Voice Label Editor の開発

Development of Voice Label Editor

青木 直史(1)　須藤 健次(2)　伊藤 博之(3)　井口 勇(3)

Naofumi Aoki(1)　Kenji Sudoh(2)　Hiroyuki Itoh(3)　Isamu Iguchi(3)

(1)北海道大学大学院情報科学研究科　(2)（資）サイクル・オブ・フィフス　(3)クリプトン・フューチャー・メディア（株）

(1) Hokkaido University　(2) Cycle of 5th, Inc.　(3) Crypton Future Media, Inc.

1. はじめに

本研究では，音声合成用音声データベースの構築を支援するエディタ「Voice Label Editor（VLE）」の開発を行なっている[1],[2]. VLE は音声データに対して対話的に音素ラベルとピッチマークを付与するためのツールである．本発表では VLE の開発について述べる.

2. 音声合成用音声データベースの構築

録音編集方式による音声合成を実現するには専用の音声データベースが必要となる．こうした音声データベースを構築するには，音声データから音素ラベルとピッチマークという 2 種類のタグ情報を抽出することが求められる．音素ラベルは所望の音韻環境から波形データを取り出すために必要となる情報であり，ピッチマークは所望のイントネーションを実現するため有声音の音高を制御する際に必要となる情報である.

こうしたタグ情報の抽出を自動化できれば，音声データベースを構築する際のコストを低減できる．しかしながら，現状では完全な自動化は困難であり，計算機の支援の下，エディタを利用してマニュアルで作業を行なうことが一般的である[3].

3. Voice Label Editor の開発

VLE は音声データに対して音素ラベルとピッチマークを付与するためのツールである．図 1 に VLE の実行画面を示す．VLE は GUI アプリケーションであり，ユーザーは音声データやスペクトログラムを参照しながら対話的にタグ付けの作業を行なう.

VLE では，音素ラベルはセグメント層，ピッチマークはイベント層として，これら 2 種類のタグ情報を階層構造で表現している.

VLE にはケプストラムの時間変化率を表示する機能が実装されており，これを音素ラベルの決定に利用することができる．また，VLE には遮断周波数 1kHz の低域通過フィルタで処理した波形データのゼロクロス点をピッチマークの候補として提示する機能が実装されている．ユーザーは提示された候補を視察により修正することで，ピッチマークを決定することができる.

VLE ではこうしたタグ情報を WAVE ファイルのチャンクデータとし，音声データと一緒に記録することができる．また，データの可読性を高め，別のアプリケーションで利用し易くするため，タグ情報を XML 形式で出力することも可能である．図 2 に XML の DTD を示す.

4. まとめ

現在，実際に音声データベースを構築しながら，VLE のバージョンアップを行なっているところである．タグ情報の抽出をできる限り簡単に行なえるようにすることが今後の課題である.

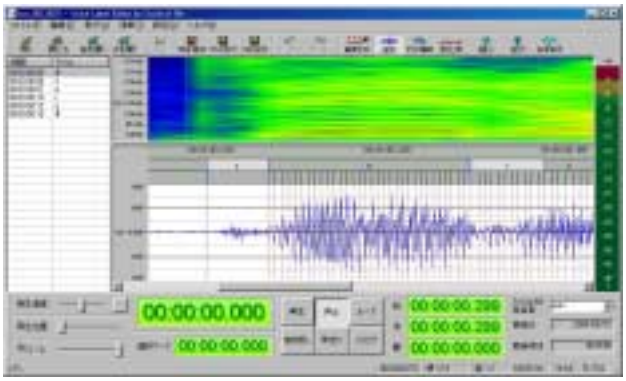

図 1. Voice Label Editor の実行画面

```
<?xml version="1.0"?>
<!ELEMENT speech (format,annotator,annotation)>
<!ELEMENT format (samplesPerSec,bitsPerSample)>
<!ELEMENT samplesPerSec (#PCDATA)>
<!ELEMENT bitsPerSample (#PCDATA)>
<!ELEMENT annotator (name,date)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT annotation (segment)*>
<!ELEMENT segment (event)*>
<!ATTLIST segment label CDATA #REQUIRED>
<!ATTLIST segment begin CDATA #REQUIRED>
<!ATTLIST segment end CDATA #REQUIRED>
<!ATTLIST event label CDATA #REQUIRED>
<!ATTLIST event begin CDATA #REQUIRED>
<!ATTLIST event end CDATA #REQUIRED>
```

図 2. XML によるタグ情報の出力

謝辞　本研究の一部は平成 16 年度文科省札幌 IT カロッツェリア創成プロジェクト研究費により行われた．ここに謝意を表する.

[1] 青木, “XML による音声データベースの構築,” 2003PC カンファレンス, 2003.

[2] 青木, 伊藤, 澤田, 須藤, “XML による音声データベースのオープンコ

ンテンツ化,” 信学総大, 2004.
[3] Dutoit, An Introduction to Text-to-Speech Synthesis, Kluwer Academic Publishers, 1997.