

# 11. HTK を使って単語を認識してみよう

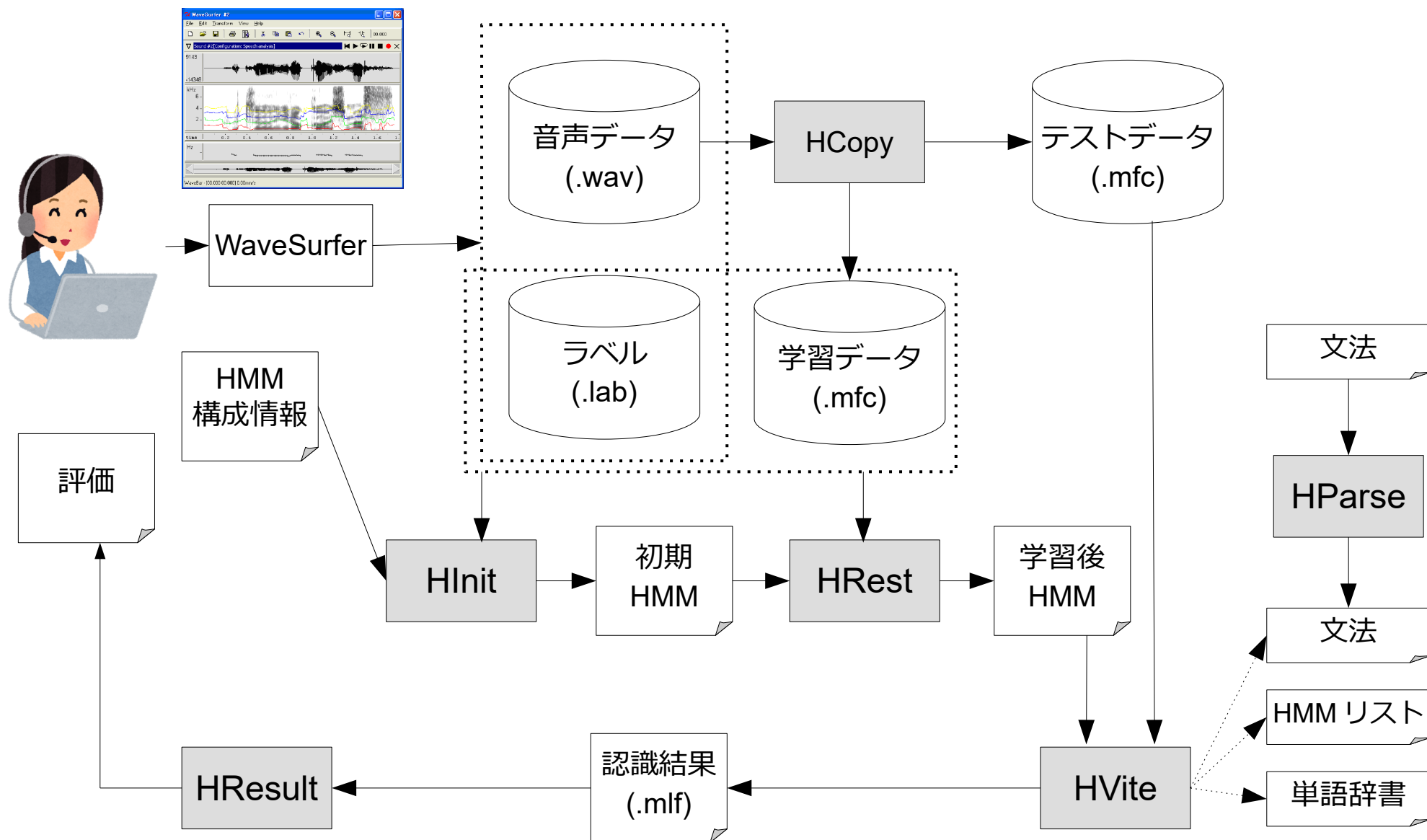
- HTK(Hidden Markov Model Toolkit) とは
  - HMM の構築・学習・認識・評価などのためのコマンドの集合からなるツールキット
  - 安定版バージョン : 3.4.1
    - GMM ベース
  - 開発版バージョン : 3.5β
    - 音響モデルにニューラルネットワークをサポート
    - リカレントネットワーク言語モデルを用いたデコードをサポート

# 11.1 HTK の構成

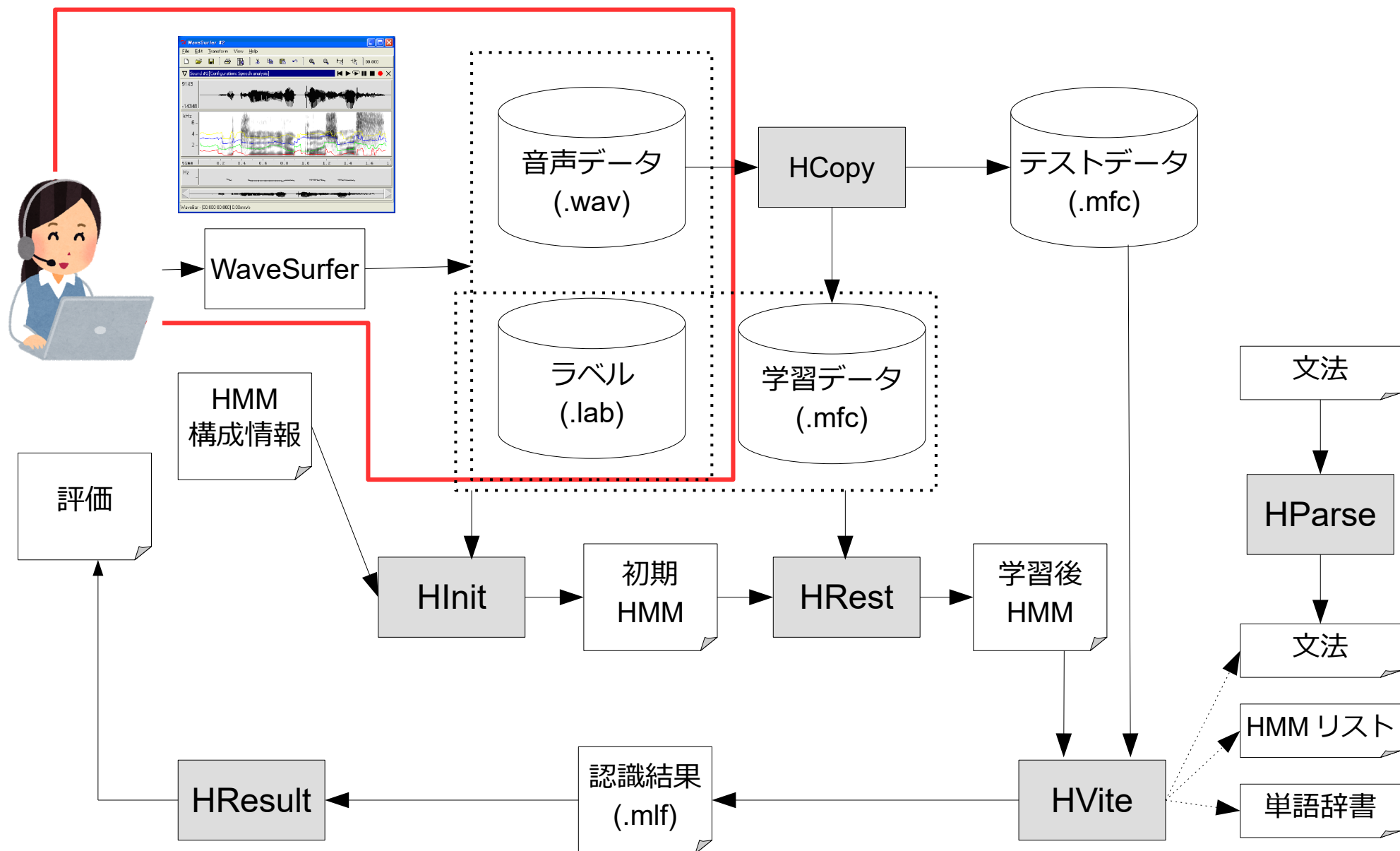
- HTK の基本コマンド

コマンド名	機能
HCopy	特徴抽出
HInit	HMM の初期化
HRest	HMM の学習
HParse	文法記述をネットワーク表現に変換
HVite	ビタビアルゴリズムによる認識
HResults	認識結果の集計

# 11.1 HTK の構成



# 11.2 音声の録音とラベル付け

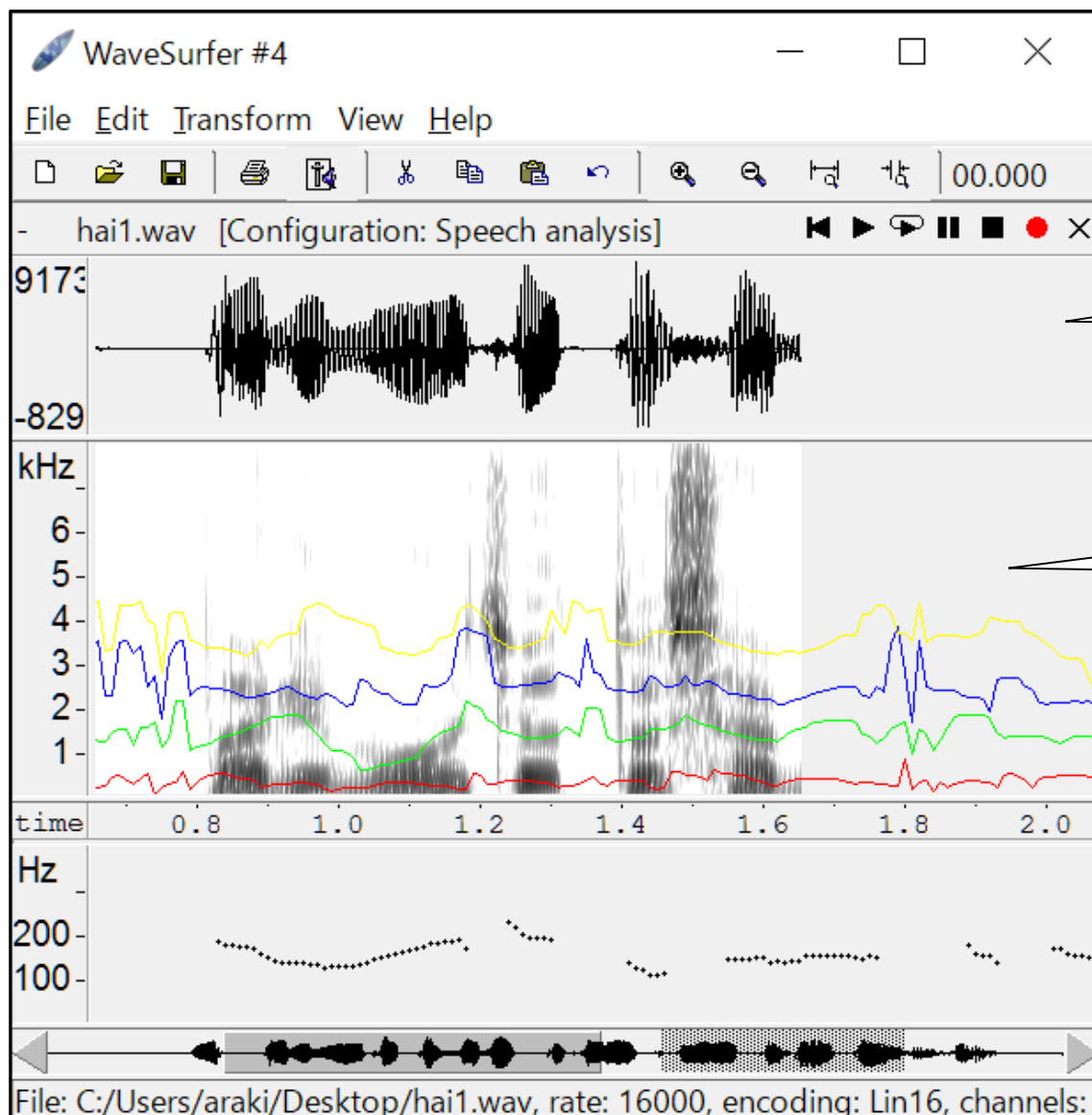


## 11.2 音声の録音とラベル付け

- 学習データを集める
  - 設定した認識タスクに合わせて音声を録音する
  - 場合によっては、性別・年齢・背景雑音などの条件を設定する
- ラベル付け
  - 正解データ：音声に対する書き起こし
- 使用可能なソフトウェア
  - HSLab, WaveSurfer, Audacity, ...

# 11.2 音声の録音とラベル付け

- WaveSurfer



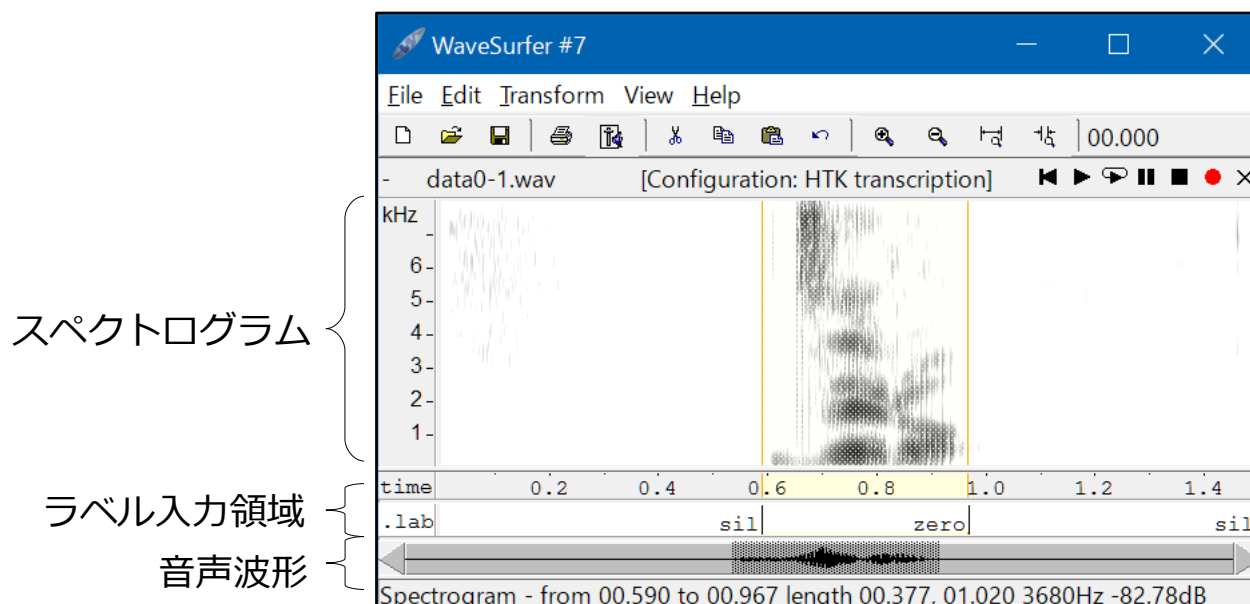
音声波形

スペクトログラム

基本周波数

# 11.2 音声の録音とラベル付け

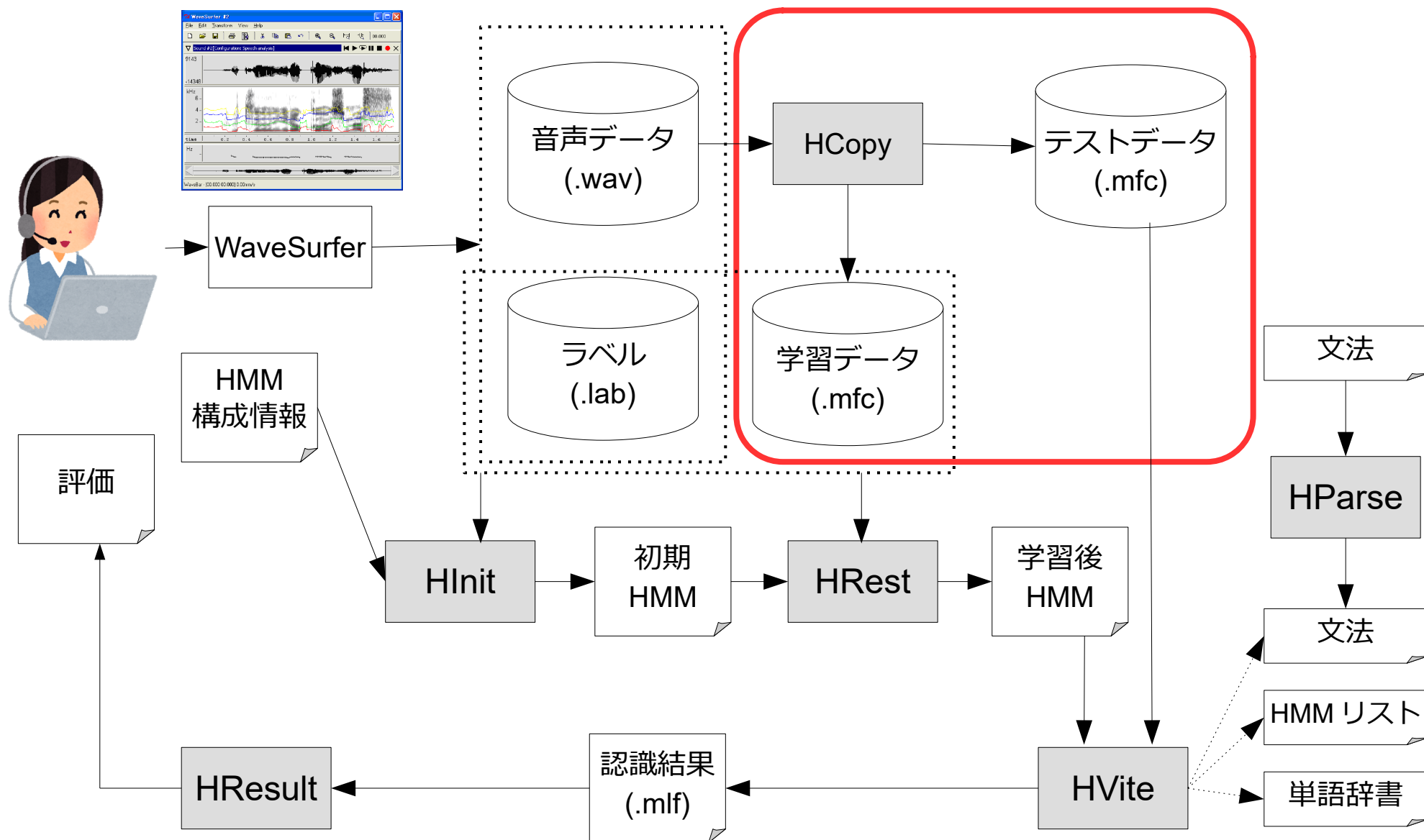
- ラベル付けの GUI



```
0 6350000 sil
6350000 8500000 hai
8500000 12125000 sil
```

ラベルファイル (0.1 マイクロ秒単位)

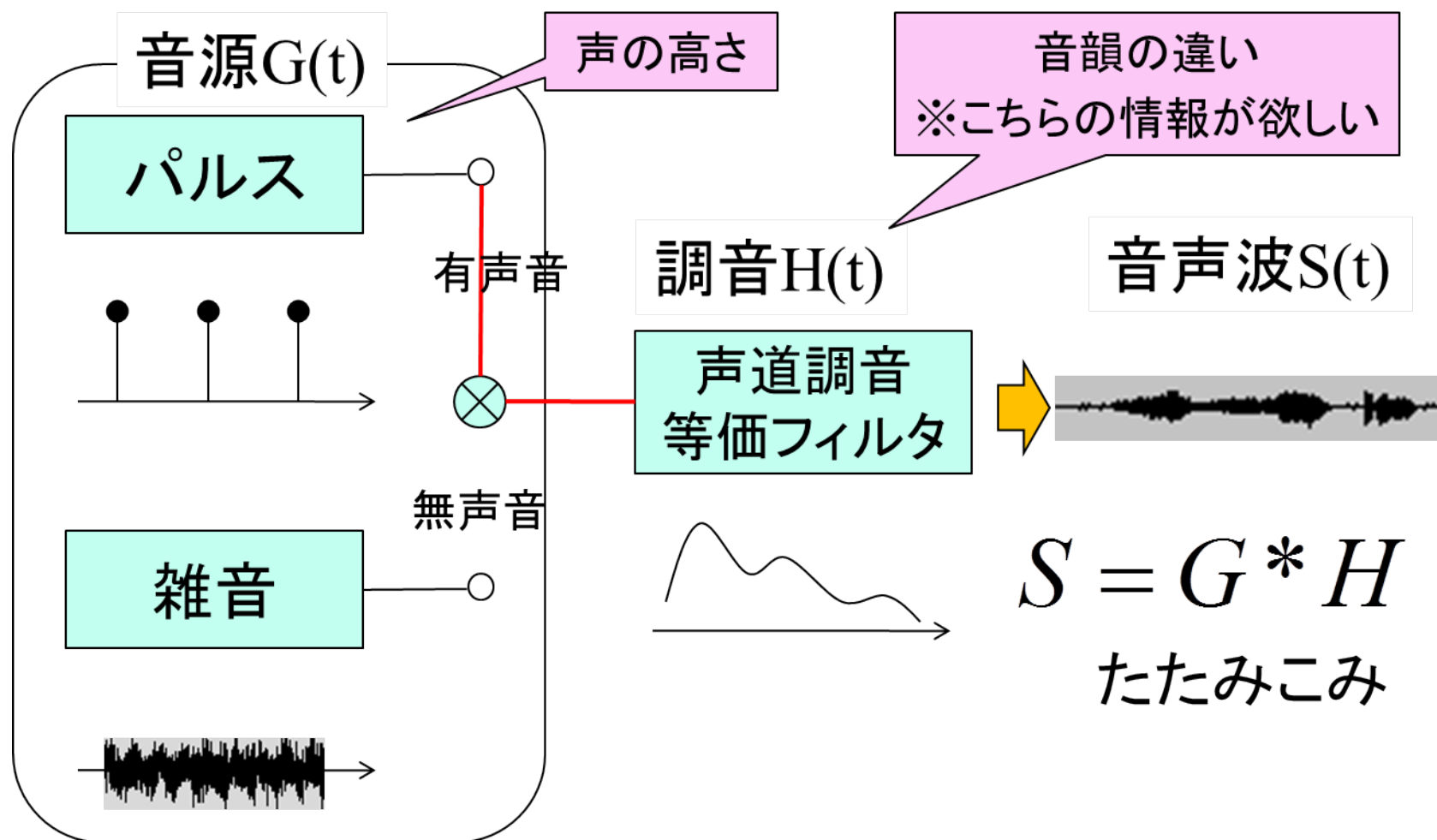
# 11.3 特徴抽出





# 11.3 特徴抽出

- 音声生成過程のモデル



# 11.3 特徴抽出

- 調音フィルタ特性の取り出し

音声信号  $S = G * H$



フーリエ変換

スペクトル  $S = G \cdot H$



絶対値・メルフィルタ・対数化

$$\log|S| = \log|G| + \log|H|$$



離散コサイン変換

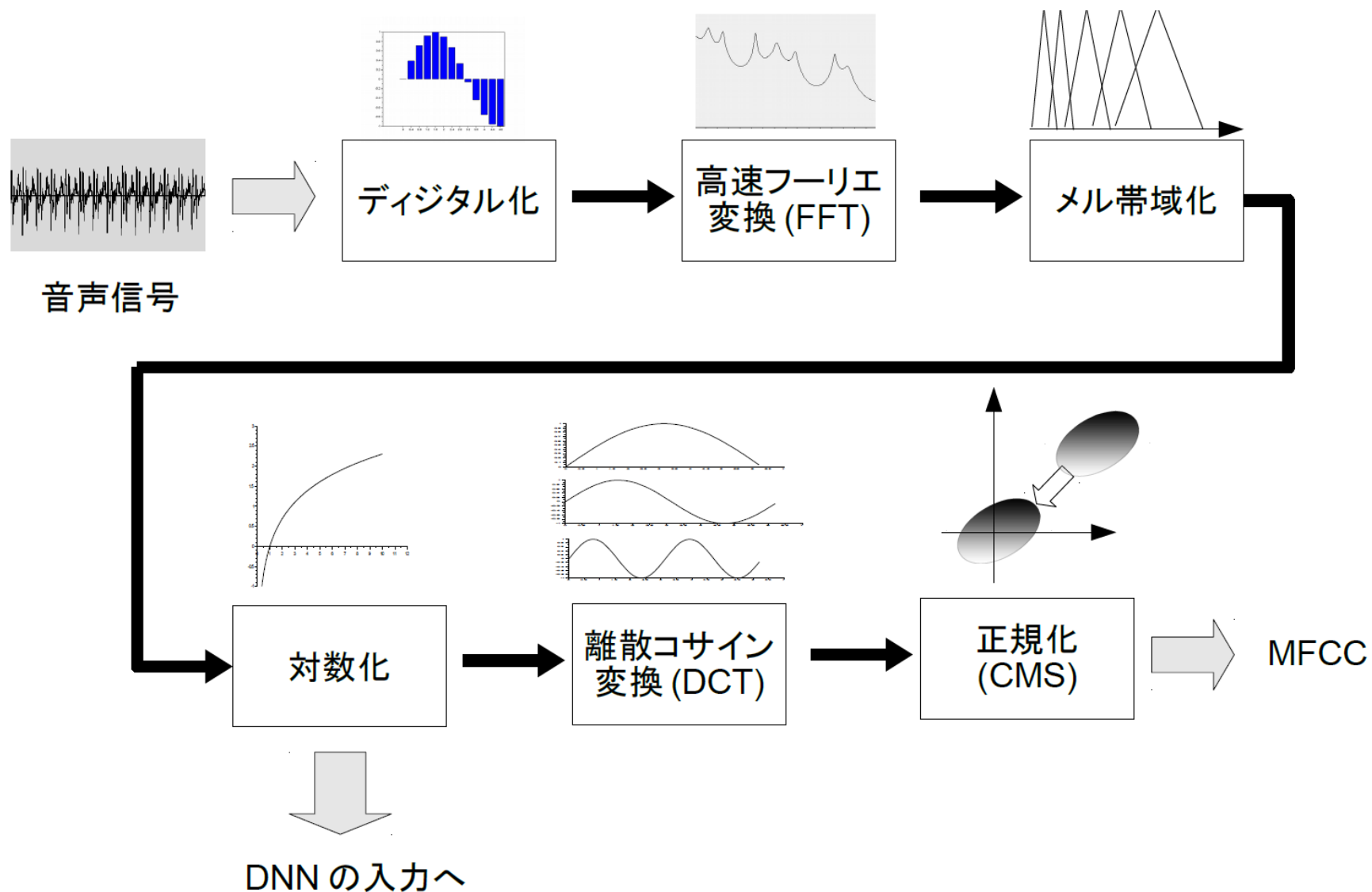
ケプストラム  $DCT \log|S| = DCT \log|G| + DCT \log|H|$

人間の感覚上の  
音の高さを反映  
(高周波ほど鈍感)

ケプストラム  
※音のスペクトルを  
信号と見なしてフー  
リエ変換したもの

# 11.3 特徴抽出

- MFCC の計算過程



## 11.3 特徴抽出

- HCopy コマンド
  - 音声データから特徴ベクトル系列を求める
    - スクリプトファイルで対応を記述することもできる
  - 抽出方法は構成ファイルで指定する

HCopy [ オプション ] 音声ファイル名 特徴ファイル名

-C configFile: 構成ファイルを指定

-S scriptFile: スクリプトファイルを指定

## 11.3 特徴抽出

- HCopy のスクリプトファイル

```
hai1.wav hai1.mfc
```

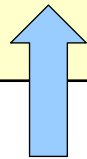
```
...
```

```
hai5.wav hai5.mfc
```

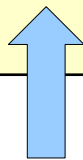
```
iei1.wav iei1.mfc
```

```
...
```

```
iei5.wav iei5.mfc
```



入力



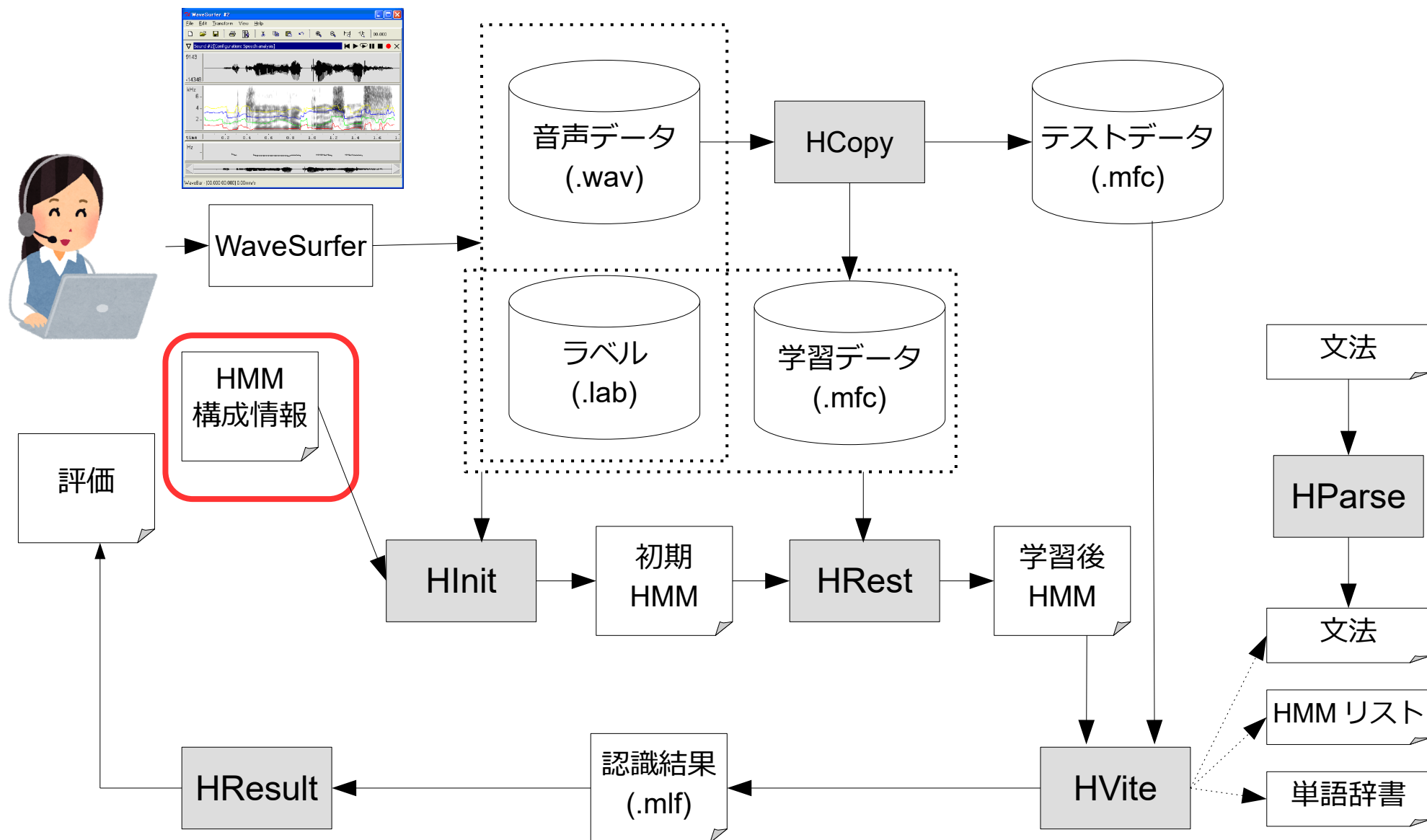
出力

# 11.3 特徴抽出

- HCopy の構成ファイル

```
SOURCEKIND = WAVEFORM      # wav 形式
SOURCERATE  = 625           # サンプリングレート 16KHz
TARGETKIND  = MFCC_0_D_A    # MFCC の指定
TARGETRATE  = 100000.0      # フレーム周期 10 [ms]
WINDOWSIZE  = 250000.0      # 窓の広さ 25 [ms]
USEHAMMING  = T             # ハミング窓の適用
PREEMCOEF   = 0.97          # highpass フィルタの適用
NUMCHANS    = 24             # filterbank のチャンネル数
NUMCEPS     = 12             # MFCC の次元数 12
```

# 11.4 初期モデルの作成



# 11.4 初期モデルの作成

- HMM の構成ファイルの作成

- 特徴ベクトルの内容

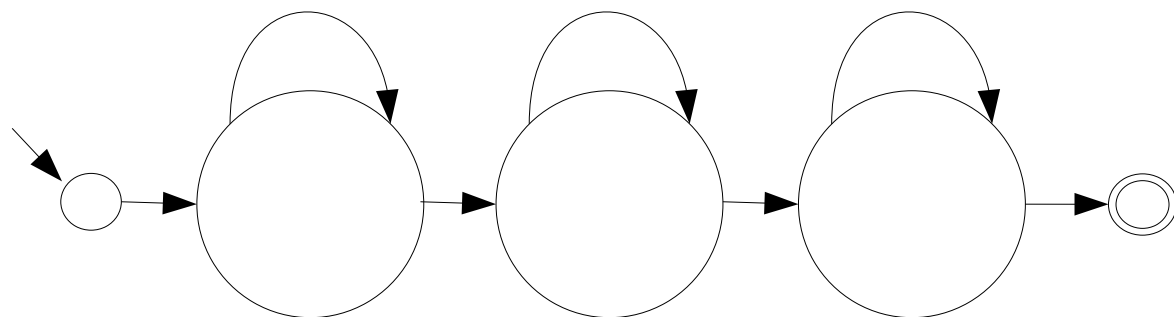
- HMM の名前

- 状態数

- 各音素 3 状態が標準

- 平均・共分散行列の次元数

- 遷移確率行列による HMM のトポロジーの指定





# 11.4 初期モデルの作成

## • HMM の構成ファイル (1/2)

```
~o <VecSize> 39 <MFCC_0_D_A>
```

```
~h "hai"
```

```
<BeginHMM>
```

```
<NumStates> 5
```

```
<State> 2
```

1: 初期状態、5: 最終状態なので、  
状態 2 から 4 までを定義

```
<Mean> 39
```

```
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
```

```
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
```

```
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
```

```
<Variance> 39
```

```
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
```

```
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
```

```
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
```

対角成分  
のみ

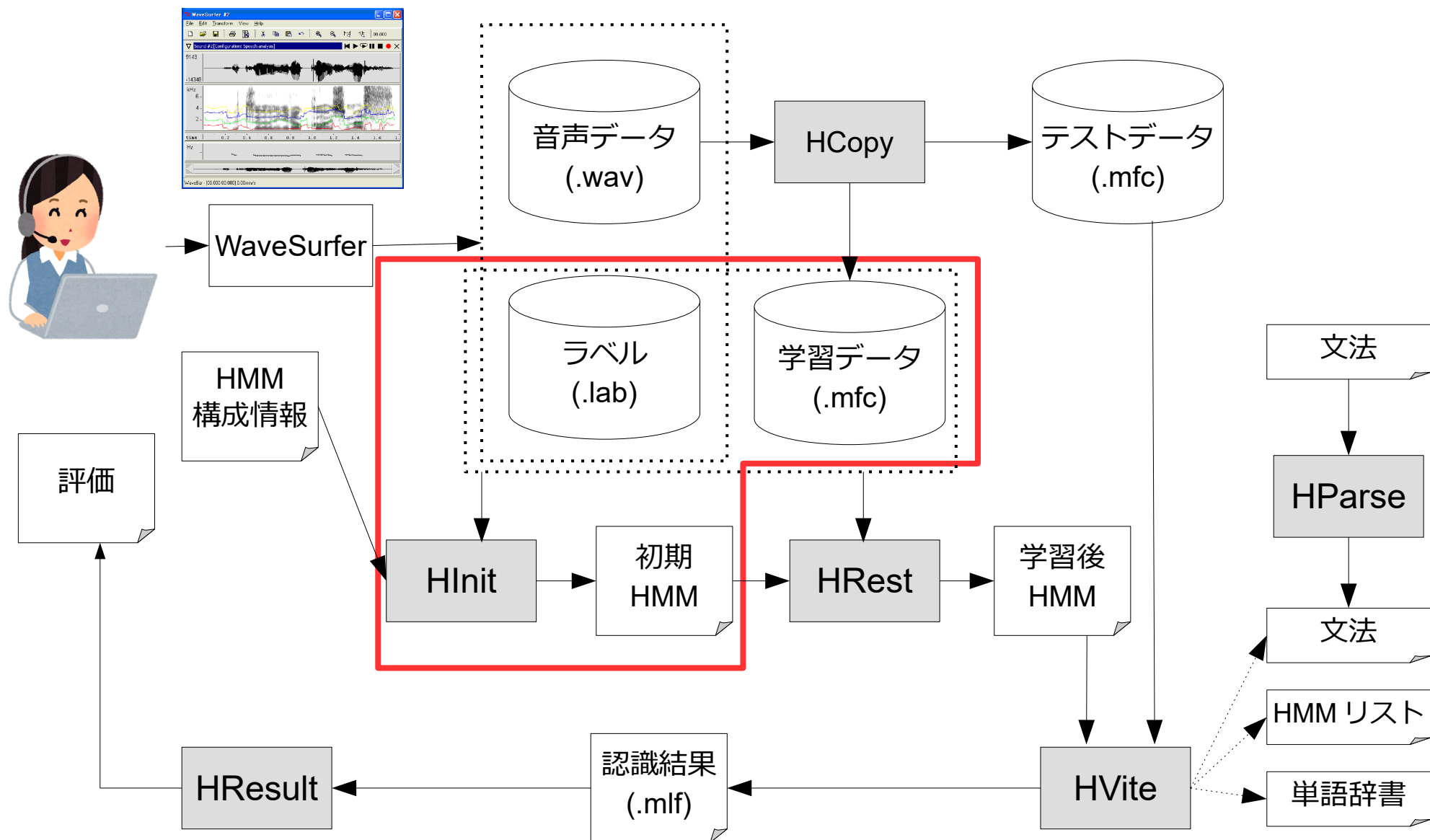
# 11.4 初期モデルの作成

- HMM の構成ファイル (2/2)

```
<State> 3
...
<State> 4
...
<TransP> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.5 0.5 0.0 0.0
0.0 0.0 0.5 0.5 0.0
0.0 0.0 0.0 0.5 0.5
0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

left-to-right 構造

# 11.5 初期値の設定



## 11.5 初期値の設定

- 適切な初期値の重要性
  - Baum-Welch アルゴリズムは局所最適解を求めるものなので、よい初期値から学習を始めることが重要

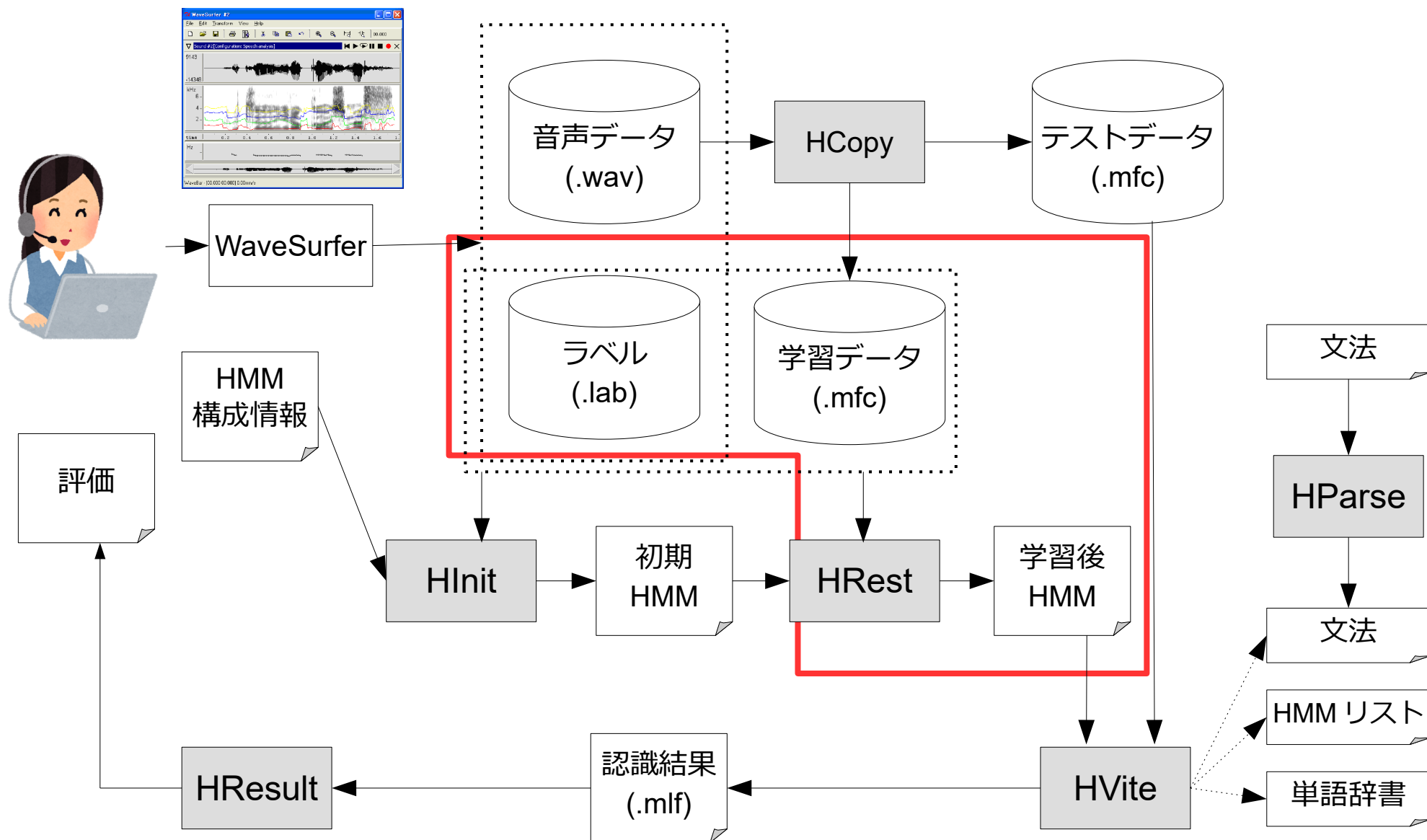
# 11.5 初期値の設定

- HInit コマンド
  - 学習データからモデルの初期値を求める
    - 強制整列から平均・分散を繰り返し計算

HInit [ オプション ] HMM 名

- T Num: 出力トレースレベル
- S listFile: 学習データリスト
- M folderName: 初期 HMM フォルダ名
- H hmmFile: HMM 定義ファイル
- I String: ラベル名
- L folderName: ラベルフォルダ名

# 11.6 HMM の学習



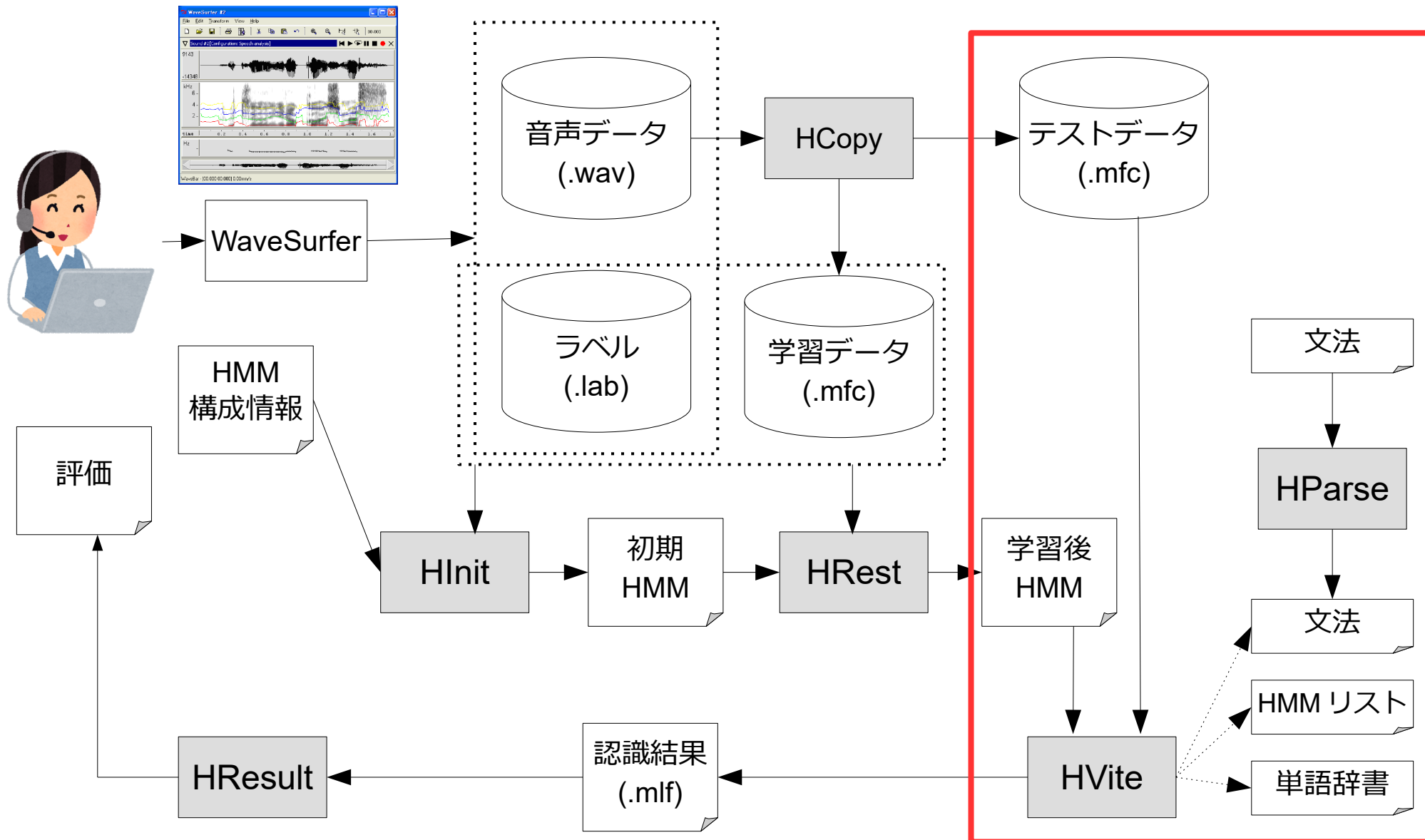
# 11.6 HMM の学習

- HRest コマンド
  - Baum-Welch アルゴリズムの実行

HRest [ オプション ] HMM 名

- T Num: 出力トレースレベル
- S listFile: 学習データリスト
- M folderName: 学習後の HMM フォルダ名
- H hmmFile: HMM 定義ファイル
- I String: ラベル名
- L folderName: ラベルフォルダ名

# 11.7 単語認識





## 11.7 単語認識

- 単語認識のための文法を記述
- 文法から HMM のネットワークを生成
- ビタビアルゴリズムによる認識の実行

## 11.7 単語認識

- 単語認識のための文法を記述
  - \$: 非終端記号
  - {}: 0 回以上の繰り返し
  - []: 0 回または 1 回の出現
  - (): 要素のまとまりを表現

```
$WORD = HAI | IIE;  
( {SIL} [$WORD] {SIL} )
```

## 11.7 単語認識

- HParse コマンド
  - 文法から HMM のネットワークを生成

HParse 文法ファイル名 ネットワークファイル名

## 11.7 単語認識

- HVite コマンド
  - ビタビアルゴリズムによる認識の実行

HVite [ オプション ] 単語辞書 HMM リスト 入力ファイル

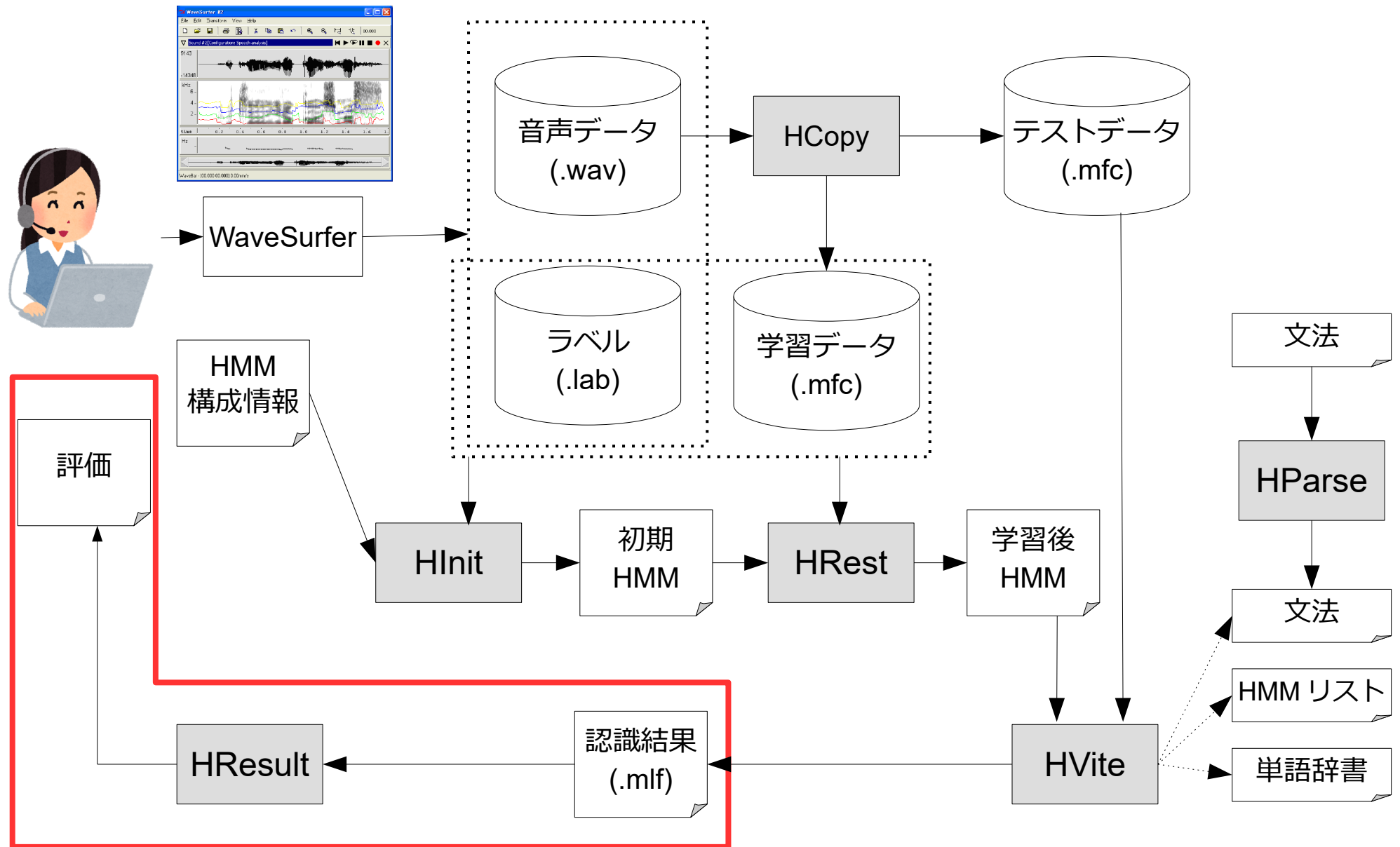
-T Num: 出力トレースレベル

-H hmmFile: HMM 定義ファイル

-i resultFile: 認識結果ファイル

-w networkFile: ネットワークファイル

# 11.8 認識率の評価



# 11.8 認識率の評価

- HResult コマンド
  - 認識率の計算

HResult [ オプション ] HMM リスト 認識結果

-e label1 label2 : label1 を label2 に置換

-I refFile: 正解ファイル

-L labelFolder: ラベルファイルのフォルダ

# 11.8 認識率の評価

- HResult コマンドの出力

```
===== HTK Results Analysis =====  
Date: Mon Aug 15 16:23:32 2015  
Ref : eval.mlf  
Rec : rec.mlf  
----- Overall Results -----  
SENT: %Correct=100 [H=10, S=0, N=10]  
WORD: %Corr=100, Acc=100 [H=10, D=0, S=0, I=0, N=10]  
=====
```

# 11.8 認識率の評価

- 評価基準
  - 系列パターン認識の評価
    - 置換・削除・挿入の誤り
    - 認識率 (Correct)

$$Correct = \frac{H}{N} \quad \text{H: 正解数、 N: データ数}$$

- 認識精度 (Accuracy)

$$Accuracy = \frac{H - I}{N}$$