

## 付録 A 数字単語音声のスペクトル 分析実例

録音した音声データから単語音声の区間を正確に切り取るためには、音声の性質についての若干の知識が必要です。音声波形の表示と聴取だけでは単語の開始および終了時刻を見誤ることが多いので、録音した音声データのスペクトログラムを表示し、そのパターンの違いに基づいて単語音声の区間を定めてください。

実習に用いる数字単語音声の特徴を説明します。説明に用いるのは、高木の声で録音した音声の振幅を正規化し、WaveSurfer の Demonstration モードで表示した図です。声には性別、年齢、個人性による特徴があります。また、全く同じ発声を繰り返すことは不可能なので、同一人物が同じ単語を発音しても、音声波形やスペクトルの形は微妙に異なります。しかし、音素による波形やスペクトルの特徴は一定の性質を示します。

0 /rei/ 日本語のラ行音は音韻環境<sup>1</sup>によって特性が異なりますが、/rei/の/r/は図 A.1 (a) のようなパターンを示します。波形は無音声区間 (SIL) の背景雑音に比べて振幅が大きく、スペクトログラムでは、350 Hz, 2300 Hz 付近にエネルギーの高い場所があります。これは後続の/e/の影響によるものです。/ei/の部分は2種類の母音がなめらかに変化していて、500 Hz および 1300 Hz~3000 Hz に強い成分が含まれています。この単語の終端位置は、スペクトログラムでフォルマントの表示が弱まって見えなくなる場所に定めます。

1 /ichi/ 子音/ch/が母音/i/に挟まれている単語です。この例では、1つ目の/i/に比べて2つ目の/i/が長く強く発音されていることがわかります(図 A.1 (b))。/ch/は無声破擦子音と呼ばれる子音の1つで、(i) 先行母音からの過渡部、(ii) 無音部 (閉鎖部)、(iii) 破裂部、(iv) 摩擦部、(v) 後続母音への過渡部からなります。図では (i) および (ii) を合わせて c1、(iii)~(v) を ch としています。音の無い c1 の部分も子音の一部であることに注意してください。母音/i/の第1フォルマントは400 Hz 付近、第2フォルマントは2300 Hz 付近にあります。

2 /ni/ ナ行の鼻子音/n/の部分では500 Hz 付近に共鳴が見られますが、波形やスペクトログラムだけでは/r/との区別は難しい場合があります。鼻音化の現象は後続の/i/にも影響しています。単語の開始点および終了点は図 A.1 (c) のように、スペクトログラム上で共鳴ピークの表示の端に定

<sup>1</sup>前後にどのような音韻 (音) があるかということ。

めてください。録音した音声には雑音が含まれているので、音声波形だけでは/n/の開始点や/i/の終了点を決めるのが難しいことがわかります。

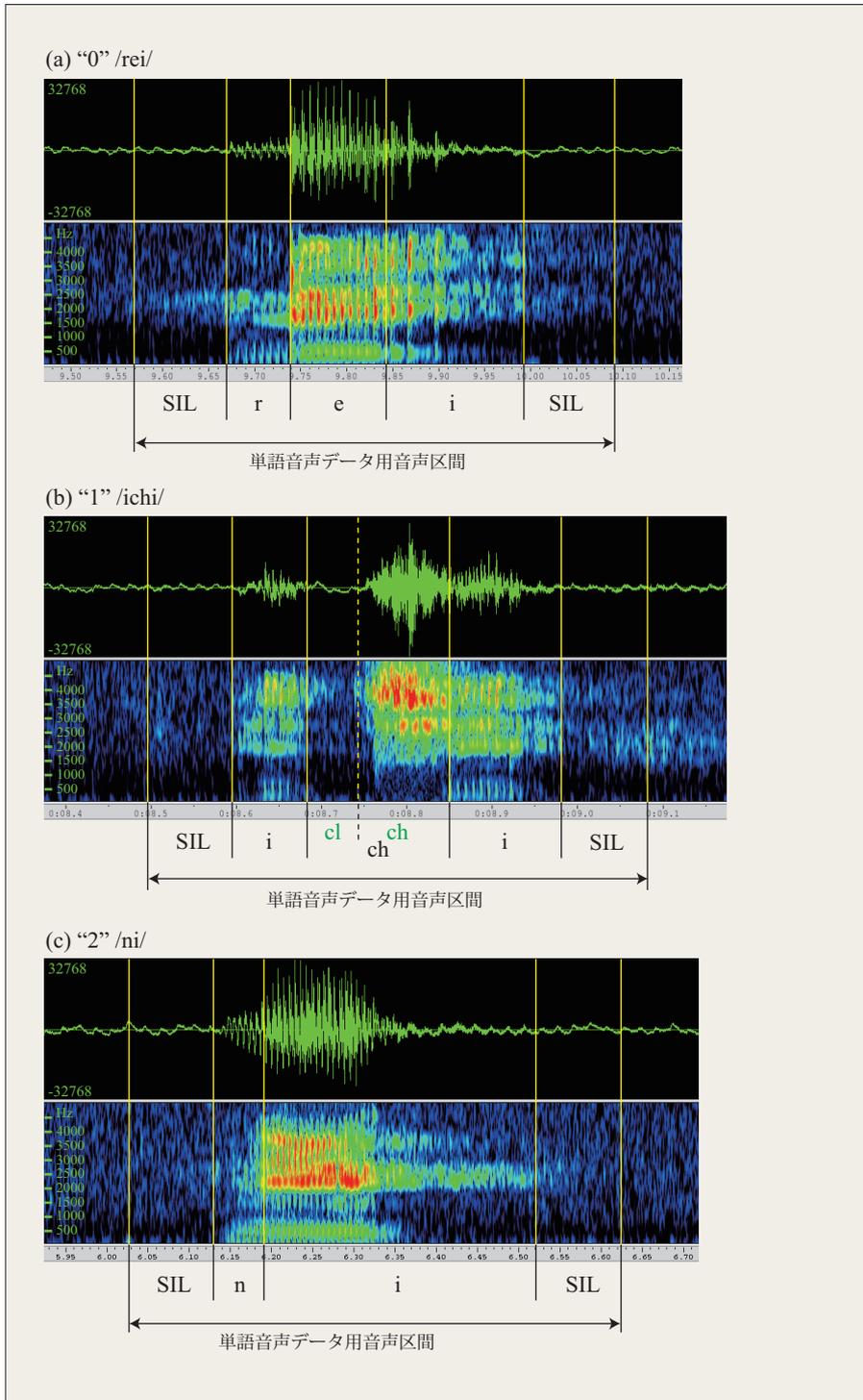


図 A.1: 単語 HMM 学習用音声データの例 (0, 1, 2).

- 3 /saN/ この単語の最初の音は無声摩擦子音という子音です。/s/はその1つで、日本語の場合「さ」「す」「せ」「そ」で使われます。定常的で比較的安定しています。波形の振動は非常に細かく、スペクトルには3000 Hz以上の高域に強いパワーの成分が現れます。母音/a/の第1フォルマントは650 Hz, 第2フォルマントは1400 Hz付近にあります。/i/のフォルマント(図 A.1 (b) (c))と見比べてみてください。この単語発声の場合(図 A.2 (a)) /N/のスペクトルパターンが長く継続しているように見えますが、波形の観察と音声聴取により無音(SIL)との境界を決めました<sup>2</sup>。
- 4 /yoN/ 最初の音は半母音と呼ばれる音素/y/です。音声波形もスペクトルも母音と同様の性質を示します。/yo/は/io/を速く発音した場合の音に似たパターンを示します(図 A.2 (b))。それは/y/の前半のフォルマントが/i/に近いことからわかります。最後の/N/は/saN/の/N/と同じパターンを示しています。
- 5 /go/ /g/は有声破裂音と呼ばれる子音の1つです。声帯が振動しているので、低い周波数帯域にエネルギーの強い部分があります(図 A.2 (c))。日本語の有声破裂音には他に/b/, /d/があり、/g/との違いは、後続母音の種類によって区別することができますが、詳しくは関連図書を見てください。単語開始部はスペクトログラム表示で/g/のエネルギーが現れ始める位置、終了部は/o/のエネルギーが消える位置に設定します。

---

<sup>2</sup>/saN/と発音して口を閉じたあとに、喉が余韻で振動していたようです。

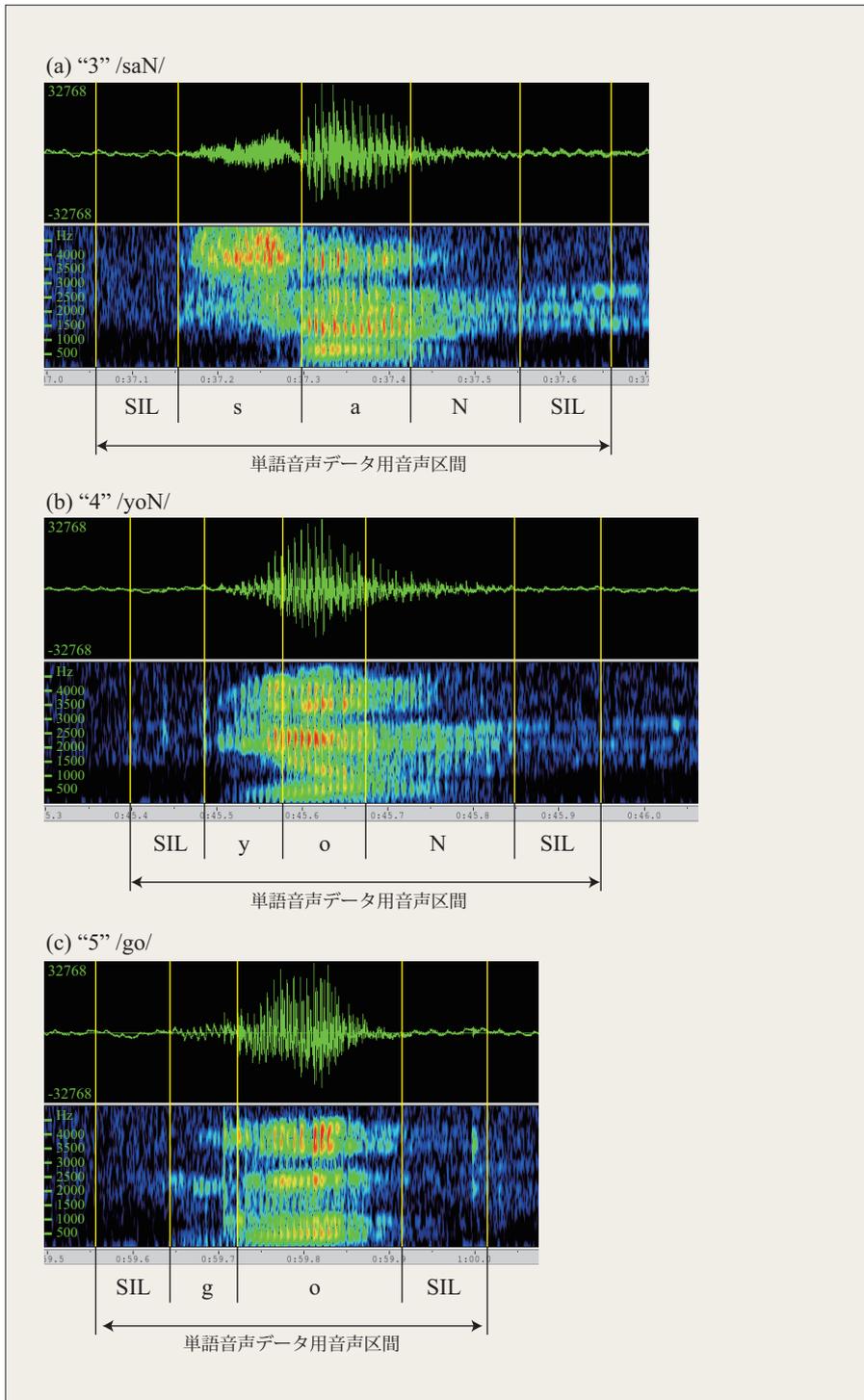


図 A.2: 単語 HMM 学習用音声データの例 (3, 4, 5).

- 6 /roku/ 最初の音素/r/は/rei/と同じパターンを示しています。この単語の母音は/o/と/u/です。ここまでで日本語の5母音の全てが揃いました。母音は音声波形にも特徴がありますが、スペクトルに現れるフォルマントの位置に特徴があります(2.2.1節)。無声破裂子音である/k/は/ch/(図 A.1 (a))と同様に無音部(c1)や破裂部(k)からなります。単語開始点と終了点は/rei/と同様に定めればよいでしょう(図 A.3 (a))。
- 7 /nana/ この単語の開始は/ni/と同じ鼻子音の/n/です。2番目の/n/は母音/a/に挟まれているため、1番目の/n/に比べて母音の共鳴の影響を強く受けています。音声に背景雑音が含まれるときには、音声波形だけで語頭/n/の開始点を決めるのは難しいことがあります。スペクトログラムを見たほうが良くわかります。
- 8 /hachi/ ハ行音の子音/h/は無声摩擦子音という種類の子音で、声帯が振動していないので低い周波数にエネルギーが無く、したがってスペクトログラムの低周波数帯にはパターンが現れていません。/ichi/で使われている無声摩擦子音がこの単語でも使われています。/ichi/とは1音だけ異なっているため、/ichi/のスペクトログラム(図 A.1 (b))ととても良く似たパターンになっています。

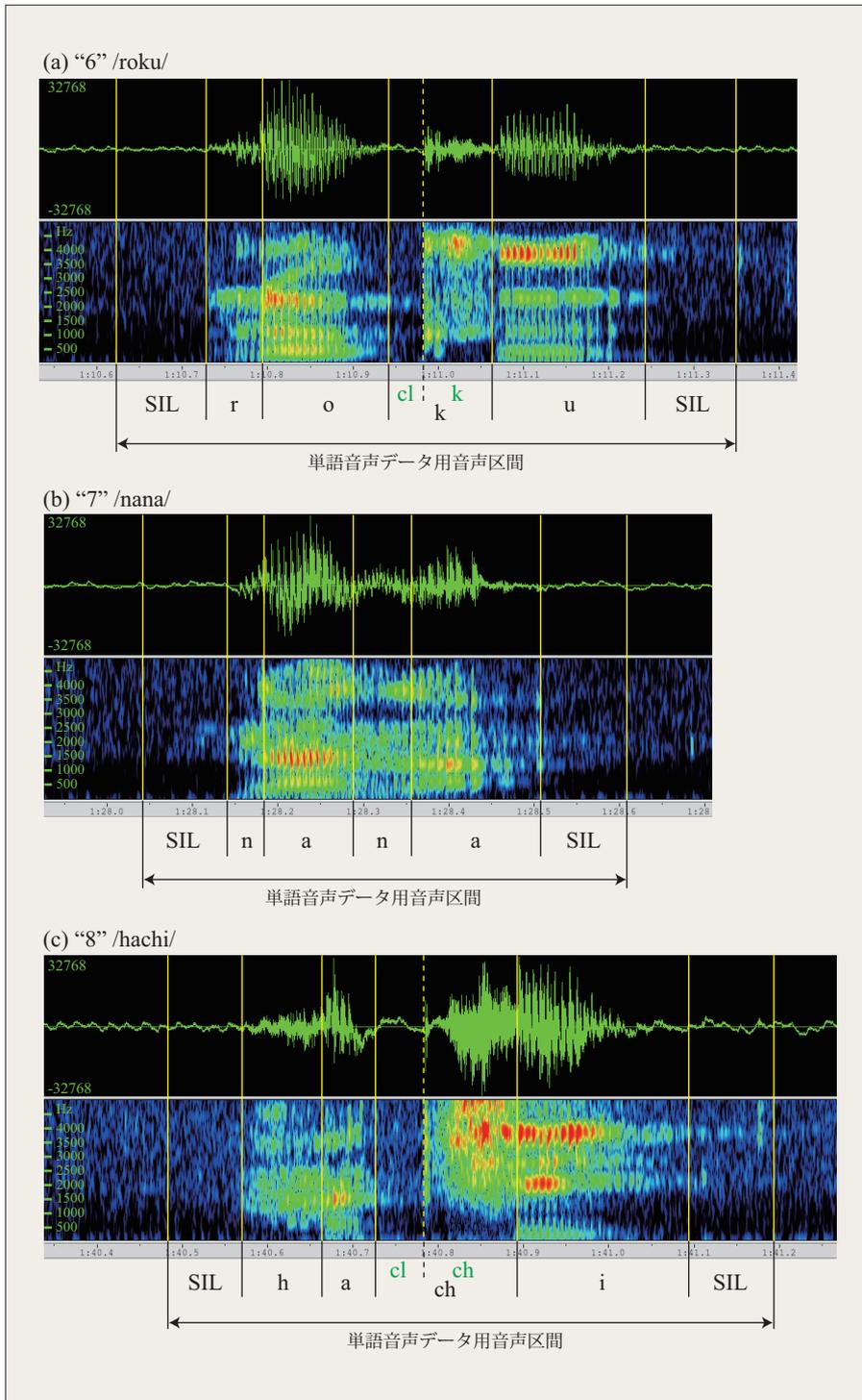


図 A.3: 単語 HMM 学習用音声データの例 (6, 7, 8).

9 /kyuu/ 無声破裂子音/k/で始まっている単語です。/roku/の場合は母音が先行していたので、閉鎖部が判別できましたが、語頭の破裂子音の閉鎖部は音声信号からはわかりません。したがってこのような場合は破裂部の開始を単語の開始とします。スペクトログラムには開始が明確に現れています。/yoN/の場合と同様に、半母音/y/は/i/に類似のスペクトルパターンを示します。/uu/は長音ですが、短音である6の/u/（図 A.3 (a)）と継続長はほとんど同じです。実は、長音と短音の区別は単に継続長で区別できるものではありません。この区別は、聴き手の母国語に強く影響されます。つまり、日本語のように音の長短が単語の意味に関わることもある言語を母国語とする人は長音と短音の違いが簡単に分かるのですが、英語や中国語のように音の長短で単語の意味を区別しない言語を母国語とする人は日本語の上級者でない限り、この2つの「う」の違いを区別するのは困難です。みなさんの音声データの6の/u/と9の/uu/ではいかがでしょう？ 継続長を比べてみてください。

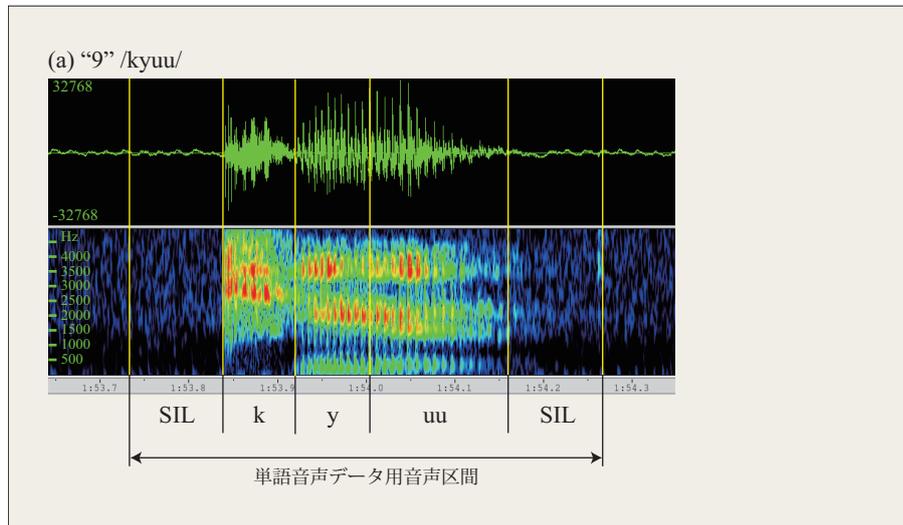


図 A.4: 単語 HMM 学習用音声データの例 (9).

## 付録B 実習用ファイル

**asr** : この実習の最上位ディレクトリ

**.wavesurfer** : WaveSurfer の環境設定ディレクトリ

**addpath** : 実習用コマンドのパス設定スクリプト

**drill** : HMM 算法の実習作業ディレクトリ

**Makefile** : ファイル生成手順、プログラム構成ファイルの関係

**backward.c** : Backward アルゴリズム関数 [穴埋め]

**baumwelch.c** : Baum-Welch アルゴリズム関数 [穴埋め]

**data1** : 信号源 (HMM<sub>1</sub>) の認識データ格納用ディレクトリ

**data2** : 信号源 (HMM<sub>2</sub>) の認識データ格納用ディレクトリ

**drawLC\_d1** : HMM<sub>1</sub> の学習曲線を描く gnuplot スクリプト

**drawLC\_d2** : HMM<sub>2</sub> の学習曲線を描く gnuplot スクリプト

**drawR** : HMM パターン認識結果の図を描く gnuplot スクリプト

**drillB.c** : Backward アルゴリズム実行

**drillF.c** : Forward アルゴリズム実行

**drillR.c** : HMM によるパターン認識シミュレーション

**drillT.c** : HMM による時系列パターン学習

**forward.c** : Forward アルゴリズム関数 [穴埋め]

**gen.c** : 信号源 (HMM<sub>1</sub> または HMM<sub>2</sub>) から時系列データを生成

**genData** : HMM によるパターン認識シミュレーションデータの生成

**paramFB.txt** : 動作確認用 HMM パラメータ

**sound** : 音声入出力機能

**Makefile** : ファイル生成手順、プログラム構成ファイル同士の関係

**\_vu.c** : 入力音声表示

**ad2txt16.c** : 音声波形振幅値表示 (波形描画用)

**audioIN** : 音声入力 (録音) シェルスクリプト

**hupAudioIN** : 音声入力プロセス終了シェルスクリプト

**vad.c** : 音声区間検出

**vu** : 入力音声レベル表示シェルスクリプト

**studentbook.pdf** : この実験指導書

**wrecog** : 単語音声認識実習作業ディレクトリ

`drawLC` : 単語 HMM 学習曲線描画 gnuplot スクリプト  
`hmm` : 学習済み HMM パラメータ格納ディレクトリ  
`lib` : 単語 HMM 学習・認識用ファイル情報格納ディレクトリ  
`log` : 単語 HMM 学習・認識過程記録格納ディレクトリ  
`mfcc` : MFCC ファイル格納ディレクトリ  
`program` : 単語認識プログラム格納ディレクトリ  
`Makefile` : ファイル生成手順、プログラム構成ファイルの関係  
`_recog.c` : On-The-Fly 単語認識  
`ad2fb.c` : フィルタバンク計算関数 [穴埋め]  
`ad2mfcc.c` : 音声波形から MFCC を計算し CMVN 処理を行う  
`countCorr` : 単語 HMM 学習検証用シェルスクリプト  
`drawspec` : 音声スペクトル描画用 Perl スクリプト  
`drillG.c` : 多次元ガウス確率密度関数計算  
`fb.c` : フィルタバンク計算  
`fb2mfcc.c` : MFCC 計算関数 [穴埋め]  
`gpdf.c` : 多次元ガウス確率密度関数 [穴埋め]  
`mfcc.c` : On-The-Fly 単語認識 MFCC 計算  
`mfccf.c` : 音声波形ファイル対象 MFCC 計算  
`prtmfcc.c` : MFCC の値の表示  
`recog` : On-The-Fly 音声認識コマンドシェルスクリプト  
`recogf.c` : 単語認識 (MFCC ファイル対象, バッチ処理用)  
`train.c` : 単語 HMM 学習  
`vtb.c` : Viterbi 計算  
`result` : 認識結果格納ディレクトリ  
`sample` : 動作確認用データ格納ディレクトリ  
`sample.fb` : フィルタバンク計算の正解 (プリント形式)  
`sample.mfcc` : MFCC 計算の正解  
`sample.mfcc.txt` : MFCC 計算の正解 (プリント形式)  
`sample.wav` : 正解計算用音声波形ファイル  
`sampleG.hmm` : 多次元ガウス確率密度関数検算用 HMM  
`sampleG.mfcc` : 多次元ガウス確率密度関数検算用 MFCC  
`vadwav` : On-The-Fly 検出された音声波形ファイル格納ディレクトリ  
`wav` : 音声波形ファイル格納ディレクトリ



