

ベトナム語の単語発音練習システムの開発

○グエン タン ミン, 村上 純 (熊本電波工業高等専門学校)

キーワード ベトナム語, 学習支援システム, 音声認識

1. はじめに

昨今の海外旅行ブームの影響で、東南アジアへの関心が高まっている。特にベトナムは、日本と同じく古くから中国の影響を受けており、外見や食生活、生活習慣など日本人と似通った点が多く^[1]、親しみやすいので、その料理や雑貨などが若い世代に流行している。ベトナムに何度も旅する人も多く、ベトナム語を学ぶ日本人も出てきている。

ベトナム語は、原則的に1語が1つの音節から成り立つ単音節言語である。発音には、日本語にない母音や声調（上がり調子、下がり調子等を示す）があって、ベトナム語を正確に、正しく発音することはかなり困難だと言われている^[2]。しかも、日本人のためのベトナム語教科書は甚少なく、練習用ソフトは見当たらないのが現状である。

そこで、ベトナム語を学ぶ日本人のために、正しい単語の発音を訓練するシステムを開発した。このシステムは、パソコンとマイクさえあれば、だれでも簡単に使用することができる。実際に数人に使用してもらい、システムの有効性を確かめた。

2. システムの設計

2.1. システム設計の概念

利用者が気軽に使用できるように、パソコンとマイクから構成される簡素なシステムとし、使いやすさを考慮した GUI(Graphical User Interface)環境を提供するようにした。

言語は Visual C++を使用してパソコン上でシステムを作成した。

2.3. システム構成

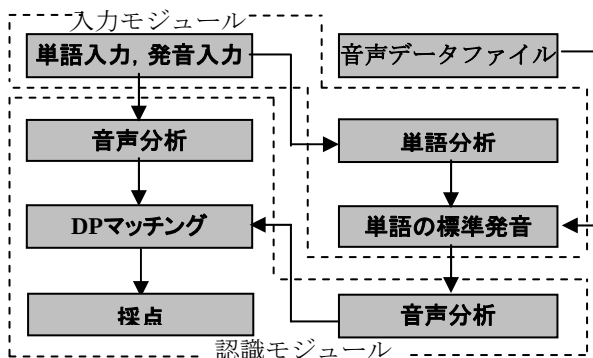


図1. 本システムの構成

本システムの構成は図1に示した通りである。利用者がベトナム語の単語をマイクから入力すると、システムはその発音と標準発音の比較を行い、採点結果を表示する。利用者は色々な単語を繰り返し発音練習することにより、その正確な発音がマスターできる。図2にユーザインターフェース画面を示した。



図2. ユーザインターフェース画面

3. システムの作成

3.1. 音声データファイルの作成

メモリの容量を節約するために、音声データファイルにベトナム語の全ての音素（数百個程度）の音声データを保存しておき、これらを合成して単語の発音データを作成することにした。音声データは22050Hzでサンプリングし、8ビットモノラルのWave形式で保存する。

具体的には、全ての音素の音声データを保存するファイルwordsound.datと、その音声データを参照するためのファイルwordpos.datを作成する。wordpos.datには、各音素ごとのデータファイルwordsound.datにおける先頭位置とサイズが保存される。

3.2. 入力モジュールの作成

3.2.1. 単語入力処理

ベトナム語では、声調を表現するために通常のアلفベットの母音に声調記号を付加して、例えばá, â, ãのように表記する。本システムでもこれらの文字の入力ができるようにした。

3.2.2. 発音入力処理

利用者は単語を入力したら、次にその単語の発音をマイクからシステムに入力する。繰り返し練習できるように、何回でも発音が録音できて、またその発音も聞くことができるようにした。

録音データは音声データファイルと同様にサンプリングして保存される。録音・再生のプログラミングはWindowsのAPI関数(Application Programming Interface Function)を利用して行った^[3]。

3.3. 標準発音再生モジュールの作成

標準発音の作成は図3のような手順で行う。利用者が、入力した単語の正しい発音を何度でも確かめられるように、標準発音は繰り返し再生するようにした。

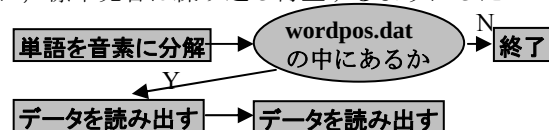


図3. 単語の標準発音の作成

3.3.1. 入力単語の音素分解

ベトナム語の単語は、1個または2個の音素に分解することができる。子音で始まる単語は2つの音素に分け

られ2音素単語に、母音で始まる単語はそのまま1音素単語となる。単語の先頭に現れる子音の組み合わせは26通りしかない。

以上の性質を利用して、26通りの子音の組み合わせの集合を用意し、単語を音素に分解した。

3.3.2. 標準発音データ作成

3.1で作成した音声データ参照ファイル wordpos.dat から、3.3.1で単語から得られた音素の音声データファイル wordsound.dat 上での先頭位置とデータサイズを読み出す。次に、wordsound.dat から音声データを読み取る。このようにして得られた音素のデータを、1音素の単語はそのまま、2音素の単語は2つの音素を繋ぎ合わせて単語の標準発音を作成する。

作成した標準音声は利用者が発音ボタンを押せば、何回でも再生するようにした。

3.4. 認識モジュールの作成

入力された発音が標準発音と比較してどの程度正しいか調べる部分であり、そのための手法としては音声波形のケプストラム^{[4],[5]}を求め、DP マッチング法^[4]によってパターン認識することにした。

3.4.1 入力発音の音声区間検出

比較を行うためには、まず入力発音の音声が存在する区間を検出する必要がある。音声区間を正確に検出するのは難しいが、ここではエネルギーと零交差回数をを用いて音声区間を検出することにした^[6]。

3.4.2. ケプストラム分析

ケプストラムとは音声波形のパワースペクトルの対数を波形とみなして逆フーリエ変換を行ったものである。ケプストラムは音声の特徴を表し、音声波形のスペクトル包絡線、音声の基本周波数などを求めることができる。

ケプストラムを求めるために、短時間スペクトル分析を行う。音声データから長さ20msのフレームを切り出し、切り出し位置を大きさ14msでシフトしながら、スペクトル分析を行う。ここで、値20msと14msは経験的に定めたものである。FFTを行う前に音声データには窓関数を乗じる。図4にケプストラム分析の手順を示す。

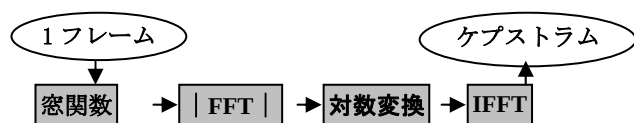


図4. ケプストラム分析の手順

3.4.3 DP マッチング法による距離計算

単語音声のケプストラム時系列は同じ人が発声したものであっても非線形の時間伸縮が起こる。このため、標準発音と入力発音のケプストラム時系列が対応するように時間軸を非線形に伸縮する時間正規化を行う必要がある。非線形の時間伸縮の方法としてDP マッチング(Dynamic Programming Matching)法がある。この方法を用いて標準発音と入力発音のケプストラムの間の距離を計算することにした。

得られた距離が小さいほど、利用者の発音が標準発音に近いことを表しているのので、得点が高くなるようにした。

4. システムの評価

実際に音声データファイルを作り、このシステムの有効性を確かめるために、評価実験を行った。

4.1 実験設備と評価方法

雑音の少ない静かな部屋で、次の条件で実験を行った。

実験設備： パソコン IBM Celeron 800MHz, 128MB

RAM, OS Windows 2000, マイク, スピーカー

実験方法： 被験者は単語を入力し、標準発音に習って5回ずつ発音練習をする。

4.2 評価結果と考察

1音素単語(o, a, ó)と2音素単語(na, ma, bó)の発音の採点結果を表1に示す。

表1. 実験結果の得点

被験者名	A		B		C	
単語 回数	o	na	a	ma	ó	bó
1回目	50	30	50	80	60	55
2回目	70	70	40	100	70	60
3回目	100	60	60	90	90	80
4回目	90	100	70	100	80	75
5回目	100	85	100	100	95	80
平均点	82	69	64	94	79	70
平均上達度	0.5	0.55	0.5	0.2	0.37	0.31

ここで、平均上達度は次式のように定義した。

$$\text{平均上達度} = \frac{1}{5} \cdot \frac{1}{\text{最高得点}} \cdot \sum_{k=1}^4 \{(i+1)\text{回目の得点} - i\text{回目の得点}\}$$

表1より、平均点では特に目立った傾向が見られなかったが、原因は被験者の息音が音声に混じることや、システムの認識精度の問題などの影響と考えられる。平均上達度はすべて正で、1回目に比べて5回目は点数が上昇し、かなり上達することが分かる。また、日本語にない母音の発音は難しく、上達度が低いことも分かった。

5. むすび

ベトナム語の学習支援を目的にシステムを開発した。評価実験の結果、どの被験者からも使いやすさについてはよい評価が得られた。発音練習の性能については、訓練を繰り返すとかなりよい発音に近づくことが確認された。しかし、認識精度や雑音の影響についてはまだ改良の余地が残されている。

また、ディスクの容量の節約のため、単語の標準発音作成は1音素ごとの音声データを合成することにより行った。これが実際の発音とは多少異なる点も問題であるが、ベトナム人にもこの標準発音が通じることから許容できると考えている。

文 献

- [1] 坪井善明, “ヴェトナムー豊かさへの夜明けー”, 岩波新書, 1994.
- [2] 宇根祥夫, ゴー・ミン・トゥイー, “語学王ベトナム語”, 三修社, 2000.
- [3] 田辺 義和, “Windows サウンドプログラミング”, 翔泳社, 2001.
- [4] 古井貞熙, “デジタル音声処理”, 東海大学出版会, 1985.
- [5] 今井聖, “音声認識”, 共立出版, 1995.
- [6] 新美康永, “音声認識”, 共立出版, 1979.

所属 : 熊本電波工業高等専門学校
氏名 : 村上 純
電話 : 096-242-
E-mail : jun@cs.knct.ac.jp