K-059

# HMMを利用した任意の音声データの検出
# Detection of Arbitrary Speech Data using HMM

佐藤 祐規† 高橋 亘† 早坂 昇‡ 宮永 喜一†
Yuki SATO Wataru TAKAHASHI Noboru HAYASAKA Yoshikazu MIYANAGA

## 1. まえがき

近年，音声認識技術を利用したアプリケーションが実現されてきているが[1],[2]，依然として音声認識の実用化範囲は限定されている．より一般的に音声認識技術を利用するため，また音声認識の精度を上げるためには数多くの学習データが必要である．しかし，従来のシステムでは手動で学習データを収集しているので，膨大な時間と人手が必要となり，実用化の範囲を限定する要因となっている．

本論文では，フレーズ音声認識において，コストを削減し実用化範囲を拡大するために，任意の音声データ検出を利用し学習データを半自動的に構築するアルゴリズムを提案する．初期段階として，僅かなサンプルデータのみを必要とし，その後も，人手がほぼ掛からないのでコスト削減を期待でき，効率よく良好な学習データの構築を期待出来る．

本文では，２章で提案手法の詳しい説明をし，３章で提案手法を使用して構築した学習データの性能の検証を行う．

## 2. 提案手法

サンプルデータから学習データを構築するプロセスと，音声データベースから単語データを取得するプロセスを経て，HMM を利用してデータベースから要求する単語音声データのみを自動で検出し，学習データを再構築する．また，再構築した学習データを利用し音声データを検出するプロセスを繰り返すことで，信頼性の高い学習データを構築する．

### 2.1 サンプルデータからの学習データ構築

提案手法は僅かのサンプルデータを元に大量のデータからなる学習データを自動構築することが目的である．まず，サンプルデータから学習データを構築する．この時必要なデータ数は 3 人 3 回発声の 9 発話程度を想定する．音声特徴量として MFCC ケプストラムを抽出し，ランニングスペクトルフィルタリング（RSF）を用いて雑音を除去，HMM を用いて学習データを構築する．

RSF はケプストラムのフレーム時間方向に FIR 型バンドパスフィルタを適用して乗法性雑音を除去し，音声部分を強調することで認識精度を向上させる手法である．[3]

### 2.2 単語区間検出を用いての単語データ取得

あらかじめ用意した音声データベース上から，単語区間検出を利用して単語データを抽出する．

†北海道大学大学院情報科学研究科
Graduate School of Information Science and Technology, Hokkaido University
‡株式会社レイトロン RayTron,INC

単語区間検出は音声の振幅をもとに音声の始端と終端を推定する．非音声区間を検出しないよう推定区間の長さや分散値によって適当でない推定区間を排除し，単語音声データとして取得する．[4]

### 2.3 任意のデータの検出、学習データ再構築

観測する単語データごとに HMM を利用して尤度計算をする．サンプルデータと同じ単語発声音声であればある程度高い尤度を持つので，条件を満たす単語データのみを取得していく．

尤度の高さは人によって異なるので，絶対的な尤度の高さで検出の判断をすると偏りのある言語モデルになる．そのような状況を防ぐため，作成したい学習データ以外に別の単語の学習データをいくつか用意する．作成したい学習データの尤度が最も高く，かつ他のモデルの尤度よりも一定以上高い値を持つ場合のみ同じ単語だと判断する．このような条件を満たす単語を自動で検出し，データを取得する．

しかし，本システムのように，非常に少ない学習データから HMM を構築し，そこから上記のような手順で必要なデータ検出を行う場合，得られる尤度の精度が十分ではなく，単語の検出にも限界がある。そのため，要求したデータのみ検出できるとは限らない。そこで，データベースの検索終了後，検出した単語から要求していない単語を手動で取り除く必要がある．

その後，検出した単語データとサンプルデータから再び学習データを構築する．この新たな学習データを追加し，再度 HMM を学習し，その HMM を用いて任意の音声単語データを検出する。このプロセスを繰り返すことで，僅かなサンプルデータから半自動的に学習データを構築する．

## 3. 性能の検証

### 3.1 実験条件

提案法により構築された学習データの性能を検証するために，男性 40 人 4 回発声，孤立単語 142 語のデータベースから提案法で構築した学習データの単語検出率と，未学習の男性 30 人に対する認識実験を行った．

表１　学習データ構築条件

| | |
|---|---|
| サンプリング | 11.025kHz 16bit 量子化 |
| フレーム長 | 23.2ms(256point) |
| フレームシフト | 11.6ms(128point) |
| 窓関数 | ハニング窓 |
| データベース | 男性 40 人，22566 発話 |
| 初期学習条件 | 男性 3 人，9 発話 |
| 再学習試行回数 | 3 回 |
| 使用学習単語モデル | 10 種類 |

データベースには目的単語データが計 160 個あり，いくつを検出したかで検出率を測定する．

また認識実験では未学習の男性 30 人が各々1 回発声した分を用いて学習回数ごとの認識率を測定した．今回は５つの単語に対して実験を行った．

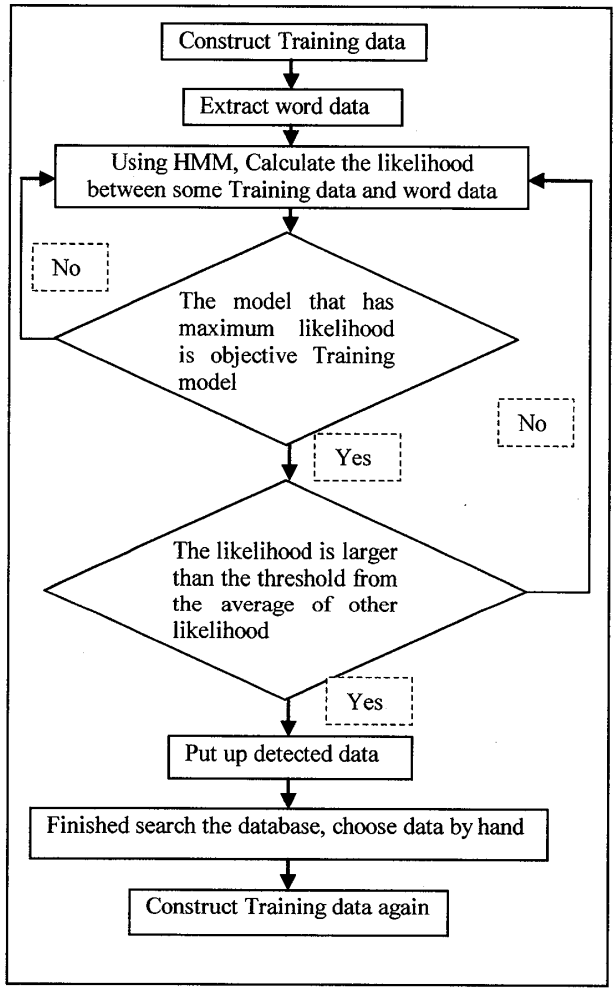


図１　学習データ構築の流れ

### 3.2 実験結果

単語検出率の実験結果を表 2 に示す．左側の数字は検出単語数であり，右側は全 160 個中の検出率である．初期学習では 3 人 3 回発声の 9 個を使用しており，検出率は全て 5.6%である．3 回ほど再学習を繰り返すと平均して 80%以上の単語データを検出でき，さらに再学習ごとに検出率が上がっている．

また表 3 に単語認識率結果を示す．従来法とは，男性 40 人 4 回発声の計 160 個の単語データを用いた学習データであり，その他の認識条件は全て提案手法と同様である．認識率も再学習を繰り返すごとに精度が高まっているのが見られる．3 回の再学習を行うと，従来の手法で構築した学習データとほぼ同等の認識率を得る．

これら 2 種類の実験より，再学習を繰り返すほど任意の音声単語データを検出でき，それにより高い認識率を持つ信頼性の高い学習データを構築できることが確かめられた．

表 2　単語検出率

| 単語 | 再学習 1 回 | 再学習 2 回 | 再学習 3 回 |
|---|---|---|---|
| ｹﾞﾝｷ | 51 31.8% | 134 83.7% | 137 85.6% |
| ｱﾘｶﾞﾄｳ | 81 50.6% | 145 90.6% | 153 95.6% |
| ｵﾝﾄﾞ | 43 26.8% | 121 75.6% | 129 80.6% |
| ｾﾞﾛ | 23 14.3% | 76 47.5% | 121 75.6% |
| ｺｳﾂｳ | 37 23.1% | 85 53.1% | 127 79.3% |

表 3　単語認識率

| 単語 | 未再学習 | 再学習 1 回 | 2 回 | 3 回 | 従来法 |
|---|---|---|---|---|---|
| ｹﾞﾝｷ | 55.0 | 91.0 | 95.0 | 95.0 | 95.0 |
| ｱﾘｶﾞﾄｳ | 78.0 | 99.0 | 100 | 100 | 100 |
| ｵﾝﾄﾞ | 48.0 | 93.0 | 99.0 | 99.0 | 99.0 |
| ｾﾞﾛ | 23.0 | 56.0 | 80.0 | 91.0 | 94.0 |
| ｺｳﾂｳ | 88.0 | 100 | 100 | 100 | 100 |
| Average | 58.4 | 87.0 | 94.8 | 97.0 | 97.6 |

## 4. むすび

本論文では，データベースから任意の音声単語データを検出し，再学習を繰り返して，学習データを半自動で構築する手法を提案した．HMM を利用して構築した学習単語モデルと観測する単語音声データ間で尤度計算を行い，条件を満たす単語データのみを検出して再学習を繰り返し信頼性の高い学習データを構築する．今回の実験では，従来の手動構築の学習データとほぼ変わらない認識率を示すことを確認した．

### 謝辞

本研究を行うに当たり多大なるご指導，ご助言を下さいました，吉澤真吾助教に心から感謝いたします．

### 参考文献

[1]鹿野清弘 他,『音声認識システム』, オーム出版局, 2000.
[2]P.C.Woodland,C.J.Leggetter, J.Odell, V.Valtchev, S.J.Young. The 1994 HTK Large Vocabulary Speec Recognition System. In IEEE Int'l Conf. on Acoustics, Speech & Signal Processing (ICASSP), Vol.1, pp. 73-76, 1995.
[3]早坂 昇, 和田 直哉, 宮永 喜一, 畑岡 信夫, "ランニングスペクトルフィルタを用いた雑音にロバストな音声認識", 信学会, 信学技報, CAS2003-6, pp31-36, 2003.
[4]高橋 亘, 大貫 和永, 吉澤 真吾, 宮永 喜一, "RSF を用いた雑音ロバスト音声区間検出の一考察", 信学会, 信学技報, SP2007-55, pp59-64, 2007.