

# メディア情報学実験・音声認識 第一週課題レポート

1510151 柳 裕太

2017 年 11 月 7 日

## 1 プログラム穴埋め

### 1.1 ad2fb.c

ソースコード 1: ad2fb 関数一部

```
1  for(j=0; j<NUMCHANS; j++) {
2  m[i_frame][j] = 0;
3  Aj = 0;
4  for(k=klo[j]; k<kc[j]; k++) {
5      /* ここから穴埋めしたエリア */
6      W_kj = ((double)(k) - klo[j]) / (kc[j] - klo[j]);
7      m[i_frame][j] += W_kj * pow(X[k], 2);
8      Aj += W_kj;
9      /* 穴埋めしたエリアおわり */
10 }
11 for(k=kc[j]; k<=khi[j]; k++) {
12     /* ここから穴埋めしたエリア */
13     W_kj = ((double)(khi[j]) - k) / (khi[j] - kc[j]);
14     m[i_frame][j] += W_kj * pow(X[k], 2);
15     Aj += W_kj;
16     /* 穴埋めしたエリアおわり */
17 }
18 /* フィルタの面積が等しくなるように正規化 */
19 m[i_frame][j] /= Aj;
```

### 1.2 fb2mfcc.c

ソースコード 2: fb2mfcc 関数一部

```
1  void fb2mfcc(int n_frame, /* フレーム数 */
2              float **m, /* フィルタバンク */
```

```

3         float **c,    /* MFCC (フレーム x 次元) */
4         int P,        /* フィルタバンクのチャンネル数 */
5         int Q)        /* MFCCの次元 */
6     {
7         int i_frame;
8         int i, j;
9         double mfnorm, pi_factor, x, sigma_p; // sigma_p変数追加
10
11
12         mfnorm = sqrt(2.0/(double)P); /* (3.3)式の $\sqrt{2/P}$  */
13         pi_factor = M_PI/(double)P; /* (3.3)式のcosの変数の $\pi/P$  */
14
15         for(i_frame=0; i_frame<n_frame; i_frame++) {
16             for(i=1; i<=Q; i++) {
17                 x = (double)i*pi_factor; /* (3.3)式のcosの変数の $\pi i/P$  */
18                 /* ここから穴埋めしたエリア */
19                 sigma_p = 0;
20                 for(j=1; j<=P; j++){
21                     sigma_p += log10(m[i_frame][j-1]) * cos(x * ((double)(j
22                                     )-0.5));
23                 }
24                 c[i_frame][i-1] += mfnorm * sigma_p;
25                 /* 穴埋めしたエリアおわり */
26             }
27         }

```

## 2 自分の名前の分析結果

自分の名前 (yanagi yuta) に対する drawspec 解析の結果は以下の図1の通りである。

中央の図を見ると、発話時に母音が”あ”の文字だった場合、下部のエリアの色が他の母音(”い”など)に比べて明るい部分が広がっているのが分かった。また、”う”を発話した際に上部のエリアが他の発話とくらべて暗めになっていることから、図の上部が子音、下部が母音の種類に対応していることが類推できた。更に最後の”た”の発話のときに、一瞬全てのエリアが暗くなっており、その後広い範囲が明るくなっていた。これは、た行そのものが一瞬声を溜めてから発話する文字であることを示唆しているのではないかと解釈した。

## 3 実験のポイント

今回の実験では、フィルタバンク分析と MFCC 分析の実装と、それによって人の音声の周波数を解析し、50 音によってどのような周波数構成の相違があるかを推察することが主な目的である

と考えている。

## 4 よくわかったこと

自ら直接音声进行分析する部分を実装し、自分の名前をサンプルにしたことで、音声分析という観点から日本語の 50 音の特徴を垣間見ることができた。

## 5 よくわからなかったこと

日本語以外の言語、特に英語の場合どのような特徴があるのか知りたかった。また、それに付随してネイティブスピーカーの英語と日本人の英語を比べることで、所謂”日本語訛り”がどの部分で確認できるか、音声分析の観点から知りたかった。

## 6 要望

今回のテキストには、説明不足な部分から実験中に少々混乱が生じてしまった。このような事態は今回限りにさせて頂けると幸いです。

## 7 感想・その他

C 言語を通じて音声分析を行うのは初めての経験だったが、非常にとっつきやすく楽しい実験だった。

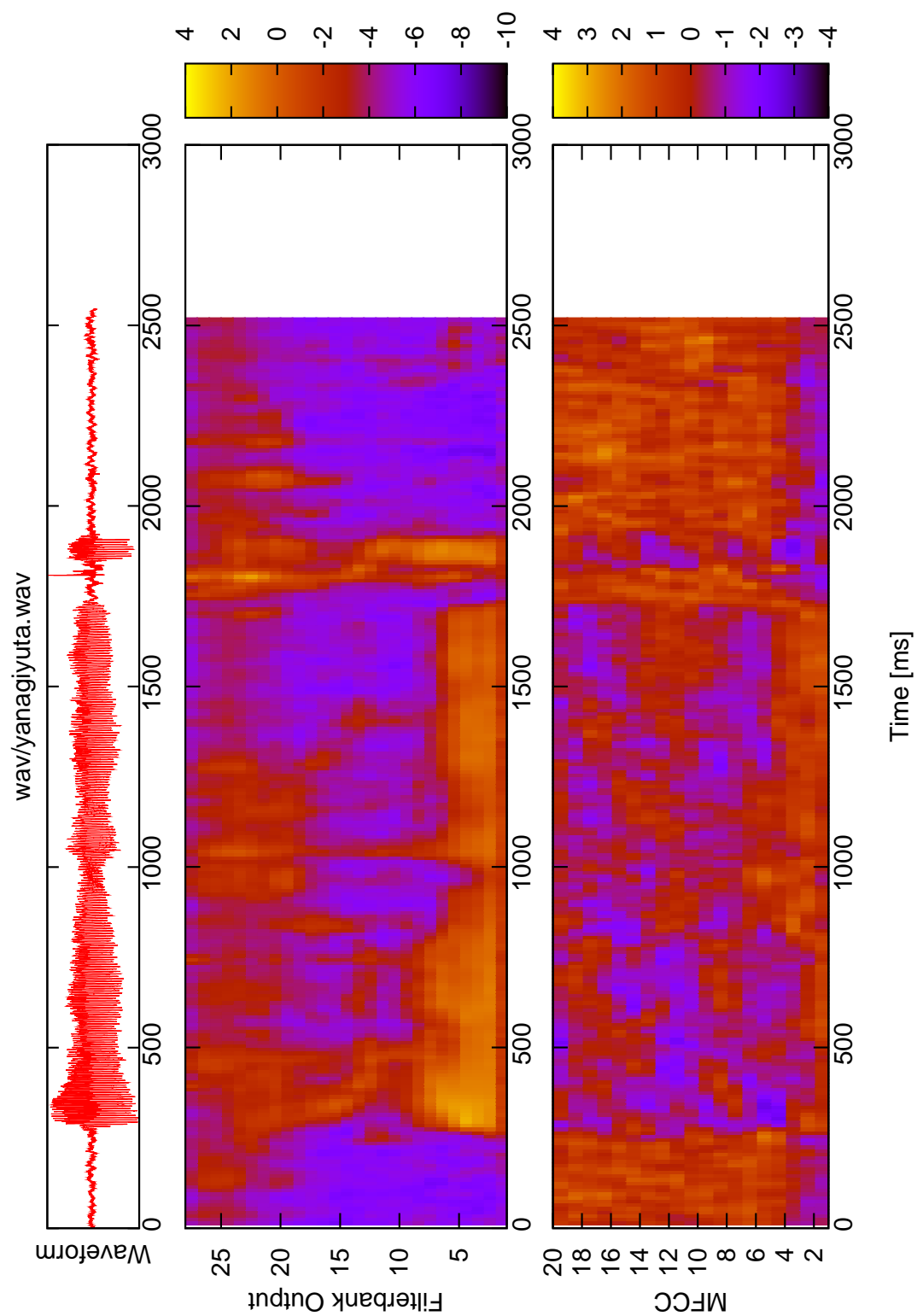


図 1: 自分の名前に分析をかけた結果