

圧縮センシングの 音声認識への応用

佐藤 達朗[†] 大木 真^{††} 兼本 大輔^{††}
[†]山梨大学大学院修士課程 ^{††}山梨大学工学部

1. 研究背景と目的

圧縮センシング[1][2]を用いると実質的なサンプリングレートを低くすることができるため、A/D 変換時の電力を減少させることができると期待されている。一方、スマートフォンは多数のセンサーを搭載しており、センサーの数は増えていくと見込まれる。それと共にスマートフォンの消費電力も増えるため、センサーの消費電力を抑えることが課題となる。そこで圧縮センシングをセンサーに用いることで、センサーの消費電力を抑え、スマートフォンの長時間駆動化を目指すことが考えられる。

本研究では、スマートフォンの機能の1つ、音声認識に圧縮センシングを応用することを考える。今回は、そのための基礎的検討として、シミュレーションにより以下の2つを確かめる。1つは圧縮センシングされた音声認識に用いることが可能かどうか、2つは圧縮センシングの圧縮率と音声の認識率との関係である。

2. 音声認識

以下の図 2.1 は音声認識の流れである。

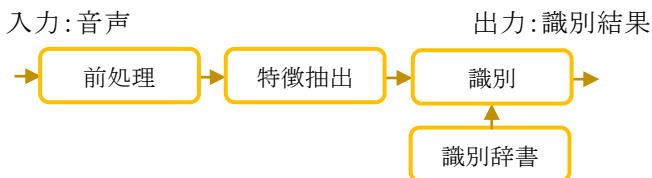


図 2.1 音声認識の流れ

- 前処理: A/D 変換, ノイズ除去など
- 特徴抽出: 認識に役立つ情報を抜き出す
- 識別: 入力と辞書をマッチングし識別
- 識別辞書: 前もって登録した特徴の情報

本研究では、前処理と特徴抽出を圧縮センシングで置き換えることを検討する。これにより、消費電力低減、回路面積削減のメリットが見込まれる。提案する音声認識の流れは図 2.2 の形になる。

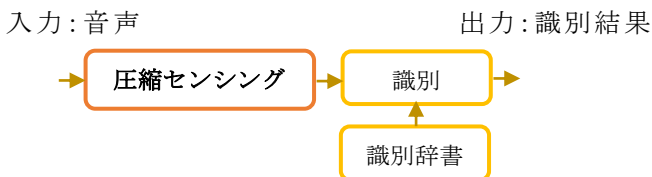


図 2.2 提案する音声認識の流れ

3. 圧縮センシング

長さ N の信号をベクトル表記して x とし、大きさ $M \times N$ の観測行列 A によって x を観測した結果を y とすると、

$$y = Ax$$

である。観測信号の長さ M が元信号の長さ N よりも小さい場合、 y から x を一意に復元することは一般にはできないが、信号 x がスパース(0成分を多く含む)な場合には精度良く x を復元できることが知られている。これを圧縮センシングと呼ぶ。信号 x そのものがスパースでない場合でも、適切な線形変換 F で変換した結果 $\alpha = Fx$ がスパースであれば、

$$y = AF^{-1}\alpha$$

と表すことができるため、 AF^{-1} を観測行列とみなすことにより同様に圧縮センシングが可能である[1][2]。

4. シミュレーション

以下の図 4.1 のような構成で日本語母音の認識率のシミュレーションを行った。音声認識は母音のみを認識できるものを Hidden Markov Model Toolkit (HTK) で構成した。母音1つに対して5つ音声を用意した。各圧縮率での各母音の認識成功回数が表 4.1 である。



図 4.1 シミュレーションの構成

表 4.1 シミュレーション結果

| 圧縮率 [%] | あ | い | う | え | お |
|---------|---|---|---|---|---|
| 0 | 5 | 5 | 5 | 5 | 5 |
| 10 | 5 | 5 | 5 | 0 | 5 |
| 20 | 5 | 5 | 5 | 0 | 5 |
| 30 | 5 | 5 | 5 | 0 | 5 |
| 40 | 5 | 5 | 5 | 0 | 5 |
| 50 | 5 | 5 | 5 | 0 | 5 |
| 60 | 5 | 5 | 4 | 0 | 5 |
| 70 | 5 | 5 | 2 | 0 | 5 |
| 80 | 5 | 5 | 0 | 0 | 3 |
| 90 | 0 | 5 | 0 | 0 | 0 |

5. まとめ

「え」の声以外は、圧縮率 50%までは圧縮率を変化させても認識に十分な特徴を確保できた。

今後は認識により最適な圧縮センシングの条件を詳しく調べる。最終的には電力のシミュレーションを行う。

参考文献

- [1] 樺島洋介, “圧縮センシングの数理”, http://www.ieice.org/~netsci/wp-content/uploads/2016/05/NetSci201605_Kabashima.pdf (2016年12月23日閲覧) .
- [2] Y.C.Eldar and G.Kutyniok, Compressed Sensing, Cambridge University Press, 2012.