2E−1

計算機上での音声データ表現の一検討

千葉　健司　上澤　功　三宅　英太
富士ゼロックス(株)システム・コミュニケーション研究所

1. はじめに

従来から文書は、情報伝達のための重要な役割を果たしている。また、近年の計算機技術の進歩により、音声などのリアルタイム情報の操作が計算機上で簡易に実現されつつある。従って、音声情報を計算機上で文書に統合することで、その表現力と文書作成能率の向上が期待される。

例えば、文書と音声波形を同一の視点で操作する試みがなされている[1]。これは、音声情報の有する構造を文書と同様に表示し、音声情報の検索・編集を効率化することを狙ったものである。しかしながら、文書操作環境の飛躍的な進歩を考えると、より高度な操作環境が望まれる。このためには、音声情報に含まれている意味内容や構造に基づく操作の実現が必要となる。

そこで、文書と音声情報の統合操作環境検討の始めとして、音声データの表現形式とそのデータ形成手法に着目する。本稿では、まず、対象とする音声情報モデルの検討から、波形属性を有するデータ表現形式を設計する。次に、実験により波形属性の抽出手法について考察し、その処理構造について検討する。

2. 音声データ形式

情報伝達を目的とした音声情報には、対話、ナレーション、キャプション等のシーンがある。所定の構造が与えられた音節の集合に、一連の意味付けがなされて文を構成する。文の構成要素を層順に文節、単語、音節とする。文毎の発声源が話者となる。無音区間で区切られた文の全体を音声情報と定義する。音声情報モデルを図1に示す。

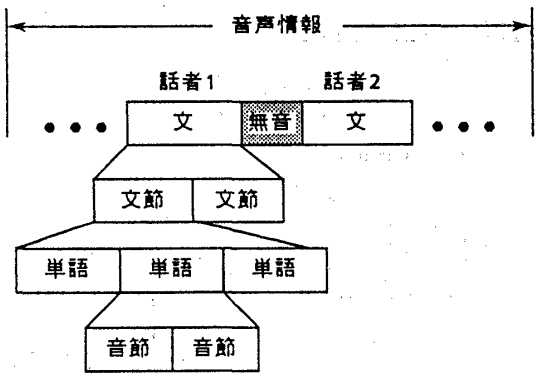


図1.　音声情報モデル

図において、文書との統合の際の基本単位を文とする。音声情報をセグメント化し、有音/無音セグメントの識別結果から、文を弁別する。弁別のため、音声データを形成するセグメントの特性を示す属性を与える。更に、文毎の話者等を表す特徴量を付加する。従って、セグメントの属性情報は、以下のようになる。

- 無音セグメント
- 有音セグメント
- 有音セグメント中のセグメント特徴量

図2に音声データの形式を示す。

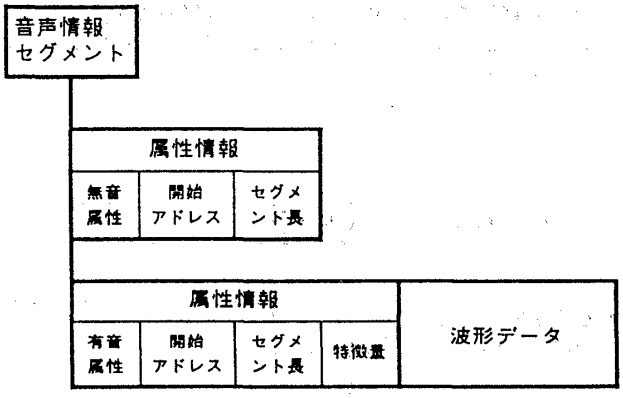


図2.　音声データの形式

3. 処理構成

2章で述べたセグメント特徴量は、有音/無音識別との整合を考えると波形特徴量といえる。一方、文書との対応付けの際には、図1に示す文節、単語、音節の弁別が必要となる。これらを弁別するためには、波形特徴量のみならず、周波数上の特徴量、更には、高度な学習手法を駆使する必要がある[2]。従って、文書統合化の処理手順は以下の2ステージに分割される。

- 前処理部：
  音声情報を波形特徴量からセグメント化し、音声データを形成する。
- 後処理部：
  音声データ中の属性に基づいて、有音セグメントの集合から、音声情報の詳細な構造を抽出する。

前処理部における波形特徴量が後処理のための重要な指標となることは、明らかである。よって、以後で波形特徴量

A Study of Speech Data Representation Scheme for Integrated Document Management Features
Takeshi CHIBA, Koh KAMIZAWA, Hidataka MIYAKE
Fuji Xerox Co., Ltd., System & Communication Lab.

とその抽出手法を実験から検討する。

4. 波形特徴量

［有音/無音の特徴量］

有音/無音を識別する波形特徴量として、音声情報のセグメンタルパワーが考えられる。この特徴量を閾値処理することで、有音/無音属性が判別できる。しかし、無音と判定されたセグメント中にも幾つか文に含まれるべき部分が存在する。このことは、文中に音声が中断する場合、ないし、パワーの小さい音韻が存在することと対応している。実験から、この部分を救済するために、セグメンタルパワーと零交差数を併せて求めることが有効であることが判った。識別結果の例を図3に示す。同図から、音声セグメント中で、*印で示した区間の音韻が救済されていることが判る。

［有音セグメントの特徴量］

セグメント内基本周波数は、有音セグメント特徴量の一つに挙げられる。基本周波数抽出は、セグメント内の自己相関係数から概略推定できることが知られている[2]。基本周波数が明確に異なる男声と女声の別は、自己相関係数のみで識別できると期待できる。同一文の男声と女声の比較結果を図4に示す。同図では、各々の平均周波数が明確に異なることが示されている。

5. 前処理の構成

4章で取り上げた波形特徴量は、属性に対応した処理構造を有する。従って、音声情報に含まれる有音セグメントの割り合いと、発声者の総数に応じて処理量が変動する。後処理に対しては、有音セグメントのみを送出すれば良く、全体の処理能率の向上と音声データの蓄積容量の低減が可能となる。図5に前処理の構成を示す。

6. おわりに

文書と音声データを統合化する音声データの表現形式とその形成手法について検討を行った。対象とした音声情報モデルから前後2ステージからなる処理手順を提案し、実験により波形属性の抽出手法を検討した。

後処理部を含めた処理系の検討と、新たな属性情報の検討が今後の課題である。

なお、実験では、(社)日本音響学会の連続音声データベースを利用している。

参考文献

[1] 林、他:視覚情報を利用した音声編集システム、NTTR&D,Vol.39 No.2,1990

[2] 古井:ディジタル音声処理、東海大学出版、1985

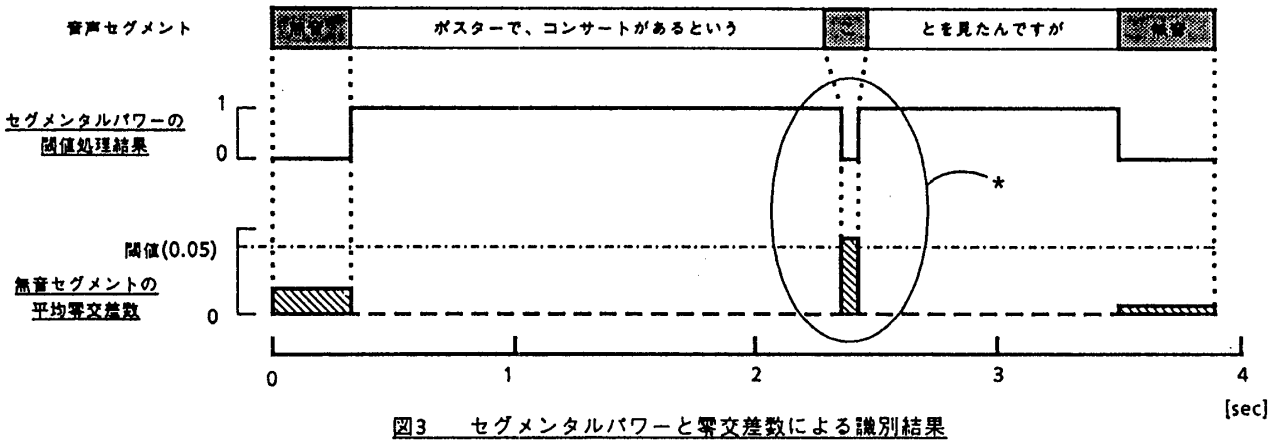
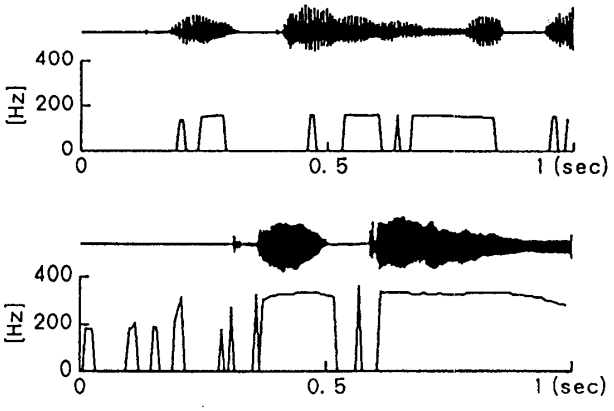


図3　セグメンタルパワーと零交差数による識別結果



上:男声　平均周波数155[Hz]　下:女声　平均周波数301[Hz]

図4　男声と女声の基本周波数の比較

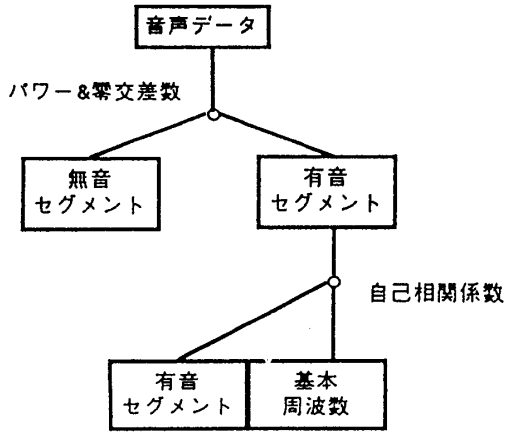


図5　前処理の構成