# 多元的音情報に基づく頑健な音声認識に関する研究

○林升柯，西田昌史，西村雅史（静岡大・情）

## 1　はじめに

雑音に頑健な音声認識システムの先行研究として、接話マイクと咽喉マイクでそれぞれ得られた結果を雑音の大きさに基づいて統合する音声認識システムが提案されている[1]。しかし接話マイクは装着負荷が大きい。また咽喉マイクは環境雑音に対して頑健であるが、接話マイクに比べ帯域が狭く一般的な音響モデルとミスマッチを起こすという問題点がある。

本研究では接話マイクの代わりに装着負荷の少ないピンマイクを咽喉マイク上に固定することで併用し、発話衝突時に頑健な音声認識手法について検討を行ったので報告する。

## 2　音声認識システム

### 2.1 提案方法

ピンマイクと咽喉マイクの 2 つのマイク入力に対してそれぞれ認識を行い、単語信頼度[2]に基づいて認識結果を選択する。なお、咽喉マイク側の音声については音響モデルとのミスマッチを解消するために、 SPLICE を使用して入力特徴量の変換[3]を行い、さらに MLLR によって音響モデルを咽喉マイクの特性に適応させる。

### 2.2 SPLICE

SPLICE は変換元の音声の特徴量を GMM でモデル化し、GMM のコンポーネント毎に線形変換を行うことで特徴量の変換を行う。学習は対応する 2 チャンネルの特徴量系列から行う。ここでは変換元を咽喉マイクの音声とし、変換先を口元で収録した通常の音声とした。評価対象話者とは異なる 6 名の計 400 文の 2 チャンネルで収録した音声を用いて学習した。

### 2.3 MLLR

一般的な音響モデルの咽喉マイクの特性への適応を目的として MLLR を利用した。学習データとしては評価対象話者とは異なる 5 名の話者の計 5 文の収録音声を用い、教師付き学習を行った。

### 2.4 単語信頼度による結果選択

2 つのマイクの認識結果より、単語の始端と終端のフレームの差がそれぞれ 5 フレーム以内の場合は対応する単語とみなし、単語信頼度のより高い方を認識結果として選択する。それより大きい場合は終端のフレームの差が 5 フレーム以内になるまで後続の単語を探索し、単語信頼度の平均値を比較する。また発話単位で信頼度の平均値に差があれば、信頼度の誤差の影響を軽減するため、発話単位での認識結果の選択を行う。

## 3　評価実験

### 3.1 実験条件

SPLICE と MLLR による咽喉マイクの認識性能の改善と、単語信頼度による認識結果の選択の性能評価を行うため、7 名の被験者の計 700 回の連続数字発声を対象として認識実験を行った。

音声認識は Julius で行い、音響モデルは Julius 付属の ASJ-JNAS コーパス(86 時間)から作成されたものを用いた。特徴量は 12 次 MFCC とパワー、それらのデルタパラメータを合わせた計 26 次元を SPLICE に用い、音声認識時にはデルタパワーを除いた計 25 次元を用いた。テストデータとしてはクリーンな環境で収録した各発話に対して、他人の発話を適宜重畳することで発話衝突を模擬したものを用いた。結果的に全発話フレームの 40%程度で発話が重畳している。またピンマイクと咽喉マイクは特性の違いにより、同じ雑音環境下で録音される雑音の入力レベルが異なる。そこで、実際に 2 つのマイクを装着した状況で雑音を収録し、ノイズレベルを測定した結果に基づいて各チャンネルの SNR を調整している。

---

* A Study on Robust Speech Recognition Using Multi-Channel Information, by LIN, Shengke and NISHIDA, Masafumi and NISHIMURA, Masafumi (Shizuoka University).

### 3.2 実験結果

表 1 に各チャンネルデータに対する SNR 別の認識結果を示す。ピンマイクにおいては MLLR によるマイク特性の適応化の効果は見られなかった。咽喉マイクにおいては外部ノイズの影響はかなり小さいが、モデルのミスマッチのため、認識性能は低い。MLLR および SPLICE の適用によって大きな改善が得られているが、比較的良好な SNR 環境下ではピンマイク側の性能に及ばなかった。一方、低 SNR 環境下では SPLICE+MLLR を適用した咽喉マイク側の性能が勝っており、適切なチャンネル選択によって、より頑健な音声認識を実現できる可能性がある。

表 1　各チャンネルの SNR 別 WER(%)

| ピンマイク (SNR) | SNR25 | SNR20 | SNR10 | SNR5 |
|---|---|---|---|---|
| BASELINE | 3.3 | 9.8 | 24.3 | 31.5 |
| MLLR | 3.3 | 12.2 | 27.2 | 34.0 |
| 咽喉マイク (SNR) | SNR43 | SNR43 | SNR42 | SNR41 |
| BASELINE | 38.0 | 38.6 | 38.3 | 38.3 |
| MLLR | 21.4 | 20.4 | 20.7 | 20.5 |
| SPLICE | 16.7 | 16.9 | 16.9 | 17.3 |
| SPLICE +MLLR | 14.7 | 14.6 | 14.7 | 15.0 |

環境に応じて適切なチャンネルを選択するため、単語信頼度を用いた提案方法を評価した。対比実験として，SNR に応じてチャンネルを選択する方法も評価した。SNR による認識結果の選択は、ピンマイクの SNR が 14db 以上の場合はピンマイクの認識結果を選択し、14db より下の場合は咽喉マイクの認識結果を選択する。結果を図 1 に示す。

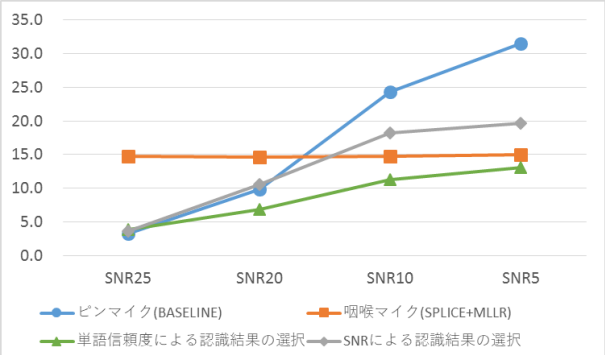


図 1　マイク選択方法による性能比較(横軸 ピンマイクの SNR，縦軸 WER(%))

図 1 に示したように SNR によって認識結果を選択した場合、高 SNR 環境下ではピンマイクに、低 SNR 環境下では咽喉マイクに比べ性能が低いが、単語信頼度によって認識結果を選択した場合、各マイク単体使用時に比べて性能が向上している。SNR に基づく方法は、測定が難しい上に，性能も単語信頼度を用いる方法よりも劣った。それに対して、単語信頼度による認識結果の選択は、単語単位で比較でき、発話衝突した単語は咽喉マイクに比べピンマイク側の信頼度が低くなることに着目することで正しい認識結果の選択ができている。このため発話の一部のみ認識を間違えている場合に、効果的に動作した。

## 4　おわりに

本稿ではピンマイクと咽喉マイクを併用した装着負荷の低い、発話衝突に頑健な音声認識手法を提案し、有効性を明らかにした。

今後の課題として咽喉マイク側の認識性能の改良や、実環境での性能評価があげられる。


### 謝辞

SPLICE についてご教授いただいた IBM 東京基礎研究所の鈴木雅之研究員に感謝いたします。


### 参考文献


[1] Stephane Dupon, *et al.*, Combined use of close-talk and throat microphones for improved speech recognition under non-stationary background noise, ISCA Tutorial and Research Workshop on Robustness Issues in Conversational Interaction, pp. 1-4, 2004.

[2] Akinobu Lee, *et al.*, Real-time word confidence scoring using local posterior probabilities on tree trellis search, ICASSP, pp.793-796, 2004.

[3] Martin Graciarena, *et al.*, Combination of standard and throat microphones for robust speech recognition in highly noisy environments, INTERSPEECH, pp. 4-8, 2004.