

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4008497号

(P4008497)

(45) 発行日 平成19年11月14日(2007.11.14)

(24) 登録日 平成19年9月7日(2007.9.7)

(51) Int. Cl. F I
H O 4 M 3/22 (2006.01) H O 4 M 3/22 Z
G 1 O L 19/00 (2006.01) G 1 O L 19/00 2 5 0

請求項の数 15 (全 19 頁)

(21) 出願番号	特願平9-530683	(73) 特許権者	ブリティッシュ・テレコミュニケーションズ・パブリック・リミテッド・カンパニー
(86) (22) 出願日	平成9年1月30日(1997.1.30)		イギリス国、イーシー1エー・7エージェイ、ロンドン、ニューゲート・ストリート
(65) 公表番号	特表2000-506327(P2000-506327A)		8 1
(43) 公表日	平成12年5月23日(2000.5.23)	(74) 代理人	弁理士 鈴江 武彦
(86) 国際出願番号	PCT/GB1997/000265		
(87) 国際公開番号	W01997/032428	(74) 代理人	弁理士 村松 貞男
(87) 国際公開日	平成9年9月4日(1997.9.4)		
審査請求日	平成15年12月11日(2003.12.11)	(74) 代理人	弁理士 橋本 良郎
(31) 優先権主張番号	96301393.3		
(32) 優先日	平成8年2月29日(1996.2.29)	(74) 代理人	弁理士 白根 俊郎
(33) 優先権主張国	欧州特許庁(EP)		
(31) 優先権主張番号	9604315.3		
(32) 優先日	平成8年2月29日(1996.2.29)		
(33) 優先権主張国	英国(GB)		

最終頁に続く

(54) 【発明の名称】 トレーニングプロセス

(57) 【特許請求の範囲】

【請求項 1】

第1の信号および第1の信号の歪んだ形式を含むトレーニングシーケンスを準備する手段(8)、トレーニングシーケンスを受取り、人間の観察者が知覚できる歪みの範囲を示す歪み知覚測度を生成する解析手段(9)、および歪み知覚測度および歪んだ信号を信号分類装置(5)に適用して、次に入力された信号に適用される分類を決定する手段を含む、信号分類装置(5)をトレーニングするためのトレーニング装置。

【請求項 2】

解析手段(9)が、歪んだ信号を複数のスペクトル成分帯域に分割し、スペクトル成分帯域がスペクトルマスキングを行うように形成され、その先行する時間の部分と後続する時間の部分の少なくとも1によって信号の時間マスキングを計算する測定手段(13、14)；各スペクトル成分信号に対して、歪んだ信号の成分信号と試験信号の対応して計算された成分との間の差の表示を形成する手段(15、16)；および前記差の表示から前記歪み知覚測度を生成する計算手段(17)とを含む請求項1に記載のトレーニング装置。

【請求項 3】

言語信号を解析するように構成された請求項1または請求項2に記載のトレーニング装置。

【請求項 4】

ビデオ信号を解析するように構成された請求項1乃至請求項3の何れか1項に記載のトレーニング装置。

【請求項 5】

人間の観察者に対して最大の知覚重要度をもつビデオ信号によって表される画像部分を識別する手段、および解析手段への入力で該画像部分に加重する手段とを含む請求項 4 に記載のトレーニング装置。

【請求項 6】

ビデオ画像内で空間周波数を解析する手段（31、32）を含む請求項 4 または請求項 5 に記載のトレーニング装置。

【請求項 7】

信号に存在する歪みの特性である信号内の所定の信号成分を検出し、歪みの存在と歪の厳格さと歪の形式の少なくとも 1 を示す出力を生成する手段をもつ信号分類装置であって、このような歪みを識別するために該信号分類装置をプログラムする請求項 1 乃至請求項 6 の何れか 1 項に記載のトレーニング装置を含み、2 以上の信号分類要素を含み、該信号分類要素の少なくとも 1 つが該トレーニング装置によってプログラムでき、少なくとも 1 つの信号分類要素を予めプログラムして所定の歪み形式を識別する信号分類装置。

10

【請求項 8】

トレーニング可能な信号解析プロセスをトレーニングする方法であり、第 1 の信号および歪んだ形態の第 1 の信号を含むトレーニングシーケンスを準備し、信号の歪みが人間の観察者に知覚可能になる程度の測度を生成し、生成された前記測度にしたがって分類動作を定義し、分類動作が知覚上重要な歪みが存在するか否かにしたがって信号を分類するようにされているトレーニング可能な信号解析プロセスをトレーニングする方法。

20

【請求項 9】

測定プロセスが歪んだ信号を複数のスペクトル成分帯域に分割し、該スペクトル成分帯域がスペクトルマスキングを行うように形成され；その先行する時間の部分と後続する時間の部分の少なくとも 1 によって信号の時間的マスキングを計算し；各スペクトル成分信号に対して、歪んだ信号の成分信号と試験信号の対応して計算された成分との間の差の表現を形成し；前記差の表現から前記歪み知覚測度を生成する請求項 8 に記載の方法。

【請求項 10】

分類動作から、存在する歪みの形式の表示を生成する段階を含む請求項 8 または請求項 9 に記載の方法。

【請求項 11】

信号が、オーディオ信号である請求項 8 乃至請求項 10 の何れか 1 項に記載の方法。

30

【請求項 12】

信号がビデオ信号である請求項 8 乃至請求項 10 の何れか 1 項に記載の方法。

【請求項 13】

請求項 12 に記載の方法であって、ビデオ信号の個々のフレームに対応するセグメントで信号を解析し、人間の観察者に対して比較的の高い知覚重要度の信号によって表示される画像の一部を識別する段階と、歪み知覚測定プロセスへの入力としてこのような画像の該部分に加重する段階と、画像要素の境界を識別する段階と、より大きな知覚重要度をもつ境界を含む画像の該部分に加重する段階と、ビデオ画像内の空間周波数を解析する段階とをさらに含む方法。

40

【請求項 14】

監視されるネットワーク全体で遠隔の場所から監視場所まで第 1 のトレーニングシーケンスを送る段階と；

監視位置で第 2 の同一のトレーニングシーケンスを生成する段階と；

請求項 8 乃至請求項 13 の何れかに記載の解析プロセスを実行して、第 2 のトレーニングシーケンスと比較することによって、遠隔の場所から監視位置で受信されるトレーニングシーケンスの歪み知覚度を測定する段階と；

前記歪み知覚度にしたがって生成された測度を分類する段階と；

生成された分類にしたがってトレーニング可能なプロセスを構成する段階とを含むトレーニング可能な信号分析装置をトレーニングする方法。

50

【請求項 15】

信号分類装置用の分類手段であって、該分類手段に記憶された分類データにしたがって、該信号分類装置が該装置への信号入力で発生する歪みを検出し分類するように構成されており、該分類手段に記憶されたデータが請求項 8 乃至 14 の何れか 1 にしたがって生成される信号分類装置用分類手段。

【発明の詳細な説明】

本発明は、トレーニング可能なプロセスを訓練するのに使用できるデータの分類に関する。遠隔通信システムによって搬送される信号の評価、例えば使用中の遠隔通信システムの状態を評価することに応用される。言語 (speech) を搬送するオーディオ信号、およびビデオ信号に対する応用を実施形態に記載することにする。

10

遠隔通信リンク上を搬送される信号は、多くの変換、例えばデジタル化、データ圧縮、データ低減、増幅、などを行うことができる。これらのプロセスの全てが信号を歪ませることができる。例えば、最大デジタル化値よりも大きい振幅をもつ波形をデジタル化するとき、波形のピークは上部が平らな形に変換されることになる (このプロセスはピーククリッピングとして知られている)。これは、望ましくない高調波を信号に付加する。外部ソースからの電磁干渉によっても歪みを生じさせることもできる。

上述の処理によってもたらされる歪みの多くは非線形であり、したがって簡単な試験信号は、言語信号のような複雑な波形と同じやり方では歪まないか、全く歪まない。データを搬送する遠隔通信リンクでは、全ての可能なデータ文字を使用してリンクを試験することができる; 例えば、二進リンク用の 2 つの文字、1 および 0、DTMF (デュアルトーンマルチフレクシイ) システムで使用される 12 のトーンの対、または QAM (クオダラチャ振幅変調) システムで使用される一定の範囲の “星座点 (constellation point)” が使われる。しかしながらアナログ信号はよく定義された信号要素の限られた数から構成されていないで、連続的に変化する信号である。例えば、言語信号の要素は言語の内容のみによってではなく、特徴、例えば音の高さ (ピッチ)、音量、特徴的な母音、などに影響を与える個々の話者の生理的および心理的特徴にもしたがって変化する。

20

遠隔通信装置を試験するのに、搬送される信号形式のサンプルを使用して、試験のシーケンスを行なうことが知られている。試験されている装置が変える試験のシーケンスと元の試験のシーケンスとの比較を使って、試験されている装置がもたらした歪みを識別することができる。しかしながら、この方法には予め (目的に合わせて) 構成された試験シーケンスの使用が必要である。したがってこの方法は活性の遠隔通信リンク、すなわち現在使用されているリンクでは使用できない。その理由は試験シーケンスは、搬送されるトラヒックと干渉し、ユーザが知覚できるからであり、加えて活性のトラヒックそれ自身 (トラヒックの内容は予め規定することはできない) が試験装置によって試験信号の歪みとして検出されないからである。

30

装置によって搬送される信号と干渉せずに (いわゆる、非侵襲的試験 (non-intrusive testing)) 使用中の装置で試験を実行するために、活性信号自体を試験信号として使用する試験を実行することが望ましい。しかしながら、活性信号を試験信号として使用することにおける問題は、測定点で元の信号のサンプルを即時に得る方法がないことである。元の信号を測定場所へ送る手段は、試験されるリンクと同様の歪みを受けることになる。

40

本発明の出願者の未決の国際特許出願第 W096/06495 号および第 W096/06496 号 (両出願とも 1996 年 2 月 29 日に公開された) は、この問題に対する 2 つの可能な解決案を提案している。第 W096/06495 号は、遠隔通信リンクが信号をどう変化させたかを判断するために、話者に依存しない言語の一定の特徴を解析を記載している。第 W096/06495 号はさらに、他の特徴に関係して変化する言語の一定の特徴 (それ自体は直接には測定できないものである) の解析も記載している。他の特徴は一定の特徴が個々の話者間で一貫しているもので、これを使用して、これらの他の特徴に関する情報を導出することができる。例えば、無声摩擦音のスペクトル内容は音量 (振幅) で変化するが、ある意味で個々の話者で独立している。したがってスペクトル内容を使用して、元の信号振幅を推定でき、元の信号振幅を受信した信号振幅と比較して、話者と測定点との間の減衰を推定することができる。

50

W096/06496号明細書では、受信した信号の内容を言語認識器によって解析し、この解析結果を言語合成器によって処理して、歪みのない言語信号を再生する。例えば国際特許出願第W094/00922号および第W095/15035号に記載されている知覚解析技術を使用して、信号を音のピッチと継続時間で正規化して、元の言語信号の推定を生成して、受信した言語信号と比較して、歪みまたは干渉を識別することができる。

通常、制限された帯域幅における言語送信は、データ整理、例えば線形予測コーデック（LPC）を使用する。このコーデックは、人間の声道に対する近似に基いており、声道モデルで等価の振舞いを励起するのに必要なパラメータとして言語波形のセグメントを表している。

1996年7月25日に出願された出願者の国際特許出願第PCT/GB96/01821号は、1995年7月25日に
10 出願された欧州特許出願第95305313.9号からの優先権を主張しており、言語を搬送する信号品質を評価する方法および装置であって、スペクトル表示モデル（不完全な声道モデルが好ましいが、その代りに可聴モデルを使用することができる）にしたがって信号を解析して、出力パラメータを生成し、この出力パラメータを所定のネットワーク定義関数にしたがって分類し、出力の分類を生成する方法および装置を開示している。予備の段階で出力値の分かっているデータから導き出されるネットワーク定義関数にしたがって、分類を行うことができる。その代りに、分類は試験されるシステムの一定の状況のもとで発生することが分かっている既知の特徴から導き出される所定の規則にしたがって導き出すことができる。

この文脈で“可聴モデル（auditory model）”という用語は、刺激に対する応答が人間の
20 可聴システム（すなわち、耳―脳の組合わせ）の応答とほぼ同じであるモデルを意味する。したがってより一般的な用語で“知覚（perceptual）”モデルの特定のカテゴリであり、すなわち刺激に対する応答が人間の感覚システム（すなわち、目―脳、耳―脳、など）の応答とほぼ同じモデルである。

この文脈で“不完全な声道モデル（imperfect vocal tract model）”という用語は声道モデル、すなわち“理想的”ではないが、人間の声道では生成できない可聴スペクトル要素に関係する係数を生成することもできる声道モデルを意味する。とくに、これは言語および歪み信号要素の両方をパラメータで表すことができるモデルを意味するが、声道モデル設計の正規の目的ではない。良いまたは悪い状態であることが分かっている言語サンプル（すなわち、それぞれこのような歪み要素を含んでいるものと含んでいないもの）を声道モデルによって解析して、次に生成された係数を、トレーニング可能なプロセス、例えばニューラルネットワークによって良い状態、または悪い状態の信号として識別することができる。この方法では、各形式の信号に関係付けられた声道パラメータに対して分類データを生成することができ、（両方の状態と関係付けられているパラメータは信頼できないので、分類データを生成するとき無視することができる）、したがって続いて未知の信号を処理するとき、未知の信号に関係するパラメータと関係付けられた前に生成された分類データを使用して、出力を生成することができる。

さらにパラメータのシーケンスと一緒に個々のパラメータを使用して、信号を特徴付けることができる。データ圧縮技術を使用して、記録したパラメータを記憶することができる。
40

上述の欧州特許出願の装置は、分類データの記憶した組を生成するトレーニング手段を含み、トレーニング手段は、モデリング手段へ言語サンプルを供給する第1の入力手段；言語サンプルに関する既知の出力情報（以下、“ラベル”と記載する）をトレーニング手段へ供給する第2の入力手段；ラベルに基いてモデリング手段から分類データを生成する手段、およびモデリング手段によって生成される分類データを記憶する記憶手段を含んでいる。

したがってトレーニングサンプルで使用する各言語セグメントは、良いまたは悪い状態としてラベルを付けなければならない。ラベルを付けることは主要な仕事である。その理由は、通常のサンプルは数時間の言語を含み、システムをトレーニングして、一定の範囲の話者、状態、および他の変数に正しく応答するのに多くのこのようなサンプルが必要だか
50

らである。個々のセグメントの継続時間は通常20ミリ秒であり、したがって全部で数百万セグメントをラベル付けしなければならない。その上知覚、集中、および他のファクタの個々の変化のために、多数の人間の解析者を使用して各サンプルを分類して、統計的に有効な結果を得ることが必要になる。さらに、人間の観察者はこの短い継続時間の個々のセグメントが良いまたは悪い状態であるか否かを正確に識別することはできない。

本発明の第1の態様にしたがって、第1の信号および第1の信号の歪んだ形式を含むトレーニングシーケンスを準備する手段、トレーニングシーケンスを受取り、人間である観察者が知覚できる歪みの範囲を示す歪み知覚測度を生成する解析手段、および歪み知覚測度および歪んだ信号を信号分類装置に適用して、次に入力された信号に適用される分類を決定する手段を含むトレーニング装置を提供する。

10

本発明は別の態様では、トレーニング可能な信号解析プロセスをトレーニングする方法であり、第1の信号および歪んだ形態の第1の信号を含むトレーニングシーケンスを準備し、信号の歪みが人間の観察者に知覚可能になる範囲を測定し、前記測定結果にしたがって分類動作を定め、分類動作が知覚上重要な歪みが存在するか否かにしたがって信号を分類するようにされているトレーニング可能な信号解析プロセスをトレーニングする方法を提供する。

好ましい構成では、測定プロセスが同じ信号の歪んだ形態と歪んでいない形態によって人間の感覚システムに生じる効果を推定し、前記効果の差を判断し、前記差に依存して前記歪み知覚測度を生成する。好ましくは測定プロセスは前記歪み知覚測度を生成して、歪み知覚測度が人間である観察者に対する前記歪みの重要度に依存し、かつ前記歪みの振幅に非線形に依存するものとする。好ましくは測定プロセスは、前記試験信号または前記歪んだ信号、あるいはその両方の複数のスペクトル成分を生成し、幾つかのスペクトル成分信号において、そのスペクトル成分信号が人間の感覚システムに対して生成することになるマスキング効果を推定する。

20

言語応用において、トレーニングシーケンスは、通常自然言語の大きなコーパスであり、個々の話者の特徴の変化を考慮に入れることができる。好ましい実施形態では、測定プロセスは、歪んだ言語信号を複数のスペクトル成分帯域に分解して、スペクトル素子帯域をスペクトルマスキングを生成するように形成する段階と；その先行するまたは後続する、あるいはその両方の信号の時間的なマスキングを計算する段階と；各スペクトル成分信号に対して、歪んだ信号の成分信号と試験信号の対応して計算された成分信号との間の差の表示を形成する段階と；前記差の測度から前記歪み知覚測度を生成する段階とを含む。

30

適切な言語解析プロセスは、国際特許明細書第W094/00922号、第W095/01011号、第W095/15035号に記載されている。セグメントを自動的にラベルを付けすることによって、歪み知覚測度を使用して、合目的的に、しかしそれにも関わらず人間である観察者に知覚可能なファクタにしたがって、分類動作を導き出すことができる。

本発明は、言語信号、またはさらにオーディオ信号にも限定されない。同じ原理を、例えばビデオ信号に応用することができる。このような場合、ビデオ信号の個々のフレームはトレーニングシーケンスの個々の要素を形成することができる。

ビデオ信号は度重なる歪みを受けるもので、それが見る人（viewer）に知覚できる効果を生むことができる。一般的に、異なる画像要素間の境界に影響を与える歪みは、1つの画像要素の一部分として観察者によって知覚される実在物本体内で起きる変化よりも知覚的により重大なものである。このような境界の歪みには、不鮮明さ（blurring）、変位（それによってオブジェクトの形状を変化させる）、境界が完全に消失すること、境界が無いはずのところに境界が出現することが含まれる。したがって好ましい構成では、このシステムは、他の形式の歪みよりも大きい知覚重要度をもつ境界の特徴に影響を与える歪みを識別する。

40

境界は画像のある特性、すなわち普通は明るさまたは色彩、あるいはその両方に、突然の変化があるとき、見る人によって知覚される。空間的に近いこのような明るさと色彩との2つの変化は単一の境界、例えば、そうでなければ明るさと色彩とが類似している2つの領域を分けるラインとして知覚することができる。したがって信号から導き出した画像の

50

スペクトル分解によって、境界を識別することができる。突然の変化は、スペクトル分解したときに高周波数の“スパイク”を生成する。カラーシステムでは色彩の変化は、画像を形成している異なる色彩の相対的な明るさの変化として識別できる。

歪み効果はビデオ信号のプロデューサによって意図的に導くことができるので、識別した効果が持続する（伝送の失敗を示唆する）か、または遷移する（慎重に導き出されたことを示唆する）か否かを識別するのに十分に長い時間、ビデオ信号を監視することが好ましい。このような効果は、言語システムで発生する確率は低い。

トレーニング可能なプロセス、例えばニューラルネットワークは、簡単な二進試験（良い／悪い；イエス／ノー）で最も効果的に機能する。したがって監視システムを構成して、このような多数のプロセスを個々に実行し、各プロセスで異なる特性、または特性の組合

10

わせを試験し、1または複数の異なるパラメータに関係付けることが効果的である。本発明を使用して、トレーニング可能なシステムをそのままトレーニング（訓練）するか、再訓練することができる。本発明ではトレーニング可能なシステムを実際のシステム上でトレーニングして、監視されるシステムが生成する新しい特徴を認識できるようにする。このそのままのトレーニングは、監視されるシステム上でトレーニングシーケンスを送り、（システムの1チャンネルを一時的に占め）、受取ったシーケンスを監視位置で生成される同じシーケンスの同一サンプルと比較することによって行うことができる。

ここで、本発明の例示的实施形態を実施形態の種々の要素の機能上の関係を示す添付の図面を参照して記載する。本発明は、汎用コンピュータで走行するソフトウェアで効果的に実現できることが分かるであろう。

20

図1は、トレーニングプロセスのために構成された、言語信号を解析するためのトレーニングシステムの機能素子を示す。

図2は、図1のトレーニング装置をより詳しく示す。

図3は、図2の一部を形成している解析装置をより詳しく示す。

図4は、図2のデータソースによって供給される最初の言語サンプルを生成できる装置を示す。

図5は、未知のデータで走行するように構成された同じシステムの機能素子を示す。

図6は、ビデオ信号を解析する装置をトレーニングするための、図3の装置に類似した解析装置を示す。

図7は、ビデオ信号を解析する装置、とくに図5に示した装置に類似し、図6の装置を使用してトレーニングされた装置のブロック図を示す。

30

図8は、図7の装置の変形であり、複数の分類装置をもち、さらに元の場所の再トレーニングプロセスを示すブロック図を示す。

ここで3つの実施形態をより詳しく記載する。第1に、図1乃至5に関係し、言語を運ぶ信号のために構成された実施形態を記載することにする。

この実施形態の理解を助けるために、ここで声道解析システムおよびトレーニング可能なプロセスの特徴を簡単に記載することが適切である。声道は非均一の音響管（non-uniform acoustic tube）であり、声門から唇部へ延在し、時間関数として形状が変化する（文献[Fant G C M, "Acoustic Theory of Speech Production", Mouton and Co., 's-grave hage, the Netherlands, 1960年]に記載されている）。時間的变化をもたらす主要な解剖学上の要素（anatomical component）は、唇、顎、舌、および軟口蓋である。計算（コンピュータ処理）を簡単にするために、このシステムのモデルは線形で、かつ時間で変化するものがないのがよい。都合の悪いことに、人間の発声機構（speech mechanism）は、これらの特性の何れも正確に満足させない。言語は継続して時間で変化するプロセスである。加えて、声門は声道から分離されず、したがって非線形の特徴をもつ（文献[Flanagan J L

40

"Source-System Interactions in the Vocal Tract", Ann. New York Acad. Sci. 155, 9-15, 1968年]に記載されている）。しかしながら合理的な仮定をたてることによって、言語イベントを記載するのに短い時間間隔で線形の時間によって変化するモデルを展開することができる（文献[Markel J D, Gray A H, "Linear Prediction of Speech", Springer-Verlag Berlin Heidelberg New York, 1976年]に記載されている）。線形予測コ

50

ーデックは言語イベントを短い時間間隔、すなわちフレームに分割し、先の言語フレームを使用して、予測パラメータの独特の組を生成して、現在のフレーム内の言語を表す（文献 [Atal B S, Hanauer S L "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave" J. Acoust. Soc. Amer., vol.50, pp.637-655, 1971年] に記載されている）。線形予測解析は、このような言語パラメータ、例えばピッチ、フォルマント、およびスペクトルを評価する方法として幅広く使用されるようになった。可聴モデル（時間／周波数／振幅スペクトログラム）は、監視される可聴音のフィーチャ（feature）に依存し、可聴モデルがどのように生成されるかを考慮しないが、一方で声道モデルは、信号が言語に類似したもの（speech-like）か否か、すなわち実際の声道がそれを生成できたか否かを識別できる。したがって非可聴音の差で可聴モデルによって認識されないものでもなお声道モデルでは認識できることになる。

10

信号品質を測定するために、生成された出力パラメータは測定される性質、すなわち知覚される言語品質に敏感でなければならない。したがってモデルは言語に類似したものにはない歪みをモデル化できなければならない、したがって理想的な声道モデルは適切でなくなる。理想的なモデルは、全ての入力信号を言語に類似した形態に変換することになる（歪みが悪いときは、元のものにする必要はない）。これは、歪んだ入力および純粋な入力の両方を言語に類似したものとして分類するので、分類プロセスを不確実にし、トレーニングプロセスを不可能にする。したがって、上述で定めた意味で、声道モデルは“不完全”であることが重要である。その理由は、人間以外の歪み要素の存在に敏感な声道モデルからの出力パラメータを信頼して、良い状態の信号と悪い状態の信号とを区別するからである。線形予測コーディングモデルは、文献（"Digital Processing of Speech Signals":Rabiner L.R.;Schafer R.W; (Prentice-Hall 1978年) 396頁）に記載されており、解析器 3として使用するのに適している。

20

声道モデルの代りにスペクトル解析を使用することができ、これには例えば“3分の1オクターブ解析（one-third octave analysis）”（R.B.Randallによる文献“Frequency Analysis”の3.6章（Bruel & Kjaer, 1987年（ISBN 87 87355 07 8）発行）に記載されている）がある。

ここでトレーニング可能なプロセス、特にニューラルネットワークの特徴について記載する。多数の入力をより少ない数の所定の結果のクラスにマップするためには、マッピングプロセスが自然系を表すときこれはとくに、一連の規則を使用することができる。しかしながら、自然系が複雑過ぎるときは、または必要なマッピングが抽象的なパラメータを操作するとき、トレーニング可能なプロセスを使用して、一連の既知の結果、すなわちトレーニングデータとよばれるものに応答して、要求されるマッピングを展開することができる。既知の結果を使用して、入力パラメータと結果のクラスとの間の関係を判断して、入力の次の未知の組み合わせを分類することができる。脳が特定の関心のある特定のタスクまたは機能を実行する方法をモデル化するように、ニューラルネットワークを設計する。ニューラルネットワークをトレーニングして、学習プロセスで有益な計算を行うことができる（文献 [Haykin S, "Neural Networks, A Comprehensive Foundtion", Macmillan IE EE Press,1994年] に記載されている）。良好な動作を達成するために、ニューラルネットワークは簡単な処理装置を大掛かりな相互接続を採用する。加重（重み付け）として知られる処理装置間接続強度を使用して、システムの知識を記憶する。これは文献 [Aleksander I, Morton H "An Introduction of Neural Computing" Chapman and Hall London, 1990年] に記載されている。学習プロセスを実行するのに使用する手続きを学習アルゴリズムと呼び、その機能は順番にネットワークの加重（重み）を変更して、所望の設計目的を達成することである。ニューラルネットワークの能力は、強力な並行分散構造およびその学習および一般化能力から導き出される；一般化は、トレーニング中に出会わない入力に対する合理的な出力を作るネットワークをいう。監督される学習はトレーニングの一形式であり、ネットワークに対するクラスの既知の例を与えて、相互接続される加重に変更を与えて、システムの所望の応答と実際の応答の間の差を最小にすることを含む。ネットワークが安定状態に到達するまで、各入力クラスからの多数の例についてトレーニングを

30

40

50

反復する。ニューラルネットワークによって実行される入力-出力マッピングと、パラメータ以外の統計的な推論によって達成される分類とは、密接に類似がある。

ここで第1の（オーディオの）実施形態のシステムの動作を記載する。図1および5に示したシステムは、トレーニングデータソース1（図1参照）および生の言語トラヒック（実際のデータ）ソース2（図5参照）を含み、両者は解析器3へ入力を供給している。トレーニングデータと関係付けられたパラメータもトレーニングデータソース1から分類ユニット5へ供給される。分類ユニット5は、トレーニング可能なプロセスとして示されており、特にこの実施形態ではニューラルネットワーク5である。他のトレーニング可能なプロセス、例えば適応クラスタリング（adaptive clustering）を使用できることが認識されるであろう。解析器3によって出力されるパラメータをニューラルネットワーク5に供給する。トレーニングプロセス中にニューラルネットワーク5はメモリ4へパラメータを供給する。これらのパラメータはネットワーク定義関数を定める。実際のデータを読取るとき、パラメータをメモリ4から呼出して、ニューラルネットワーク5で使用して、声道解析器3によって生成される値でネットワーク定義関数を実行し、分類データを生成して、出力6へ供給する。通常出力データは、解析器3によって生成される値に基づいて分類されて、ニューラルネットワーク5へ入力され、ニューラルネットワーク5はネットワーク定義関数にしたがって動作し、識別された歪みの程度を示す。多数の出力クラスを設定することによって、幾つかの品質レベルを定めることができる。実際の目的では、信号は時間フレームのシーケンスとして解析される。第1の時間フレームに関係するデータから導き出されたパラメータは、次の時間フレームの解析で使うことができる。このために声道解析3の出力はバッファメモリ7に記憶され、後でニューラルネットワーク5の次の動作で使用する。

本発明のトレーニングプロセスに関して記載する前に、図5を参照して一旦トレーニングしたシステムを使用して、生のトラヒックを監視することを記載する。実際のデータは、ソース2から声道解析システム3へ供給することができる。歪みおよび干渉は、元の信号の個々の時間フレームを歪ませたり、纏めて無くしてしまったりすることができる。例えば所定のフレームが可能なフレームの小さいサブセットの1つに続いてのみ現れることができる。すると、該サブセットのメンバでないフレームに後続するフレームの出現は、従属フレームまたはその前のフレームの何れか（あるいは両者）がコンテキストに適していた某かの元のフレームから歪ませられたことを示す。各個々のフレームのパラメータは“許容”できる（すなわち、パラメータが予測される範囲に入る）が、パラメータのシーケンスは、全体を考慮すると、無効とされ、歪んでいることを示す。メモリ4に記憶したパラメータは、このシーケンスでトレーニングしたネットワーク定義関数を定める。声道解析によって生成されるパラメータは入力としてニューラルネットワーク5へ供給され、ネットワーク定義関数によって定義され、声道解析によって生成されるデータを分類して、出力6を生成する。ネットワーク定義関数はメモリ4に記憶されたパラメータによって定められ、ソース2に供給された信号品質の分類を導き出す。

時間に依存する特性に関するパラメータを含む、例えば、サンプルの瞬間的な特徴が人間の声道の能力内であるか否かのみを識別するのではなく、時間にしたがって変化する特性も人間の声道の能力内にあるかどうかを識別するために、声道解析からの出力をバッファメモリ7に記憶する。次のサンプルをニューラルネットワーク5へ供給するとき、“過去”のデータとして記憶したパラメータをニューラルネットワーク5の入力として供給し、この時間に依存するサンプルの特徴を測定する。

多くの個々の遠隔通信リンクをソース2として継続的に接続して、多数のリンクの信号品質を監視することができる。本発明は、非侵襲的測定プロセスに特に適しているが、いわゆる“侵襲的”測定にも使用でき、この場合は、試験信号を生でなくソースとして使用する。

出力6はユーザにとって適切な形で表示することができる。例えば、不具合な性質を表わす分類が生じたソース2がネットワークマネージャに示されて、ソース2によって代表された遠隔通信リンクがサービスに使われないようにでき、もし必要であれば修理し、可能

10

20

30

40

50

であれば別のルート設定によってリンクを再び確立することができる。1つの可能な構成では、このような動作が自動的に制御されるか、または人間の制御者に任して、出力6によって供給される指示について実行することができる。

各時間フレームに記録されたパラメータは、パラメータを表す短いコードとして記憶することができる。短いコードが占めるメモリはより小さくなり、さらに処理時間を著しく短縮することができる。連続する時間フレームのコードのシーケンスは、コードが表すパラメータのように、実際の言語音に対応する多数の認識されたシーケンスの1つにしたがわなければならない。1組のパラメータを時間フレームに対して識別し、この時間フレームが、シーケンス内の前のメンバにしたがわないコードをもつか、または全くコード化されないときは、歪みが存在することを示している。

10

メモリ4に記憶したパラメータを生成するために、トレーニングデータを使用して、ニューラルネットワーク5を最初にトレーニングして、ネットワーク定義関数を確立しなければならない。このプロセスは図1乃至4に記載されている。試験データはトレーニング装置1から声道解析器3へ供給される。トレーニング装置1はさらに、試験データに関する分類パラメータをニューラルネットワーク5へ供給し、ネットワーク定義関数を定めるラベルを生成することができる。

ここでトレーニング装置1をより詳しく記載した図2を参照して、これらのラベルの生成を記載することにする。人間のオペレータによって個々に正確にアクセスするには短すぎる言語セグメントを使用して、ニューラルネットワークをトレーニングするために必要なデータ量を生成するために、このような信号を生成する自動的な方法を案出した。このプロセスは知覚解析モデルの使用に依存しており、すなわち信号の歪みが人間の観察者にとって重要であるかどうかを評価するプロセスである。最初に試験信号ソース8が用意され、2つの関係付けられたメモリ(8a,8b)をもっている。第1のメモリ8aは“良好な”信号サンプルをもつ。完全なサンプルは、一般に数時間の長さである。第2のメモリ8bは、同じサンプルの対応する形式であり、以下に記載する手段によって歪まされている。第2のメモリ8bに記憶したサンプルは歪みの程度および形式が変わる。歪んだ信号は短いセグメントに分割される。メモリ8aからの良好な信号、およびそれに対応するメモリ8bからの歪んだ信号は、それぞれ第1および第2の入力11、12を介して解析ユニット9に供給され、解析ユニット9はラベルのシーケンスを含む出力を供給し、この出力はニューラルネットワーク5へ送られる(図1参照)。歪んだ形式の信号はさらにセグメンタ(segmentor)10へ送られ、セグメント10はこの信号をラベルに対応する個々のセグメント(通常、20ミリ秒)に分割する。次にこれらのセグメントは声道解析器3へ送られる(図1参照)。解析ユニット9は、“良好な”サンプルを歪んだサンプルと比較して、ラベルのシーケンスを生成する。ラベルのシーケンスは、各セグメント内に存在する歪みが、このモデルによって人間の聞き手が知覚できると考えられる程度を表す。この解析プロセスはここで概略的に記載するが、使用する解析技術は、公開された国際特許出願第W094/00922号、第95/01011号、および第95/15035号に記載されたものが特に適している。

20

30

図3は、解析ユニット9をより詳しく示している。それぞれが“良好な”信号と良好な信号の歪んだ形式を帯している第1および第2のメモリ(8a,8b)からの入力11および12は、可聴モデル(それぞれ13、14)へ供給されそこを通り、可聴モデルの出力はコンパレータ15で比較される。当業者に明らかなように、その代りに、良好な信号と歪んだ信号の対応する流れを同じ可聴モデルを通すように交番に供給し、良好な信号と歪んだ信号をこの可聴モデルの出力間で比較することができる。何れにしても同じプロセスを両方の信号に適用することが重要である。このモデルは、個々の信号セグメントの特徴についての知覚重要度に関係する多数のパラメータを生成する。このプロセスは、サンプルを種々の重なり合うスペクトル成分に分割することも含み、重なり合うフィルタを使用して、同時マスキング現象をモデル化し、そこではある音でこの音に近い周波数をもつより静かな同期音をマスクすることになり、また、各セグメントを1または複数の前または後のセグメントと比較して、時間的なマスキング現象をモデル化し、より大きな音のすぐ前または直ぐ後の静かな音をそのより大きな音が存在しないときよりも知覚できにくくすることを含ん

40

50

でいる。

上述の特許明細書に記載されているように、可聴モデルプロセスは、サンプルの各スペクトルおよび時間成分の知覚重要度の一連の値を生成する。サンプルはスペクトルおよび時間の両方で解析するので、この一連の値を表面として視覚化するのに便利である。この表面では、時間およびピッチ軸を定め、これらの軸によって定められる面上の表面の高さによって各時間／スペクトルの座標の対に対して知覚重要度を表すことによって、各スペクトル／時間成分の知覚重要度を表す。本明細書ではこの表面を“可聴表面”と記載する。この表面を定義する値は、もちろんデジタルで記憶され処理される。

次に“良好な”サンプルおよび歪んだサンプルに対応する2つの可聴表面をコンパレータ15で比較して、一連のエラー値を生成し、エラー表面生成ユニット16でエラー値をコンパイルして、エラー表面を形成する。上述の公開された国際特許明細書に詳細に記載されているように、エラー表面は本質的に多数の時間セグメントおよび周波数またはピッチ帯域（例えば信号をバーク目盛（Bark scale）に合わせることによって、等しい知覚重要度をもつように、帯域の個々の範囲が選択されている）にわたる測度であり、知覚された音信号の大きさは、ピッチおよび時間軸の両方に垂直な座標軸上に表される。異なる加重を正および負の値に適用してよく、例えば、付加されたノイズに例えられる信号損失により生ずる減損の差を説明する。歪みが全く存在しないときは、エラー表面は全表面でゼロの値をもつことになる。上述の例のように、エラー表面上の値は可聴モデル出力間の差（上述のように加重された結果である場合が多い）の絶対値として判断され、したがってエラー表面の全ての値は正である。

上述の特許出願に記載されているように、エラー表面の特徴を使用して、その上を運ばれるエラーの知覚重要度に対する値を導き出すことができる。とくに国際特許出願第W095/15035号に記載されているように、この値はエラー表面上で集計されるエラーの絶対値であってもよい。“聞く努力（listening effort）” Y_{LE} に対する最後の加重値を導き出すことができる：

エラー活性度,
$$E_A = 10 \log \sum_{i=1}^n \sum_{j=1}^m |c(i, j)|$$

なお、 $c(i, j)$ は、解析されるエラー表面における i 番目の n 時間セグメントのおよび j 番目の m ピッチ帯域のエラー値である。これは、存在する歪みの絶対量の表示を与える。

個々のセグメントに対するエラー活性度 E_A に対する適切な閾値を使用して、特定のセグメントを“良い状態”または“悪い状態”としてラベルを付けるか否かを判断することができる。このように生成されたエラー表面の特性を使用して、エラー表面発生器16によって定められるエラー表面の特徴に適したラベル発生器17内でラベルを求める。これらのラベルは、セグメンタ10内の信号のセグメント化と同期して生成される。ラベルは、ニューラルネットワーク5へ出力される（図1参照）。

メモリ8で使用する歪んだ信号と“良好な”信号のソースは、予め生成されたメモリから供給することができる。適切なデータの種々のコーパスは既に使用可能であるが、別のデータを容易に生成することができる。このようなデータの生成は比較的簡単であり、これを図4に示した。ここでソース18からの最初の試験信号は、異なる話者を使用して、表示選択を確実にすることができる実際の言語の幾つかのサンプルを含み、“良好な”メモリ8aに供給される。同じ信号は歪み発生器19にも供給される。生成された歪んだ信号は“歪んだ”信号メモリ8bに記憶される。歪みの種々のソースを適用することができる。異なる試験信号および歪んだ形式の信号の種々の入換え（permutation）を使用することによって、試験データの大きな表示コーパスを生成して、トレーニングデータソース1によって供給されるトレーニングデータとして供給することができる。

典型的な形式の歪みを歪み発生器19によって試験信号に与えて、トレーニングプロセスへこのような信号の代表的な選択を与える。これらの歪みを生成して、種々の効果をシミュレートすることができる。この種々の効果は、試験装置または実際のシステム、例えば遠

10

20

30

40

50

隔通信ネットワークの何れかにおいて、アルゴリズムで（すなわち、サンプルの数学的操作、例えばプロトタイプシステムをエミュレートすることによって）または元の信号を実際の装置に送ることによって生成することができる。

トレーニング装置 1 によってニューラルネットワーク 5 に供給されるラベルは、送られているトレーニング信号の特性に関する情報をネットワークに与え、したがってこれらの特徴をもつデータに関係する、メモリ 4 に記憶した種々のパラメータに適切な加重を加えることができる。異なる形式の歪んだ信号および歪んでいない信号の例をトレーニング装置 1 によって供給し、したがって出力 6 は、知覚可能な歪みが存在するだけでなく、歪みによって生じる減損の範囲、すなわち歪みの存在をどのように聞き手に分配しているかも識別することができる。

10

ネットワーク定義が正確であることを保証するために、分類が予め分かっているような試験データを入力 2 に供給し、ニューラルネットワーク 5 内のネットワーク定義関数によって生成された分類データを（示されていないデータによって）既知の分類データと比較してもよい。

上述のオーディオシステムを言語信号と関係させて記載したが、他のオーディオ信号の適切なサンプルを使用することもできる。他の形式の信号、例えばビデオ信号を同じやり方で解析することもでき、ここで記載することにする。

一般に、ビデオ信号は音響チャンネルと 3 つの主要な視覚成分を含む。幾つかの特定の応用で、これらの成分は表示される実際の赤、緑、および青の画像成分である。しかしながら、モノクロム（“黒”および“白”）とカラーシステムの両立を可能にするために、大抵のシステムでは視覚成分は輝度（“明るさ”）信号、（モノクロムおよびカラー受信機の両方で使用される）および 2 つの“色差”信号（カラー受信機のみで使用）される。2 つの色差信号は、全輝度のどれくらいの量が画像のそれぞれ青色および赤色の成分によって与えられるかを示す。第 3 の（緑色）の成分はルミナンスおよび色差信号から導き出すことができ、したがって全ての輝度のバランスを形成している。ルミナンス信号および色差信号を使用して、3 単色画像（赤色、緑色、青色）、すなわち色を重ね合せると、全色画像を生成する画像の個々の発生器に指令を発生する。

20

図 6 は、図 3 に示した装置に類似した装置を示しているが、ビデオ用のトレーニングに適用して構成されている。元の信号および品質を下げた信号ソースは各人間の視覚フィルタ 31、32、さらに明るさ、活性度、および明るさマスキング装置 33、34 に送られる。次に比較装置 35 でこれらの信号を比較する。

30

比較ユニット 35 からの出力はエラー表面生成ユニット 36 へ送られ、画像内の異なる点におけるエラー知覚度に対して一連の値を生成する。したがって、“エラー明るさマップ”が生成され、エラー知覚度が画像全体でどのように変化するかを示す。エラー表面生成ユニット 36 からの出力は、ラベル生成ユニット 37 へ入力される。素子 31 乃至 37 の全ては、図 3 に示した言語解析システム内の素子に相当し、各素子は図 3 の等価の素子と同じ最終デジットをもつ。加えて、各信号に関係する出力も、マスキングユニット 33、34 から各画像分解ユニット 38 a、38 b へ送られる。エラー表面生成ユニット 36 からの出力は、ラベル生成ユニット 37 へ送られる前に、画像分解ユニット 38 a、38 b の出力にしたがって加重値を供給してエラー主観性ユニット（error subjectivity unit）39 で変更する。

40

図 7 は、トレーニングされたユニットの基本素子を示し、この基本素子は生のデータを解析するように構成されている。これらの素子は全て図 5 に示した言語解析システムに類似しており、等価の素子は同じ参照番号をもつが、図 2 では参照番号“2”に予め定められている。基本素子は図 5 に示された基本素子に類似しているが、ビデオ信号で使用するためのものである。サンプリングユニット 22 はビデオ信号のソースに接続されている。次にサンプリングユニット 22 は信号をパラメータ抽出ユニット 23 へ送り、パラメータ抽出ユニット 23 は生成されたパラメータをバッファメモリ 27 を介して分類ユニット 25 へ送る。分類ユニット 25 はネットワーク定義関数メモリ 24 からの入力も備えている。分類ユニット 25 は、ネットワーク定義関数 24 にしたがって決定された入力パラメータの分類を生成し、出力 26 へ送る。ビデオ信号のサンプルはデータ流 20（例えば、2 Mbit/秒）からサンプリング

50

ユニット22によってとられる。次に各サンプルはパラメータ抽出ユニット23によって処理される。パラメータ抽出ユニット23は多数の機能を実行して、ビデオ信号が良い状態であるかまたは悪い状態であることを示す特徴を識別する。これらのパラメータは分類ユニット25（通常はニューラルネットワーク）へ送られ、元のビデオ信号が良い状態であるか、または悪い状態であることを表す出力26を生成する。

パラメータ抽出ユニットによって実行されるプロセスは、通常スペクトル解析、境界検出および解析、並びに時間的に隣接するフレームとの相関を含み、スペクトル成分、境界の位置、および相関内の空間オフセットを含む相関度に関するパラメータを生成する。

次にこれらのパラメータをニューラルネットワーク25によって処理して、特定の形式の歪みの知覚的に関係するパターンの特徴を識別する。抽出されたパラメータは時間的に隣接するフレーム（先行または後続のフレームの何れか）との相関を必要とすることができ、このために先のサンプルのバッファメモリ27は、一番最近に受取ったサンプルとの比較のために記憶される。未知のビデオ信号によって生成される画像のパラメータはニューラルネットワーク25によって解析され信号の全体的な品質を示すラベルを出力26に戻す。その代りに、トレーニングを受けた特定の形式の歪みのパラメータの特徴を示すラベルを生成するように、ニューラルネットワークをプログラム作成をすることができる。

好ましい実施形態では、全体的な品質および歪みの形式の両方の測度を適用する。これによりユーザは、どのような修正行為が必要であることを識別すること、およびその修正行為を優先させることができる。

この実施形態では、第1の実施形態の可聴モデル（13,14）は、同じ原理で動作するモデルに置き換えられるが、これは相関するパラメータを人間の聴覚の知覚特性ではなく、視覚的知覚特性によって判断する。歪みの知覚的重要度は、歪みが現れる視覚のコンテキストに依存する。マスキング効果は、人間の視的知覚上の残像特性などに依存して、所定のビデオフレーム内、または連続するフレーム間で重要なことがある。セグメント化時間はフレームレート（通常、テレビジョンシステムで25 Hzのフレームレートを使用して40ミリ秒：各フレームが2つの走査によって形成される典型的な“インターレース”システム）によって都合よく定められる。

一般にビデオを送る各フレームは先行するフレームに非常に似ているので、マスキングおよび類似の効果を測定して、画像の同じ部分に関するビデオ信号のセグメントを比較することが適切であろう。解析プロセスそれ自身は、上述のオーディオサンプルに類似している。連続するフレームを解析し、人間の視覚フィルタおよびマスキングモデルを使用して、関係する基準知覚フレームに変換し、エラーを量子化し、ラベルを生成する。人間の視覚フィルタモデルに含まれるファクタは、空間的および時間的周波数分解能を含む。モデル化されるマスキング効果には、一般に活性度マスキング（大きな動作が小さい変動をマスクすること）および明度のマスキングがある。

ビデオ信号の特性的特徴を生成する一定の形のデグラデーションがあり、このような特徴の存在はニューラルネットワーク25によって識別され、歪みが存在する可能性があるという指標として使用される。しかしながら、問題は言語のときよりも複雑である。その理由はビデオ信号は言語信号よりも複雑であり、元の信号が信号ソースの生理学的特性によって言語信号が受取るようには制約されないからである。言語信号の特徴は“言語に類似していないもの（non-speech-like）”であると識別され、したがって歪みを加えた合理的な確度で識別することができる。ビデオ信号はこのようなやり方で制限されないが、受信した信号が歪んでいるか否かを確実に識別することは容易ではない。したがって、この実施形態では特徴の検出によって、調査される潜在的な問題を示すことのみができる。とくに、ビデオ画像のプロデューサによって意図的に歪ませることができる。例えば“ブロックネス”効果（“blockiness” effect）はMPEGシステムにおけるデータ圧縮によって生じる効果に類似しており、例えばスクリーン上に描かれた個人の匿名を残すために、ビデオ画像のプロデューサによって意図的に生成することができる。

ビデオサンプルを抽出した後で、パラメータ抽出ユニット23によって画像を解析して、特徴の境界を検出する。通常境界は領域（特徴）間で知覚され、各領域で画像の特性、普通

10

20

30

40

50

は色彩または明るさは一定のままであるか、または次第に変化する。例えば各特徴は、それ自身全体で次第に明暗を付けてもよいが、特徴の境界は特徴の突然に変化によって判断することができる。通常、類似の色彩をもつ2つのオブジェクトがあり、かつ各特徴のシェーディング (shading) が特徴全体で変化しても、2つのオブジェクト間の境界はシェーディングの突然の変化によって検出することができる。信号のある特質の突然の変化は、画像のスペクトル分解における継続時間の短い、広帯域幅の成分として検出することができる。例えば境界を不明瞭にするか、または境界を移動して、オブジェクトの形状を変更することによって、異なる画像要素間の境界に影響を与える歪みは、1つの画像要素の一部として観察者によって知覚される実存物本体内で起こる変化よりも知覚的に重要である。このような知覚的に重要な境界の変化は、ある1つの境界の完全な消失、または1つであってはならない場合にある1つの境界が出現することを含む。

10

ニューラルネットワーク25を制御するのに使用するネットワーク定義関数を生成するために、デグラデーションかつ人間である看る人にとっての重要度を示す画像特性とを識別しなければならない。これを行うために、図6にブロック図で示されているように、知覚解析ラベル生成ユニット29でトレーニングプロセスを実行する。これは、図2および3に示したユニット9の内容に類似している。

元の (品質の落ちていない; デグラデーションしていない) 信号および同じ信号のデグラデーションした形式の両方は、最初に各HVF (ヒューマンビジュアルフィルタ; 人間の視覚フィルタ) 31、32へ送られ、画像を人間の目/脳システムによって知覚可能なものに適合させる。ヒューマンビジュアルフィルタ31、32は、人間の光知覚システムの既知の応答にしたがって一定の空間または時間周波数をもつ信号のパワー (または振幅) を変更し、知覚重要度の低い周波数は、視覚重要度の高い周波数よりもパワーが低くなるようにする。人間の光知覚システムは、ある空間および時間周波数に対して他のものよりも応答性が高い。例えば、規則的なパターンのストライプ (縞) は距離において解像するのに困難である。対照的に、単一のストライプが視界の大部分の範囲を定めるように近付いていると、全体的なパターンは知覚的に重要ではない。ある中間の位置では、末端部でよりもパターンが知覚的により重要である。重要なのは絶対的な距離ではなく、各要素によって目で範囲を定められる角度であることに注意されたい。これは、目で範囲を定められる単位角度当たりの線 (ライン) という項で測定することができる。この値はもちろん、観察者と、画像が表示されるスクリーンとの間の距離、およびその大きさそれ自身に依存するが、何れの場合でも理想的な視覚距離は画像の大きさによって決まるので、画像の大きさがどんなものであっても、画像の要素によって目で範囲を定められる角度が著しく異なるとは考えられない。同様に、時間周波数は画像の知覚可能性に影響を与える。緩慢な変化は知覚できないが、一方で高周波数の画像は中間の明暗または明るさの連続的な信号 (“残像” として知られている現象) として知覚される。実際、空間 (ピクセル) および時間 (フレーム) 内の両方の小さい要素からビデオ画像が構成されているので、ビデオ画像は人間の光システムが高周波数の空間および時間の変化を解像できないことに依存する。フィルタによってモデル化される人間の知覚的特性は、例えばA.N.NetravaliおよびB.G.Haske lによる文献 ("Digital Pictures", 1988年, Plenum Press, New Yorkより刊行: ISBN 0-306-42791-5、とくに図4-3-12参照) に十分に記載されている。

20

30

40

画像は、次にマスキングプロセス (それぞれ33、34) に進む。一方の画像要素が他方の画像要素に対してもっているマスキング効果は複雑である。その理由はマスキング効果が空間周波数、強度、および互いにかつ観察者の目に関係している、マスクをかけかつマスクされる特性の配向に依存しているからである。画像の一定の明るさまたは迅速に移動する部分は、画像の他の部分の知覚重要度をマスクするかまたは高めることができる。その結果のマスクされた画像の明るさは各マスキングプロセス33、34によって出力され、次に差生成器35によって比較される。これは、画像の各点に対する明るさエラーの値を生成し、エラー表面を作る。明るさエラーは、元の信号と品質を落とした信号との間の差の大きさであり (元の信号は品質を落とした信号の明るさよりも明るいまたは暗い)、マスキングまたは他の知覚効果に対して調節される。差の大きさが画像全体または時間全体、ある

50

いはその両方のゼロ以外の平均値を判断するので、それを使用するのが便利である。この実施形態ではエラー表面の座標系は、上述のオーディオの実施形態の時間およびピッチ軸ではなく、画像それ自身の“x”および“y”座標を使用することに気付くと思う。時間に依存するか、または空間の周波数に依存する性質は、さらに次元を付加することによってシステムに含めることができる。生成されたデータをグラフィックで表すことは簡単ではないが、それにもかかわらずこのエラー表面のパラメータを生成して、デジタルで操作することができる。本明細書では“エラー表面”という用語を使用して、知覚できるエラーレベルが、1または複数の独立可変パラメータに関してどのように変化するかを記載したデータを意味するものとした。

ビデオ画像を生成する方法のために、画像の“x”および“y”軸は両者とも時間軸であると考えることができ、その目盛はxおよびy方向の走査レート（通常の625ラインのビデオ画像で、1ライン当り32マイクロ秒および1フレーム当り40マイクロ秒）によって決定されることに注意されたい。

活性度および明るさのマスキングシステム33および34からの出力は、さらに各画像分解ユニット38a、38bへ送られる。これらは画像の異なる要素間の境界を検出する。上述のように、画像要素の境界に影響を与える品質低下（境界の完全な消失、またはスプリアス境界の存在を含む）は知覚重要度が最も高い。スペクトル分解において高い空間的周波数を含むかなり局在する構成要素をもつ画像領域を検出することによって、画像要素の境界を識別することができる。エラー主観性生成ユニット39で画像分解ユニット38A、38Bの出力を使用して、エラー表面生成ユニット36によって生成されたエラーの明るさマップに加重する。例えば上述の実施形態のエラー活性度の値を導出す処理に類似するやり方でそれらを加算することによって、これらの加重値をエラー主観性生成ユニット39においてアルゴリズムで処理する。全体的なエラー主観性値をラベル生成ユニット37に供給して、全体的なエラー主観性値にしたがって（例えば1または複数の閾値を参照することによって）出力を生成する。

図8の構成は図7のシステムの変形を示し、監視されるシステムにおいて送られる試験信号によって、このシステムをトレーニングすることができる。これにより、監視されるシステムが発展すると、実際の動作状態にトレーニング可能なプロセスを細かく同調させることと、さらにシステムの別のトレーニングでもこのプロセスを適合させることができる。図8も監視システムの多数の分類器のアーキテクチャを示している。オンライントレーニングと多数の分類器アーキテクチャの両方は、ビデオ信号を処理するために示されているが、それらはオーディオの実施形態で使用するのにも適している。

図8には、トラヒック搬送通信システム20が示されており、搬送中の信号トラヒックのサンプルをサンプリングユニット22によってとることができる。サンプルを取られたデータは、多数のパラメータ抽出ユニット23A、23B（2つ示されている）によって解析され、各パラメータ抽出ユニットは個々の性質、例えば空間周波数、相関、全体的なコントラストレベルなどを測定するようにされている。それによって抽出されるパラメータは、個々の分類ユニット25A、25B、25C、25D、25E（一般的にニューラルネットワーク）に送られる。図示されるように、各分類ユニットは、パラメータ抽出ユニット23A、23B、などの1つ以上から抽出されるパラメータを使用する。各分類ユニット25A、などは、出力を出力調整ユニット26へ送り、ここで種々の分類ユニット25A、などの出力を処理して、ディスプレイを生成する。この構成によって、個々に構成される各分類器（ニューラルネットワーク）は、特定の特性の存在を識別することができ、それによって簡単な二進（イエス／ノー）出力を生成することができる。この分類器の出力を結合することによって、複雑な出力を生成することができ、例えば所定数の分類器が出力を生成するときに警報をトリガする。

システムはさらに2つのトレーニングサンプルソース1、1Aを準備し、一方は通信システム20内のどこかに配置され、他方はパラメータ解析ユニット23A、23Bにローカルに配置されて、ユニット23A、23Bへの直接的な入力を準備している。

さらに予めプログラムされた特性識別ユニット28が準備されており、図において該ユニッ

10

20

30

40

50

ト28はパラメータ抽出ユニット23Aからの入力を備えている。

分類ユニット25A乃至25Eは並列の分類ユニットとして配置され、各分類ユニットはサンプリングユニット22によって受信される信号の特定の特性を示すように構成されており、出力整合ユニット26への出力を生成して、その性質が存在するか否かを示す。パラメータ抽出ユニット23A、23Bによって識別される1または複数のパラメータが存在するか否かにしたがって、その性質それ自身が測定される。例えば1つの分類ユニット25Aは特定の形式の歪みが存在するか否かを識別することができ、一方で別の分類ユニットは第2の形式の歪みを識別する。別の分類ユニットでは全体的な歪みから知覚の厳格さを識別することができる。

遠隔通信ネットワーク20が発展するのにしたがって、ネットワーク内で新しいタイプの信号処理方法を使用することができる。この信号処理方法は種々の特性をもち、この特性により一定の環境で、新しい特性をもった歪みを生成することができる。解析システムを保持して、このような歪みを識別し、既存のネットワークの変化に適合させるために、再構成プロセスを採用する。したがって1つのチャンネルが一時的に使用できず、トレーニングサンプル生成ユニット1によって準備されたトレーニングサンプルを搬送することができる。トレーニングサンプルはサンプリングユニット22によって抽出され、他のサンプル、すなわち監視点でソース1Aによって生成され、遠隔通信リンク上で受信されるサンプルと同じサンプルと比較される。

知覚モデルラベル生成ユニット29（上述で図6を参照して詳しく示されている）は、第2の（局部的に生成された）トレーニングデータのソース1Aと関係付けられ、トレーニングソース生成ユニット1から通信リンク20上で受信される信号と、ローカルに生成されたトレーニング信号とを比較して、品質ラベルを生成し、トレーニング可能な分類ユニット25Eへの入力される。これらのラベルはパラメータ抽出ユニット23Bによって生成される対応するパラメータと関係付けられて、分類ユニット25Eのためのネットワーク定義関数を生成する。このトレーニングプロセスに続いて、生成されたネットワーク定義関数を使用して、サンプリングユニット22によって通信リンク20から抽出される未知のサンプルに対応する未知のパラメータパターンを分類する。この方法では、通信システム20が発展するとき、分類ユニットを元のシステム、または再びプログラム作成される既存のシステムに付加することができる。

全ての分類ユニットは必ずしもトレーニングプロセスによってプログラム作成されない。特性が既に分かっているときは、分類ユニット28を予めプログラム作成して、パラメータ抽出ユニット23A、23B、などによって抽出される適切なパラメータを識別することができる。解析ユニット28は、トレーニング可能なユニット25A乃至25Eと同様の方法で動作するが、特定の形式の信号のデグラデーションの既知の特性を認識するように予めプログラム作成されている。例えば、データ圧縮によって境界の不鮮明、ブロックネス（blockiness）、ぼやけ（fuzziness）、ジャーキネス（ピクピク動きの具合、すなわち滑らかでない動き, jerkiness）、色収差を生じる可能性があることが分かっている。とくに、非常に素早い動きを含むビデオ画像では、MPEG信号圧縮システムはピクセルの分解能を低減し、その結果生じる“ブロックネス”によって、通常8×8のピクセルの特徴的な直線で囲まれた境界で、画像の迅速な動きから生ずるオーバーロードを取り扱う。マルチパスインターフェイスは2つの境界を生成し、この2つの境界は固定された水平方向の距離を移動する（“ゴースティング（ghosting）”として知られている）。ぼやけは、縁部それ自身の高空間周波数成分を広げる傾向がある。色彩が不鮮明になることによって、画像の異なる色の成分によって定義される縁部が一致なくなることがある。画像全体で1つの色の成分のコントラストレベルが低いことは、色収差を示している（画像は問題とされている色によって色付けされるか、補色で色付けされる）。全体的に画像のコントラストが低いことは、信号クリッピングを示している。画像の3色成分間の完全相関はモノクローム画像を示し、すなわち色の情報を運ぶ信号帯域の損失を示唆する。

連続するフレーム間の相関度から別の形式のデグラデーションを明らかにすることができる。例えば各連続するフレーム間の大きなランダムな差は、干渉信号の存在を示している

10

20

30

40

50

。ビデオ画像上に明るい斑点と暗い斑点がランダムに現れる効果（“スノー”として知られている）は、このような特徴の例である。スノーは、各斑点の場所がランダムであるので、1つのサンプルから検出することは難しい。このサンプルが先のサンプルと完全に異なるときは、情景の変化（写真撮影の変化）が生じたことを示し、このような場合は有益な時間的相関が得られないことを示している可能性が高い。しかしながら、情景が大きくは変化していないとき、すなわち各フレームが先のフレームと強力に相関があるが、先のフレームとは異なっており、フレーム間の差は互いに相関していないか、もっと先のフレームからの同じような差もないときには、これはホワイトノイズ（“スノー”）が信号内に存在していることを示している。

識別できる別の時間に依存する相関は、信号圧縮によって生じるジャーキーな画像である。画像が“例外によって”処理されるときにこれがとくにありそうなことである、すなわち各画像が、先のものと同じであるが一定の差があり、送られる差だけがあるときにありそうである。迅速に変化する画像では、このデータレートは各フレームにとって必要な変化の全てを定めるのには緩慢過ぎることがある。画像の中での特徴の動きは通常滑らかである。特徴が段階的に動くときは、これはジャーキーな画像を示している。ジャーキネスが縁部を生成し、この縁部の位置が特徴的なやり方で1つの画像から次の画像への相関を有し、継続する一群の画像の各画像内の1または複数の要素は同じ位置にあるが、継続する群の画像の要素は互いに変位されている。

全ての縁部の位置の完全かつ非過渡的な変化によって識別可能な情景の変化、または平均の明るさの変化のような幾つかの特性の変化、あるいはその両方は、オーディオチャンネル（例えば、そのラウドネス）の全体的な特性の変化と相関があることを期待することができる。このような多数の変化についてこのような相関が存在しないことは、オーディオチャンネルの干渉、もしくはオーディオ信号の実際の損失を示すことができる。

分類ユニット28では、パラメータユニット23Aから要求されるパラメータを解析して、出力調整ユニット26へラベルを出力し、このラベルで特定の形式の歪みの特徴を備えたパラメータの存在を示す。したがって、例えば8×8のピクセルブロックを識別するパラメータはMPEGコーディングアルゴリズムのオーバーロードを示すことになる。信号のスペクトルに明白に定められた高周波数の内容がないことは、明白に定められた縁部がないことを示し、ぼやけまたは焦点のずれた画像を示唆している。空間的に密接している縁部の特徴間強い相関は、ゴースト発生（マルチパス干渉）などを示唆している。1つの画像から次の画像への相関がない特徴は、ノイズ信号（“スノー”）を示している。

トレーニング可能分類ユニット25A乃至25Eに関して記載したように、幾つかの予めプログラム作成した分類ユニット28も提供することができ、各ユニットは特定の特性のみを識別する。

10

20

30

Figure 7

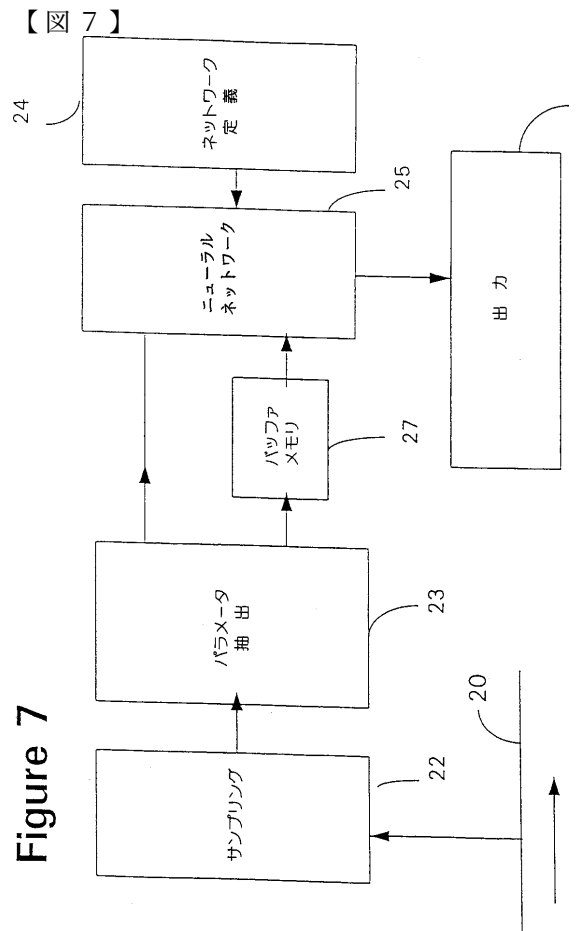


Figure 5

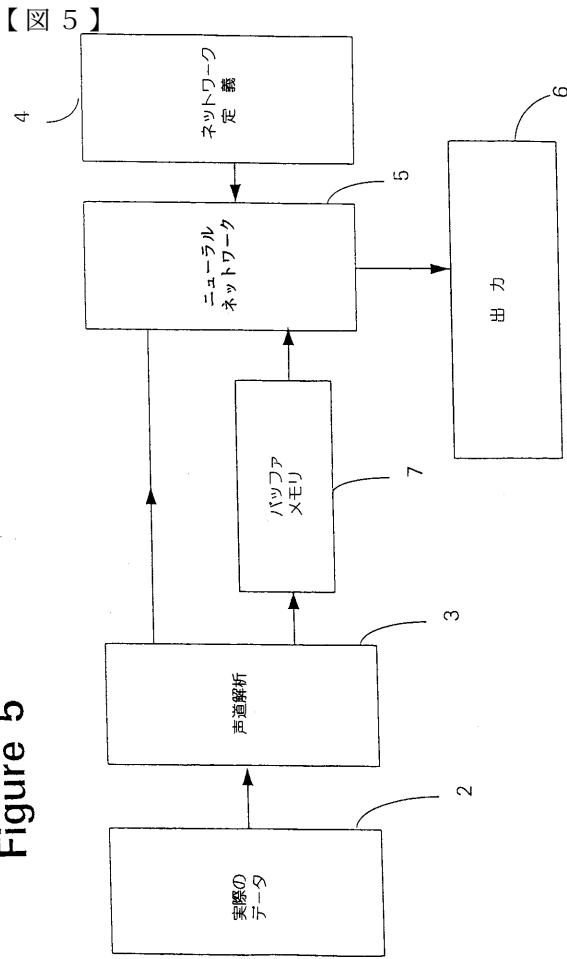


Figure 8

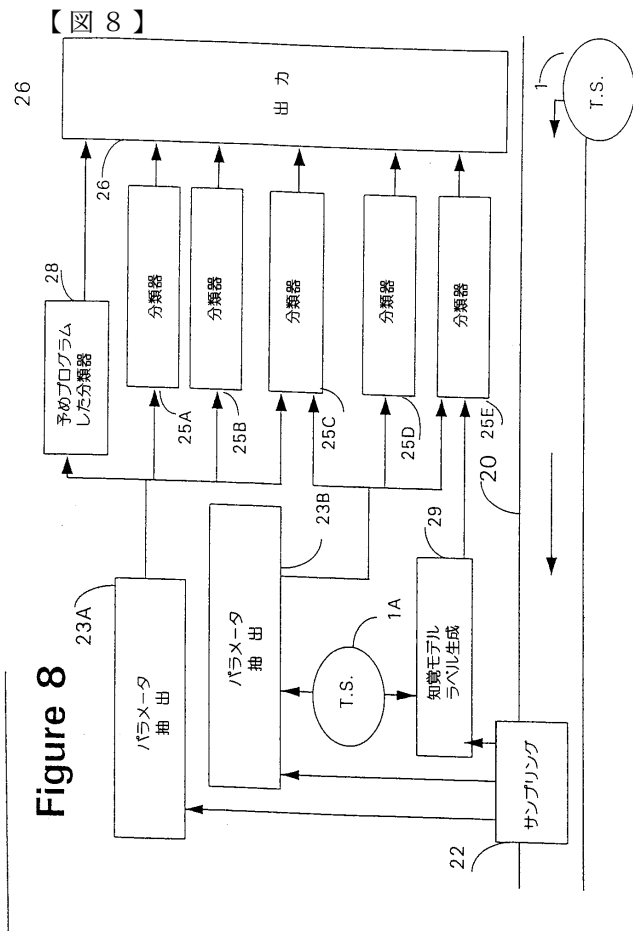
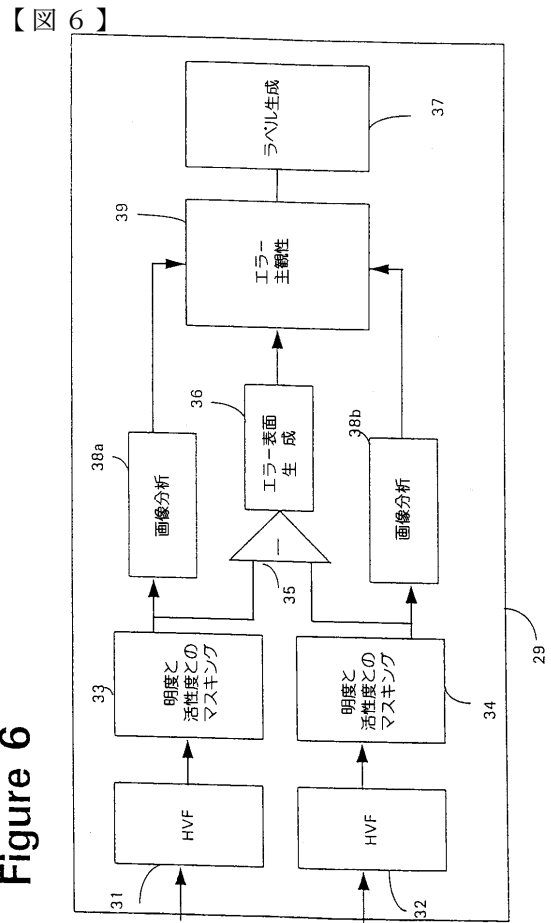


Figure 6



フロントページの続き

(72)発明者 ホリアー、マイケル・ピーター
イギリス国、アイピー５・７エックスエックス、イプスウィッチ、ケスグレイブ、グランジ・ファ
ーム、ハーバート・ロード １５

(72)発明者 グレイ、フィリップ
イギリス国、アイピー４・５エイチピー、サフォーク、イプスウィッチ、クラブ・ストリート
１５

審査官 甲斐 哲雄

(56)参考文献 特表平１１－５１０３３４（ＪＰ，Ａ）
特開平０４－３４５３２７（ＪＰ，Ａ）

(58)調査した分野(Int.Cl., DB名)
H04M 3/00
G10L 19/00