar peak frequency-components to peak frequency-components of the current-frame from the detected sound information;

(f2_4) if it is determined that the current pitches of the
30 performance-error-information are contained in next notes in the

score-information, adding the current pitches of the performance-error-information to the expected-performance-value and moving the current pitches of the performance-error-information into the monophonic-pitches-information; and

5 (f2_5) removing the frequency-components of the sound-information detected as the performance-error-information or the monophonic-pitches-information from the digital-sounds in the current-frame.

10 24. The method of claim 23, wherein the step (f2_3) comprises detecting the pitch and strength of a corresponding performed note as the performance-error-information.

25. The method of claim 22, wherein the step (f3) comprises
15 the steps of:

(f3_1) selecting the sound-information of the lowest peak frequency-components which is not compared with frequency-components contained in the current-frame of the digital-sound-signals from the sound-information corresponding to the
20 expected-performance-value which has not undergone comparison;

(f3_2) if it is determined that the frequency-components of the selected sound-information are included in frequency-components contained in the current-frame of the digital-sound-signals, detecting the selected sound-information as monophonic-pitches-information and
25 removing the frequency-components of the selected sound-information from the current-frame of the digital-sound-signals; and

(f3_3) if it is determined that the frequency-components of the selected sound-information are not included in the frequency-components contained in the current-frame of the
30 digital-sound-signals, adjusting the expected-performance-value.

26. The method of claim 23, wherein the step (f3_3) comprises removing an expected-performance-value corresponding to the selected sound-information whose frequency-components are included in the digital-sound-signals at one or more time points but are not included in at least a predetermined number (N) of consecutive previous frames.

27. The method of claim 12, further comprising the step of (h) estimating performance accuracy based on the performance-error-information detected in step (f).

28. The method of claim 12, further comprising the step of (i) adding the individual notes of the performance-error-information to the existing score-information based on the performance-error-information detected in step (f).

29. The method of claim 12, wherein the step (b) comprises generating and storing at least one kind of information selected from the group consisting of pitch-information, note-length-information, speed-information, tempo-information, note-strength-information, detailed performance-information including staccato, staccatissimo, and pralltriller, and discrimination-information for performance using two-hands or performance using a plurality of instruments, based on the score to be performed.

WO 02/47064

PCT/KR01/02081

1/40

FIG.1

ADAGIO CANTABILE



FIRST TWO MEASURES OF SECOND MOVEMENT
IN BEETHOVEN'S PIANO SONATA NO. 8

FIG.2

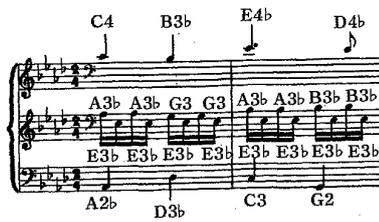


FIG.3

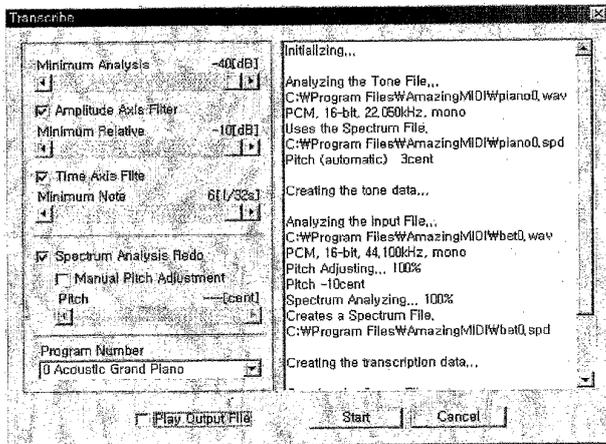


FIG.5



FIG. 6

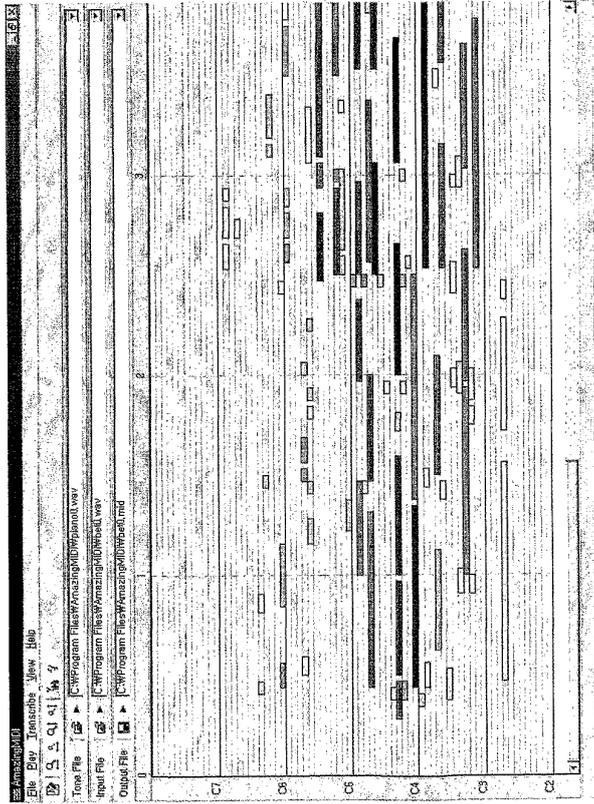


FIG. 7

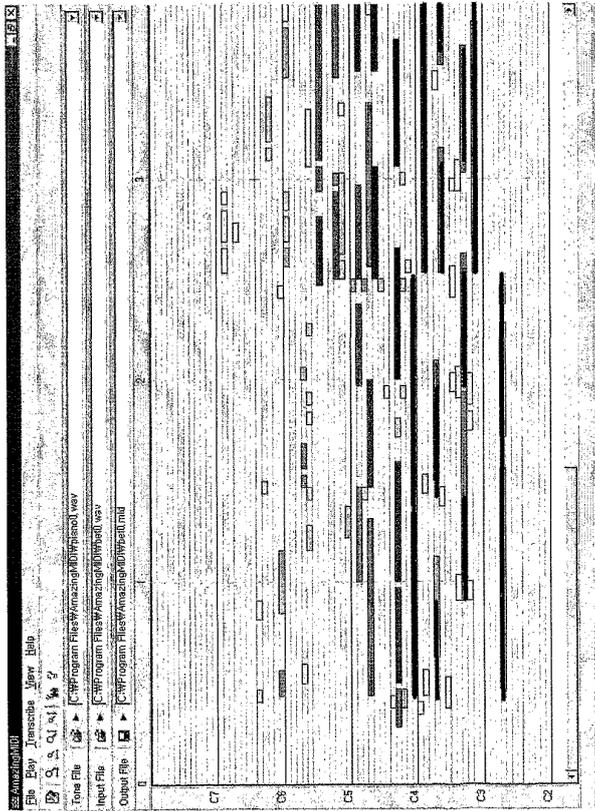


FIG.8

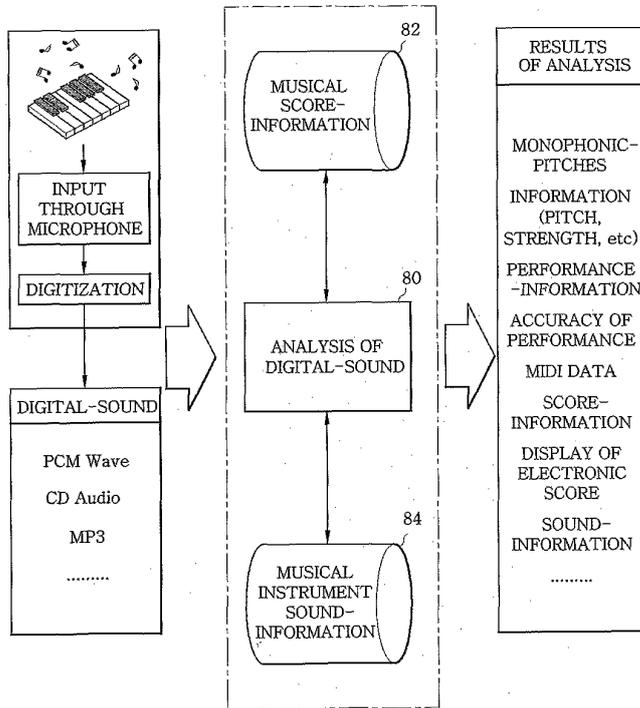


FIG.9A

PITCH	A0	B0 _b	B0	C1	D1 _b	...	C4	...	A7	B7 _b	B7	C8
-------	----	-----------------	----	----	-----------------	-----	----	-----	----	-----------------	----	----

FIG.9B

NOTE STRENGTH	-∞	...	-7dB	...	0
---------------	----	-----	------	-----	---

FIG.9C

PEDAL TABLE	000	001	010	100	101	110	011	111
-------------	-----	-----	-----	-----	-----	-----	-----	-----

FIG.9D

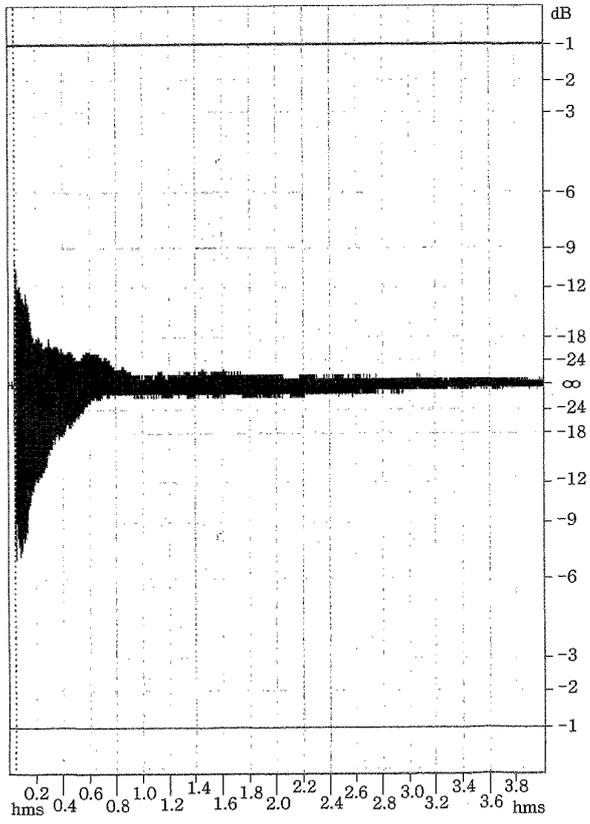


FIG.9E

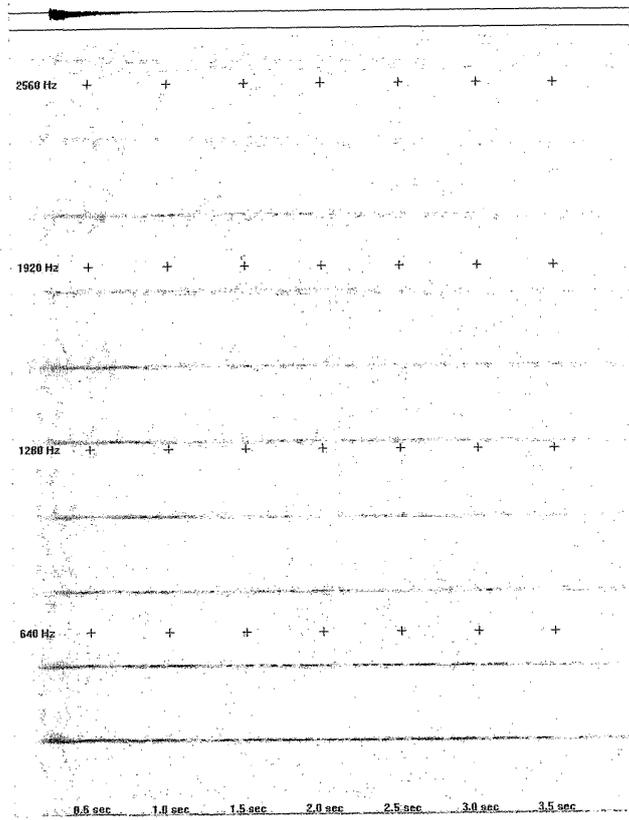


FIG.10

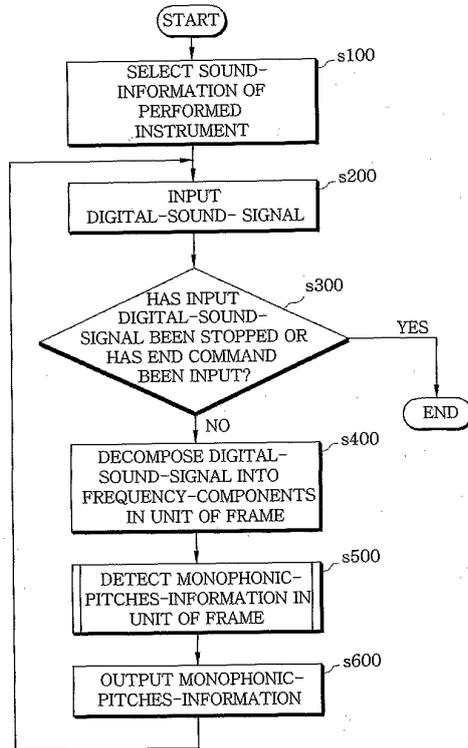


FIG.10A

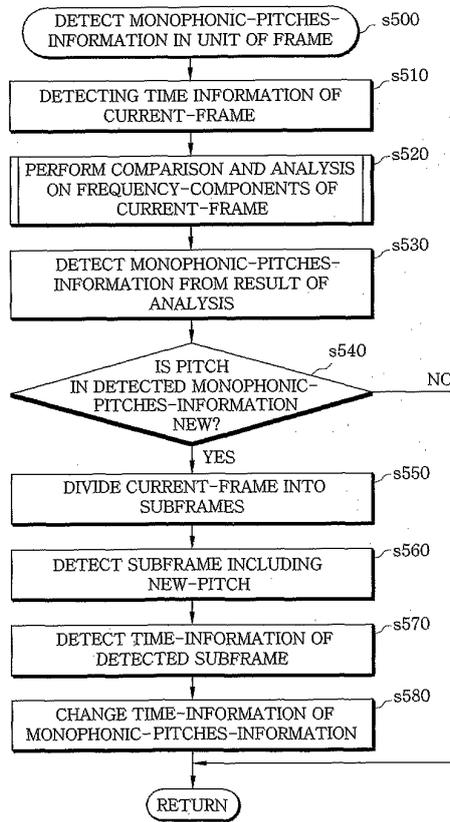


FIG.10B

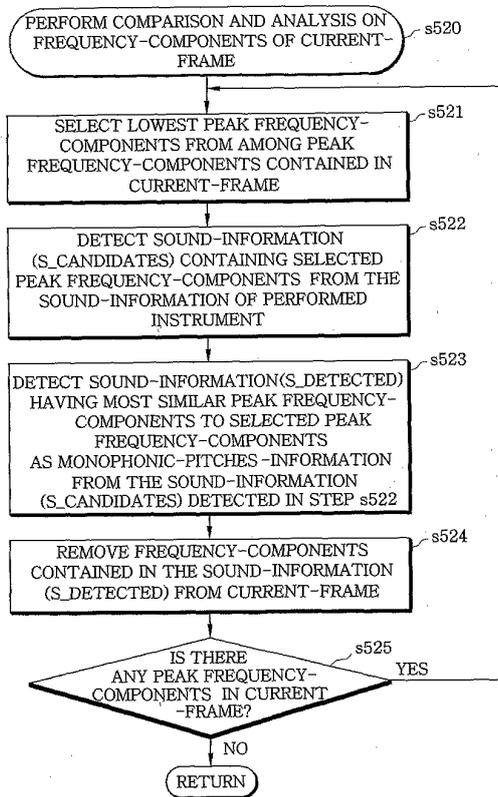
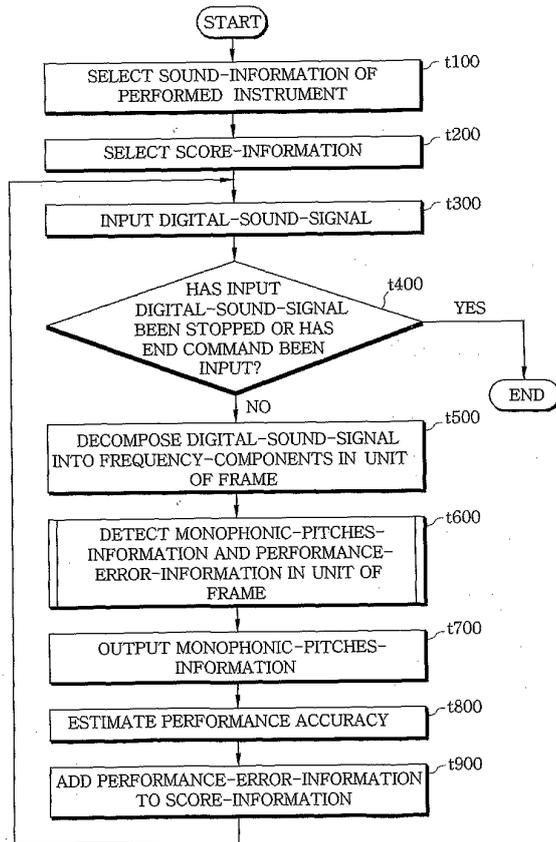


FIG.11



15/40

FIG.11A

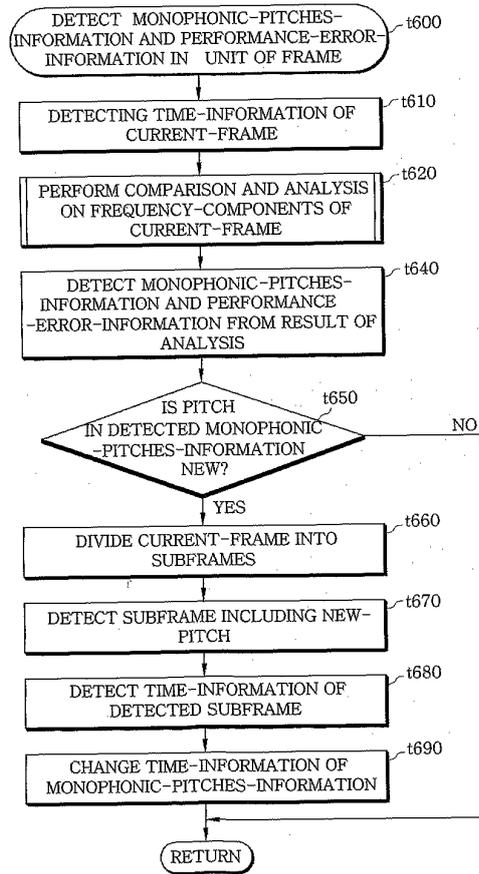


FIG.11B

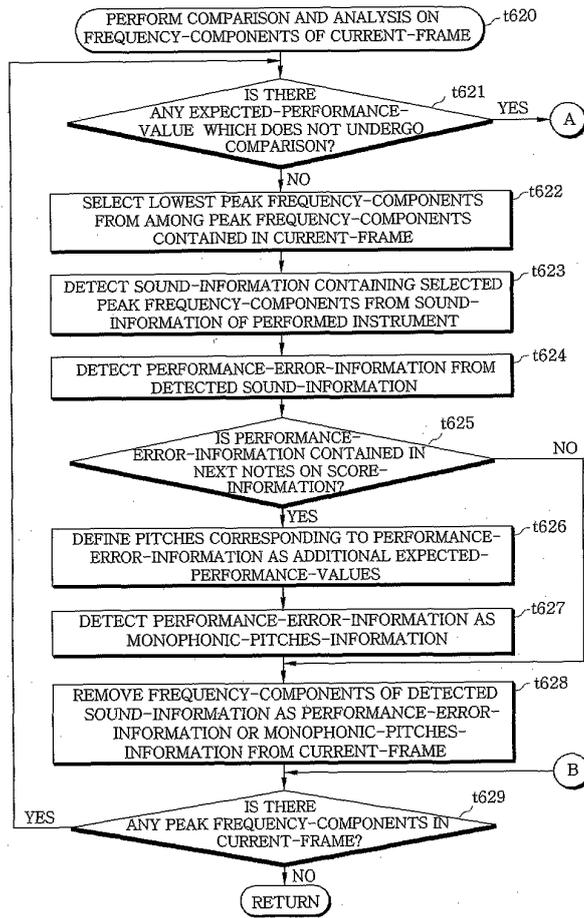


FIG.11C

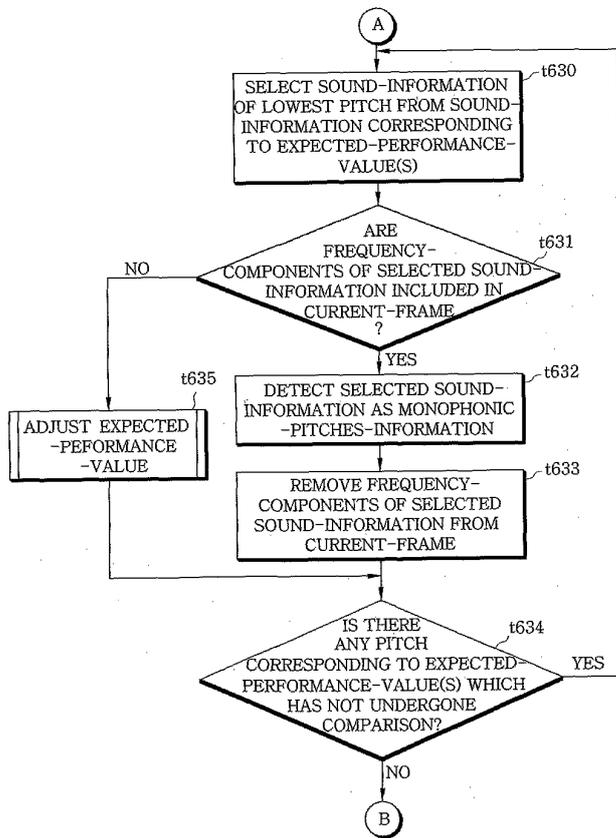


FIG.11D

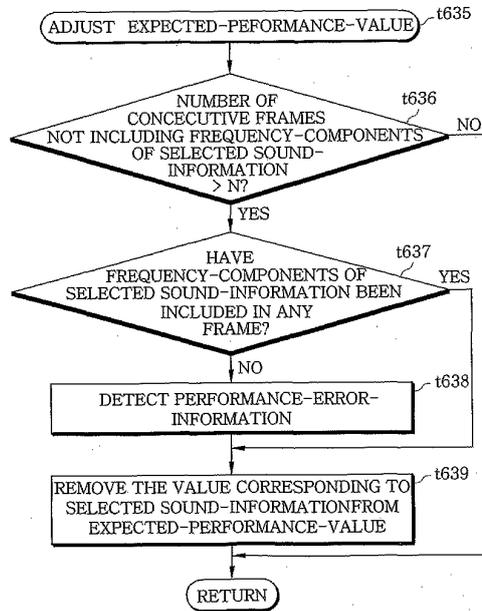


FIG.12

19/40

100

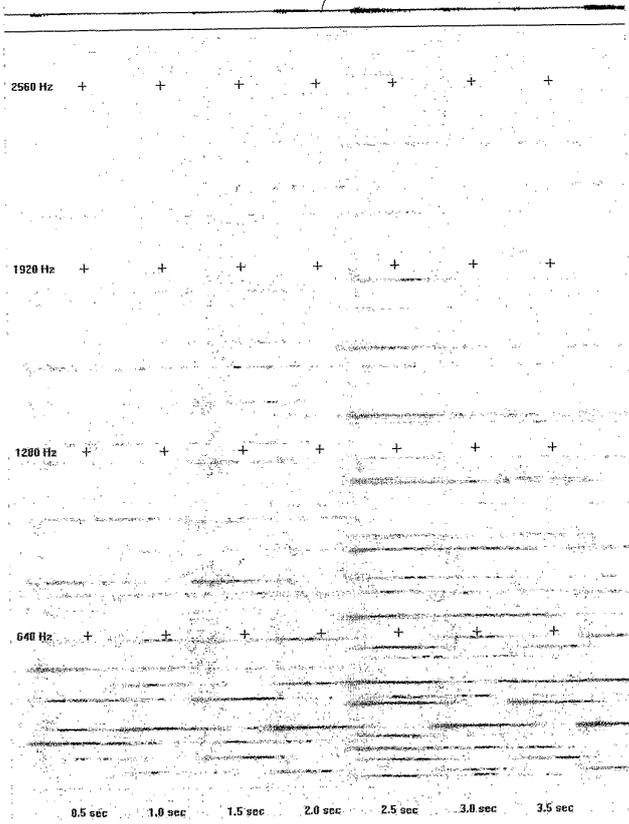


FIG.13A

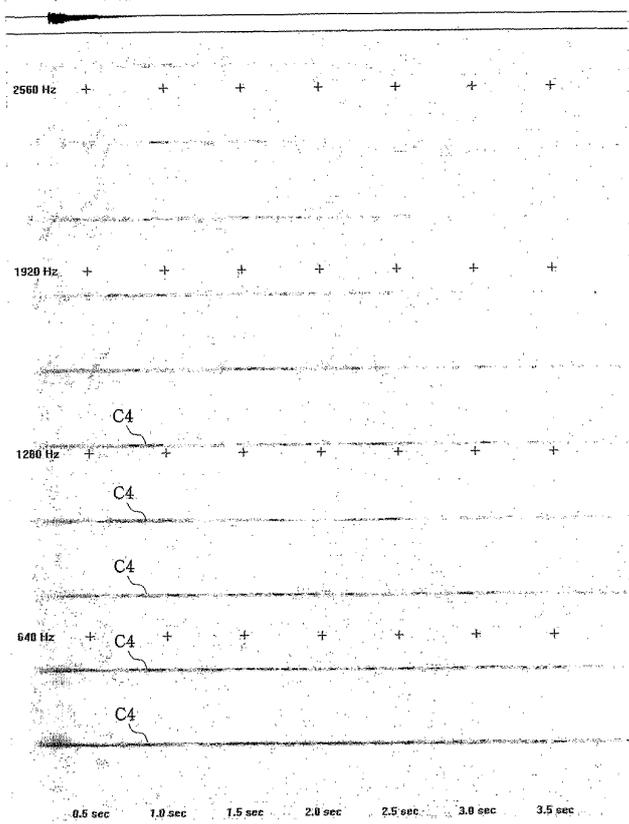


FIG.13B

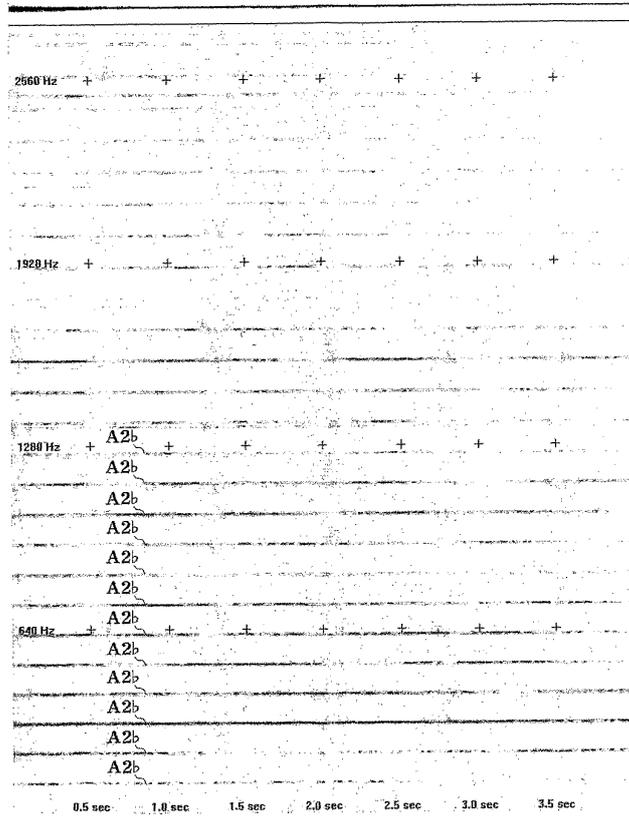


FIG.13C

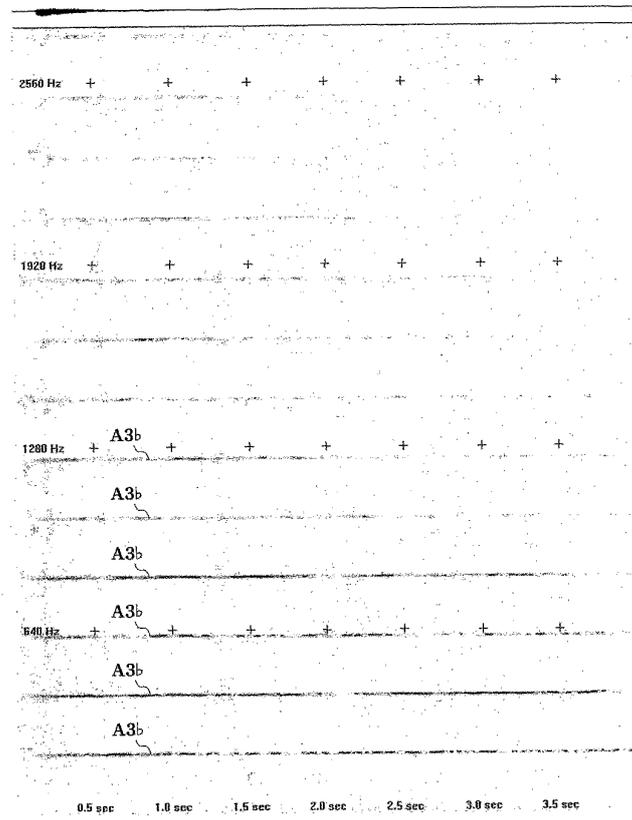


FIG.13D

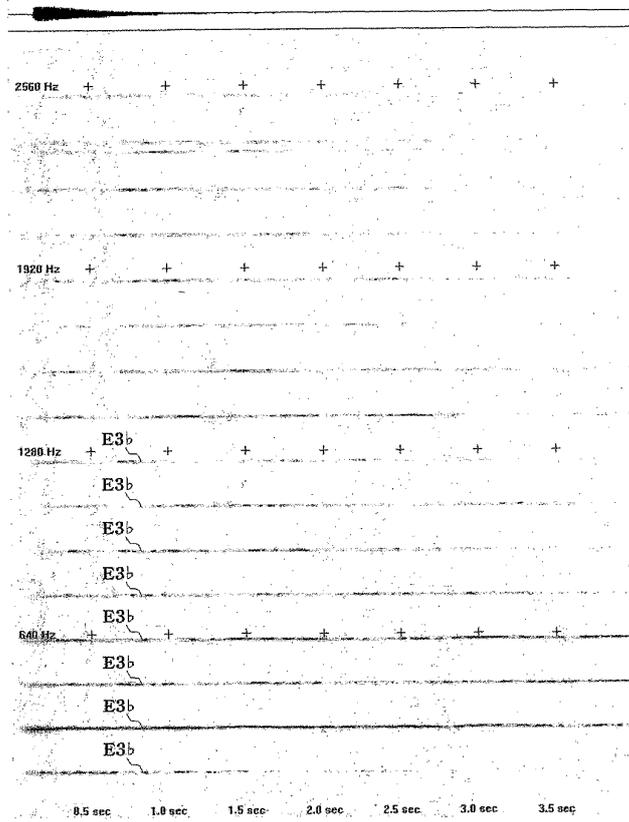


FIG. 13E

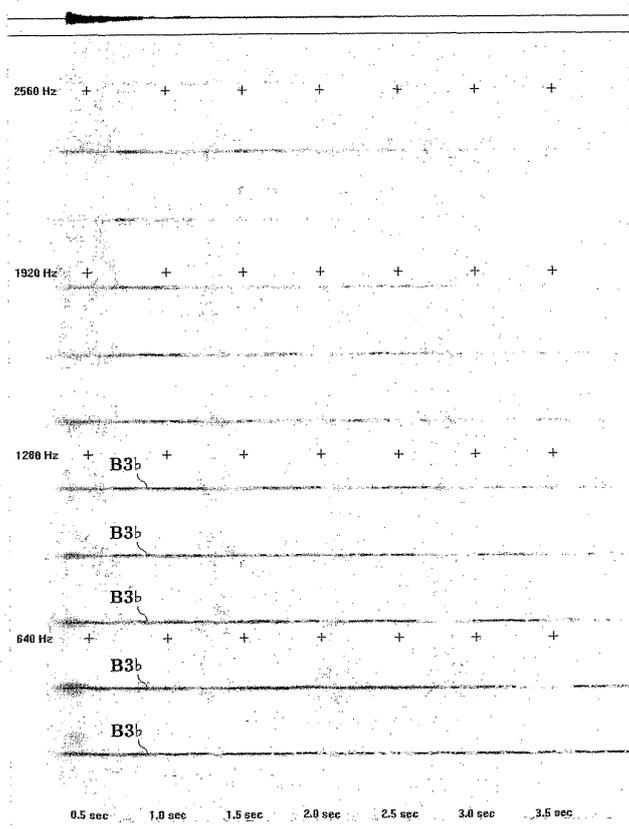


FIG.13F

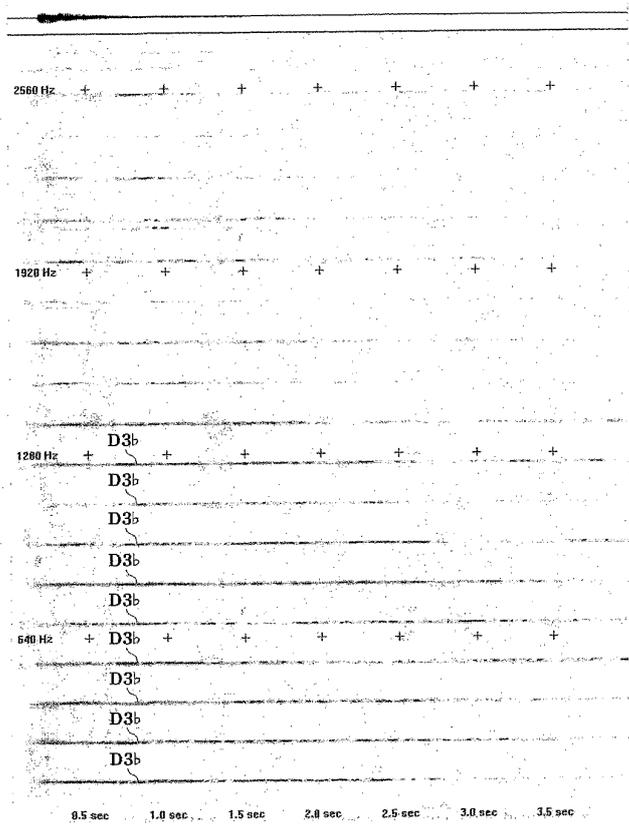


FIG.13G

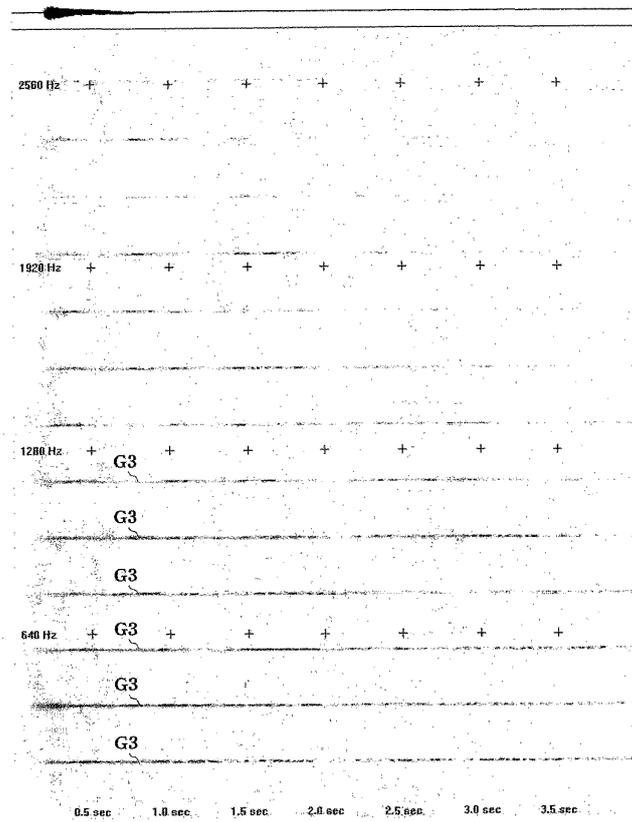


FIG.14A

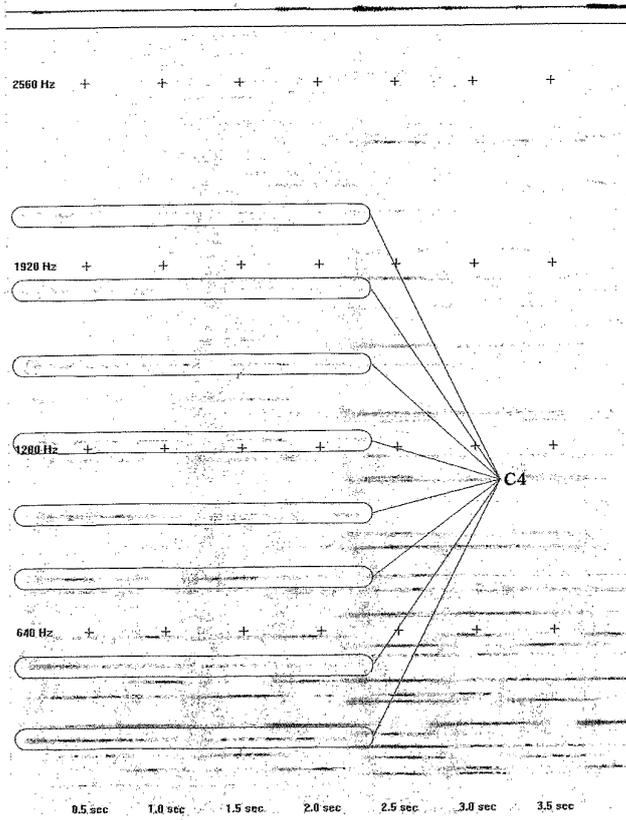


FIG.14B

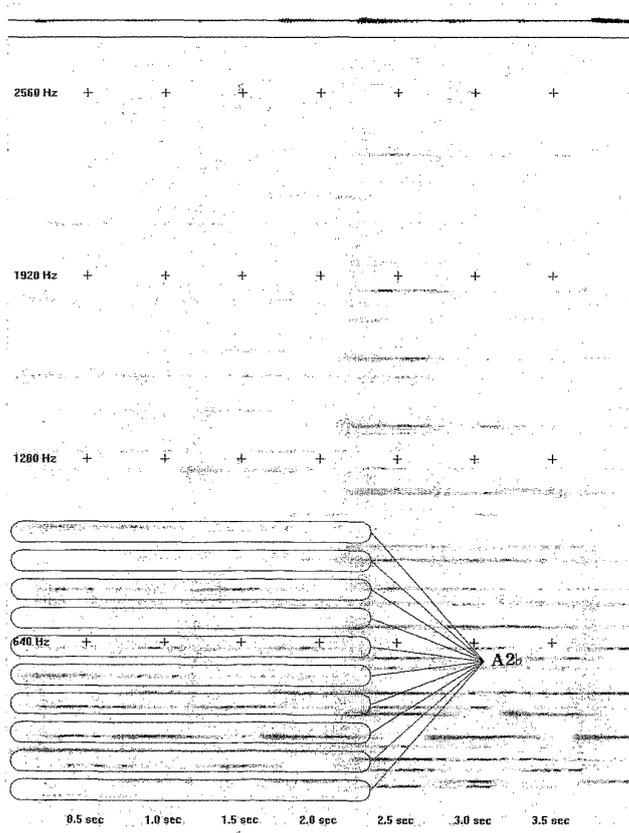


FIG.14C

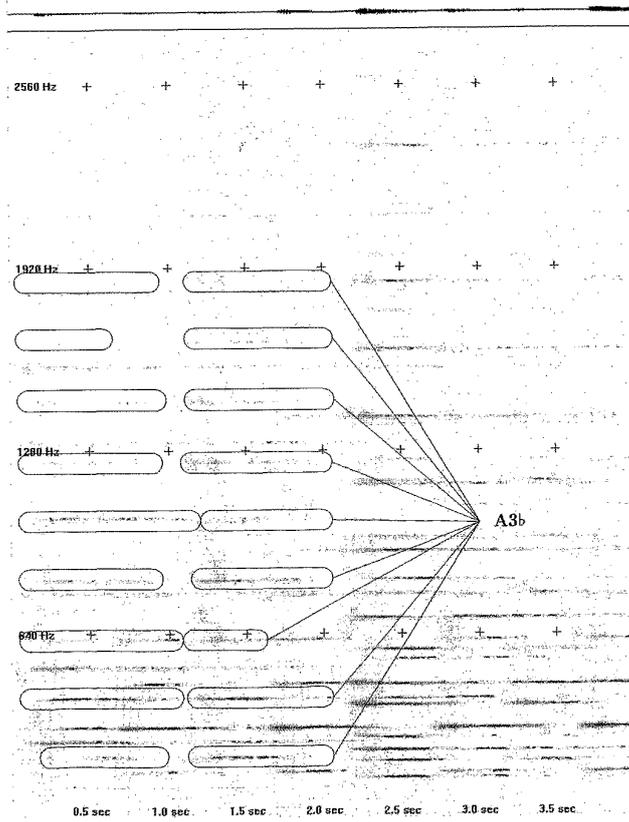


FIG.14D

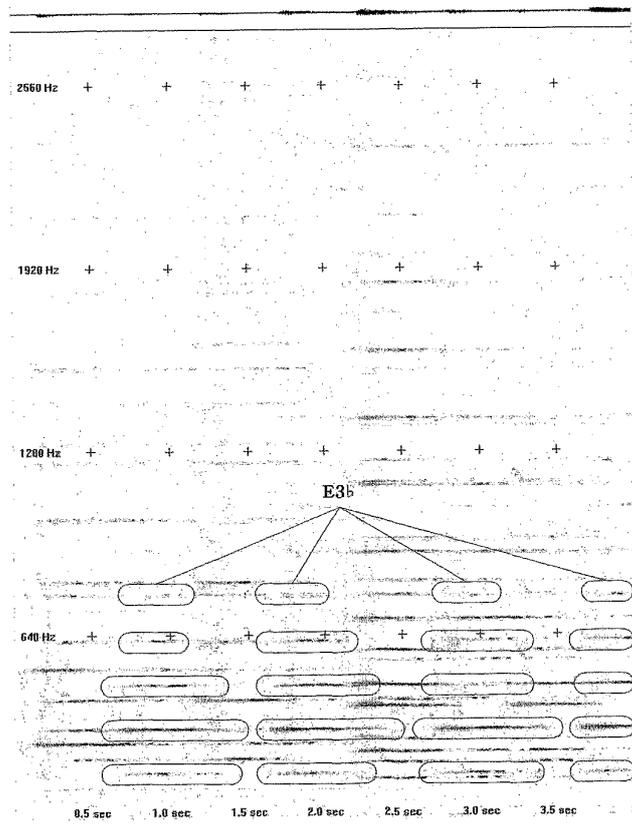


FIG.14E

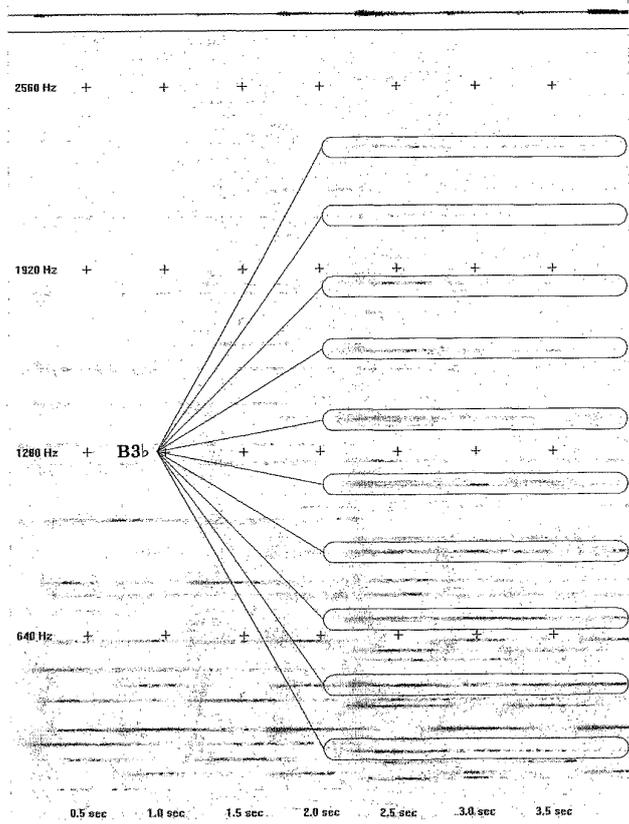


FIG.14F

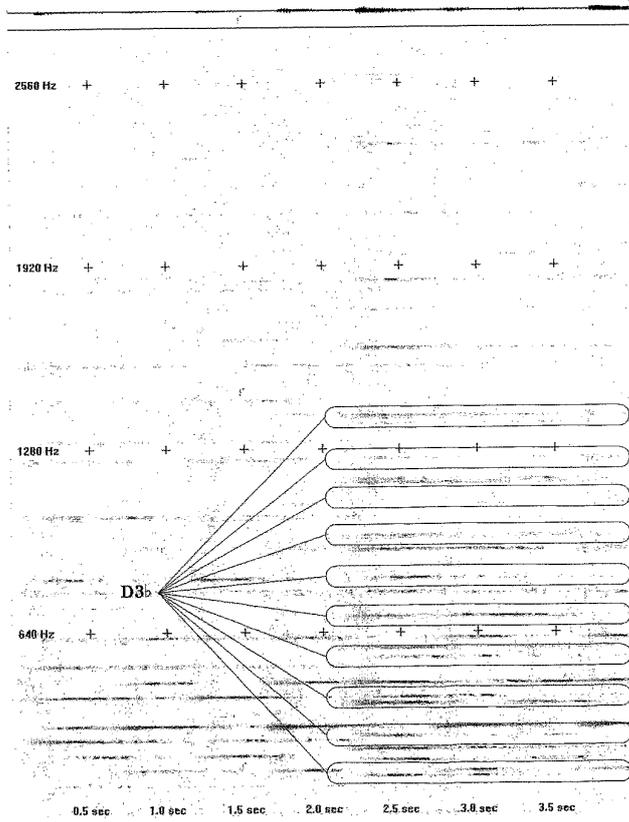


FIG.14G

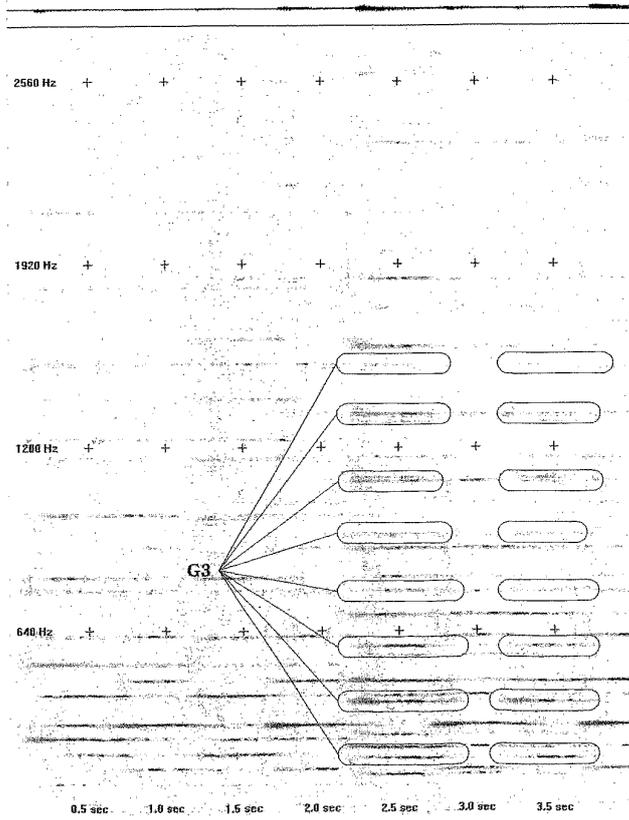


FIG.15

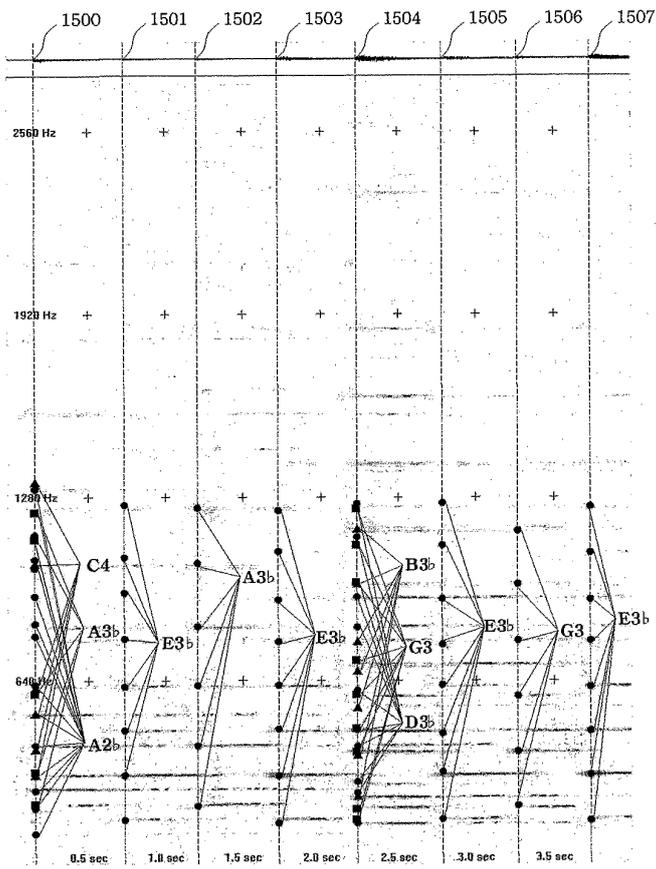


FIG.16A

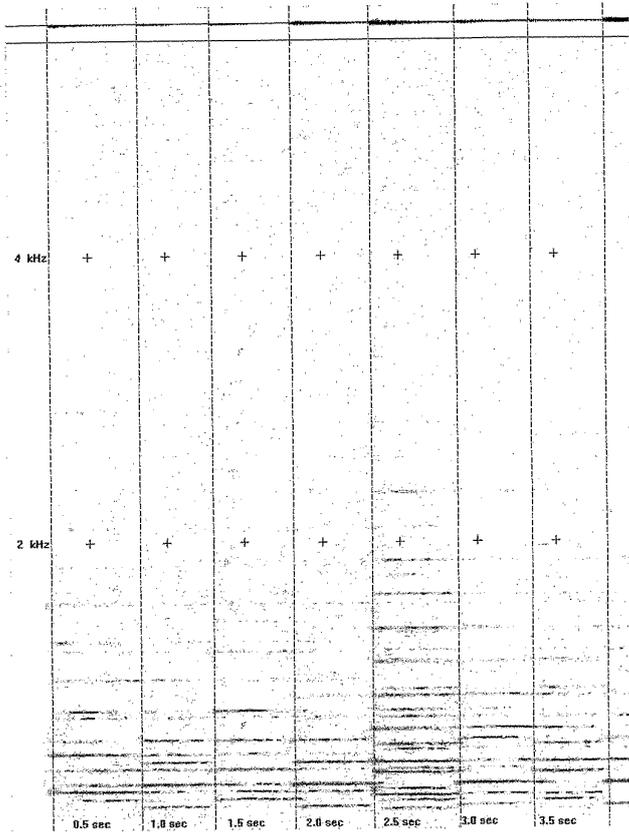
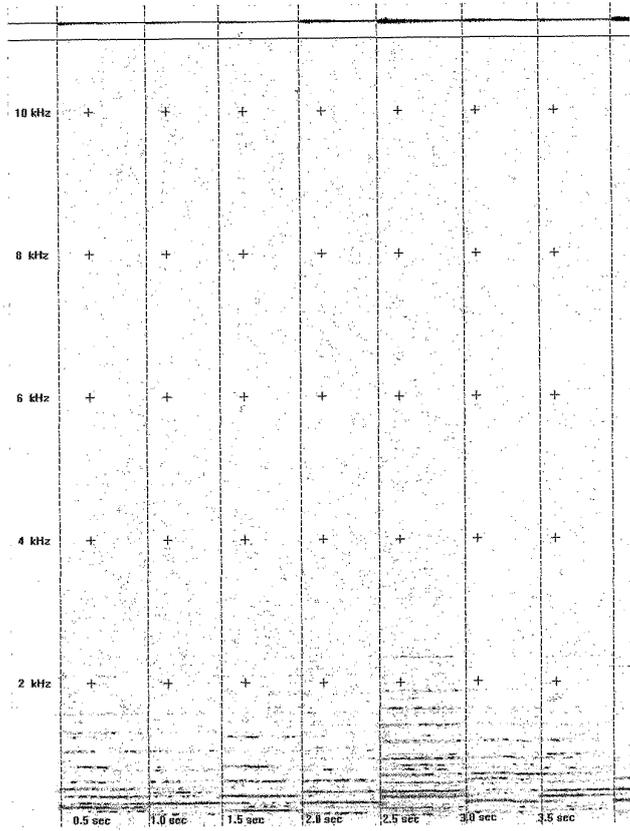


FIG.16B



WO 02/47064

PCT/KR01/02081

37/40

FIG.16C

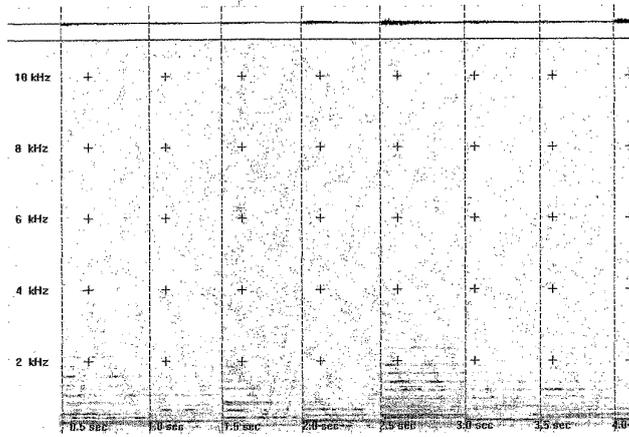


FIG.17A

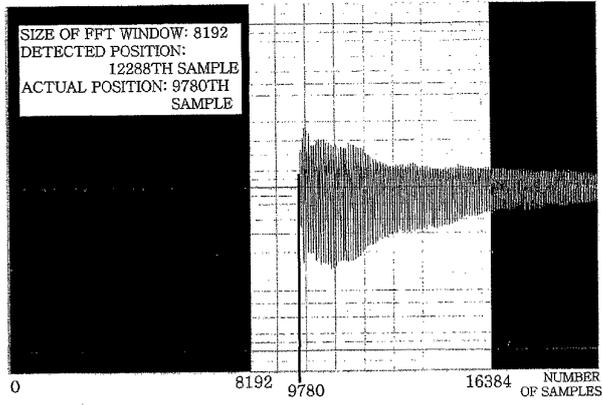


FIG.17B

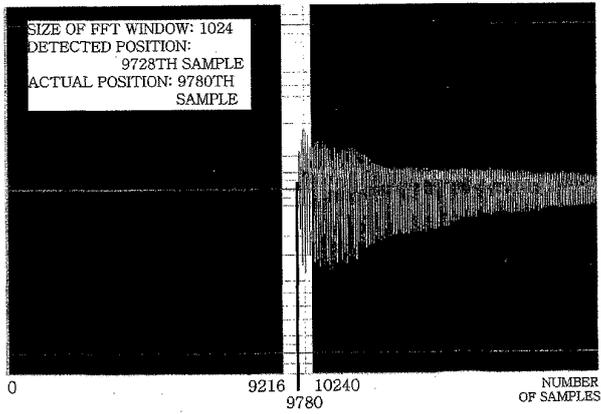
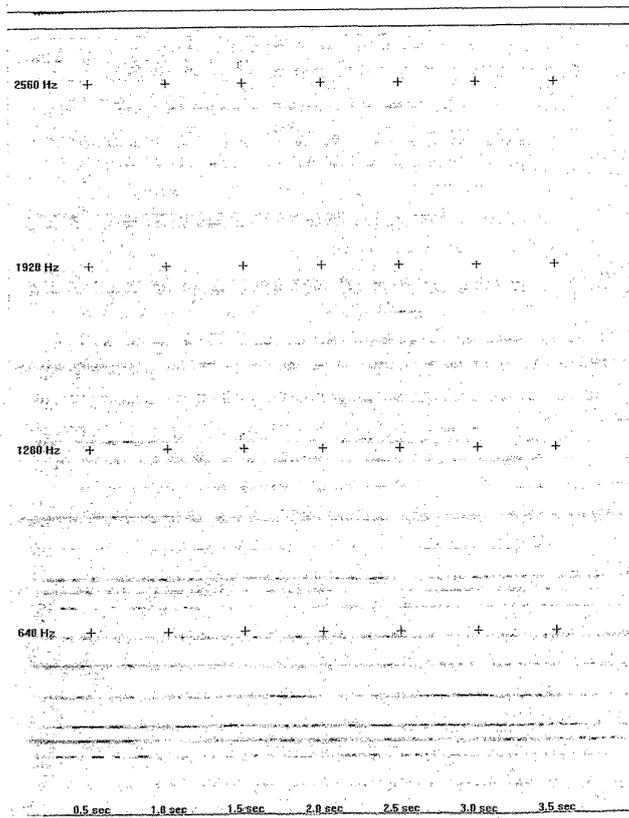


FIG. 18



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/KR01/02081
A. CLASSIFICATION OF SUBJECT MATTER IPC7 G10H 1/00 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) IPC7 G10H 1/00 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean Patents and applications for inventions since 1975 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) NPS : *music, instrument, frequency, analyze, digital, information, compare*		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, A	KR 2001-0016009 A (HANSUL SOFT, Inc.)05 March 2001 "the whole document"	1, 12
A	US 5,276,629 A (Reynolds Software, Inc.)04 January 1994 "abstract"	1, 12
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "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 "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 "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 "&" document member of the same patent family		
Date of the actual completion of the international search 21 FEBRUARY 2002 (21.02.2002)	Date of mailing of the international search report 22 FEBRUARY 2002 (22.02.2002)	
Name and mailing address of the ISA/KR Korean Intellectual Property Office Government Complex-Daedeon, 920 Dunsan-dong, Seo-gu, Daejeon Metropolitan City 302-701, Republic of Korea Facsimile No. 82-42-472-7140	Authorized officer NAM, In Ho Telephone No. 82-42-481-5761	

Form PCT/ISA/210 (second sheet) (July 1998)



INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR01/02081

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 2001-0016009 A	05. 03. 2001	None	
US 5,276,629 A	04. 01. 1994	None	

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT, BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN, TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE, GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PH,PL,PT,R O,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ジョン, ドイル

大韓民国, ソウル 156-710, トンジャク - ク, シンデバン - 2 - ドン, ボラメ - ナサン -
スイート 3103

Fターム(参考) 5D082 AA04 AA05 AA18 BB13 BB29
5D378 KK01 KK02 KK03 KK07