

# 音声認識におけるノイズ重畳モデルを用いたオンライン適応 On Line Mixture Weight Adaptation for Noisy Acoustic Model

政田 悠佑†  
Yusuke Masada

山本 博史‡  
Hirofumi Yamamoto

## 1. はじめに

現在、音声認識システムは普及しており、携帯電話など様々な場面で利用されている。しかし、音声認識システムが利用されている場面では、ノイズが混入することがほとんどであり、音声認識システムを利用するためには、ノイズ対策が必須となる。このノイズには大きく分けて時間経過により変化しない定常雑音と変化する非定常雑音の2種類がある。ノイズ対策も大きく分けて音声から直接的にノイズを除去する方法と音響モデルでノイズに対応する2つの方法がとられており、後者の一手法としてノイズ重畳モデルがあげられる。このモデルは予めモデルにノイズを重畳することにより、ノイズ対策を行うので、定常雑音に有効な手法の1つである。しかし、重畳すべきノイズは一般に既知ではないため、様々なノイズを重畳したモデルを複数用意しておき、それらのモデルを適切な混合比で混ぜることにより、未知のノイズに対する重畳モデルを作成する。

しかしながら、定常雑音といえども時間変化に従いその性格が変化して行くことが考えられる。そこで、本稿では、この定常雑音の時間変化に追従することを目的とし、ノイズ重畳モデルにおける混合比をリアルタイムで更新することを試みる。

## 2. 研究内容

### 2.1 ノイズ重畳モデル

ノイズ対策としては、ノイズを直接的に除去する方法の他に、ノイズを含む音声に対応した音響モデルを用いる方法があり、本研究では後者の方法を用いてノイズ対策を行う。音響モデルとは英語の発音記号にあたる音素と、それに対応する音声の特徴量の関係を統計量で表現したもので、統計をとった時の音声の特徴量と認識をする時の音声の特徴量を比較することにより、音素を推定する。従って、統計をとった時と認識をする時の条件が同一であることが必要となる。しかし、認識中にノイズが混入した場合、ノイズの混入していない音響モデルとノイズの混入した音声の比較となり、条件が同一でないため、正しく認識されない可能性がある。そこで、音響モデルを作成する際に予め特定のノイズ $\alpha$ を学習データに重畳させることにより、ノイズ $\alpha$ に対しては頑健な音響モデルを作成することが可能である。しかし、実環境においては様々なノイズが発生するため、それら全てに対して対策を行うことは困難である。そこで、いくつかの代表的なノイズに対して対策を行うことで、その他のノイズに対しても有効な音響モデルを作成することができると考えられるので、予め代表的なノイズに対応した音響モデルを作成し、それらを混合することに

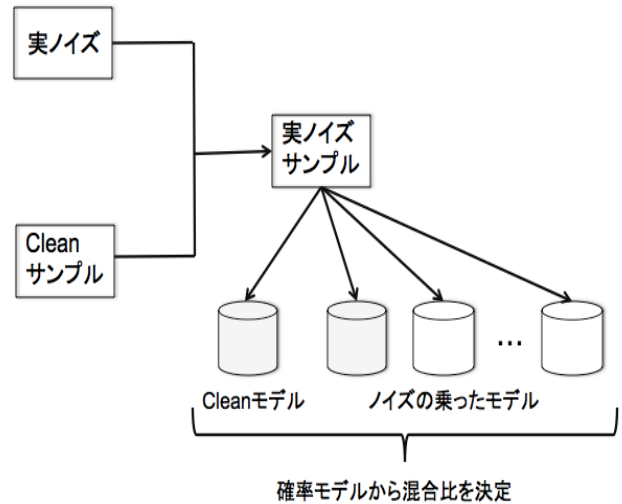


図1: 混合比の決定方法の概要

より新たに音響モデルを作成する。その際に、ノイズの混入していないクリーンな音響モデルを作成し、混合することにより、ノイズの発生していない状況にも対応できる音響モデルが作成できるため、クリーンモデルも混合の際に使用する。これにより作成されたモデルがノイズ重畳モデルである。

### 2.2 混合比の決定方法

ノイズ重畳モデルは、 $n$ 種類のノイズに対応した $n$ 個の音響モデルを作成しておき、それらを混合することにより作成される。従って、各モデルをどの程度混合するかというモデルの混合比がノイズ重畳モデルの性能に大きく関わっている。そこで、ノイズ重畳モデルでは、混合比の決定方法が重要となる。混合比の決定手法は様々な手法が考えられるが、本研究では確率モデルを用いて混合比を動的に決定する。実ノイズ入りの音声サンプルを $n$ 種類のノイズ対応音響モデルで認識させ、そこで得られた認識結果の確率値から混合比を動的に決定することとする。つまり、 $i$ 番目のモデルの混合比 $\alpha_i$ をモデル $M_i$ の選択確率 $P(M_i)$ とすると、あるノイズの入った音声 $A$ をモデル $M_i$ を用いて音声認識を行った時の確率値は $P(W|M_i, A)$ で表される。ここで、十分大量な $W, A$ を用いれば、 $P(M_i)$ は $P(W|M_i, A)$ に比例するため、モデル $M_i$ の混合比 $\alpha_i$ は次式で表される。

$$\alpha_i = \frac{P(M_i)}{\sum_{i=1}^n P(M_i)} = \frac{P(W|M_i, A)}{\sum_{i=1}^n P(W|M_i, A)} \quad (1)$$

となるので、(1)式を用いて、混合比 $\alpha_i$ を決定する。

†近畿大学大学院総合理工学研究科, Graduate School of Science and Engineering, Kinki University

‡近畿大学理工学部, Department of Science and Engineering, Kinki University

## 2.3 混合比の動的決定方法

前節で与えられたノイズに対する混合比の決定方法を述べたが、実環境では認識を行う直前までノイズが未知である場合も多い。そこで、次に混合比を動的に計算する方法について述べる。図 1 に示すように、得られた実ノイズを予め用意しておいた発声内容が既知のクリーン音声に重畳する。これによって、実ノイズ入りの音声を得ることができる。次にこの実ノイズ入りの音声を  $n$  種類のノイズ対応音響モデルを用いて、それぞれ認識を行う。その結果、各音響モデルに対する認識尤度が得られ、これから確率値の比率を得ることができ、(1)式に基づき混合比を動的に計算することができる。

## 2.4 混合比のリアルタイム更新

前節の動的決定方法を用いることにより、逐次的に混合比を計算していくことにより、ノイズ重畳モデルを更新することができる。これにより、時間変化に従いノイズが変化する場合においても、時間ごとに発生しているノイズを取得しサンプルを作成することで、そのノイズに適応したノイズ重畳モデルを作成することができる。

## 3. 評価実験

### 3.1 実験に用いるデータ

以下のデータを用いて実験を行う。

- 代表的なノイズ 20 種類  
ボイラー室の音、バスターミナルの音、クーラーの音、公園の音、一般道の音、高速道路の音、建設現場の音、デパート内の音、お祭りの音、飛行機内の音、バス内の音、電車内の音、駅のホームの音、田んぼ付近の音、駅のコンコースの音、駅前音、木々の揺れる音  
以上の 17 種類のノイズを学習データに混入させ、音響モデルを作成する。  
駅の改札付近の音、掃除機の音、空港のロビーの音  
以上の 3 種類のノイズを評価用音声に混入させ、評価実験時に用いる。
- 音響モデル作成時の学習データ  
サンプリングレート 16kHz  
男性 57 名、女性 40 名、計 97 名  
発話数各話者 100 発話  
計 9700 発話
- 評価データ  
学習データに含まれない男女各 100 発話

### 3.2 実験手順

以下の実験手順により評価実験を行う。

- ノイズの混入した音声データを用いて 17 種類の音響モデル、ノイズの混入していないクリーンモデルをそれぞれ作成する。
- 作成した音響モデルに対し、実ノイズを混入した実ノイズサンプルを用いて音声認識を行う。
- 音声認識の結果より得られた確率値から、(1)式を用いて混合比  $\alpha_i$  を求める。
- 3 で得られた混合比  $\alpha_i$  に基づいてノイズ重畳モデルを更新する。
- 4 で更新したノイズ重畳モデルを用いて、3 種類の評価用音声で音声認識を行う。

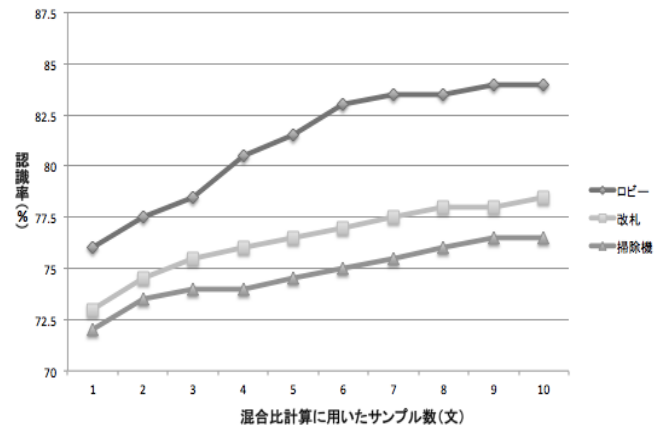


図 2: 実験結果

- 3 で使用する実ノイズサンプル数を変更しながら 3~5 を繰り返し行い、評価する。

## 4. 実験結果

評価実験により得られた、サンプル数と認識率の関係を図 2 に示す。

なお、評価実験は SN 比 10 で行い、認識率は、(認識率) = (正解文数)/(総文数)で計算を行っている。

図 2 より、サンプル数の増加に従い、認識率の向上が見られる。これにより、ノイズ重畳モデルが各ノイズに対して、適応していくのが分かる。

## 5. 考察

実験結果より用いる実ノイズサンプル数により、音声認識率の変化が見られ、サンプル数が増えるにつれて認識率が向上している。また、サンプル数が 8 文程度で、認識率の変化が少なくなるため、8 文程度でノイズ重畳モデルが各ノイズに対して適応したと考えられる。

## 6. 今後の課題

混合比をリアルタイムで変更することにより、ノイズ重畳モデルを適応させることができたが、モデル数やノイズパターンによって、適応するまでに違いがあるか、また、モデル数と適切な更新間隔の関係などが今後の課題となる。

## 7. むすび

ノイズ重畳モデルにおいて、混合比をリアルタイムで変更することにより、ノイズに適応でき、ノイズが変化した場合にも高い認識率となるノイズ重畳モデルを作成することができると考えられる。これにより、時間変化に従いノイズが変化する場所での音声認識システム利用へ貢献できるものと考えられる。

## 参考文献

- 鹿野清宏、伊藤克亘、河原達也、武田一哉、山本幹雄：音声認識システム、オーム社(2001)
- 中村哲：実音響環境に頑健な音声認識を目指して、信学技報(SP2002-12)
- 杵村昭宏、山本博史：ノイズ重畳モデルをベースとした音響モデルのノイズ適応、情報処理学会関西支部支部大会講演論文集(2011)