世界知的所有権機関
国際事務局

PCT

特許協力条約に基づいて公開された国際出願

(51) 国際特許分類6 **G10L 9/00, H04N 5/76, G11B 20/10**	**A1**	(11) 国際公開番号	**WO98/41978**
		(43) 国際公開日	1998年9月24日(24.09.98)

(21) 国際出願番号 PCT/JP97/00905

(22) 国際出願日 1997年3月19日(19.03.97)

(71) 出願人（米国を除くすべての指定国について）
株式会社 日立製作所(HITACHI, LTD.)[JP/JP]
〒101 東京都千代田区神田駿河台四丁目6番地 Tokyo, (JP)
(72) 発明者；および
(75) 発明者／出願人（米国についてのみ）
藤田武洋(FUJITA, Takehiro)[JP/JP]
〒185 東京都国分寺市西恋ケ窪三丁目8番1号 Tokyo, (JP)
宮武孝文(MIYATAKE, Takafumi)[JP/JP]
〒193 東京都八王子市館町1585-10 Tokyo, (JP)
長坂晃朗(NAGASAKA, Akio)[JP/JP]
〒185 東京都国分寺市東恋ケ窪6丁目13番地8号 Tokyo, (JP)
(74) 代理人
弁理士 小川勝男(OGAWA, Katsuo)
〒100 東京都千代田区丸の内一丁目5番1号
株式会社 日立製作所内 Tokyo, (JP)

(81) 指定国 CN, JP, KR, US, 欧州特許 (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

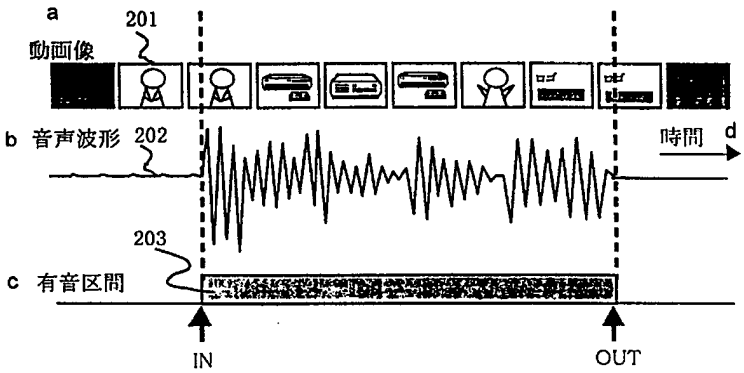
添付公開書類
国際調査報告書

(54)Title: METHOD AND DEVICE FOR DETECTING STARTING AND ENDING POINTS OF SOUND SECTION IN VIDEO

(54)発明の名称 映像中の有音区間の始終点検出方法及び装置

(57) Abstract

A device for detecting starting and ending points of a sound section in video is provided with an envelope calculating means which calculates the envelope of the waveform of a sound signal from time-sequentially inputted moving image signals, an acoustic level threshold setting means which presets an acoustic level threshold for the value of the envelope, and a starting and ending point detecting means which detects the points at which the threshold level intersects the envelope as the starting and ending points of a sound section and calculates the envelope of the sound waveform (202) in a video and detects the point where the value of the envelope becomes higher than the acoustic level threshold as the starting point of a sound section (203) and the point where the value of the envelope becomes lower than the threshold as the ending point of the section (203). The device then registers the video section corresponding to the starting and end points by use of the number of the frame constituting the moving image. Therefore, the labor of work for detecting the starting and ending points of the sound section can be saved, because the section is quantitavely and automatically detected by calculating the envelope of the waveform of sounds. In addition, since the envelope is calculated through filtering process requiring a small amount of calculation, the processing ability of the device can be low, and hence the detecting device is inexpensive.



a ... moving image

b ... wave form of sounds

c ... sound section

d ... time

（５７）要約

時系列に入力した動画像信号から音声信号波形の包絡線を算出する包絡線算出手段と、該包絡線の値に対して音響レベルの閾値を予め設定する音響レベルの閾値設定手段と、該閾値のレベルと該包絡線が交差する点を有音区間の始終点として検出する始終点検出手段とを設け、映像中の音声波形２０２の包絡線を算出し、該包絡線の値が該音響レベルの閾値を上回った点を有音区間２０３の開始点、下回った点を終了点として検出する。そして、その開始点と終了点とに対応する映像の区間を動画像を構成するフレームの番号によって登録する。

本発明によれば、音声波形の包絡線を算出して定量的かつ自動的に有音区間を検出しているので検出作業を省力化できる。また計算量の少ないフィルタリング処理で音声波形の包絡線を算出しているので検出装置の処理能力は低くてよく、安価な検出装置を作成できる。

ＰＣＴに基づいて公開される国際出願のパンフレット第一頁に掲載されたＰＣＴ加盟国を同定するために使用されるコード（参考情報）

AL アルバニア
AM アルメニア
AT オーストリア
AU オーストラリア
AZ アゼルバイジャン
BA ボスニア・ヘルツェゴビナ
BB バルバドス
BE ベルギー
BF ブルキナ・ファソ
BG ブルガリア
BJ ベナン
BR ブラジル
BY ベラルーシ
CA カナダ
CF 中央アフリカ
CG コンゴー共和国
CH スイス
CI コートジボアール
CM カメルーン
CN 中国
CU キューバ
CY キプロス
CZ チェッコ
DE ドイツ
DK デンマーク
EE エストニア
ES スペイン
FI フィンランド
FR フランス
GA ガボン
GB 英国
GE グルジア
GH ガーナ
GM ガンビア
GN ギニア
GW ギニア・ビサオ
GR ギリシャ
HU ハンガリー
ID インドネシア
IE アイルランド
IL イスラエル
IS アイスランド
IT イタリア
JP 日本
KE ケニア
KG キルギス
KP 北朝鮮
KR 韓国
KZ カザフスタン
LC セントルシア
LI リヒテンシュタイン
LK スリランカ
LR リベリア
LS レソト
LT リトアニア
LU ルクセンブルグ
LV ラトヴィア
MC モナコ
MD モルドヴァ
MG マダガスカル
MK マケドニア旧ユーゴスラヴィア共和国
ML マリ
MN モンゴル
MR モーリタニア
MW マラウイ
MX メキシコ
NE ニジェール
NL オランダ
NO ノールウェー
NZ ニュー・ジーランド
PL ポーランド
PT ポルトガル
RO ルーマニア
RU ロシア
SD スーダン
SE スウェーデン
SG シンガポール
SI スロヴェニア
SK スロヴァキア
SL シエラ・レオーネ
SN セネガル
SZ スワジランド
TD チャード
TG トーゴー
TJ タジキスタン
TM トルクメニスタン
TR トルコ
TT トリニダッド・トバゴ
UA ウクライナ
UG ウガンダ
US 米国
UZ ウズベキスタン
VN ヴィェトナム
YU ユーゴースラヴィア
ZW ジンバブエ

明　細　書

映像中の有音区間の始終点検出方法及び装置

技術分野

本発明はビデオテープやディスクに格納されたビデオ等の映像中に含まれる音声データの有音区間を検出する方法及び装置に係り、映像中の音声の頭出しを簡単にする方法及び装置に関わる。

背景技術

近年、計算機の高速化と記憶装置の大容量化を背景にして、映像中の動画像、音声情報をディジタル化して扱うことができるようになってきた。特に映像編集や映像管理の分野においては、テレビ放送制作、ビデオ制作のための撮影装置や編集装置、管理装置でディジタル化された映像を扱えるようになってきた。このような装置の一つに、数千種類のコマーシャル(以下CM)映像を管理して、任意のCM映像を放送順に準備するCM管理装置（通称CMバンク）がある。従来は複数のCM素材を放送前にビデオテープに一本化していたが、最近ではビデオテープを用いずにCM素材のファイルを直接に放送するCM管理装置も利用されている。CM管理装置には、広告代理店等の素材制作元から供給されたCM素材が登録される。従来、CM素材はCM毎に個別にビデオテープで供給され、素材の映像にはCM以外にも制作元の名称や制作日時を記した画像等が撮影されている。また、CMの前後には送出のタイミングを合わせるための遊びの映像が数秒間入っている。そのため、CM素材を管理装置に登録する際には、製作もと～供給されたマザー素材を別のテープやディスク等の記録媒体にコピーして格納するのに加えて、放送すべきCMの始まりと終わりを登録する必要がある。

CM映像の始まりと終わりをチェックする作業は、現状、すべて人手で行わ

れ、作業者にとって大きな負担であった。遊びの映像はCM映像本体の始まりと終わりに連続して撮影されているため、画像を見ただけではどこからどこまでが放送すべきCM映像なのかわからない場合が多い。しかし、音と映像の組み合わせで構成されるCM映像等において、遊びの部分には音声は録音されないため、作業者は映像中の遊びの区間における音を聞いて始まりと終わりを判断している。このような判断を行う際、従来技術では、作業者が映像の再生、停止、逆再生等の操作を繰り返しながら、聴覚で音の有無を判断するしか方法がなかった。これらの操作は映像再生装置のジョグ、シャトル等のダイヤルあるいは計算機画面のスクロールバーを利用することで改善されているが、依然として手間のかかる作業であった。

本発明は、CM素材を管理装置に登録する際、CMの始まりと終わりを音の有無によって聴覚で判断する作業を自動化し、登録操作を自動化して簡単にする方法および装置を提供することを目的としている。

本発明の他の目的は、CM映像本体の始終点を実時間で検出し、その位置を登録する方法および装置を提供することにある。

発明の開示

本発明は、映像管理装置への対話的登録処理において、時系列に入力した音声信号波形の包絡線を算出する包絡線算出手段と、該包絡線の値に対して音響レベルの閾値を予め設定する音響レベルの閾値設定手段と、該閾値のレベルと該包絡線が交差する時点を有音区間の開始点もしくは終了点として検出する始終点検出手段とを設け、従来聴覚によって判断していた音の有無を定量的かつ自動的に判断することができるようにする。ここで、該始終点検出手段には、無音状態の経過時間長の下限を予め設定する無音時間長の下限設定手段と、音声信号波形の包絡線の値が音響レベルの閾値を下回った経過時間を算出する無音時間長算出手段と、該無音時間長が該下限を超えたことを判定する無音時間長判定手段とを設け、発話中の句の継ぎ目のような微少

時間の音の途切れを検出しないようにする。同じく該始終点検出手段には、有音状態の経過時間長の下限を予め設定する有音時間長の下限設定手段と、音声信号波形の包絡線の値が音響レベルの閾値を上回った経過時間を算出する有音時間長算出手段と、該有音時間長が該下限を超えたことを判定する有音時間長判定手段とを設け、ノイズや単発的な音を検出しないようにする。また、該包絡線算出手段には、時系列に入力した音声信号に対して一定の時間幅を有するフィルタリング処理を施すフィルタリング手段を設ける。該フィルタリング手段には、時系列に入力した音声信号に対して一定の時間幅の最大値を順次求める最大値フィルタと最小値を順次求める最小値フィルタを用いる。

また本発明は、素材映像を再生する映像再生手段と、再生映像の音声トラックに記録された音声信号をディジタル信号として時系列に入力する音声入力手段と、入力した音声信号から有音区間の始終点を検出する音声処理手段と、検出結果を表示するディスプレイ手段を設け、素材映像中の有音区間の始終点の位置を作業者に提示できるようにする。該音声処理手段には、上記包絡線算出手段、音響レベルの閾値設定手段、始終点検出手段に加えて、有音区間の始終点を検出した時点の映像のフレーム位置を決めるフレーム位置決定手段を設ける。該フレーム位置決定手段には、検出処理開始からの時間経過を計数するタイマー手段と、映像のフレーム位置を読み出す手段と、該始終点を検出した時点の経過時間とフレーム位置を読み出した時点の経過時間を記憶する経過時間記憶手段と、両者の経過時間の差分から読み出したフレーム位置を補正するフレーム位置補正手段とを設け、始終点検出からフレーム位置読み出しの時間遅れを補正して検出時点でのフレーム位置を決定できるようにする。さらに該音声処理手段には、検出した始終点で映像の再生を停止させる手段を設け、該始終点のフレーム位置で映像の再生を一時停止できるようにする。ここで、映像再生手段には計算機によって映像再生を制御できる映像再生装置を用いる。例えば、VISCA (Video System Control

Architecture：ビスカ）端子付ビデオデッキや一般にプロ用の編集でよく使われるビデオデッキ等を用いる。これによって、検出した有音区間へ効率的に頭だしできるようにする。

さらに本発明は、上記音声処理手段に有音区間の開始点および終了点のフレーム位置を個別に記憶するフレーム位置記憶手段と、該開始点および終了点のフレーム位置を個別に表示するディスプレイ手段とを設け、素材映像中の有音区間の開始点と終了点の位置を個別に作業者に提示できるようにする。また、一定時間長単位で時系列に入力した音声信号を蓄積するバッファメモリ手段と、入力した音声信号を再生する再生手段とを設け、検出した有音区間を視覚的にも聴覚的にも作業者が確認できるようにする。さらに処理対象の映像をCM素材として、CM本体は一本あたり１５秒または３０秒の長さを持つという一般的なルールを利用して、これら一定時間の有音区間の時間長の上限と１または２秒の許容範囲を予め設定する時間長設定手段と、検出した有音区間の開始点から終了点までの検出時間長と該設定時間長を比較する時間長比較手段とを設け、CM映像の一定の時間長の有音区間だけを検出できるようにする。また、検出した有音区間前後にマージンを設定するマージン設定手段を設け、CM素材から一定時間長の放送用のCM映像をCM管理装置に登録できるようにする。

図面の簡単な説明

第１図は、本発明の実施例を実現するためのシステム構成図であり、第２図は、本発明の有音区間検出方法の概念図であり、第３図は、本発明の有音区間検出方法のフローチャートであり、第４図は、本発明の有音区間の始終点判定条件を示した図であり、第５図は、本発明を実現するための操作画面例を示した図であり、第６図は、全体の処理の流れを示したフローチャートであり、第７図は、本発明の有音区間検出の制御方式を示した図であり、第８図は、本発明のフィルタリング処理の入出力データの位置関係を示した図

であり、第９図は、CMの時間長ルールを用いた有音区間検出処理の流れを示したフローチャートであり、第１０図は、本発明の有音区間検出を実現するためのデータ構造例を示した図である。

発明を実施するための最良の形態

以下、本発明の実施例を図面をもちいて説明する。

第１図は、本発明を実現するためのシステム構成図の一例である。１０１はＣＲＴ等のディスプレイ装置であり、音声処理装置１０４の出力画面を表示する。音声処理装置１０４に対する命令および閾値等の設定は、マウス等のポインティング・デバイスおよびテンキー等の数値入力デバイスからなる入力装置１０５を使って行う。映像再生装置１１０は、ビデオテープや光ディスク等に記録された映像を再生する装置である。映像再生装置１１０から再生出力される映像信号中の音声信号は、逐次、音声入力装置１０３によってディジタル信号に変換され、音声処理装置１０４に入力される。ディジタル信号に変換する際のサンプリング周波数、サンプリング・ビット数、およびモノラル又はステレオを示すチャンネル数（モノラルは１、ステレオは２）といった情報も音声入力装置１０３から音声処理装置１０４に受け渡される。もちろん、これらの情報は１０４で設定した数値を１０３に与えるようにしてもよい。音声処理装置１０４は受け取った信号を処理して映像再生装置１１０を制御する。音声処理装置１０４と映像再生装置１１０は通信線１０２を通じて制御命令や応答の送受信をする。映像再生装置１１０の扱う映像の各フレームに映像の先頭から順にフレーム番号（タイムコード）が割り付けられている場合には、該フレーム番号及び検索命令を音声処理装置１０４から映像再生装置１１０に送信することで、当該フレーム番号の画像を検索できる。同様に音声処理装置１０４は映像再生装置１１０にフレーム番号を要求して映像の現在のフレーム番号を受け取ることもできる。音声処理装置１０４内部では、音声のディジタル信号はインタフェース１０８を介して一旦

メモリ１０９に入力され、メモリ１０９に格納された処理プログラムに従って、ＣＰＵ１０７によって処理される。処理プログラムは補助記憶装置１０６に格納されており、ＣＰＵ１０７の命令に応じてメモリ１０９に適宜転送される。メモリ１０９には、以下に説明する処理によって作成される各種のデータが蓄積され、必要に応じて参照される。また音声のディジタル信号や処理結果等の各種情報は１０６に格納することもできる。スピーカ１１１は音声信号入力装置１０３から音声処理装置１０４に入力された音声信号を入力に合わせて再生したり、メモリ１０９に蓄積された音声信号をユーザの要求に応じて再生したりする。

以下では、まず最初に、ユーザが映像を確認しながら映像中の音声の有音区間を簡便に検出できる映像中の有音区間検出方法について説明する。次いで、この方法を用いて構成する有音区間検出装置について説明し、CM素材映像から一定時間長の放送用CM映像を見つけ出す方法について説明する。

第２図は本発明の映像中の有音区間の検出方法の概要を示す模式図である。

動画像２０１と音声波形２０２は、映像中に含まれる画像および音声の信号を示している。ここでは簡単のため音声波形２０２はモノラルで示しているが、ステレオであっても構わない。対象とする映像がCM素材の場合、素材の映像にはCM本体の前後に遊びの映像が数秒間入っている。遊びの映像はCM本体の前後に同じ映像が連続して撮影されているため、動画像２０１を見ただけではどこからどこまでが放送すべきCM映像なのかわからない場合が多い。但し、遊びの部分には音声は録音されない。そのため従来は、作業者が映像の再生、停止、逆再生等の操作を繰り返しながら映像中の音の有無を判断してCM映像の先頭や末尾を決めている。本発明では有音区間を検出することによってこのような聴覚による判断を自動化する。

音声波形２０２の振幅値は正負の値が交互に頻繁に現れ、瞬間的に大きさが零となっていることも非常に多い。そのため、単にある瞬間の振幅の大きさを調べたのではその近傍での音の有無は必ずしもわからない。ここでは音

の大きさを音声波形２０２の包絡線の値で判定する。包絡線の値には近傍の音の有無が反映される。包絡線の値が予め定めた音響レベルの閾値を上回った点を有音区間２０３の開始点（IN）、下回った点を終了点（OUT）として検出する。該開始点から該終了点までの音声データをメモリ１０９や補助記憶装置１０６に蓄え該データを再生することで、有音区間２０３の音声内容を確認することも容易である。これらの検出点の映像中での位置はフレーム番号として求める。有音区間２０３の該開始点、該終了点といった変化点を検出したとき、映像再生装置１１０は既に該変化点以降の映像を再生している。そのため映像再生装置１１０から検出時点でのフレーム番号を読み取り、読み取った時刻と該変化点の時刻の差分を用いて該フレーム番号を換算し、該変化点のフレーム番号を算出する。該換算方法については後ほど第７図で説明する。包絡線を利用して有音区間を検出し、フレーム番号を利用して元の映像と有音区間を対応づけることによって、音が連続してある一定の音響レベルを上回っている映像区間を抽出できる。また、該開始点のフレーム番号と検索要求を映像再生装置１１０に送信することによって、音の立ち上がりフレームの頭だしが容易に実現できる。さらに、該開始点から該終了点までの時間長がわかるため、放送用CM映像に仕上げるのに必要なマージンを抽出した映像区間の前後に設定することも容易である。これによって、時間長にばらつきのない品質の良いCM映像をCM管理装置に登録できる。

本発明によれば、ユーザは第１図に示したシステムを使って、素材の入ったビデオテープ等を映像再生装置１１０にセットし、ディスプレイ１０１に表示される音声処理装置１０４のコンソールのボタンを操作するだけでよい。コンソールの画面例は後ほど第５図で説明する。ユーザはジョグ、シャトル等を使った手作業で映像中の有音区間の先頭や末尾を見つけ出さなくともよくなる。それゆえ、操作の簡単化の効果がある。

次に第３図、第４図を用いて、有音区間の検出方法の詳細について説明する。

第３図は本発明の映像中の有音区間の始終点検出方法のフローチャートである。

３０１から３０６はそれぞれプログラムステップを、３１１から３１６は各ステップの出力データを示している。これらプログラムやデータはすべてメモリ１０９に置かれ、ＣＰＵ１０７によって処理する。ここでは説明を簡単にするために音声波形をモノラル（チャンネル数１）で示しているが、ステレオ（チャンネル数２）の場合も同様に扱うことができる。ステレオの場合は左右各チャンネルの音声波形に対して以下に示すモノラルの処理を実行し、両チャンネルの結果のAND（論理積）を取って重なりを判定したりOR（論理和）を取って全体を判定すればよい。

まずステップ３０１では映像中の音声データを音声入力装置１０３から受け取る。３１１は受け取った音声データの波形である。ステップ３０２では、３１１の各データの絶対値を求め、音声波形の折り返し処理を行う。ここでは、音声の意味内容に関わらず、音響レベルだけを扱うためである。３１２は３１１を正に折り返した音声波形である。次にステップ３０３、３０４では最大最小型フィルタリングによって波形３１２の包絡線を求める。各フィルタリングではフィルタサイズ３２１および３２２のフィルタを準備し、入力データを順次フィルタに読み込んで、フィルタ内の最大値、最小値を求めて出力する。ステップ３０３では波形３１２に対してフィルタ中の最大値を各データ毎に出力する。ステップ３０４では最大値波形３１３に対してフィルタ中の最小値を各データ毎に出力する。３１４は結果として得られた包絡線である。ステップ３０５では、包絡線３１４の各データを予め定めた音響レベルの閾値３２３と比較する閾値処理を行う。包絡線３１４が閾値３２３を上回るときには有音として１を、下回るときには無音として０を出力する。３１５はステップ３０５で出力した有音、無音の２値データである。最後にステップ３０６では該２値データ３１５の有音、無音の連続性を調べて有音区間３２４を検出し、有音区間の始終点３１６を出力する。すなわち該有音

区間の立ち上がり点を音の開始点３２５（IN）として、また該有音区間の立ち下がり点を音の終了点３２６（OUT）として出力する。このステップ３０６については第４図でタイミングチャートを用いながら説明する。

最大最小型フィルタリングによって包絡線を算出する方法は、音声波形のパワースペクトルを計算して０次のパワーを包絡線とする方法よりも計算量を格段に少なくすることができる。それゆえＣＰＵの能力があまり高くなくても実現できる。

ステップ３０３、３０４に記した１次元の最大最小型フィルタリングについては、例えば「最大最小型画像フィルタリングの高速演算手法」（電子情報通信学会論文誌D-II, Vol.J78-D-II, No.11, pp.1598-1607, 1995年11月）に記載されたフィルタリング手法を用いればよい。フィルタサイズnに対して、n+1個のデータが格納できるリングバッファを利用したデータ順次処理方式である。この手法では、データ性質やフィルタサイズに関わらず、一つのデータあたり平均演算回数３回程度で最大値や最小値を求めることができる。それゆえ、今回のように大量のデータを高速に処理したい場合に適している。

第４図は有音区間の始終点判定方法を示した図である。

有音区間の始終点の判定にあたっては、始終点判定条件を、

開始点：無音状態がTn以上続いた後、有音状態がTs以上続いたときの状態変化点

終了点：有音状態がTs以上続いた後、無音状態がTn以上続いたときの状態変化点

と定める。ここでTs[msec]は有音状態の経過時間長の下限であり、Tn[msec]は無音状態の経過時間長の下限である。TsやTnの値は、音声中の一音節の時間長や発話文の間の休止の時間長を基準にして予め設定しておく。これによって、Ts未満の有音状態やTn未満の無音状態を検出しないようにできるので、単発的なノイズや発話中の句の継ぎ目のような微少時間の音の途切れに影響されない、安定した有音区間の検出方法を実現できる。

４０１は、ステップ３０６にて入力データ３１５から有音区間の始終点３１６が得られるまでを示したタイミングチャートである。状態を判定するフラグとして無音フラグ、有音フラグ、開始フラグ、終了フラグの４つのフラグを設ける。

ステップ３０６では、有音、無音の２値状態を示す入力データ３１５を順次調べ、各状態の経過時間として０（無音）と１（有音）のデータ個数を計数する。音声信号のディジタル化の際のサンプリング周波数は音声入力装置１０３から音声処理装置１０４に渡されているので、時間条件TsおよびTnをデータ個数の条件に置き換えることは容易である。なお、有音状態のデータ個数は無音フラグがONになった時点でクリアし、無音状態のデータ個数は有音フラグがONになった時点でクリアする。最初、フラグはすべてOFFに設定し、各状態のデータ個数を０に設定する。まず無音状態がTnだけ続いた時点で無音フラグをONにする（４０２）。無音フラグがONであるとき、無音から有音に変化した点はすべて開始点の候補とし、そのデータ位置をメモリ１０９に格納する。まず有音状態４０３の立ち上がりを開始点の候補とするが、有音状態４０３は経過時間がTsに満たないため４０３のデータ個数は無音状態のデータ個数（経過時間）に組み込み、単発的なノイズとしてキャンセルする。次に有音状態４０４の立ち上がりを開始点の候補とし、有音状態がTsだけ続いた時点で有音フラグをONにする（４０５）。これによって無音フラグ、有音フラグともONとなって開始点の条件を満たすので、開始フラグをONにして開始点３２５（IN）に決定する。ONになった開始フラグはセンスした時点でOFFに戻す。以上時間軸上４２０までが開始点検出である。

開始点を検出すると引き続いて終了点の検出を開始する。まず無音フラグはOFFにする（４０６）。有音フラグがONであるとき、有音から無音に変化した点はすべて終了点の候補とし、そのデータ位置をメモリ１０９に格納する。無音状態４０７は経過時間がTnに満たないため４０７のデータ個数は有音状態のデータ個数に組み込み、微少時間の途切れとしてキャンセルする。次に

無音状態４０８がTnだけ続いた時点で無音フラグをONにする（４０９）。これによって有音フラグ、無音フラグともONとなって終了点の条件を満たすので、終了フラグをONにして終了点３２６（OUT）に決定する。ONになった終了フラグはセンスした時点でOFFに戻す。また、有音フラグも次の開始点検出に備えてOFFにする（４１０）。以上時間軸上４２１までが終了点検出である。

以上、第４図のようにフラグを操作することによって継続的に有音区間の始終点を検出することができるので、一本の映像に複数の有音区間が含まれている場合でも各々を検出できる。それゆえ、本発明の有音区間検出方法はCM素材やビデオ映像に限らずTV放送映像やアーカイブ映像等の一般的な映像に対しても適用可能である。また処理映像がCM素材の場合には「CMは一本あたり１５秒もしくは３０秒の時間長で構成される」という一般的なCMの時間長ルールを利用できるので、複数の有音区間を検出してもルールによって区間をひとまとめにすることで、適切なCM本体の始終点を決定できる。CMのルールを用いた始終点の検出方法については後ほど第９図で説明する。

次いで、以上の有音区間検出方法を用いて構成する有音区間検出装置について説明する。

第５図は本発明を実現する有音区間検出装置の操作画面の一例である。操作ウィンドウ５０１は音声処理装置１０４のコンソールとしてディスプレイ装置１０１に表示され、ユーザに操作環境を提供する。５０１にはQUITボタン５０２、DETECTボタン５０３、検出結果表示パネル５０４、音声波形モニタ５０５、有音区間表示パネル５０６、PLAYボタン５０９、映像再生装置操作パネル５１０、パラメータ設定パネル５１３を配置する。ユーザは５０１上に配置された命令ボタンを入力装置１０５のマウスでクリックして自分の要求を音声処理装置１０４に伝える。QUITボタン５０２は操作処理を終了し操作ウィンドウ５０１を閉じるための命令ボタンである。

DETECTボタン５０３は有音区間検出処理を実行するための命令ボタンである。ユーザが５０３をクリックすると、音声処理装置１０４は検出結果表示

パネル５０４をクリアして、プログラム３００に従って有音区間の検出を開始し、処理途中の結果を音声波形モニタ５０５に表示する。モニタ５０５には算出した包絡線３１４や音響レベルの閾値３２３を表示する。有音区間の始終点が検出されると、検出したフレーム番号をタイムコード形式でパネル５０４に表示する。タイムコードはhh:mm:ss:ff（hh：時間、mm：分、ss：秒、ff：フレーム）という構成で位置や長さを直観的に把握しやすいため、ユーザにとって都合がよい。

有音区間表示パネル５０６には有音区間検出の始終点を検出するまでに入力した音声データの波形５０７および有音区間５０８を表示する。有音区間５０８は検出結果表示パネル５０４のINフレームからOUTフレームまでに対応している。一般にCM映像は一本あたり最長３０秒なので、ここでは４０秒分の音声波形を表示するようにしている。PLAYボタン５０９は有音区間５０８の音声データを再生するためのボタンである。ユーザは音声データ波形５０７によって、映像中の音声信号を視覚的に確認できる。また５０９をクリックして音声を再生することで聴覚によっても確認することができる。これによって、ユーザは有音区間検出後すぐに検出状況を確認することができ、確認作業を省力化することができる。

ユーザが有音区間にマージンを設ける場合、有音区間５０８の端をドラッグして区間を広げることによって設定すればよい。検出結果表示パネル５０４に表示しているように有音区間の始終点は既にわかっているので区間長は計算できる。ユーザは該区間長の前後にマージンを設定して全体の時間長が所望の長さになるようにする。システムは設定されたマージンの長さに応じてパネル５０４のフレーム番号を変更し、変更後のフレーム番号をCM管理装置に登録するCM映像の始終点にする。これによって、ユーザは簡単にCM管理装置への登録作業を進めることができる。また、該映像の始終点に挟まれた映像を切り出すことで、ユーザは所望の長さの放送用のCM映像を作成できる。

映像再生装置操作パネル５１０には映像再生装置操作ボタン群５１１を配

置する。操作ボタン群５１１は映像の早送り、巻き戻し、再生、コマ送りや一時停止を実行するための命令ボタンである。ユーザが操作ボタン群５１１から必要な命令ボタンをクリックすると、音声処理装置１０４は該当する操作命令を映像再生装置１１０に送信する。映像のフレーム位置はタイムコードとして映像位置表示ボックス５１２に表示する。

パラメータ設定パネル５１３には有音区間検出のためのパラメータを設定するパラメータ設定ボックス５１４を配置する。パネル５１３には変更可能なパラメータとして、音響レベルの閾値（Threshhold Value）、フィルタの時間長（Filterlength）、有音状態の経過時間長の下限（Noise Limt）、無音状態の経過時間長の下限（Silence）の４つを配置している。ユーザがパラメータを変更したい場合、５１４をクリックして入力装置１０５から数値を入力する。このとき、音響レベルの閾値（図中のThreshold Value）の設定にあたっては、入力装置１０５から数値を入力する他に、別方式として次のように設定するようにしてもよい。まず該音響レベルの閾値のパラメータ設定ボックスがクリックされると、映像再生装置１１０は停止もしくは一時停止状態にしておき、その状態で音声入力装置１０３から音声処理装置１０４に音声データを数秒間入力する。次いで、数秒間入力した該音声データの音響レベルの最大値を音響レベルの閾値とする。このとき、数秒間入力することによって、映像再生装置１１０や音声入力装置１０３で生じる音声信号のランダムノイズが音声処理装置１０４に入力できる。また最大値を音響レベルの閾値とすることによって、１１０や１０３で生じるノイズが再生映像中の音声信号に影響するのを押さえることができるようになる。

第６図は処理の全体の流れを示したフローチャートである。ユーザからプログラムの起動要求があるとＣＰＵ１０７は補助記憶装置１０６からプログラム６００を読み出し、メモリ１０９に格納して実行する。このときメモリ１０９には音声データや処理データも各種格納されるが、これらデータの構造については後ほど第１０図で説明する。

ステップ６０１は処理開始にあたっての初期化処理である。まず、ＣＰＵ１０７はメモリ１０９上に処理に必要なメモリ領域を確保してクリアし、音響レベルの閾値等のパラメータのデフォルト値をセットする。次いで、ディスプレイ装置１０１に音声処理装置１０４の操作ウィンドウ５０１を表示する。また、映像再生装置１１０との通信の設定を初期化し、通信ポートを開く。次に映像再生装置１１０に制御命令を送信して、映像再生装置１１０の再生を一時停止状態（STAND BY ON）にする。映像再生装置１１０を停止状態ではなく一時停止状態にしておくことで、別の制御命令が送られたときに映像再生装置１１０を即座に動作させることができ、音声信号やフレーム番号を素早く読み出すことができるようになる。

ステップ６０２ではユーザの終了要求の有無を判定し、要求がない間、ステップ６０３の画面制御を繰り返して実行する。

ステップ６０３ではユーザに指定された命令ボタンに対応して処理を分岐させる。例えば、ユーザが操作ウィンドウ５０１のDETECTボタン５０３をクリックした場合にはステップ６０８と６０９を実行してユーザの入力を待つ。操作ウィンドウ５０１に配置される命令ボタンの個数、種類の増減に応じて分岐数、分岐判断を増減し、常に最適な処理が選択できるようにすればよい。

ステップ６０４から６０９は各命令ボタンに応じた処理である。

ステップ６０４は映像再生装置操作ボタン群５１１が指定されたときの処理である。この制御処理は、操作ボタン群５１１がクリックされた時ばかりでなく、映像再生装置１１０を制御する処理として利用可能である。まず映像再生装置１１０に制御命令を送信し、映像再生装置１１０の応答ステータスを受信する。次に応答ステータスを判定し、エラーが発生した場合にはエラーメッセージをディスプレイ装置１０１に表示して処理を中断する。正常に制御できた場合はフレーム番号を読み出して表示ボックス５１２に表示してステップ６０３に戻る。

ステップ６０５はパラメータ設定ボックス５１４が指定されたときのパラ

メータ設定処理である。ユーザが入力装置１０５から数値を入力して設定パラメータを変更するとメモリ１０９に格納された該当パラメータを書き換える。また時間長に関するパラメータが変更されると、音声データのサンプリング周波数に従って変更後の時間長をデータ個数に換算する。

ステップ６０６は検出した有音区間５０８の入力音声データをスピーカ１１１で再生する音声再生処理である。検出結果表示パネル５０４に有音区間の始終点がセットされていれば、５０４のINフレームからOUTフレームまでの音声データを再生する。すなわち音声蓄積データバッファ１０５０の音声データをデータ位置１０５２からデータ位置１０５３までで再生する。これによって、ユーザは検出結果を聴覚で確認できる。

ステップ６０７は検出した有音区間にマージンを設けるマージン設定処理である。ユーザは有音区間５０８の端をドラッグして区間を広げることによって設定する。まず、検出結果表示パネル５０４のINフレームからOUTフレームまでの有音区間の時間長を計算する。一本あたりのCM映像の時間長を一定長に定めておくと、該有音区間の時間長からマージンの上限は一意に決まる。該上限を越えないようにユーザの操作を監視しながらマージンを決定し、始終点のフレーム番号を補正する。これによって、時間長にばらつきのない品質の良いCM映像を管理装置に登録できるようになる。別の方式としては、該上限を満たす適切なマージンを区間の前後に自動的に付加してもよい。時間長を限定しないのであれば、ユーザの要求通りにマージンを付加する。また逆に有音区間長を短くすることも可能である。

ステップ６０８は有音区間の始終点検出処理である。DETECTボタン５０３が指定されると、映像再生装置１１０の映像を再生し、音声入力装置１０３から音声データを入力し、有音区間の始終点を検出して検出結果表示パネル５０４に表示する。詳細については後ほどプログラム９００（第９図）で説明する。プログラム９００はプログラム３００で示した有音区間の始終点検出方法を有音区間検出装置に適用した例である。このとき、別の方式として、

有音区間の始終点を検出した後に、映像再生装置１１０の映像を有音区間の開始点に頭出しするようにしてもよい。該頭出しは音声処理装置１０４から映像再生装置１１０に有音区間の開始点のフレーム番号と検索命令を送信することによって実現できる。

ステップ６０９では波形５０７と有音区間５０８をパネル５０６に表示する。有音区間検出の始終点を検出するまでに入力した音声データを波形５０７として表示し、検出結果表示パネル５０４のINフレームからOUTフレームまでを有音区間５０８として表示する。すなわち、音声蓄積リングバッファ１０５０の音声データをオフセット１０５４からリングバッファを一巡して波形表示する。また１０５２と１０５３に挟まれるデータ区間を５０８として表示する。これによって、ユーザは検出結果を視覚的に確認できる。

ステップ６１０は終了処理である。まず映像再生装置１１０に制御命令を送信して映像再生装置１１０を停止状態（STAND BY OFF)にした後に通信ポートを閉じる。次いでディスプレイ装置１０１上の操作ウィンドウ５０１を閉じる。最後に確保していたメモリ領域を開放し、処理を終了する。

ここで、プログラム３００に示した有音区間の始終点検出方法を有音区間検出装置に適用する際の制御方式とフィルタリング処理方式を開示する。

プログラム３００では、映像全体の音声データを入力してから始終点を検出することもできるが、長時間の音声データを一度に入力すると検出までのタイムラグが長くなり、検出のリアルタイム性が損なわれる。検出のリアルタイム性を保持するためには、音声データを短い時間毎に分割して入力処理および検出処理を行った方がよい。

まずリアルタイム検出を実現するための制御方式を示す。第７図は本発明の有音区間検出装置の制御方式を示した図であり、有音区間の開始点を検出するまでを示している。図中の各矩形は制御対象の処理を示し、該矩形の幅は処理時間長を示す。

７０２は音声入力装置１０３での音声データ入力処理である。入力音声は

一定時間長の音声バッファが一杯になるまで音声入力装置１０３に蓄積し、一杯になった時点で音声処理装置１０４に該音声バッファが一杯になったことを示す割り込み信号を送る。７０２の時間長は音声バッファの大きさを示している。１０４では該割り込み信号を受け取ると該音声バッファのデータをメモリ１０９に転送する。７０３は音声処理装置１０４での音響解析処理であり、プログラム３００を実行する。１０４は該割り込み信号が届いた時点から７０３を開始し、次の該割り込み信号が届くまでに実行する。例えば該音声バッファの時間長を１秒に設定すると、７０３の処理には最大１秒まで時間をかけることができるので処理時間として十分である。また、Tsを200msec、Tnを500msecに設定すると、音の開始点や終了点は最大２回分の音声データを処理することで検出できる。このとき、１０３への入力開始から１０４での検出までのタイムラグは最大３秒程度に抑えることができ、ほぼリアルタイムに検出できる。ここでTsやTnは第４図で説明した有音／無音状態の経過時間長の下限であり、これらの数値は音声中の一音節の時間長や、発話文の間の休止の時間長を基準に設定する。メモリ１０９に転送されるデータ量は、サンプリング周波数を１１kHz、サンプリングビット数を８ビット、チャンネル数を１（モノラル）に設定すれば１秒のバッファでは１１kバイトであるため、転送時間は問題にならない。

開始点検出までの処理の流れを示す。DETECTボタン５０３がクリックされると、まず全体制御処理によって映像再生装置１１０で映像を再生し、音声入力処理７０２を起動し、有音区間検出処理を準備し、処理の経過時間をタイマによって数え始める（７０１）。７０２の音声入力処理によって音声データを入力すると、７０３の音響解析処理ではデータ到着時刻T1をメモリ１０９に記録する（７０４）。また音響解析処理によって音の開始点を検出するとメモリ１０９上の検出フラグをONに設定する（７０５）。７０３の音響解析処理が済むと、全体制御処理では検出フラグをセンスする。検出フラグがOFFの場合、途中結果を音声波形モニタ５０５に表示する（７０６）。一方、

ONの場合、映像再生装置１１０から現在のフレーム番号を読み出し、タイマからフレーム番号読み取り時刻T2を取得し、該フレーム番号および該読み取り時間をメモリ１０９に格納する。またT1、T2を利用して該フレーム番号を音の開始点の時点のフレーム番号に換算してメモリ１０９に格納する（７０７）。音の終了点を引き続き検出する場合は、７０２から７０７の処理を終了点検出まで繰り返し実行する。７０２から７０７の処理は何度でも繰り返し実行できるので、一本の映像に複数の有音区間が含まれている場合でも各々を検出できる。

次に処理７０７での該開始点のフレーム番号の換算方法について説明する。７０８の音声入力処理で得られた音声データ中の位置Xに音の開始点が含まれていたとする。このとき、データ到着時刻T1、フレーム番号取得時刻T2、フレーム番号TC2から音の開始点の時刻T0を推定し、TC2を開始点のフレーム番号TC0に換算する。この換算方法は次式で表せる。

$$T0 = T1 - dT(L-X) / L \ [msec] \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

（数１）

$$TC0 = TC2 - 1000(T2 - T0) / 30 \ [フレーム] \quad \cdots\cdots\cdots\cdots\cdots$$

（数２）

ここで、Lは音声バッファサイズ（データ個数）であり、dTは音声バッファの時間長である。音声データが８ビット、モノラルの場合、Lは音声バッファのバイト数そのものとなる。数２で分母に３０とあるのは、NTSCの映像信号が１秒間に３０フレームで構成されるためである。音の終了点についても同様に計算できる。

以上の制御方式によって、ほぼリアルタイムに有音区間の始終点を検出できる。

次に分割入力した音声データを連続的にフィルタリング処理する方式について説明する。第８図はステップ３０３もしくは３０４のフィルタリング処理における入出力データの位置関係を示した図である。図中の矩形はそれぞ

れデータ配列を示す。８０１は入力データ配列（データ個数L[個]）、８０２はフィルタバッファ（データ個数Lf[個]）である。８０２はステップ３０３では３２１のフィルタに該当し、ステップ３０４では３２２のフィルタに該当する。

ステップ３０３と３０４の各フィルタリング処理では、フィルタバッファ８０２に８０１のデータを順次読み込み、８０２の全データから最大値または最小値を求め、フィルタサイズの中央の位置のデータとして出力する。このとき、８０１全体の入力データから８０３の部分の出力データが得られる。L個の入力データのうち、フィルタサイズ分のLf個は８０２の初期化に使うため、出力データ配列の最初８０４と最後８０５の部分については出力データが得られない。第７図の制御方式で音声入力装置１０３からデータを受け取る度に８０２を初期化したのでは、フィルタリング結果の包絡線が途切れ途切れになってしまう。

ここではフィルタバッファ８０２は７０１で一回だけ初期化し、その後は途中でクリアしないで、次に入力データを読み込む位置やデータ内容を絶えず保持しておく。これによって、（ｎ＋１）回目の音響解析処理に、ｎ回目から継承した８０２のデータLf個と（ｎ＋１）回目の入力データ８０６のL個を利用できるようになり、８０５の部分と８０７の部分の合計L個の出力データが得られるようになる。すなわちL個の入力データに対してL個の出力データが得られるようになるので、分割入力した音声データを連続的にフィルタリング処理できるようになる。

但しこのとき、ｎ回目の８０５の部分の出力データが得られるのは、（ｎ＋１）回目の８０６が入力されてからになる。第７図の制御方式では音の始終点のフレーム番号を算出するとき、数１に示したように始終点のデータ位置Xとタイマから読み取った入力データ到着時刻T1を使う。そのため、ｎ回目と（ｎ＋１）回目の２回分のデータ到着時刻をメモリ１０９に記録しておく。音の始終点が８０５で見つかった場合にはｎ回目の到着時刻を、８０７で見

つかった場合には（ｎ＋１）回目の到着時刻を使うようにする。

なお、フィルタサイズLfはL-Lfが正になるような値を設定しておけばよい。人間の音声の基本周波数は一般に100Hz以上であるから、その逆数の10msec以上の時間長（例えば１フレーム時間33mec）に含まれるデータ個数を設定しておけば包絡線を算出するのに問題はない。該データ個数は時間長にサンプリング周波数を乗ずることで算出できる。

以上によって、処理する音声データに途切れを生ずることなく検出処理を実行できる。

以上の制御方式、フィルタリング方式を反映した有音区間の始終点検出処理のフローチャートを第９図に示し、メモリ１０９に格納する音声データや制御データのデータ構造を第１図０に示す。

第９図はCMの時間長ルールを用いた有音区間検出処理の流れを示したフローチャートである。プログラム９００は有音区間の始終点を一組検出する処理プログラムであり、ステップ６０８で実行する。９００は大きく分けて次の４つの処理から構成される。(1)有音区間の開始点検出処理、(2)有音区間の終了点検出処理、(3)CMの時間長ルールを用いた判定処理、(4)規定時間を経過すると検出を打ち切る検出時間制限処理、である。(1)の処理はステップ９０２から９０４であり、(2)の処理はステップ９０６、９０７と９１０である。これらによって、第７図に示した処理７０３から７０７の制御を実行する。(3)の処理はステップ９０５と９１１から９１５である。これらによって、予め定めた一定時間の有音区間のみをふるい分ける。(4)の処理はステップ９０８と９０９である。これらによって、検出処理の実行時間に上限を設け終了点が見つからない場合のエラー処理を実行する。但し、ここで有音区間の検出に最低限必要なのは(1)の処理と(2)の処理であり、(3)と(4)の処理は省略可能である。

以下、各ステップを順に説明する。

ステップ９０１は初期化処理である。メモリ１０９に格納する音声データ

や制御データを初期化し、第７図の制御処理７０１を実行する。すなわち音声バッファ１０３０や音声蓄積バッファ１０５０、制御パラメータ１０１０を初期化し、フィルタバッファ１０４０の空フラグ１０４２をTRUEにする。

ステップ９０２では有音区間の開始点の検出状況を判定する。制御パラメータ１０１０の開始点フラグflagIN１０１７がTRUEになるまでステップ９０３を実行する。

ステップ９０３では有音区間の開始点を検出する。プログラム３００を実行し、途中結果を音声波形モニタ５０５に表示する。開始点を検出するとflagIN１０１７をTRUEにし、映像再生装置１１０から現在のフレーム番号を読み取り、フレーム番号取得時刻T2をタイマから読み取る。

ステップ９０４では検出した開始点のフレーム番号を算出する。数１で開始点の時刻T0を算出し、数２で開始点のフレーム番号TC0を算出する。開始点のTC0は検出結果表示パネル５０４に表示し、またflagINをFALSEに戻す。

ステップ９０５では有音区間の検出状況を判定する。一定時間長の有音区間を検出するまで以下のステップを実行する。

ステップ９０６では有音区間の終了点の検出状況を判定する。制御パラメータ１０１０の終了点フラグflagOUT１０１８がTRUEになるまでステップ９０７から９０９を実行する。

ステップ９０７では有音区間の終了点を検出する。プログラム３００を実行し、途中結果を音声波形モニタ５０５に表示する。終了点を検出するとflagOUT１０１８をTRUEにし、映像再生装置１１０から現在のフレーム番号を読み取り、フレーム番号取得時刻T2をタイマから読み取る。この時の終了点のフレーム番号はステップ９１０で算出する。

ステップ９０８では検出処理の経過時間を判定する。ステップ９０３で開始点を検出してからの経過時間が規定の検出時間よりも長くなると、処理中の映像には適切な時間長の映像が含まれていないと判定してステップ９０９を実行する。該規定の検出時間としては、例えばCMの時間長３０秒の２倍と

して６０秒にすればよい。現在の入力データ到着時刻T1１０２２がステップ９０３で取得したT2に対してT1 > T2 + 60[sec]となれば、適切な時間長の映像はないと判定する。

ステップ９０９では検出結果を棄却して、検出処理を中断する。先に検出した開始点をキャンセルする。また音声入力装置１０３からデータ入力するのを止め、映像再生装置１１０での映像再生を一時停止し、音声バッファ１０３０やフィルタバッファ１０４０をクリアする。

ステップ９１０では検出した終了点のフレーム番号を算出する。数１で終了点の時刻T0を算出し、数２で終了点のフレーム番号TC0を算出する。終了点のTC0は検出結果表示パネル５０４に表示し、またflagOUTをFALSEに戻す。

ステップ９１１では有音区間の時間長Tを算出する。ステップ９０４で求めた開始点の時刻とステップ９１０で検出した終了点の時刻の差分としてTを求める。

ステップ９１２はCMの時間長ルールを用いた判定処理である。検出した有音区間の時間長が予め定めた一定時間長を満足すれば、ステップ９１３、９１４を実行し、該一定時間長を超過するとステップ９１５を実行する。該一定時間長に満たない場合は、次の有音区間の終了点検出に戻る。これによって、一定時間長の有音区間を含む映像のみ検出できるようになる。ここでは「CMは一本あたり１５秒もしくは３０秒の時間長で構成される」という一般的なルールを用いて一定時間長を１５秒と３０秒とし、許容範囲を１５秒に対しては１秒、３０秒に対しては２秒に設定しているが、これらの値は用途に応じて適当に変更しても構わない。

ステップ９１３、９１４では検出した始終点を有音区間の始終点として採用する。また音声入力装置１０３からデータ入力するのを止め、映像再生装置１１０での映像再生を一時停止し、音声バッファ１０３０やフィルタバッファ１０４０をクリアする。

ステップ９１５では検出結果を棄却して、検出処理を中断する。検出した

始終点をキャンセルし、パネル５０４の表示をクリアする。また音声入力装置１０３からデータ入力するのを止め、映像再生装置１１０での映像再生を一時停止し、音声バッファ１０３０やフィルタバッファ１０４０をクリアする。

以上の方法によって、一定時間長の有音区間のみを検出できる。

最後に、メモリ１０９に格納する音声データや制御データのデータ構造を説明する。 第１図０は本発明の有音区間検出を実現するためのデータ構造例を示した図である。処理データはメモリ１０９に格納し、必要に応じてＣＰＵ１０７に読み出す。

１０００は音声信号情報である。音声入力装置１０３で音声信号をディジタル化する際に使用しているサンプリング周波数１００１、サンプリング・ビット数１００２、およびチャンネル数１００３（モノラルは１、ステレオは２）を格納する。

１０１０は制御用パラメータである。有音区間検出処理で用いる各種パラメータやフラグを格納する。１０１１から１０１４はパラメータ設定パネル５１３で変更可能な可変パラメータである。１０１５から１０１８は第４図で説明した有音区間始終点判定時の状態を示す４つのフラグであり、１０１９と１０２０は有音／無音状態を計数するためのカウンタである。開始点フラグ１０１７と終了点フラグ１０１８は始終点が未検出ならFALSE、検出済みならTRUEにする。１０２１は第７図で示した入力音声データ中の始終点のデータ位置Xである。１０２２と１０２３は第８図で説明したデータ到着時間T1と前回のデータ到着時刻である。フラグ１０１７、１０１８がONになったことを検出した時点のフレーム番号を読み出して数１、数２に従って始終点のフレーム番号は算出できる。該始終点のフレーム番号もメモリ１０９に格納する。別方式としては、該算出したフレーム番号を補助記憶装置１０６に順次書き込むようにしてもよい。１０６の容量が許す限り有音区間を検出するようにできる。

音声バッファ１０３０は、プログラム３００の各ステップ間で受け渡される処理データ３１１から３１５を格納するバッファのデータ構造である。メモリ１０９には入力用、作業用、出力用に３つ用意する。各々のバッファサイズ１０３１にはすべて同じ値を設定する。データ個数１０３２は該バッファ１０３０に格納されるデータの個数である。第８図で説明したように最初の入力バッファデータだけでは８０４、８０５部分の出力データが得られないので出力バッファのデータ個数が減少する。そのため１０３１とは別に１０３２を設ける。１０３３は処理データである。

フィルタバッファ１０４０は、ステップ３０３と３０４の最大最小型フィルタリングに用いるリングバッファのデータ構造である。メモリ１０９にはMAXフィルタリング用とMINフィルタリング用に２つ用意する。バッファサイズ１０４１は１０１２のTLfから算出する。空フラグ１０４２はフィルタバッファの初期化状況を示し、バッファが空である初期状態でTRUEにし、いったんデータで満たされるとFALSEにする。入力音声バッファ１０３０を処理するとき、１０４２がTRUEならば入力データをサイズ１０４１分だけ複写することで初期化する。FALSEならば初期化しない。これによって途切れなく包絡線を算出することができる。１０４３は次の入力データを読み込む位置を指し示すオフセットである。１０４４は読み込んだ入力データであり、フィルタリング処理の対象データである。

１０５０は音声入力装置１０３から入力される音声データを複写して絶えず過去何秒分か保持する音声蓄積リングバッファである。１０５０の蓄積データは、音声データ波形５０７を表示したり、PLAYボタン５０９で音声を再生したりするのに利用する。１０５１はバッファサイズである。バッファサイズ１０５１は１０３１の整数倍にすると複写が容易になる。１０５２は第７図の有音区間の開始点のデータ位置Ｘに対応するリングバッファ上のデータ位置である。１０５３は同様に終了点に対応するリングバッファ上のデータ位置である。１０５２と１０５３には最初、負の値を設定し、始終点の検

出に応じてデータ位置の値と置き換える。１０５４は次の入力データ複写位置の先頭を指し示すオフセットである。１０５５は音声データである。

ここで、有音区間検出処理に用いるデータのメモリ量を見積もる。例えば音声信号情報１０００が１１kHz、８ビット、モノラルの音声データを扱う場合、入力バッファに記録できる時間長を１秒にすると音声バッファ１０３０の必要容量は１１kバイト程度になり、３つのバッファの合計は３３kバイト程度になる。音声を蓄積する時間長を４０秒とすると音声蓄積リングバッファ１０５０の必要容量は４４０kバイト程度である。フィルタ時間長を３０msecとするとフィルタバッファ１０４０の必要容量は０．３kバイト程度になり、２つでも１kバイトに満たない。その他のパラメータと合わせてもデータに関して必要なメモリ量は全部で５００kバイト程度である。それゆえ、メモリ量が比較的少ない安価な計算機でも本発明の方法は十分に実装可能である。

本発明の構成は、従来聴覚によって判断していた音の有無を定量的かつ自動的に検出できるので、有音区間の検出作業の省力化効果がある。作業者はCM素材を映像再生機にセットして、音声処理装置画面のボタンを操作するだけでよい。この操作には、映像の再生、停止、逆再生を頻繁に繰り返すといった煩雑な操作は一切不要であるので、操作の簡単化の効果がある。また音声信号を短時間毎に分割入力するようにしているためリアルタイムに有音区間を検出でき、作業性能向上の効果がある。確認作業の観点からは、検出した有音区間の音声を波形表示したり再生したりするようにしているので、検出結果を視覚および聴覚ですぐに確認することができ、確認作業の省力化の効果がある。CM映像の時間長ルールを利用して有音区間を検出しているので、長すぎたり短すぎたりする不適当な素材は却下でき、CM映像の時間長を別途検査する必要がなくなる。さらに検出した有音区間にマージンを設定できるようにしているので、時間長にばらつきのない品質の良いCM映像を管理装置に登録でき、登録映像の品質向上の効果がある。

また本発明の包絡線の算出に用いたフィルタリング処理はパワースペクト

ルの計算よりも計算量が少ないので、パソコン等の小型の計算機で実現でき、音声信号入力のサンプリング・レートが高くても計算が速く行える効果がある。

このような映像中の有音区間の検出方法を実現する装置は、パソコンのような小型の計算機で実現でき、安価な検出装置が達成できる。

産業上の利用可能性

以上のように本発明に係る有音区間の検出方法及び装置は、映像と音声とで構成されるCM画像の開始点と終了点を検出してCM画像を登録するCM登録装置に利用するのに適している。

また、映画やＴＶ番組に挿入されたCM映像の区間を検出するCM検出装置に利用することができる。

請 求 の 範 囲

1．映像に記録された音声信号を時系列に入力し、該音声信号の波形の包絡線を算出し、該予め設定された音響レベルの閾値と該包絡線が交差する時点を有音区間の開始点もしくは終了点として検出することを特徴とする映像中の有音区間の始終点の検出方法。

2．請求の範囲第1項において、無音状態の経過時間長の下限を予め設定し、上記音声信号の波形の包絡線の値が上記音響レベルの閾値を下回った経過時間が該下限よりも長ければ、該音響レベルの閾値と該包絡線が交差する時点を有音区間の開始点もしくは終了点として検出することを特徴とする映像中の有音区間の始終点の検出方法。

3．請求の範囲第1項において、有音状態の経過時間長の下限を予め設定し、上記音声信号の波形の包絡線の値が上記音響レベルの閾値を上回った経過時間が該下限よりも長ければ、該音響レベルの閾値と該包絡線が交差する時点を有音区間の開始点もしくは終了点として検出することを特徴とする映像中の有音区間の始終点の検出方法。

4．請求の範囲第1項において、上記時系列に入力した音声信号に対して一定の時間幅を有するフィルタリング処理を施して上記包絡線を算出することを特徴とする映像中の有音区間の始終点の検出方法。

5．請求の範囲第4項において上記フィルタリング処理では、時系列に入力した音声信号に対して一定の時間幅の最大値を順次求める最大値フィルタと最小値を順次求める最小値フィルタとが用いられることを特徴とする映像中の有音区間の始終点の検出方法。

6．請求の範囲第1項において、上記音響レベルの閾値は、映像を再生しないで無音の音声信号を数秒間入力し、その際に生じるノイズの音響レベルの最大値に設定されることを特徴とする映像中の有音区間の始終点の検出方法。

7．ユーザが指定した所望の位置で映像を停止させる映像再生装置と、映像

の音声トラックに記録された音声信号をディジタル信号として時系列に入力する音声入力装置と、入力した音声信号から有音区間の始終点を検出する音声処理装置とから成り、該音声処理装置には該音声信号の波形の包絡線を算出する包絡線算出手段と、該包絡線の値に対して音響レベルの閾値を予め設定する閾値設定手段と、該音響レベルの閾値と該包絡線が交差する時点を有音区間の開始点もしくは終了点として検出する始終点検出手段と、該有音区間の開始点もしくは終了点を検出した時点の映像のフレーム位置を決めるフレーム位置決定手段と、該フレーム位置を表示するディスプレイ手段とを設け、該有音区間の開始点もしくは終了点のフレーム位置を表示させるようにしたことを特徴とする映像中の有音区間の始終点の検出装置。

８．請求の範囲第７項において、上記フレーム位置決定手段は、検出処理開始からの時間経過を計数するタイマー手段と、映像のフレーム位置を読み出す手段と、始終点を検出した時点の経過時間とフレーム位置を読み出した時点の経過時間を記憶する経過時間記憶手段と、両者の経過時間の差分を用いて読み出したフレーム位置を補正するフレーム位置補正手段とを有することを特徴とする映像中の有音区間の始終点の検出装置。

９．請求の範囲第７項において、上記音声処理装置は、更に、該有音区間の始終点のフレーム位置で映像の再生を停止させる手段を有することを特徴とする映像中の有音区間の始終点の検出装置。

１０．ユーザが指定した所望の位置で映像を停止させる映像再生装置と、映像の音声トラックに記録された音声信号をディジタル信号として時系列に入力する音声入力装置と、入力した音声信号から有音区間の始終点を検出する音声処理装置とから成り、該音声処理装置には該音声信号の波形の包絡線を算出する包絡線算出手段と、該包絡線の値に対して閾値のレベルを予め設定する閾値設定手段と、該包絡線が該閾値のレベルを初めて上回る時点を有音区間の開始点として検出する開始点検出手段と、該包絡線が該閾値のレベルを初めて下回る時点を有音区間の終了点として検出する終了点検出手段と、

該開始点および終了点を検出した時点の映像のフレーム位置を決めるフレーム位置決定手段と、該開始点および終了点のフレーム位置を個別に記憶するフレーム位置記憶手段と、該開始点および終了点のフレーム位置を個別に表示するディスプレイ手段とを設け、該開始点および終了点のフレーム位置を表示させるようにしたことを特徴とする映像中の有音区間の始終点の検出装置。

１１．請求の範囲第１０項において、上記音声処理装置は、時系列に入力した音声信号を蓄積するバッファメモリ手段を有し、有音区間の開始点および終了点を検出したときに該区間の音声波形が表示されることを特徴とする映像中の有音区間の始終点の検出装置。

１２．請求の範囲第１０項において、上記音声処理装置には、入力した音声信号及び有音区間の開始点および終了点を検出したときの該区間の音声信号を再生する再生手段を有することを特徴とする映像中の有音区間の始終点の検出装置。

１３．請求の範囲第１０項において、上記音声処理装置は、予め定めた有音区間の時間長の上限と許容範囲を設定する時間長設定手段と、検出した有音区間の開始点から終了点までの検出時間長と該設定時間長を比較する時間長比較手段とを有し、該検出時間長が該設定時間長に比べて短い場合は有音区間の開始点を保持したまま次の有音区間の終了点を検出させ、該設定時間長に比べて長い場合は検出を終了して検出結果を却下させ、該設定時間長の許容範囲である場合は検出を終了して検出結果を保持させ、該設定時間長の２倍の時間を超えても終了点が検出されない場合は検出を終了させることを特徴とする映像中の有音区間の始終点の検出装置。

１４．請求の範囲第１３項において、上記有音区間の時間長の上限を１５秒または３０秒とし、該許容範囲は１または２秒とし、検出処理対象の映像をコマーシャル映像としたことを特徴とする映像中の有音区間の始終点の検出装置。

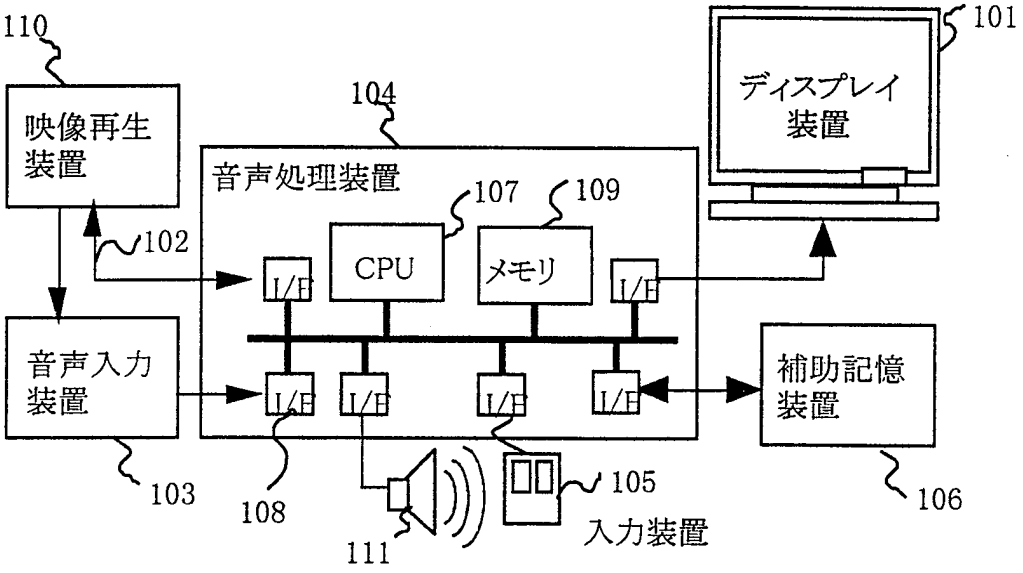
１５．請求の範囲第１３項において、上記音声処理装置は、有音区間の開始点の前方および終了点の後方に対してマージンを設定するマージン設定手段を有し、有音区間の検出時間長が該設定時間長の該許容範囲内である場合に、検出した開始点および終了点を前方および後方に時間をずらせた結果を有音区間の開始点および終了点とさせるようにしたことを特徴とする映像中の有音区間の始終点の検出装置。

１６．音声信号と映像信号と～なる動画像信号を準備し、上記動画像信号を再生して上記音声信号と映像信号とを別々に入力し、上記音声信号の音声波形において無音区間の連続性に基づき有音区間の始点を検出し、上記有音区間のたち下がり点を終了点として検出し、上記有音区間の始点と終了点とにより示される区間に対応する上記映像信号の映像フレーム区間を抽出する有音区間の映像の始終点検出方法。

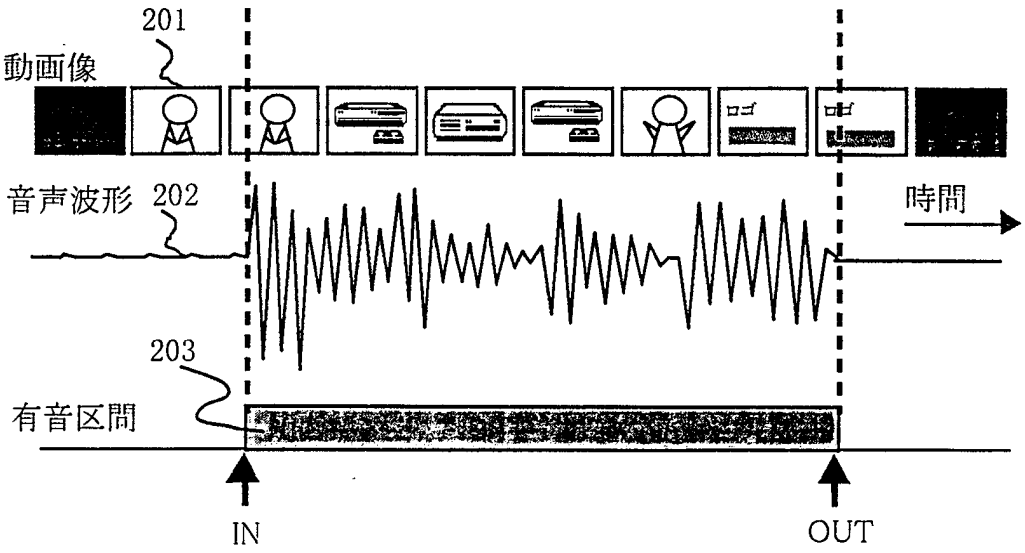
１７．請求の範囲第１６項において、上記映像信号から動画像を構成するフレームを所定時間間隔で時系列に表示し、上記フレームの時系列の表示に合わせて上記音声信号が示す音声波形と上記映像フレーム区間を表す表示バーとを視覚的に表示し、上記映像フレーム区間バーを表示上の時間軸上変形することにより、上記映像フレーム区間の始点又は終了点のフレーム番号を再設定することを特徴とする有音区間の映像の始終点検出方法。

１８．請求の範囲第１７項において、上記音声は刑の包絡線を算出し、予め設定された音響ﾚﾍﾞﾙの閾値と該包絡線が交差する時点を上記有音区間の始点又は終了点とすることを特徴とする有音区間の映像の始終点検出方法。

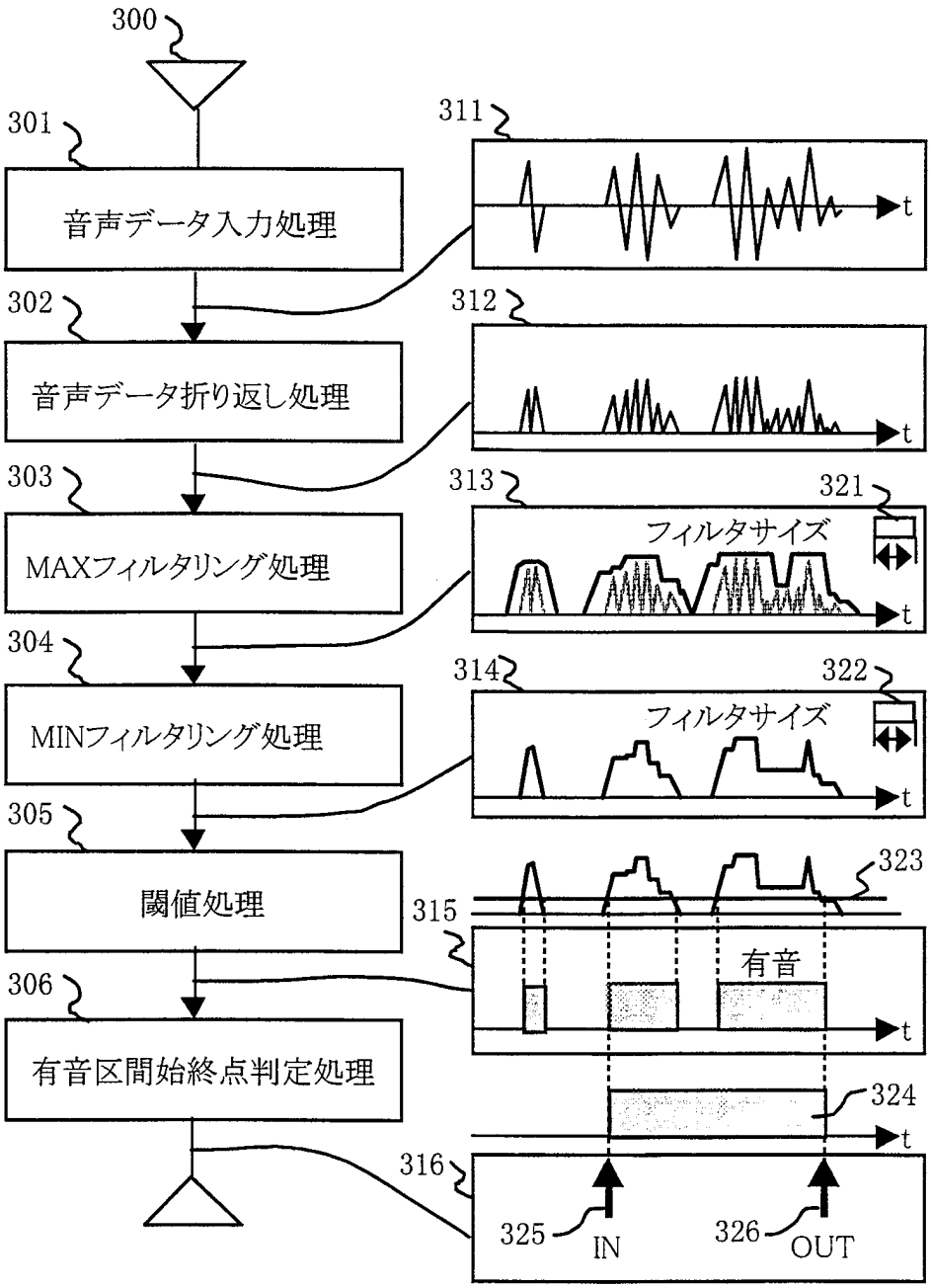
第1図


110
映像再生装置
104
音声処理装置
107
109
CPU
メモリ
I/F
102
101
ディスプレイ装置
音声入力装置
補助記憶装置
103
108
111
105
入力装置
106

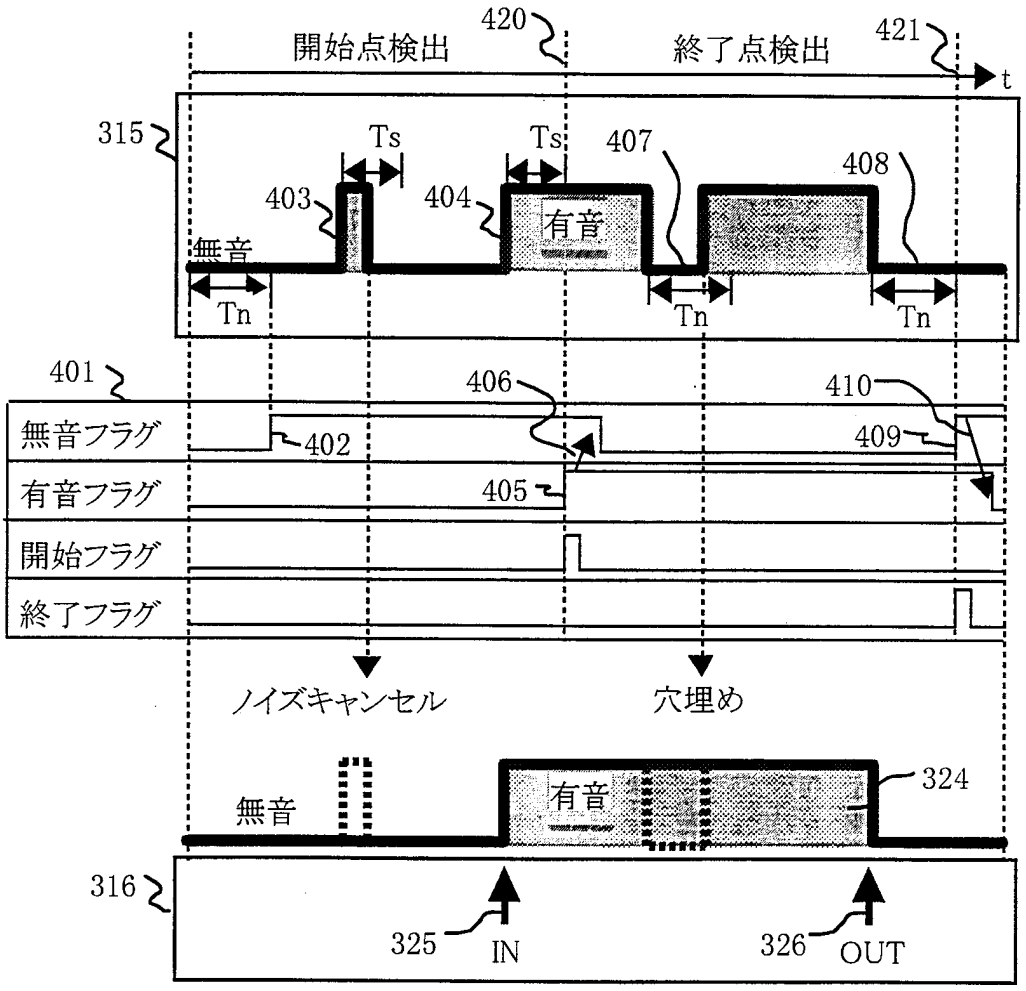
第2図


201
動画像
音声波形
202
時間
203
有音区間
IN
OUT

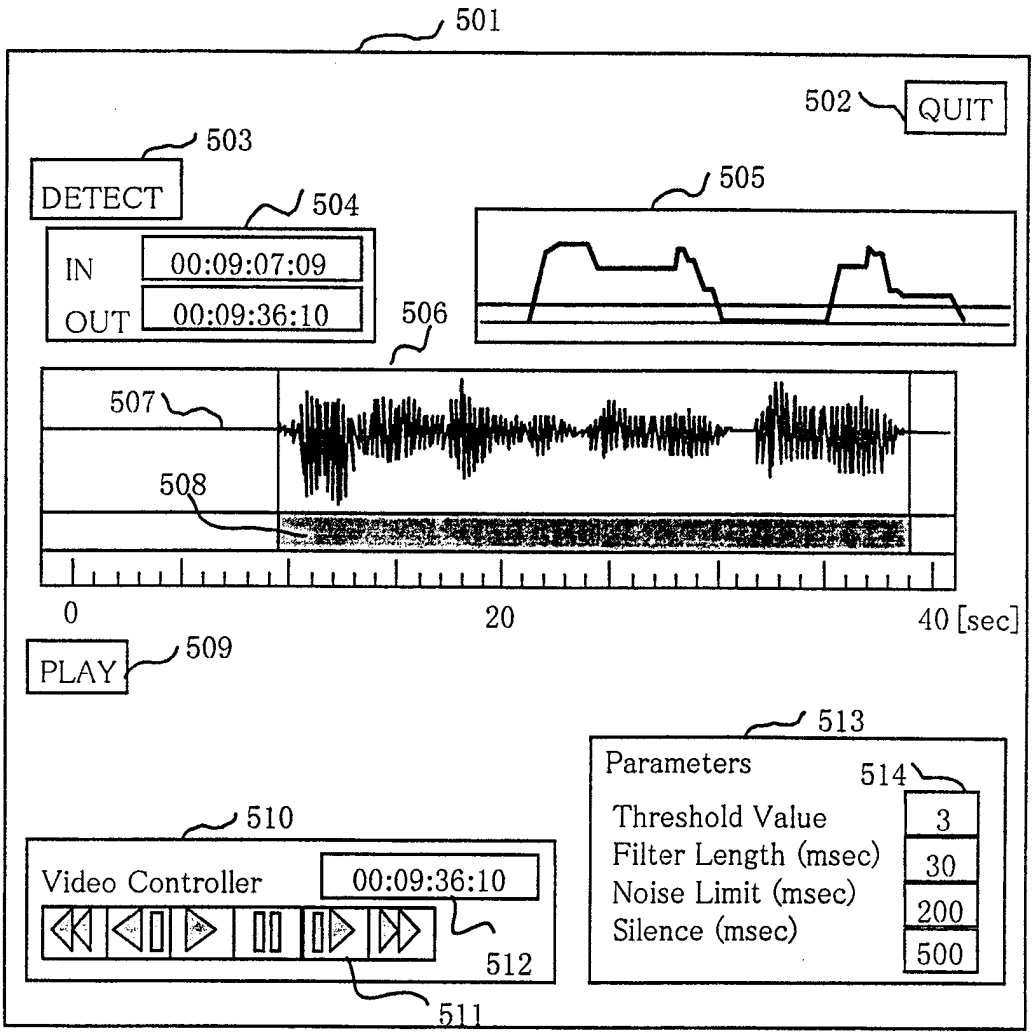
第3図


300
301
音声データ入力処理
302
音声データ折り返し処理
303
MAXフィルタリング処理
304
MINフィルタリング処理
305
閾値処理
306
有音区間始終点判定処理
311
312
313
321
フィルタサイズ
314
322
フィルタサイズ
323
315
有音
324
316
325
IN
326
OUT
t

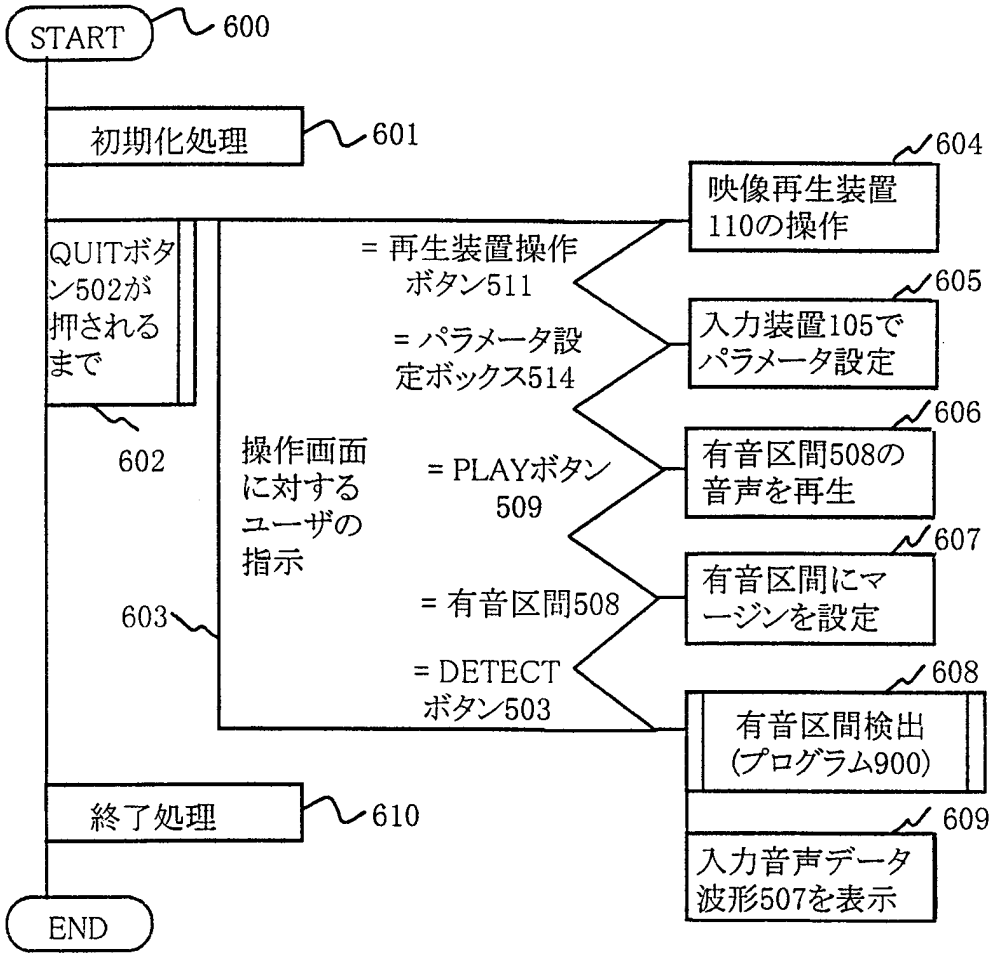
第4図


420
開始点検出
終了点検出
421
t
315
Ts
Ts
407
408
403
404
有音
無音
Tn
Tn
Tn
401
406
410
無音フラグ
402
409
有音フラグ
405
開始フラグ
終了フラグ
ノイズキャンセル
穴埋め
無音
有音
324
316
325
IN
326
OUT

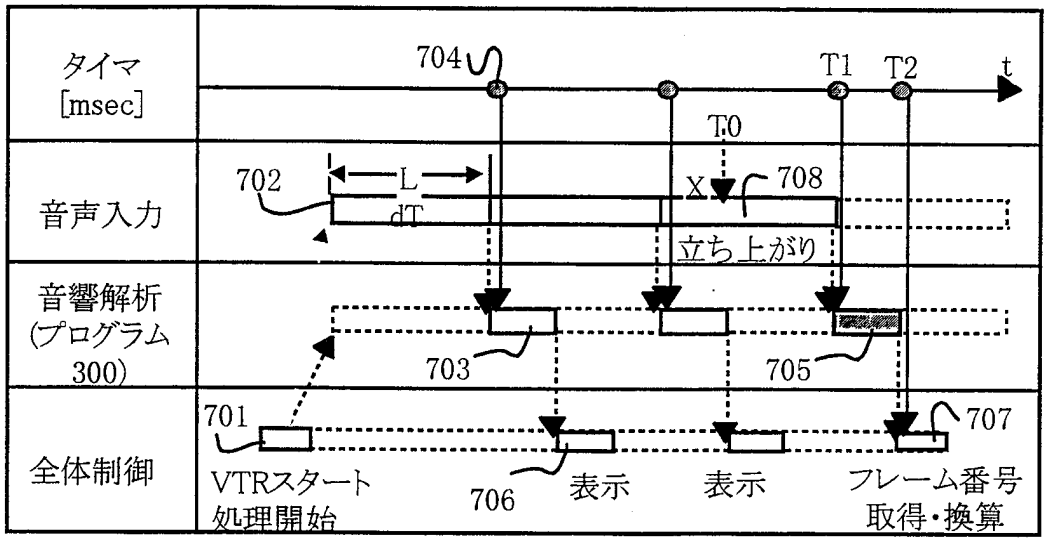
第5図


501
502
QUIT
503
DETECT
504
IN
00:09:07:09
OUT
00:09:36:10
505
506
507
508
0
20
40 [sec]
PLAY
509
513
Parameters
514
Threshold Value
Filter Length (msec)
Noise Limit (msec)
Silence (msec)
3
30
200
500
510
Video Controller
00:09:36:10
512
511

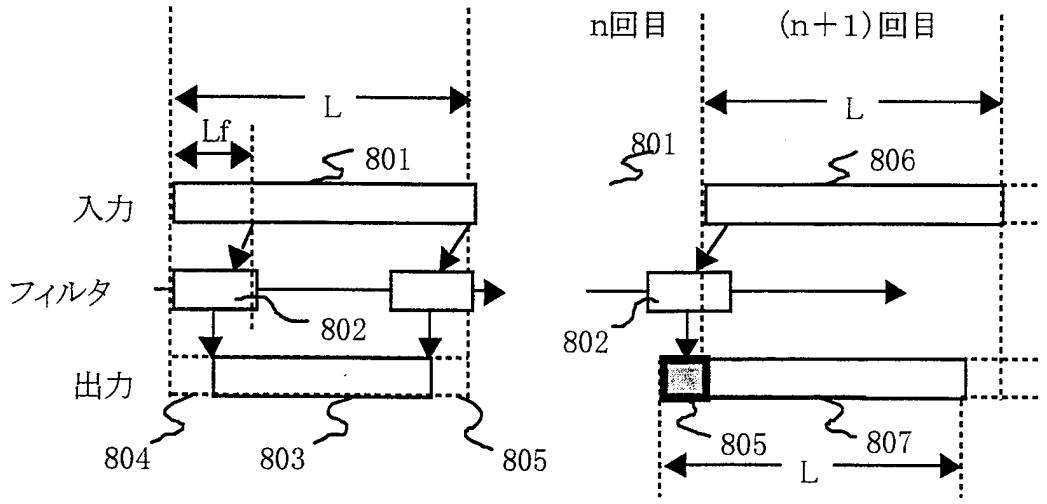
第6図


START
600
初期化処理
601
QUITボタン502が押されるまで
602
603
操作画面に対するユーザの指示
= 再生装置操作ボタン511
= パラメータ設定ボックス514
= PLAYボタン509
= 有音区間508
= DETECTボタン503
604
映像再生装置110の操作
605
入力装置105でパラメータ設定
606
有音区間508の音声を再生
607
有音区間にマージンを設定
608
有音区間検出(プログラム900)
609
入力音声データ波形507を表示
終了処理
610
END

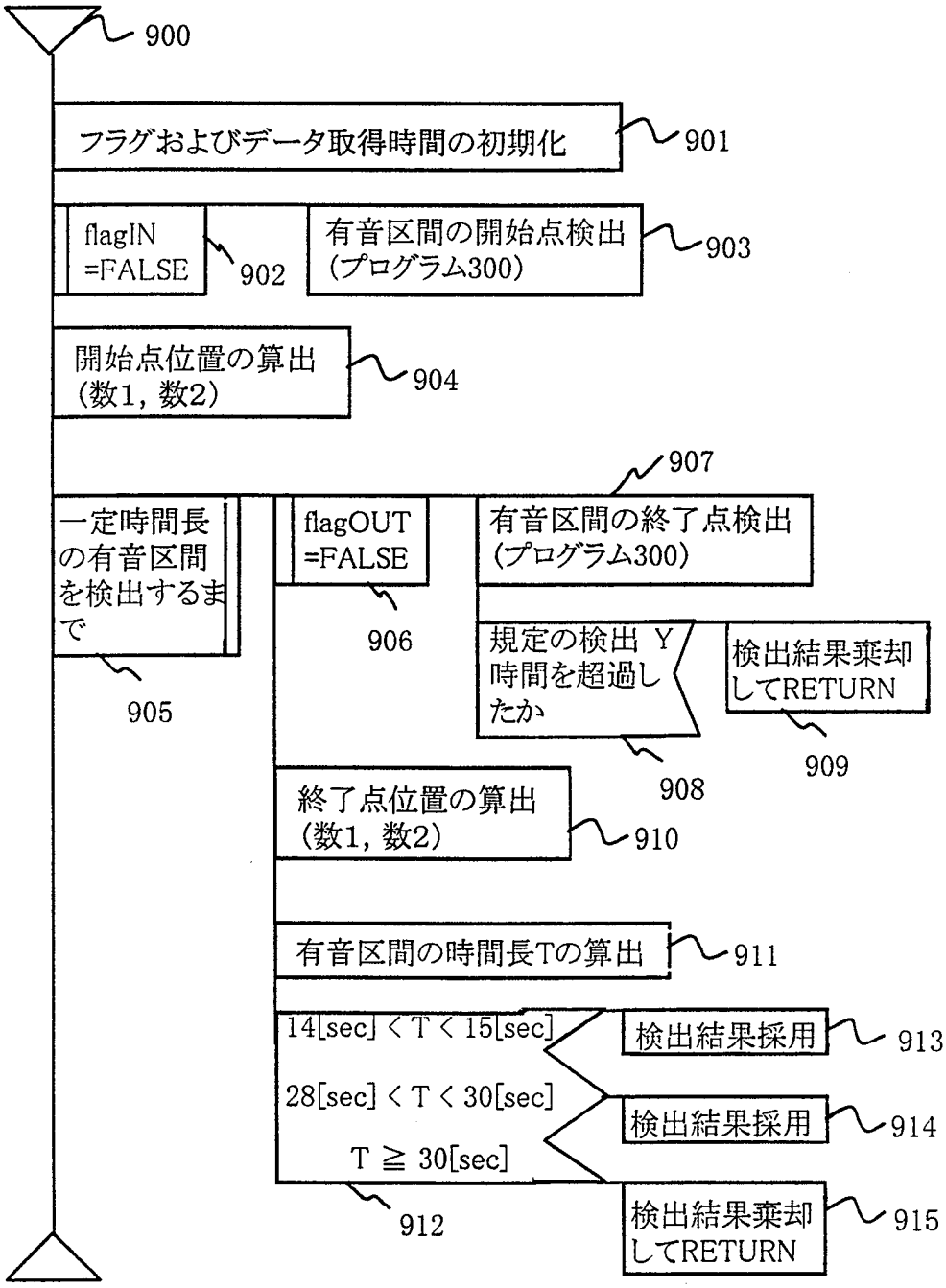
第7図


タイマ
[msec]
704
T1
T2
t
T0
音声入力
702
L
dT
X
708
立ち上がり
音響解析
(プログラム
300)
703
705
全体制御
701
707
VTRスタート
処理開始
706
表示
表示
フレーム番号
取得・換算

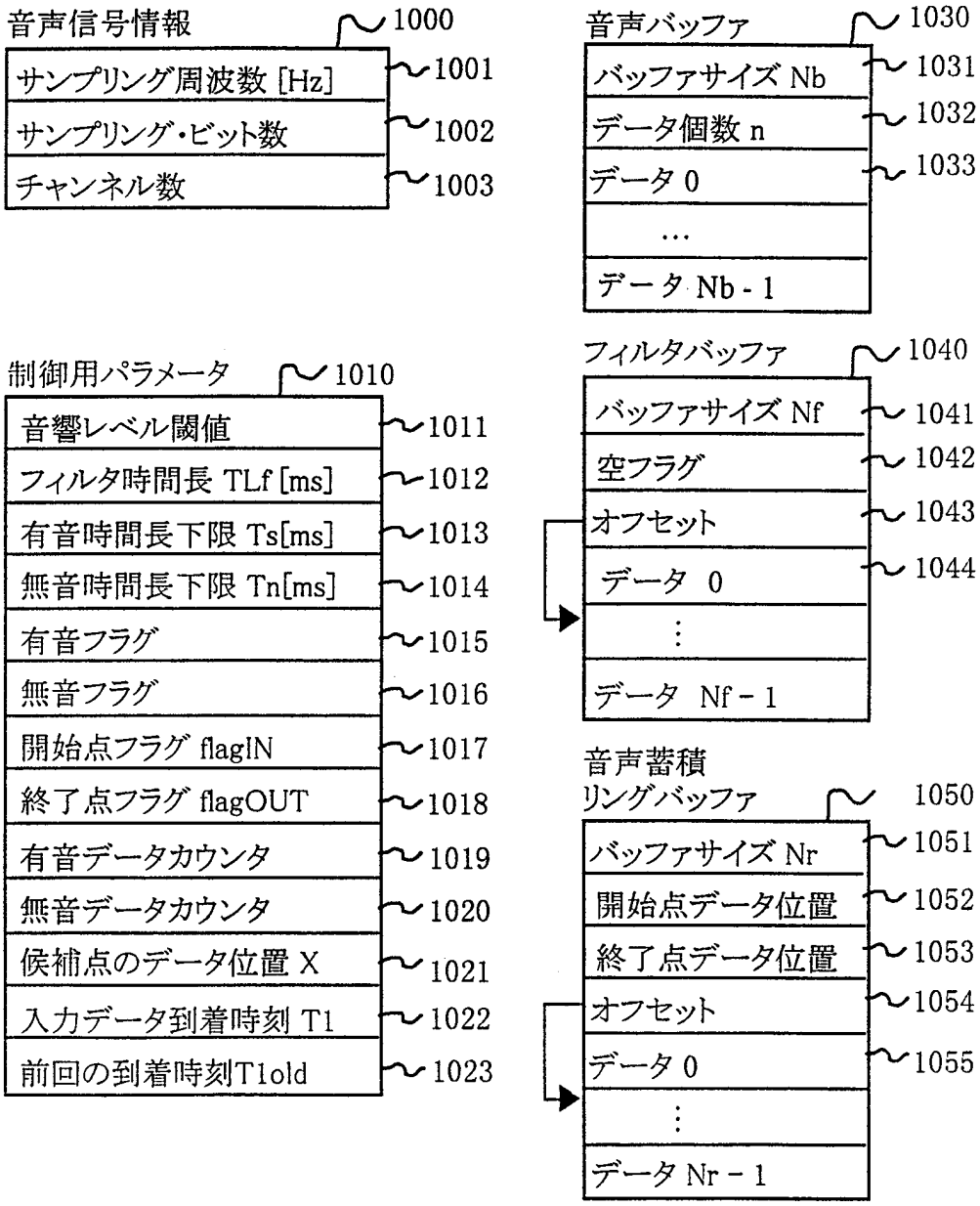
第8図


L
Lf
801
入力
フィルタ
802
出力
804
803
805
n回目
(n+1)回目
L
801
806
802
805
807
L

第9図


900
フラグおよびデータ取得時間の初期化
901
flagIN
=FALSE
902
有音区間の開始点検出
(プログラム300)
903
開始点位置の算出
(数1, 数2)
904
907
一定時間長
の有音区間
を検出するま
で
905
flagOUT
=FALSE
906
有音区間の終了点検出
(プログラム300)
規定の検出
時間を超過し
たか
Y
908
検出結果棄却
してRETURN
909
終了点位置の算出
(数1, 数2)
910
有音区間の時間長Tの算出
911
14[sec] < T < 15[sec]
28[sec] < T < 30[sec]
T ≧ 30[sec]
912
検出結果採用
913
検出結果採用
914
検出結果棄却
してRETURN
915

第10図


音声信号情報
1000
サンプリング周波数 [Hz]
1001
サンプリング・ビット数
1002
チャンネル数
1003
音声バッファ
1030
バッファサイズ Nb
1031
データ個数 n
1032
データ 0
1033
…
データ Nb - 1
制御用パラメータ
1010
音響レベル閾値
1011
フィルタ時間長 TLf [ms]
1012
有音時間長下限 Ts[ms]
1013
無音時間長下限 Tn[ms]
1014
有音フラグ
1015
無音フラグ
1016
開始点フラグ flagIN
1017
終了点フラグ flagOUT
1018
有音データカウンタ
1019
無音データカウンタ
1020
候補点のデータ位置 X
1021
入力データ到着時刻 T1
1022
前回の到着時刻T1old
1023
フィルタバッファ
1040
バッファサイズ Nf
1041
空フラグ
1042
オフセット
1043
データ 0
1044
⋮
データ Nf - 1
音声蓄積
リングバッファ
1050
バッファサイズ Nr
1051
開始点データ位置
1052
終了点データ位置
1053
オフセット
1054
データ 0
1055
⋮
データ Nr - 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP97/00905

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl^6 G10L9/00, H04N5/76, G11B20/10

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)

Int. Cl^6 G10L3/00, G10L9/00, H04N5/76, H04N5/91, G11B20/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1926 - 1997
Toroku Jitsuyo Shinan Koho 1994 - 1997

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP, 8-279962, A (NEC Engineering K.K.), October 22, 1996 (22. 10. 96)(Family: none)	16 1-4, 6, 7, 10-12, 17, 18
Y	JP, 60-29800, A (Toshiba Corp.), February 15, 1985 (15. 02. 85)(Family: none)	1-4, 6, 7, 10-12, 17, 18
Y	JP, 6-302160, A (Sony Corp.), October 28, 1994 (28. 10. 94) & EP, 625781, A	7, 10, 11, 12
Y	JP, 60-498, A (Casio Computer Co., Ltd.), January 5, 1985 (05. 01. 85)(Family: none)	6
Y	JP, 5-28717, A (Sony Corp.), February 5, 1993 (05. 02. 93)(Family: none)	11
Y	JP, 8-205076, A (Canon Inc.), August 9, 1996 (09. 08. 96)(Family: none)	17
A	Takafumi Miyatake, Hitoshi Matsushiama,	5

[X] Further documents are listed in the continuation of Box C. [ ] 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 document 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 another 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	Date of mailing of the international search report
June 17, 1997 (17. 06. 97)	July 1, 1997 (01. 07. 97)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP97/00905

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Masakazu Ejiri, "High-speed computing method for maximum/minimum image filtering (in Japanese)" The transaction of IEICE, Vol. J78 D-II, No. 11 November 25, 1995 (25. 11. 95), p. 1598-1607	

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ９７／００９０５

A.　発明の属する分野の分類（国際特許分類（ＩＰＣ））

IntCl[6]　Ｇ１０Ｌ９／００　Ｈ０４Ｎ５／７６　Ｇ１１Ｂ２０／１０

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

IntCl[6]　Ｇ１０Ｌ３／００　Ｇ１０Ｌ９／００　Ｈ０４Ｎ５／７６　Ｈ０４Ｎ５／９１　Ｇ１１Ｂ２０／１０

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報　　１９２６－１９９７年
日本国登録実用新案公報　　１９９４－１９９７年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
X Y	ＪＰ，８－２７９９６２，Ａ（日本電気エンジニアリング株式会社）２２．１０月１９９６（２２．１０．９６），ファミリー無し	１６ 1-4, 6, 7, 10-12, 17, 18
Y	ＪＰ，６０－２９８００，Ａ（株式会社東芝）１５．２月．１９８５（１５．０２．８５），ファミリー無し	1-4, 6, 7, 10-12, 17, 18
Y	ＪＰ，６－３０２１６０，Ａ（ソニー株式会社）２８．１０月．１９９４（２８．１０．９４），&ＥＰ，６２５７８１，Ａ	7, 10, 11, 12
Y	ＪＰ，６０－４９８，Ａ（カシオ計算機株式会社）５．１月．１９８５（０５．０１．８５），ファミリー無し	6

☒ Ｃ欄の続きにも文献が列挙されている。　　☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」先行文献ではあるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
１７．０６．９７	01.07.97

国際調査機関の名称及びあて先	特許庁審査官（権限のある職員）		
日本国特許庁（ＩＳＡ／ＪＰ） 郵便番号１００ 東京都千代田区霞が関三丁目４番３号	南　義明　印	５Ｈ	９３８１
	電話番号　０３－３５８１－１１０１　内線　３５３３		

様式ＰＣＴ／ＩＳＡ／２１０（第２ページ）（１９９２年７月）

Ｃ（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
Ｙ	ＪＰ，５－２８７１７，Ａ（ソニー株式会社）５．２月．１９９３（０５．０２．９３），ファミリー無し	１１
Ｙ	ＪＰ，８－２０５０７６，Ａ（キャノン株式会社）９．８月．１９９６（０９．０８．９６），ファミリー無し	１７
Ａ	宮武　孝文，松島　整，江尻　正員，最大最小型画像フィルタリングの高速演算手法 電子情報通信学会論文誌，　vol.Ｊ７８　Ｄ－Ⅱ　No.１１ ２５．１１月１９９５（２５．１１．９５），第１５９８－１６０７頁	５