

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-79188

(P2010-79188A)

(43) 公開日 平成22年4月8日(2010.4.8)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 11/00 (2006.01)	G 1 O L 11/00 4 O 2 Z	5 D O 1 5
G 1 O L 15/10 (2006.01)	G 1 O L 11/00 1 O 1 Z	
	G 1 O L 15/10 5 O O Z	

審査請求 未請求 請求項の数 4 O L (全 25 頁)

(21) 出願番号 特願2008-250360 (P2008-250360)
 (22) 出願日 平成20年9月29日 (2008.9.29)

(出願人による申告) 平成19年度、文部科学省、財団法人岐阜県研究開発財団 (知的創造による地域産学官連携強化プログラム「知的クラスター創成事業」) 委託研究、産業技術力強化法第19条の適用を受ける特許出願

(71) 出願人 304021277
 国立大学法人 名古屋工業大学
 愛知県名古屋市昭和区御器所町字木市29番
 (74) 代理人 100076473
 弁理士 飯田 昭夫
 (74) 代理人 100112900
 弁理士 江間 路子
 (74) 代理人 100136995
 弁理士 上田 千織
 (74) 代理人 100150935
 弁理士 村松 孝哉
 (72) 発明者 岩田 彰
 愛知県名古屋市昭和区御器所町 名古屋工業大学内

最終頁に続く

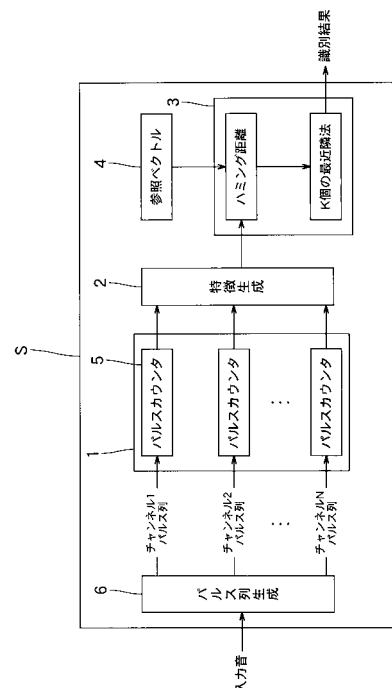
(54) 【発明の名称】 音源同定方法及び音源同定装置

(57) 【要約】

【課題】より少ない回路数でハードウェア化可能な音源同定方法を提供する。

【解決手段】入力音から生成された音圧に応じたパルス頻度を持つ周波数帯域別のパルス列において、時間軸方向に所定幅を有するカウント範囲内のパルス数を数えて、パルス数ベクトルを生成する第1ステップと、パルス数ベクトルの各要素のうちの大きい方からN個を1とし、残りを0とした特徴ベクトルを生成する第2ステップと、特徴ベクトルと同様に生成されて元の音源を示す音源カテゴリに分類された複数の参照ベクトルから、特徴ベクトルにハミング距離で近い方からk個の参照ベクトルを検索し、それらのうちの最も多くの参照ベクトルが属する音源カテゴリを決定する第3ステップとを、カウント範囲をずらしつつ繰り返す音源同定方法。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

入力音を、複数の周波数帯域別に、音圧に応じたパルス頻度を持ち時間軸方向にパルスが並んだパルス列に変換し、前記各周波数帯域のパルス列を用いて入力音の音源を識別する音源同定方法であって、

前記各周波数帯域のパルス列において、時間軸方向に所定幅を有するカウント範囲内のパルス数を数えて、前記各パルス列における前記パルス数を要素とするパルス数ベクトルを生成する第 1 ステップと、

前記パルス数ベクトルの各要素のうちの大きい方から N (N : 正整数) 個の要素を 1 とし、残りの要素を 0 とした特徴ベクトルを生成する第 2 ステップと、

音源が分かっている音から前記特徴ベクトルと同様に生成され、それぞれ元の音源を示す音源カテゴリに分類されて記憶されている複数の参照ベクトルから、前記特徴ベクトルにハミング距離で近い方から k (k : 正整数) 個の参照ベクトルを検索し、前記 k 個の参照ベクトルのうちの最も多くの参照ベクトルが属する音源カテゴリを決定し、該音源カテゴリを示すカテゴリ情報を出力する第 3 ステップとを、

前記カウント範囲を時間軸方向に重ならないように移動させつつ、繰り返し、

出力された前記カテゴリ情報に基づいて前記入力音の音源を識別することを特徴とする音源同定方法。

【請求項 2】

前記第 1 ステップと、

前記第 2 ステップと、

前記第 3 ステップと、

前記複数の参照ベクトルが分類される複数の音源カテゴリのうち、前記第 3 ステップで出力された前記カテゴリ情報が示す音源カテゴリについてはポテンシャル値を上げて、他の音源カテゴリについてはポテンシャル値を下げる第 4 ステップとを、

前記カウント範囲を時間軸方向に重ならないように移動させつつ、繰り返し、

前記複数の音源カテゴリのうち、前記ポテンシャル値が最大になった音源カテゴリを、前記入力音の音源と判定することを特徴とする請求項 1 記載の音源同定方法。

【請求項 3】

入力音を、複数の周波数帯域別に、音圧に応じたパルス頻度を持ち時間軸方向にパルスが並んだパルス列に変換するパルス列生成手段と、

時間軸方向に所定幅を有し時間軸方向に重ならないように設定されるカウント範囲毎に、前記各周波数帯域のパルス列における前記カウント範囲内のパルス数を数えて、前記各パルス列における前記パルス数を要素とするパルス数ベクトルを生成するパルス数ベクトル生成手段と、

前記パルス数ベクトルの各要素のうちの大きい方から N (N : 正整数) 個の要素を 1 とし、残りの要素を 0 とした特徴ベクトルを生成する特徴ベクトル生成手段と、

音源が分かっている音から前記特徴ベクトルと同様に生成され、それぞれ元の音源を示す音源カテゴリに分類された複数の参照ベクトルを記憶した参照ベクトル記憶手段と、

前記参照ベクトル記憶手段に記憶されている参照ベクトルから、前記特徴ベクトルにハミング距離で近い方から k (k : 正整数) 個の参照ベクトルを検索し、前記 k 個の参照ベクトルのうちの最も多くの参照ベクトルが属する音源カテゴリを決定し、該音源カテゴリを示すカテゴリ情報を出力する音源カテゴリ識別手段と、

を有し、出力された前記カテゴリ情報に基づいて前記入力音の音源を識別することを特徴とする音源同定装置。

【請求項 4】

前記複数の参照ベクトルが分類される複数の音源カテゴリのうち、前記音源カテゴリ識別手段によって出力された前記カテゴリ情報が示す音源カテゴリについてはポテンシャル値を上げて、他の音源カテゴリについてはポテンシャル値を下げるポテンシャル値処理手段を有し、

10

20

30

40

50

前記複数の音源カテゴリのうち、前記ポテンシャル値が最大になった音源カテゴリを、前記入力音の音源と判定することを特徴とする請求項3記載の音源同定装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音源の種類を同定する音源同定方法及び音源同定装置に関する。

【背景技術】

【0002】

音による周辺環境把握の基本機能は、音源方向を識別する音源定位と、音源種類を識別する音源同定（音源認識）であり、パルスニューロンモデルを用いた音源同定装置には、
下記非特許文献1に記載されたものがある。また、出願人による音源同定に関する出願に、
下記特許文献1、2があり、特許文献2には、処理の高速化のため、音源定位・同定装置をFPGA（Field Programmable Gate Array）に実装した例が示されている。

【特許文献1】特開2008-77177号公報

【特許文献2】特開2008-85472号公報

【非特許文献1】坂口晋也、黒柳奨、岩田彰、「環境把握のための音源同定システム」、電子情報通信学会NC研究会技術研究報告、社団法人電子情報通信学会、1999年12月、NC99-70、p.61-68

【発明の開示】

【発明が解決しようとする課題】

【0003】

しかし、上記特許文献2の表1から分かるように、特許文献2記載の装置では、音源同定用周波数パターン検出部のみでも約5,000ALUTsの回路数が必要であり、更なる処理の高速化と装置のコンパクト化のためには、より少ない回路数でハードウェア化可能な音源同定方法が望まれていた。

【0004】

この発明は、上述した問題を解決するものであり、より少ない回路数でハードウェア化可能な音源同定方法及び音源同定装置を提供することを目的とする。

【課題を解決するための手段】

【0005】

本発明の音源同定方法は、入力音を、複数の周波数帯域別に、音圧に応じたパルス頻度を持ち時間軸方向にパルスが並んだパルス列に変換し、前記各周波数帯域のパルス列を用いて入力音の音源を識別する音源同定方法であって、前記各周波数帯域のパルス列において、時間軸方向に所定幅を有するカウント範囲内のパルス数を数えて、前記各パルス列における前記パルス数を要素とするパルス数ベクトルを生成する第1ステップと、前記パルス数ベクトルの各要素のうちの大きい方からN（N：正整数）個の要素を1とし、残りの要素を0とした特徴ベクトルを生成する第2ステップと、音源が分かっている音から前記特徴ベクトルと同様に生成され、それぞれ元の音源を示す音源カテゴリに分類されて記憶されている複数の参照ベクトルから、前記特徴ベクトルにハミング距離で近い方からk（k：正整数）個の参照ベクトルを検索し、前記k個の参照ベクトルのうちの最も多くの参照ベクトルが属する音源カテゴリを決定し、該音源カテゴリを示すカテゴリ情報を出力する第3ステップとを、前記カウント範囲を時間軸方向に重ならないように移動させつつ、繰り返し、出力された前記カテゴリ情報に基づいて前記入力音の音源を識別することを特徴とする。

【0006】

なお、好ましくは、前記第1ステップと、前記第2ステップと、前記第3ステップと、前記複数の参照ベクトルが分類される複数の音源カテゴリのうち、前記第3ステップで出力された前記カテゴリ情報が示す音源カテゴリについてはポテンシャル値を上げて、他の音源カテゴリについてはポテンシャル値を下げる第4ステップとを、前記カウント範囲を時間軸方向に重ならないように移動させつつ、繰り返し、前記複数の音源カテゴリのうち

10

20

30

40

50

、前記ポテンシャル値が最大になった音源カテゴリを、前記入力音の音源と判定する。

【0007】

本発明の音源同定装置は、入力音を、複数の周波数帯域別に、音圧に応じたパルス頻度を持ち時間軸方向にパルスが並んだパルス列に変換するパルス列生成手段と、時間軸方向に所定幅を有し時間軸方向に重ならないように設定されるカウント範囲毎に、前記各周波数帯域のパルス列における前記カウント範囲内のパルス数を数えて、前記各パルス列における前記パルス数を要素とするパルス数ベクトルを生成するパルス数ベクトル生成手段と、前記パルス数ベクトルの各要素のうちの大きい方から N (N : 正整数) 個の要素を1とし、残りの要素を0とした特徴ベクトルを生成する特徴ベクトル生成手段と、音源が分かっている音から前記特徴ベクトルと同様に生成され、それぞれ元の音源を示す音源カテゴリに分類された複数の参照ベクトルを記憶した参照ベクトル記憶手段と、前記参照ベクトル記憶手段に記憶されている参照ベクトルから、前記特徴ベクトルにハミング距離で近い方から k (k : 正整数) 個の参照ベクトルを検索し、前記 k 個の参照ベクトルのうちの最も多くの参照ベクトルが属する音源カテゴリを決定し、該音源カテゴリを示すカテゴリ情報を出力する音源カテゴリ識別手段と、を有し、出力された前記カテゴリ情報に基づいて前記入力音の音源を識別することを特徴とする。

10

【0008】

なお、好ましくは、前記複数の参照ベクトルが分類される複数の音源カテゴリのうち、前記音源カテゴリ識別手段によって出力された前記カテゴリ情報が示す音源カテゴリについてはポテンシャル値を上げて、他の音源カテゴリについてはポテンシャル値を下げるポテンシャル値処理手段を有し、前記複数の音源カテゴリのうち、前記ポテンシャル値が最大になった音源カテゴリを、前記入力音の音源と判定する。

20

【発明の効果】

【0009】

本発明の音源同定方法及び音源同定装置は、入力音の特徴を0、1で表した特徴ベクトルを生成し、音源が分かっている音から特徴ベクトルと同様に生成された参照ベクトルを用いて、特徴ベクトルと参照ベクトルとの遠近(類似度)をハミング距離で決定するという、シンプルで演算容易なロジックを用いているので、より少ない回路数でハードウェア化可能である。

【発明を実施するための最良の形態】

30

【0010】

以下、本発明の一実施形態について図面に基づいて説明する。

【0011】

音源同定装置5は、図1に示すように、マイクロホン(図示せず。)に接続されたパルス列生成手段6を備えている。パルス列生成手段6は、AD変換部14と、人の聴覚系の蝸牛に相当する周波数分解部15と、有毛細胞に相当する非線形変換部16と、蝸牛神経に相当するパルス変換部17とを備えている。AD変換部14は、マイクロホンから入力された信号(入力音)をAD変換する。周波数分解部15は、バンドパスフィルタ(BPF)群により構成され、AD変換された信号を所定の周波数範囲について対数スケールで複数の周波数帯域(以下、「チャンネル」ともいう。)別の信号に分解する。非線形変換部16は、周波数分解部15から入力された各周波数帯域の信号に対して、それぞれ、非線形変換を行うことによりその正の成分だけを取り出すとともに、ローパスフィルタ(LPF)によりエンベロープ検出を行う。パルス変換部17は、非線形変換部16から入力された各周波数帯域の信号を、それぞれ、信号強度(すなわち、音圧)に比例したパルス頻度を持つパルス列に変換する。これらの処理により、パルス列生成手段6は、入力音を、チャンネル別に、音圧に応じたパルス頻度を持ち時間軸方向にパルスが並んだパルス列に変換する。

40

【0012】

また、音源同定装置5は、図2に示すように、パルス数ベクトル生成手段1と、特徴ベクトル生成手段2と、音源カテゴリ識別手段3と、参照ベクトル記憶手段4とを備えてい

50

る。

【 0 0 1 3 】

パルス数ベクトル生成手段 1 には、パルス列生成手段 6 で生成された各チャンネルのパルス列が入力される。実施形態では、チャンネル数は 4 3 である。パルス数ベクトル生成手段 1 は、チャンネル数分のパルスカウンタ 5 を備え、各チャンネルのパルス列におけるカウント範囲内のパルス数を、各パルスカウンタ 5 により数え、それらのパルス数を要素とするパルス数ベクトルを生成する。パルス数ベクトルの要素数は、チャンネル数すなわち 4 3 となる。カウント範囲は、実施形態ではパルス 1000 個分の幅とする。実施形態では、入力信号を 4 8 kHz でサンプリングしてパルス列を生成するため、1 sec に 48000 個のパルスを発生可能であり、パルス 1000 個分の幅とは $1000 \div 48000 = 0.02 \text{sec}$ の幅となる。すな

10

【 0 0 1 4 】

特徴ベクトル生成手段 2 には、パルス数ベクトルが入力される。特徴ベクトル生成手段 2 は、パルス数ベクトルの各要素のうちの大きい方から N 個の要素を 1 とし、残りの要素を 0 とした特徴ベクトルを生成するものである。すなわち、特徴ベクトルは、音圧の強い部分を 1 とし残りの部分を 0 とすることにより音の特徴を表すバイナリーベクトルである。なお、N は、パルス数ベクトルの要素数以下の正整数であり、実施形態では $N = 9$ とする。

【 0 0 1 5 】

音源カテゴリ識別手段 3 には、特徴ベクトルが入力される。音源カテゴリ識別手段 3 は、参照ベクトル記憶手段 4 に記憶されている参照ベクトルから、特徴ベクトルにハミング距離で近い方から k 個の参照ベクトルを検索し、それら k 個の参照ベクトルのうちの最も多くの参照ベクトルが属する音源カテゴリを決定し、その音源カテゴリを示すカテゴリ情報を出力するものである。k は、参照ベクトルの総数以下の正整数であり、実施形態では $k = 1 2$ とする。

20

【 0 0 1 6 】

参照ベクトル記憶手段 4 は、複数の参照ベクトルを記憶したもの（実施形態では ROM）である。各参照ベクトルは、音源が分かっている入力音から上記特徴ベクトルと同様に生成したバイナリーベクトルである。すなわち、所定時間長さを持つ入力音からチャンネル別のパルス列を生成し、時間軸方向に所定幅（パルス 1000 個分）を有するカウント範囲を、時間軸方向に重ならないように、かつ、間隔が開かないように移動させつつ、各パルス列におけるカウント範囲内のパルス数をカウントして、各パルス列におけるパルス数を要素とするパルス数ベクトルを生成する。そして、パルス数ベクトルの要素のうち大きい方から $N (= 9)$ 個を 1 に、残りの要素を 0 にした特徴ベクトルを生成し、その特徴ベクトルを参照ベクトルとする。各参照ベクトルは、元の音源を示す音源カテゴリ（音源種類）に、例えばその音源カテゴリを示すカテゴリ情報に関連付けることにより、分類されている。なお、参照ベクトルの総数を少なくするために、K 平均法（K-means clustering）を用いて、各音源カテゴリについて、その音源カテゴリに属する特徴ベクトルを複数のクラスタに分けるクラスタリングを行い、各クラスタの代表（中心）をその音源カテゴリの参照ベクトルとする。実施形態では、各音源カテゴリについて参照ベクトルを 1,000

30

40

【 0 0 1 7 】

音源同定装置 S において実行される音源同定方法について、次に説明する。

【 0 0 1 8 】

音源同定装置 S では、まず、パルス列生成手段 6 により入力音をチャンネル別のパルス列に変換する。図 3 は、入力音（入力信号）からパルス列が生成される様子を示したイメージ図であるが、図の最上段が入力信号（audio signal）であり、図の 2 段目以降に、この入力信号をチャンネル毎に周波数分解したものを破線で、周波数分解された信号から生成されたパルス列を実線で示している。なお、図 3 では、チャンネル 1（channel 1）、チャンネル 4（channel 4）、チャンネル 7（channel 7）、チャンネル 10（channel 10）

50

)、チャンネル 13 (channel 13) のみを図示している。

【0019】

また、図 4 - 1 ~ 図 6 - 2 は、音源が分かっている入力音、具体的には、図 4 - 1 は目覚まし時計のアラーム音、図 4 - 2 はインターホンの呼び出し音、図 4 - 3 は笛吹きケトルの沸騰音、図 5 - 1 は救急車のサイレン音、図 5 - 2 はパトカーのサイレン音、図 5 - 3 は電話のベル音、図 6 - 1 は消防車のサイレン音、図 6 - 2 は人の声から生成されたパルス列を示す。なお、縦軸はチャンネル、横軸は時間を示し、濃淡が濃い程パルス頻度大であることを示している。

【0020】

第 1 ステップ 次に、音源同定装置 S は、パルス数ベクトル生成手段 1 により、チャンネル別のパルス列からパルス数ベクトルを生成する第 1 のステップを実行する。第 1 のステップでは、パルス数ベクトル生成手段 1 は、図 7 のステップ S 0 1 に示すように、各ウィンドウ (時間窓) についてパルスカウンタ 5 によりパルス数を数える。ここで、ウィンドウとは、カウント範囲を示すものであり、図 8 のウィンドウ A、ウィンドウ B に示すように、すべてのチャンネルを通してパルス列を所定幅 (実施形態ではパルス 1000 個分の幅) で言わば覗いて数えるためのものである。

【0021】

パルス数ベクトル生成手段 1 は、まず、最初のウィンドウ内に存在するパルス数を数える。最初のウィンドウは、時刻 0 から約 20msec の範囲となる。パルス数を数えている間は、図 7 に示すように、待機状態 (アイドリング状態) となる。そのウィンドウについてパルス数を数え終わると、パルス数ベクトル生成手段 1 は、パルス数をチャンネル順に並べたパルス数ベクトルを、特徴ベクトル生成手段 2 に渡す。次にステップ S 0 1 を行うときは、パルス数ベクトル生成手段 1 は、ウィンドウを、最初のウィンドウに重ならないように、かつ、最初のウィンドウとの間に間隔が開かないように移動させ、そのウィンドウ内に存在するパルス数を数える。そして、そのウィンドウについてパルス数を数え終わると、パルス数をチャンネル順に並べたパルス数ベクトルを、特徴ベクトル生成手段 2 に渡す。パルス数ベクトル生成手段 1 は、かかる処理を、ウィンドウを時間軸方向に (すなわち、時間の早い方から遅い方に向かって)、前のウィンドウと後のウィンドウとが重ならないように、かつ、前のウィンドウと後のウィンドウとで間隔が開かないように移動させつつ、繰り返す。

【0022】

かかる処理により、例えば、図 8 に示すように、ウィンドウ A についてはパルス数ベクトル $A_1 = (3, 2, 4, 3, \dots, 5, 2, 4, 5)$ が、ウィンドウ B についてはパルス数ベクトル $B_1 = (2, 3, 5, 3, \dots, 3, 5, 9, 10)$ が生成される。

【0023】

第 2 ステップ パルス数ベクトル生成手段 1 からパルス数ベクトルを受け取った特徴ベクトル生成手段 2 は、パルス数ベクトルから特徴ベクトルを生成する第 2 のステップを実行する (図 7 の S 0 2)。

【0024】

詳しくは、図 9 に示すように、特徴ベクトル生成手段 2 は、パルス数ベクトルの各要素から $N (= 9)$ 個の要素を任意に選択し、それらの要素を所定領域に格納する (S 1 0 1)。次に、所定領域に格納された 9 個の要素の中で最小値の要素を見つける (S 1 0 2)。最小値の要素を見つけたら、所定領域に格納されなかった残りの要素とその最小値の要素とを比較する (S 1 0 3)。そして、残りの要素の中から、その最小値の要素より大きい値の要素を見つけたら、その最小値の要素の代わりに、見つかったより大きい値の要素を所定領域に格納し、ステップ S 1 0 2 に戻って、所定領域に格納されている要素の中で最小値の要素を見つける。

【0025】

ステップ S 1 0 2、1 0 3 は、所定領域中の最小値の要素より大きい値の要素が、残りの要素から見つからなくなるまで繰り返される。ステップ S 1 0 3 で、所定領域中の最小

10

20

30

40

50

値の要素より大きい値の要素が、残りの要素から見つからなかったら、所定領域にはパルス数ベクトルの各要素のうちの大きい方から 9 個の要素が格納されているので、パルス数ベクトル中のそれら 9 個の要素を 1 とし、それら 9 個以外の要素を 0 とすることにより、特徴ベクトルを生成する (S 1 0 4)。

【0 0 2 6】

かかる処理により、例えば、図 8 に示すように、パルス数ベクトル A 1 については特徴ベクトル A 2 が、パルス数ベクトル B 1 については特徴ベクトル B 2 が生成される。生成された特徴ベクトルは音源カテゴリ識別手段 3 に渡される。

【0 0 2 7】

第 3 ステップ 特徴ベクトル生成手段 2 から特徴ベクトルを受け取った音源カテゴリ識別手段 3 は、参照ベクトルと特徴ベクトルとの距離を調べ、特徴ベクトルに近い参照ベクトルが多く属する音源カテゴリのカテゴリ情報を出力する第 3 のステップを実行する (図 7 の S 0 3)。

【0 0 2 8】

詳しくは、図 10 に示すように、音源カテゴリ識別手段 3 は、まず、参照ベクトル記憶手段 4 から参照ベクトルをすべて作業領域に読み出す (S 2 0 1)。次に、作業領域中の参照ベクトルから任意の k ($= 1 2$) 個の参照ベクトルを選択し、それらの参照ベクトルを所定領域に格納する (S 2 0 2)。

【0 0 2 9】

そして、所定領域に格納された 1 2 個の参照ベクトルの中から、特徴ベクトルとの距離が最大のものを見つける (S 2 0 3)。なお、距離はハミング距離とする。最大距離の参照ベクトルを見つけたら、所定領域に格納されなかった残りの参照ベクトルと特徴ベクトルとの距離をそれぞれ調べ、それらの距離と見つけた最大距離とを比較する (S 2 0 4)。そして、残りの参照ベクトルの中から、その最大距離より小さい距離の参照ベクトルを見つけたら、その最大距離の参照ベクトルの代わりに、見つかったより小さい距離の参照ベクトルを所定領域に格納し、ステップ S 2 0 3 に戻って、所定領域に格納されている参照ベクトルの中で最大距離の参照ベクトルを見つける。

【0 0 3 0】

ステップ S 2 0 3、2 0 4 は、所定領域中の最大距離の参照ベクトルより小さい距離の参照ベクトルが、残りの参照ベクトルから見つからなくなるまで繰り返される。ステップ S 2 0 4 で、最大距離の参照ベクトルより小さい距離の参照ベクトルが、残りの参照ベクトルから見つからなかったら、所定領域には、すべての参照ベクトルのうち、特徴ベクトルとの距離が小さい方から k ($= 1 2$) 個の参照ベクトル (k -Nearest Neighbor) が格納されている (k 最近隣法)。各参照ベクトルはその参照ベクトルが属する音源カテゴリに分類されているので、所定領域の 1 2 個の参照ベクトルが属する音源カテゴリをそれぞれ調べて、例えば、救急車のサイレン音であれば救急車のサイレン音の投票数に 1 を加える等、属する音源カテゴリに投票を行う (S 2 0 5)。そして、所定領域の 1 2 個の参照ベクトルの中で最も多くの参照ベクトルが属する音源カテゴリを決定して、その音源カテゴリを示すカテゴリ情報を、識別結果として出力する (S 2 0 6)。

【0 0 3 1】

この出力されたカテゴリ情報によって示される音源カテゴリは、特徴ベクトルに最も近い方から 1 2 個の参照ベクトルのうち、最も多くの参照ベクトルが属するものであるので、特徴ベクトルが属する音源カテゴリであると判断できる。

【0 0 3 2】

図 7 に示すように、カテゴリ情報の出力を終えると (すなわち、クラス分けが終了すると)、音源同定装置 5 は、ステップ S 0 1 に戻って、次のカウント範囲についてパルス数ベクトルの生成を行い、以下、ステップ S 0 1 ~ S 0 3 の処理をパルス列が終了するまで繰り返す。なお、ステップ S 0 1 ~ 0 3 の処理は、例えば、前のパルス数ベクトルについてステップ S 0 2、S 0 3 の処理が行われているときに、次のパルス数ベクトルの生成をステップ S 0 1 で行う等、平行して行ってもよい。

10

20

30

40

50

【 0 0 3 3 】

図 1 1 - 1 ~ 図 1 3 - 2 は、各種の入力音を音源同定装置 S で識別したときの識別結果を示す図であり、出力されたカテゴリ情報が示す音源カテゴリを黒の棒線で示している。図 1 1 - 1 は目覚まし時計のアラーム音、図 1 1 - 2 はインターホンの呼び出し音、図 1 1 - 3 は笛吹きケトルの沸騰音、図 1 2 - 1 は救急車のサイレン音、図 1 2 - 2 はパトカーのサイレン音、図 1 2 - 3 は電話のベル音、図 1 3 - 1 は消防車のサイレン音、図 1 3 - 2 は人の声をそれぞれ入力したときの識別結果を表す。なお、縦軸は音源カテゴリで、上から順に、不明 (Unknown)、目覚まし時計のアラーム音 (Alarm)、インターホンの呼び出し音 (Interphone)、笛吹きケトルの沸騰音 (Kettle)、救急車のサイレン音 (Ambulance)、パトカーのサイレン音 (Police)、電話のベル音 (Phone)、消防車のサイレン音 (Fire)、人の声 (Voice) であり、横軸は時間である。これらの図から、音源同定装置 S は、かなり正確に音源を識別していることが分かる。

10

【 0 0 3 4 】

音源同定装置 S において、パルス数ベクトルをそのまま識別に用いずに特徴ベクトルに変換したのは、パルス数ベクトルをそのまま用いると、音圧 (音の強さ) の影響を強く受けるため、ノイズに弱くなってしまうからである。1つのウィンドウ内の各チャンネルのパルス数は、そのウィンドウに相当する時間内の各チャンネルの平均エネルギーに比例するので、特徴ベクトルは、その時間内の音のエネルギーの強い部分を表すことになるが、強い部分を「1」でその他の部分を「0」で表しているため、音圧の影響は小さくなって、ノイズに強い。

20

【 0 0 3 5 】

また、図 1 4 に示すように、パルス数ベクトルをそのまま用いるとともに、参照ベクトルとしてパルス数ベクトルと同様のベクトルを用いて、両者の距離をマンハッタン距離で計測すると、幾つかの要素が入れ替わっただけの殆ど同じベクトルでも、距離が遠くなってしまうことがある。図 1 4 では、参照ベクトル (reference) と殆ど同じベクトル A (vector A) が、参照ベクトルに似ていないベクトル B (vector B) よりも、参照ベクトルから遠くなってしまっている。音源同定装置 S では、特徴ベクトルと参照ベクトルとをいづれもバイナリーベクトルとし、両者の距離をハミング距離で計測しているので、似ているもの同士は距離が近く、似ていないもの同士は距離が遠くなり、識別の正確さが向上する。また、ハミング距離は排他的論理和を用いて容易に演算可能である。

30

【 0 0 3 6 】

なお、図 1 1 - 1 ~ 図 1 3 - 2 から分かるように、音源同定装置 S は、単純な音については正しく分類できるが、複雑な音については分類ミスを生じている。この分類ミスを除去するには、音源カテゴリ識別手段 3 がカテゴリ情報を出力する度に、そのカテゴリ情報が示す音源カテゴリのポテンシャル値を上げ、他の音源カテゴリのポテンシャル値を下げるポテンシャル値処理手段を、音源カテゴリ識別手段 3 の後段に設け、ポテンシャル値が最大になった音源カテゴリを、入力音の音源と判定することが好ましい。以下、ポテンシャル値処理手段が行う第 4 のステップについて説明する。第 4 のステップは、第 3 のステップの次に実行される。

【 0 0 3 7 】

40

第 4 のステップ ポテンシャル値処理手段は、各音源カテゴリ i のポテンシャル値 $P_i(t)$ を記憶している。なお、 i は、上記 8 種類の音源カテゴリにそれぞれ付されたインデックスで $i = 0 \sim 7$ である。例えば、 $i = 0$ は目覚まし時計のアラーム音 (Alarm) に、 $i = 1$ はインターホンの呼び出し音 (Interphone) に付されたインデックスである。音源カテゴリ識別手段 3 は、カテゴリ情報として、かかるインデックスを出力するものとする。また、 t は時刻であり、 $P_i(0) = 0$ ($i = 0 \sim 7$) とする。

【 0 0 3 8 】

ポテンシャル値処理手段は、音源カテゴリ識別手段 3 から時刻 t におけるカテゴリ情報 $y(t)$ を受け取ると、次の数式 (1) (2) に従って、 $P_i(t)$ ($i = 0 \sim 7$) を増減する。

50

【 0 0 3 9 】

$i = y(t)$ に対しては、 $P_i(t) = \min(P_{\max}, P_i(t-1) + \gamma) \dots (1)$

$i \neq y(t)$ に対しては、 $P_i(t) = \max(0, P_i(t-1) - 1) \dots (2)$

すなわち、カテゴリ情報 $y(t)$ で示された音源カテゴリに対しては、そのポテンシャル値を γ 上昇させ、それ以外の音源カテゴリに対しては、そのポテンシャル値を 1 下降させる。なお、1 回あたりの上昇幅は 1 回あたりの下降幅よりも大きいもの（すなわち、 $\gamma > 1$ ）とし、ここでは $\gamma = 2$ とする。また、 $P_{\max}$ はポテンシャル値の上限であり、ポテンシャル値の下限は 0 とする。

【 0 0 4 0 】

このように、時間情報を加えれば、音源は一般に急変することはないので、時間の経過と共にその音源が識別されるようになる。以下に実験例を示す。

10

【 0 0 4 1 】

実験例

上記 8 種類の音源をマイクの周囲に並べ、インデックスが小さい音源から順に音を発してマイクで集音し、サンプリング周波数 48 kHz で 3 つの音信号ファイルを作った。そのうち 2 つのファイルをトレーニング（すなわち、参照ベクトルの作成）に用い、1 つのファイルをテストに用いた。各パラメータは、次のように定めた。

【 0 0 4 2 】

ウィンドウ（カウント範囲）の幅 = 1000（パルス 1000 個分）

$N = 9$

$k = 12$

$P_{\max} = 192$

$\gamma = 2$

20

テスト・ファイルの音信号を音源同定装置 S に入力したときの出力結果を、図 15 に示す。図 15 の最上段（Original Labels）は入力音を示し、2 段目（k-Nearest Neighbor Classification Result）は音源カテゴリ識別手段 3 の出力結果を × 印で示す。なお、× 印が多数密集している部分は棒状に見える。また、4 段目（Time Potentials）のグラフは、ポテンシャル値処理手段によって処理されたポテンシャル値を示す。なお、このグラフにおいて、符号 P_0 が付された線は $P_0(t)$ 、符号 P_1 が付された線は $P_1(t)$ 、符号 P_2 が付された線は $P_2(t)$ 、符号 P_3 が付された線は $P_3(t)$ 、符号 P_4 が付された線は $P_4(t)$ 、符号 P_5 が付された線は $P_5(t)$ 、符号 P_6 が付された線は $P_6(t)$ 、符号 P_7 が付された線は $P_7(t)$ を表している。4 段目のグラフからは、時間の経過と共に正しい音源がポテンシャル値によって示されることが分かる。なお、3 段目（Time Potentials Classification Result）は、4 段目のポテンシャル値のうち最大となったものの音源カテゴリを示している。

30

【 0 0 4 3 】

このように、時間情報を加えたポテンシャル値によって音源を判定すれば、複雑な音であっても入力音の音源を正しく識別できることが分かる。

【 0 0 4 4 】

本発明の音源同定方法は、一般のコンピュータでソフトウェアにより実行させることもできる（すなわち、音源同定装置 S を一般のコンピュータで実現することもできる）が、処理の高速化のためには処理ロジックをハードウェア化することが好ましい。実施形態の音源同定方法（但し、ポテンシャル値の処理を行う第 4 ステップを除く。）をコーディングして FPGAs に書き込みハードウェア化した場合、回路数は約 2,300 ALUTs となり、上記特許文献 2 記載の装置に比して回路数が大幅に少なくなった。これは、処理ロジックがシンプルでステップ数が少ないためである。このように、本発明の音源同定方法は、少ない回路数でハードウェア化可能であるので、音源同定装置 S のコンパクト化と処理の高速化が可能である。

40

【 0 0 4 5 】

また、本発明の音源同定方法では、パラメータ数が、従来のパルスニューロンモデルを

50

用いた手法に比して少ない。しかも、図 16 に、パラメータ N (Number of Features) の値を 6 ~ 12 の範囲で変更するとともにパラメータ k (Nearest Neighbor) の値を 1 ~ 17 の範囲で変更して、識別の正確さ (正しく識別された割合) の変化を調べた結果を示すが、図 16 から分かるように、パラメータ N、k は、ある程度大きいところからは値を変えても識別の正確さを維持できる。すなわち、これらのパラメータはいずれも臨界的 (critical) でないため、調整が容易であり、新しい音を学習させる (すなわち、新しい音により参照ベクトルを生成する) ときも、パラメータの調整が容易である。

【0046】

また、参照ベクトルがバイナリーベクトルであるため、参照ベクトルを記憶するメモリの容量が少なく済む。

10

【0047】

なお、上記実施形態では、カウント範囲を、時間軸方向に重ならないように、かつ、間隔が開かないように移動させつつ、各パルス列におけるカウント範囲内のパルス数をカウントしたが、カウント範囲同士の間隔を開けるように構成してもよい。適宜間隔を開けつつカウント範囲を移動させてパルス数をカウントしても、入力音の特徴を抽出でき、音源同定が可能であるとともに、データ量を減少させることができるからである。

【図面の簡単な説明】

【0048】

【図 1】本発明の一実施形態に係る音源同定装置のパルス列生成手段のブロック構成図である。

20

【図 2】同実施形態に係る音源同定装置のブロック構成図である。

【図 3】入力信号からパルス列を生成する様子を示したイメージ図である。

【図 4 - 1】目覚まし時計のアラーム音から生成されたパルス列を示す図である。

【図 4 - 2】インターホンの呼び出し音から生成されたパルス列を示す図である。

【図 4 - 3】笛吹きケトルの沸騰音から生成されたパルス列を示す図である。

【図 5 - 1】救急車のサイレン音から生成されたパルス列を示す図である。

【図 5 - 2】パトカーのサイレン音から生成されたパルス列を示す図である。

【図 5 - 3】電話のベル音から生成されたパルス列を示す図である。

【図 6 - 1】消防車のサイレン音から生成されたパルス列を示す図である。

【図 6 - 2】人の声から生成されたパルス列を示す図である。

30

【図 7】同実施形態に係る音源同定方法を示すフローチャートである。

【図 8】パルス列からパルス数ベクトル、特徴ベクトルが生成される様子を示したイメージ図である。

【図 9】同実施形態に係る音源同定方法の第 2 ステップのフローチャートである。

【図 10】同実施形態に係る音源同定方法の第 3 ステップのフローチャートである。

【図 11 - 1】目覚まし時計のアラーム音を入力したときの識別結果を示す図である。

【図 11 - 2】インターホンの呼び出し音を入力したときの識別結果を示す図である。

【図 11 - 3】笛吹きケトルの沸騰音を入力したときの識別結果を示す図である。

【図 12 - 1】救急車のサイレン音を入力したときの識別結果を示す図である。

【図 12 - 2】パトカーのサイレン音を入力したときの識別結果を示す図である。

40

【図 12 - 3】電話のベル音を入力したときの識別結果を示す図である。

【図 13 - 1】消防車のサイレン音を入力したときの識別結果を示す図である。

【図 13 - 2】人の声を入力したときの識別結果を示す図である。

【図 14】パルス数ベクトルと、パルス数ベクトルと同様の参照ベクトルとの距離をマンハッタン距離で計測した例である。

【図 15】テスト・ファイルの音信号を入力したときの出力結果を示す図である。

【図 16】パラメータ N、k の値を変えて、識別の正確さの変化を調べた結果を示す図である。

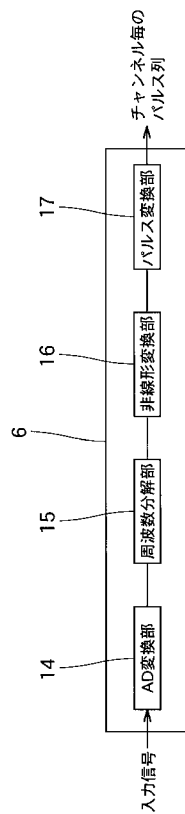
【符号の説明】

【0049】

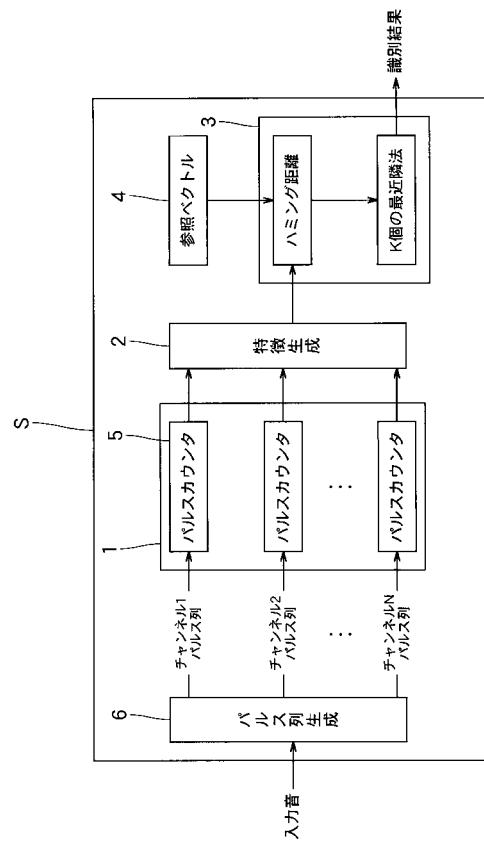
50

- S ... 音源同定装置
 1 ... パルス数ベクトル生成手段
 2 ... 特徴ベクトル生成手段
 3 ... 音源カテゴリ識別手段
 4 ... 参照ベクトル記憶手段
 6 ... パルス列生成手段

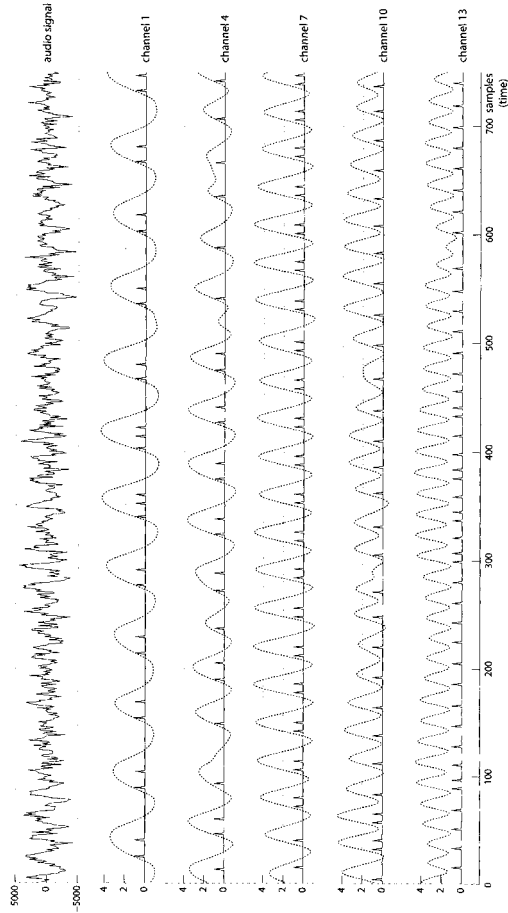
【図 1】



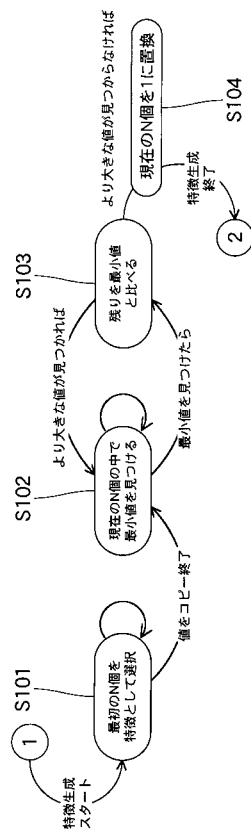
【図 2】



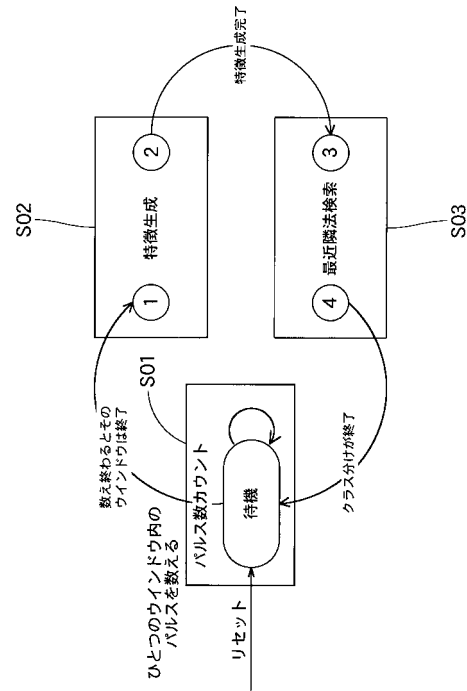
【図 3】



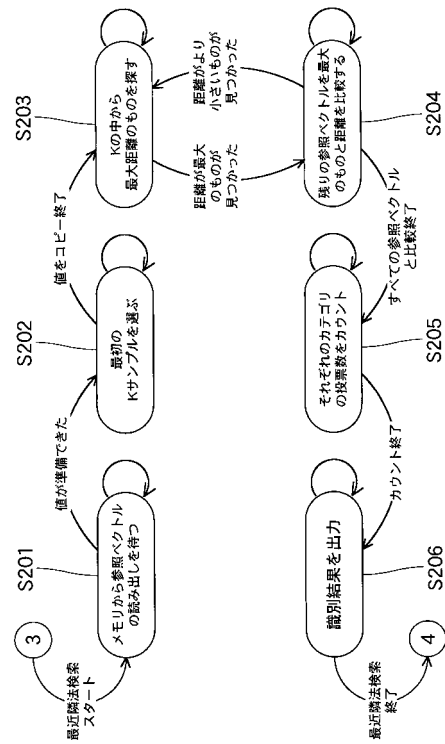
【図 9】



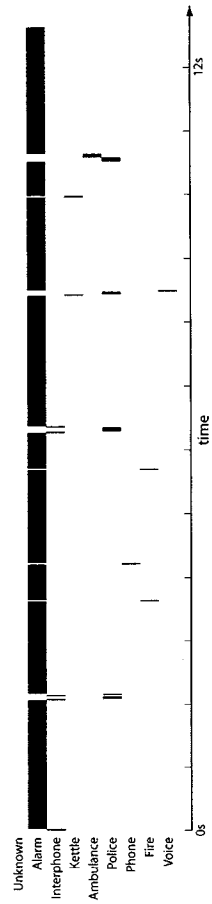
【図 7】



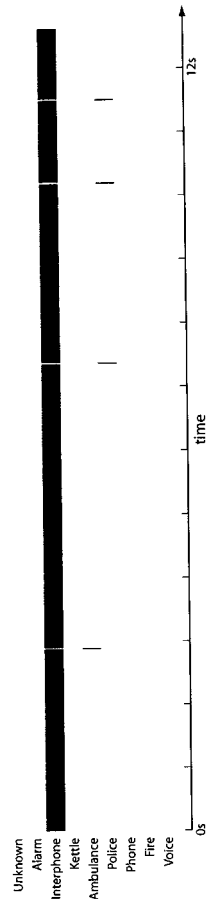
【図 10】



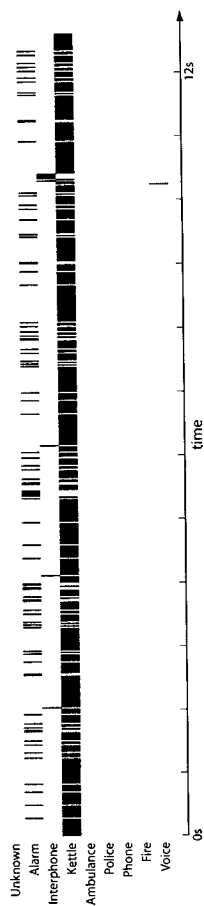
【図 1 1 - 1】



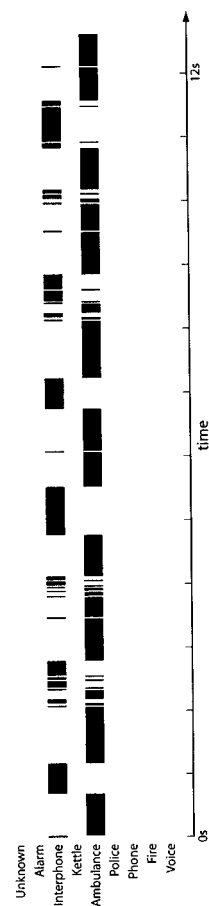
【図 1 1 - 2】



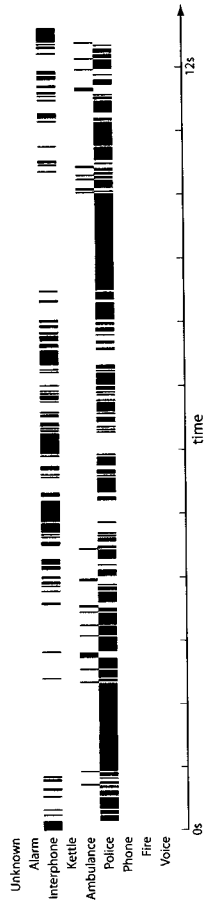
【図 1 1 - 3】



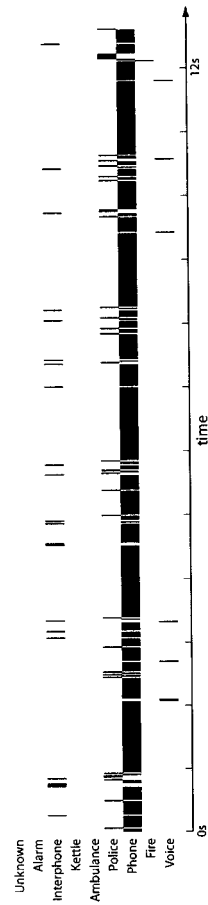
【図 1 2 - 1】



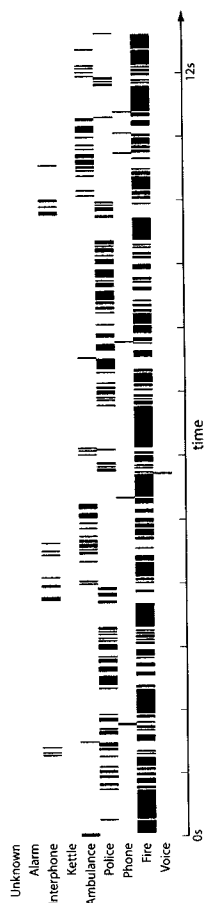
【図 1 2 - 2】



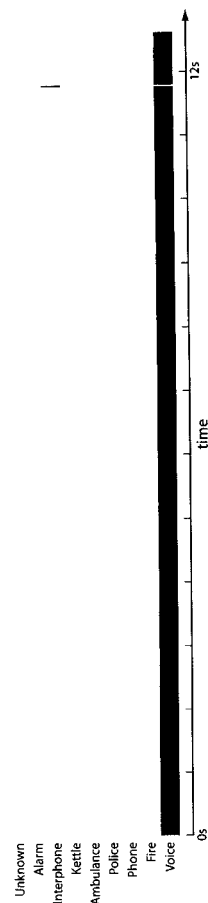
【図 1 2 - 3】



【図 1 3 - 1】



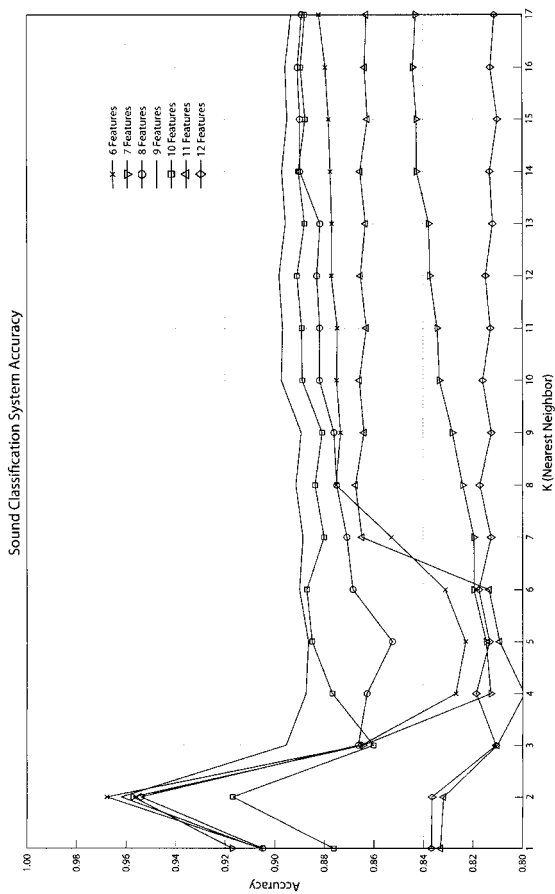
【図 1 3 - 2】



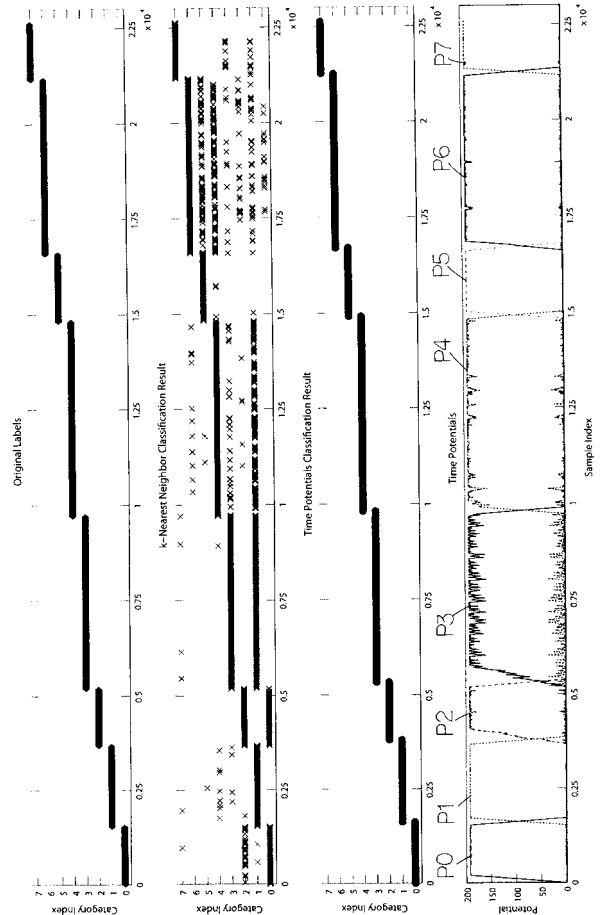
【図 14】

reference	12	35	8	32	11	
vector A	35	12	8	11	32	distance = 88
vector B	9	24	14	38	7	distance = 30

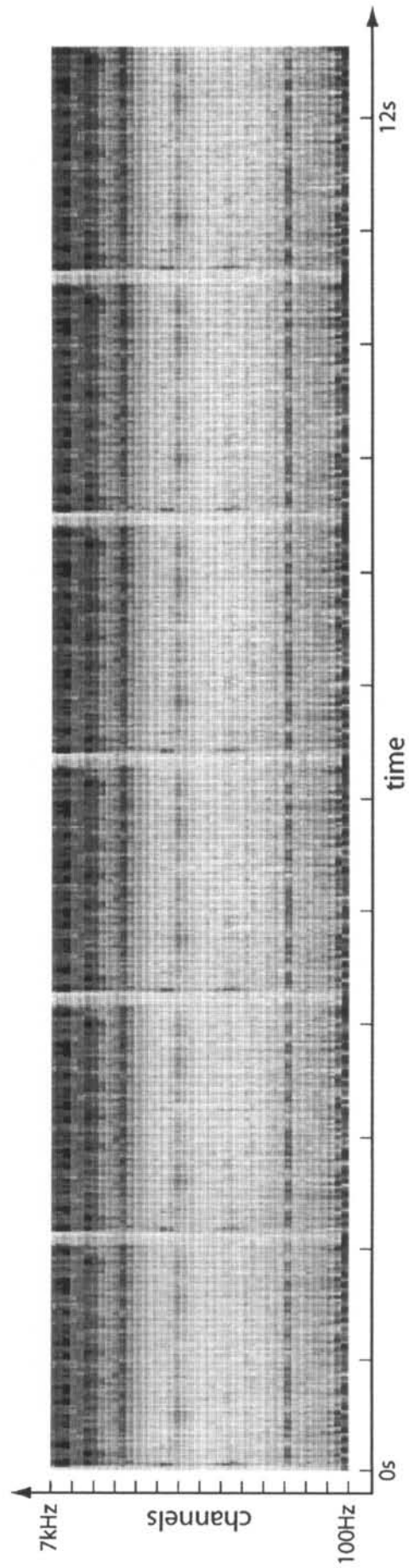
【図 16】



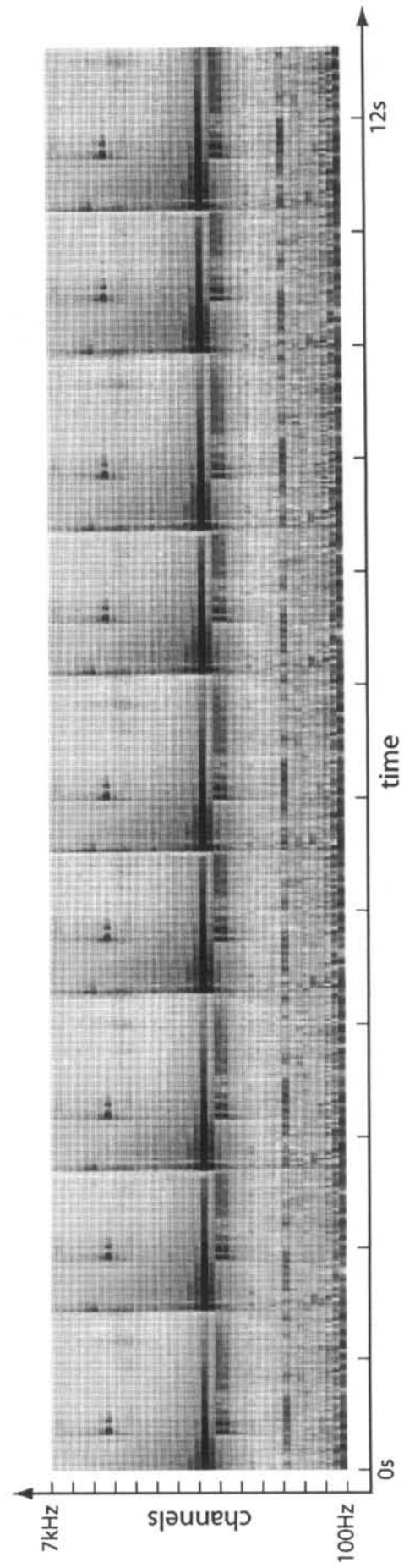
【図 15】



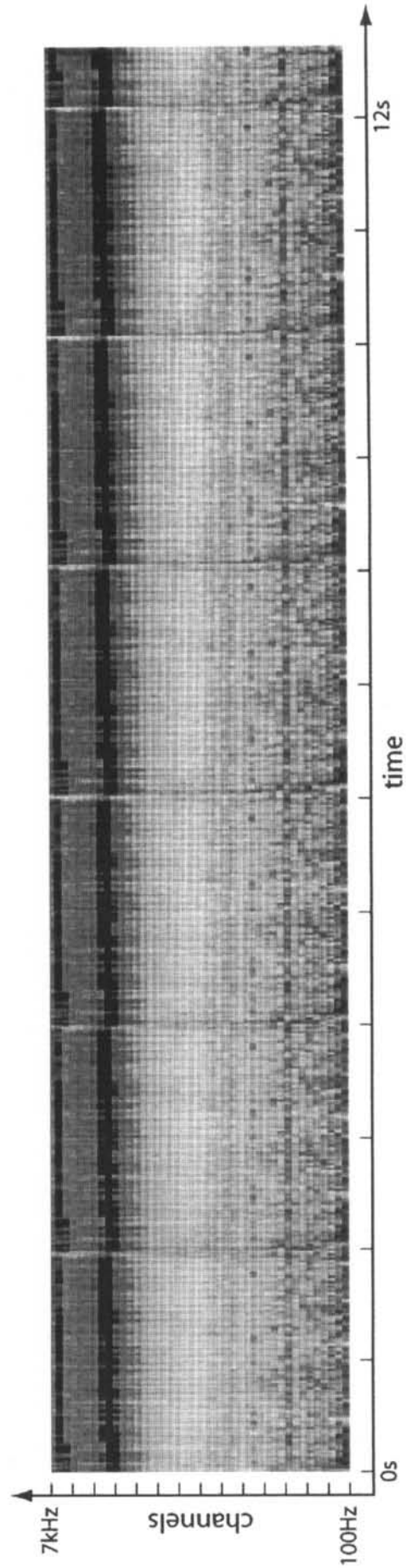
【 図 4 - 1 】



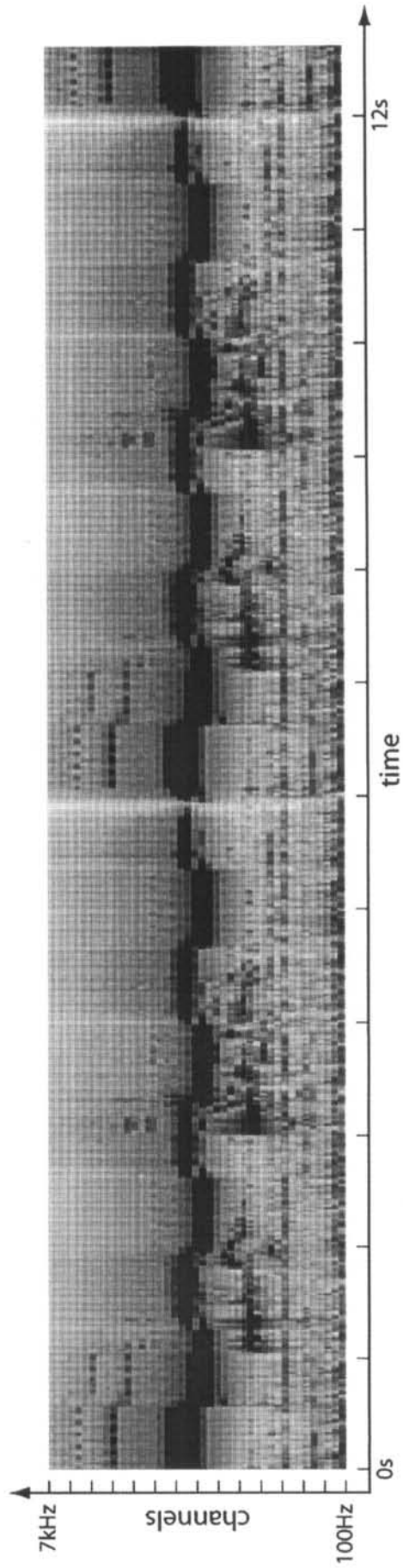
【図 4 - 2】



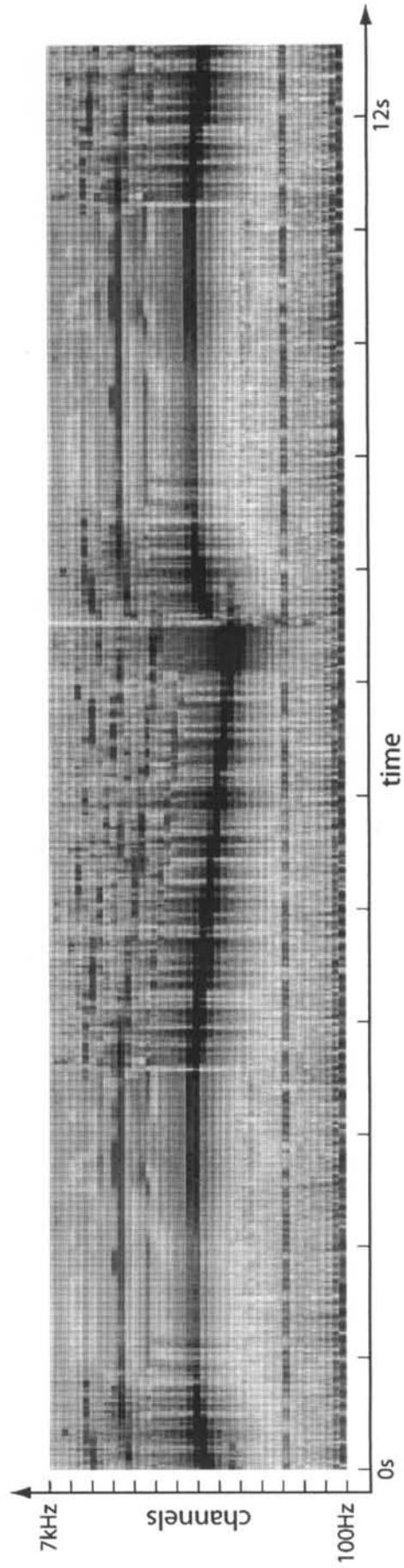
【 図 4 - 3 】



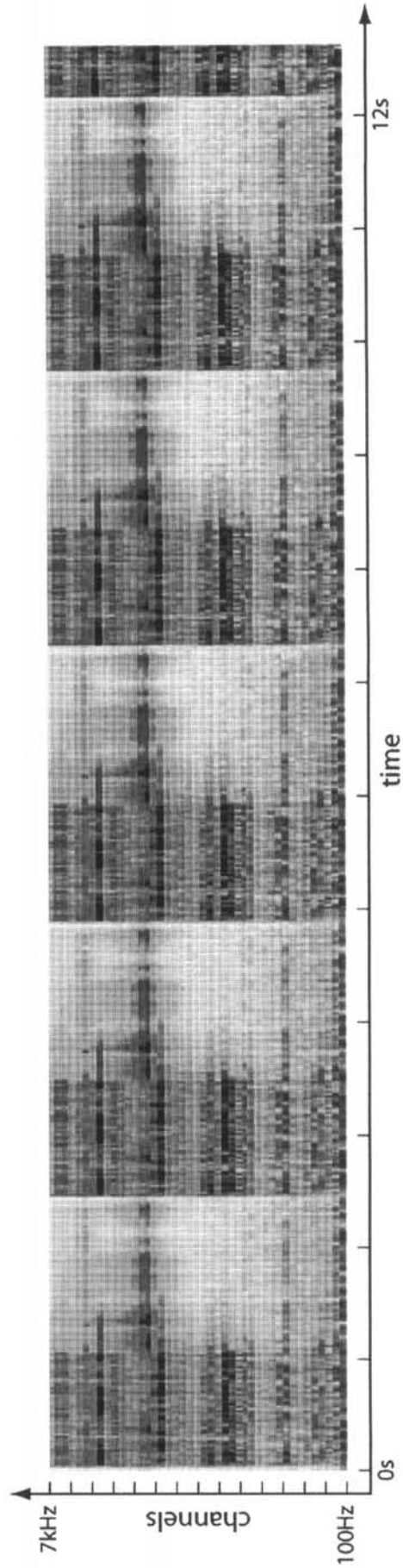
【図 5 - 1】



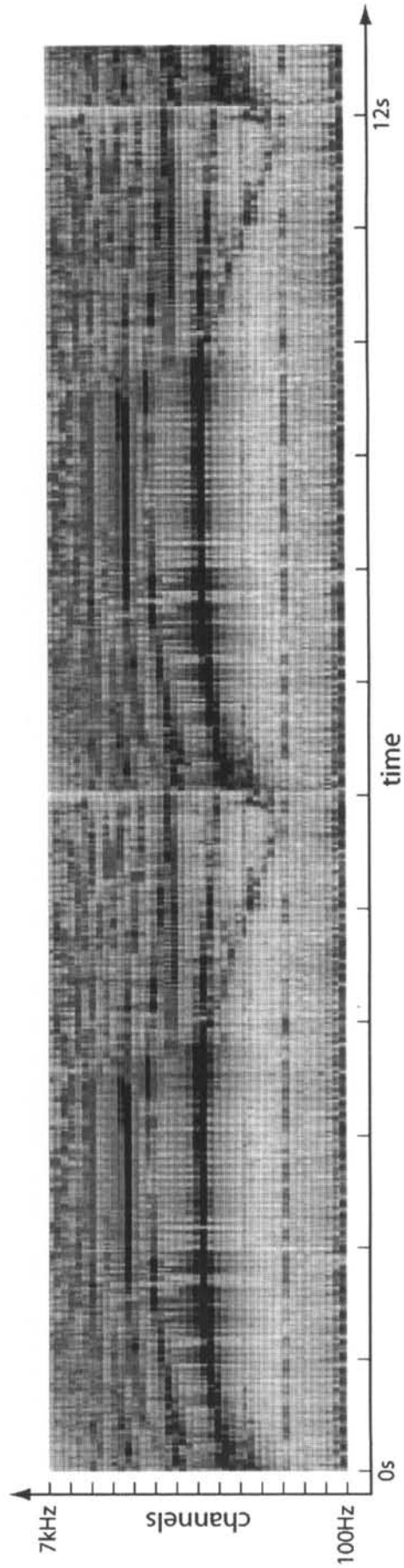
【図 5 - 2】



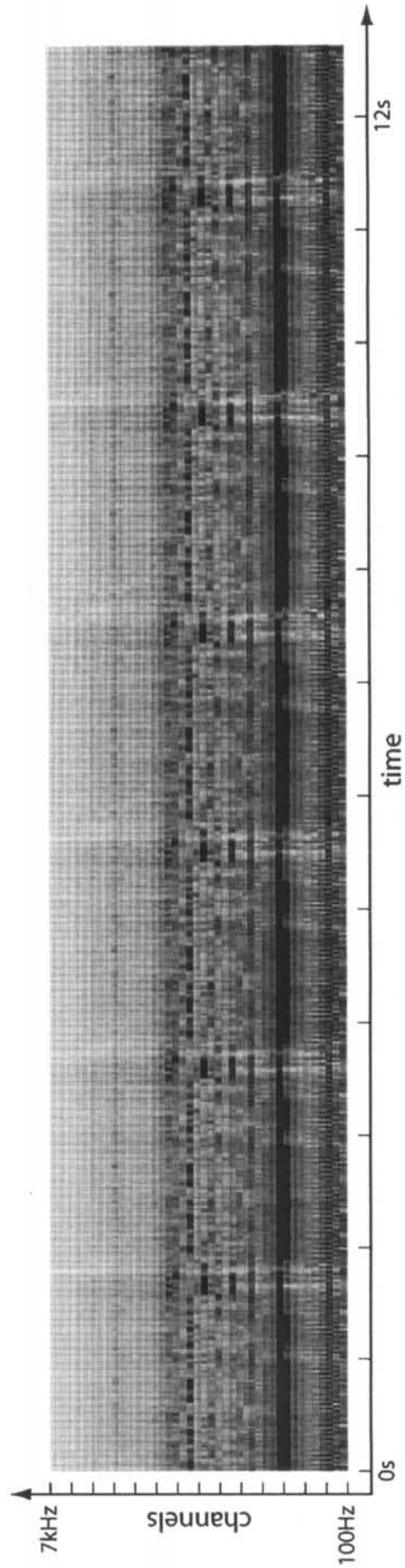
【図 5 - 3】



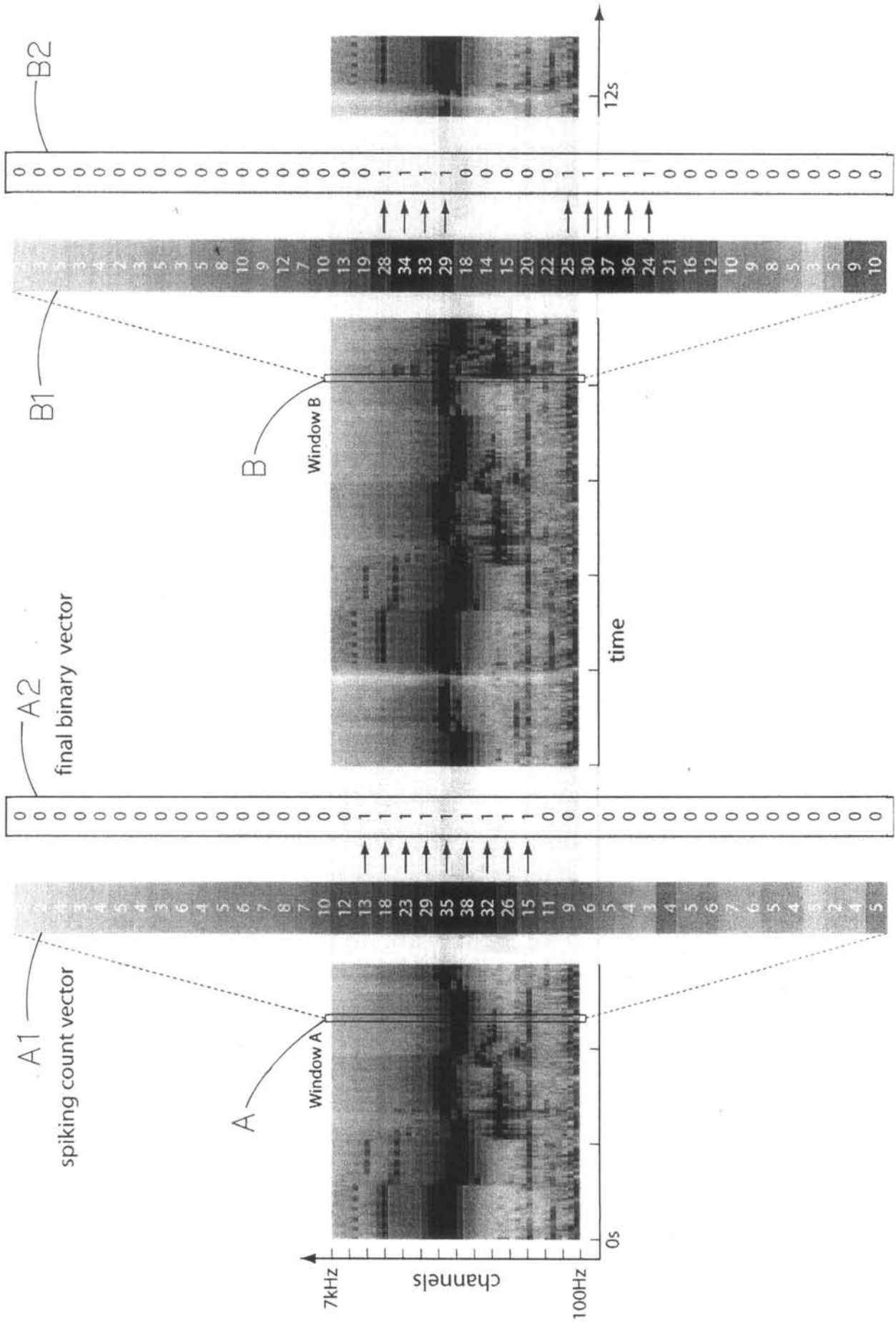
【図 6 - 1】



【図 6 - 2】



【図 8】



フロントページの続き

(72)発明者 マウリシオ クグレ
愛知県名古屋市昭和区御器所町 名古屋工業大学内
Fターム(参考) 5D015 AA06