

音声認識・音声対話技術講習会

実習2 HMMによる音響モデル

担当：南條浩輝

(龍谷大学理工学部情報メディア学科)

実習2の内容

- 孤立単語認識システムの構築（10数字）
 1. 音素HMMの学習と認識（60分）-- 講習会テキスト演習課題3に相当
 2. 話者適応(MLLR)と認識（30分）-- 講義2のプリント最後に相当
 3. 単語HMMの学習と認識（30分）-- 講習会テキスト演習課題2に相当
- ツール
HTK3.2 (Cambridge University)
フリー(要登録, <http://htk.eng.cam.ac.uk/>)
- 準備
 - ~/exercise/work で作業

```
% cd ~/exercise/work
```
 - この資料(pdf)を表示

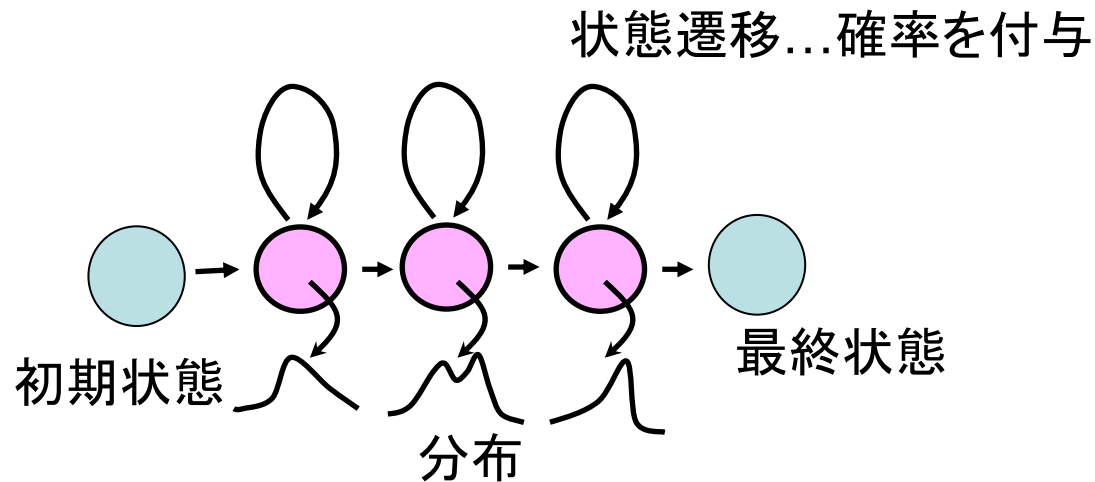
```
% acroread /n/media/sp/doc/excecize2.pdf
```

音響モデルのモデル化単位

- ワード(単語)
 - 大量の学習データ(各単語に対して)が必要
 - 単語内での音素のつながりをモデル化できる
 - サブワード...音素など
 - 40程度の音素の組み合わせですべての単語を表現できる
 - 前後の音素のつながりのモデル化が必要
- 実習ではサブワード(音素)によるモデル化を行った後、ワード(単語)によるモデル化を行う

HMM(隠れマルコフモデル)

- 複数の定常信号源(状態)で非定常信号(音声)をモデル化



3状態HMM

HTKにおける音素HMM表現

確認

```
% less ../model/phone/IPA97-0/hmmdefs.male
```

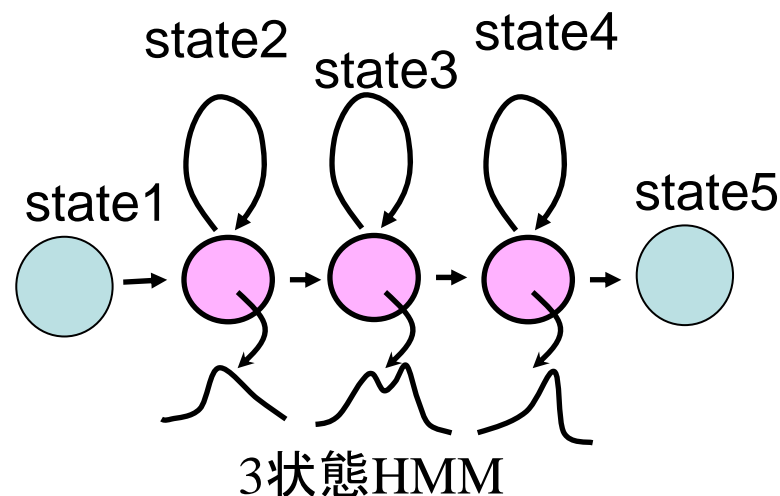
```
~0
<STREAMINFO> 1 25
<VECSIZE>25<NULLD><MFCC_E_D_N_Z>
~h "N"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 25
.....
<VARIANCE> 25
.....
<STATE> 3
.....
<TRANSP> 5
0.0 1.0 .....
.....
```

ベクトルは25次元
MFCC+Energy, そのdelta
(Energyの絶対値は使用しない)

モデル名(音素名)はN
状態数は5

第2状態の
平均(25次元)
分散(25次元)

状態遷移確率
(行列形式)



1. 音素HMMの学習と認識

講習会テキスト演習課題3に相当

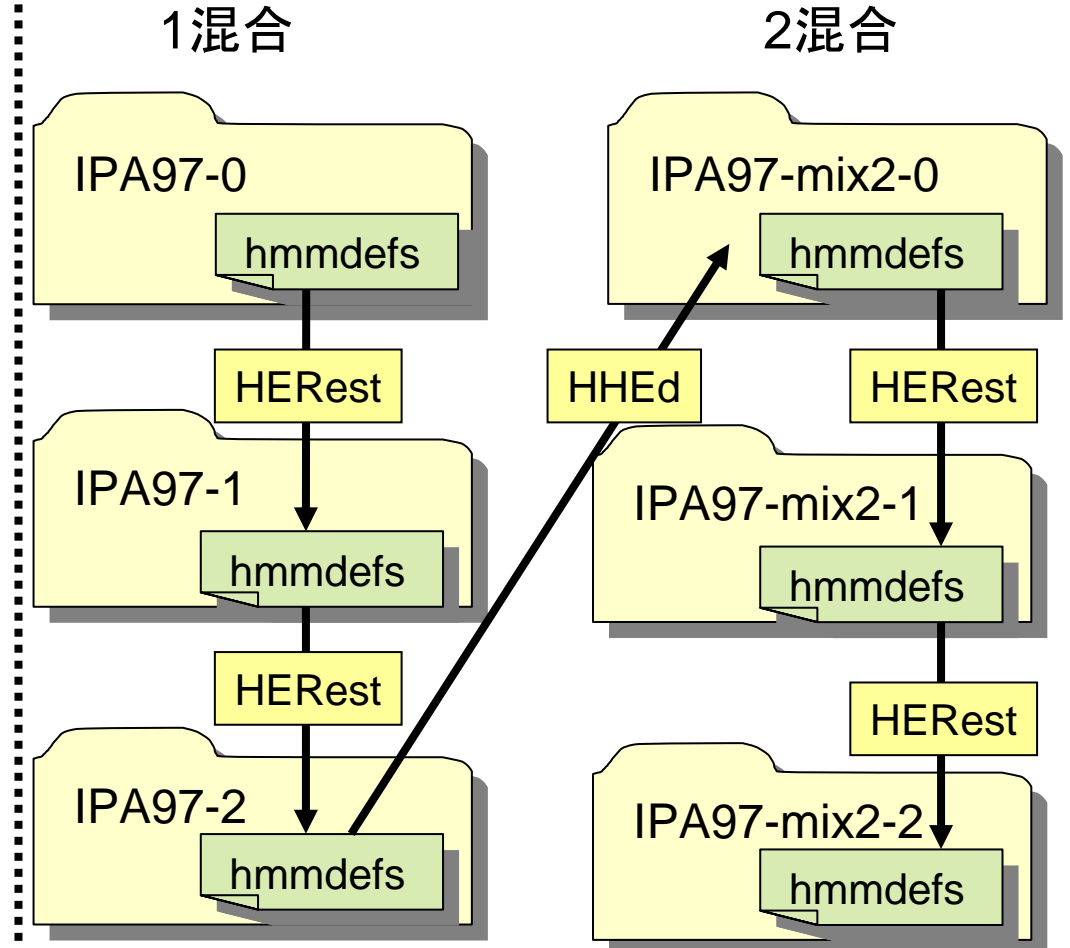
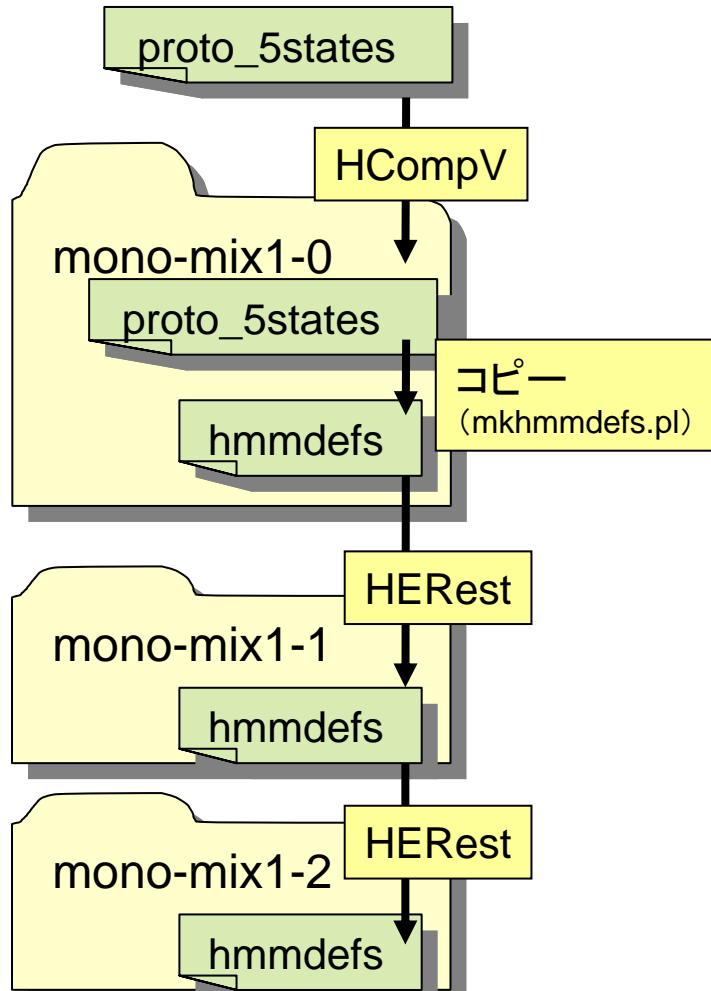
概要

1. 初期モデルの作成(HCompV)
2. 連結学習 (HERest)
3. 混合数の増加 (HHEd)
4. 認識 (HVite)
5. 認識率の計算 (HResults)

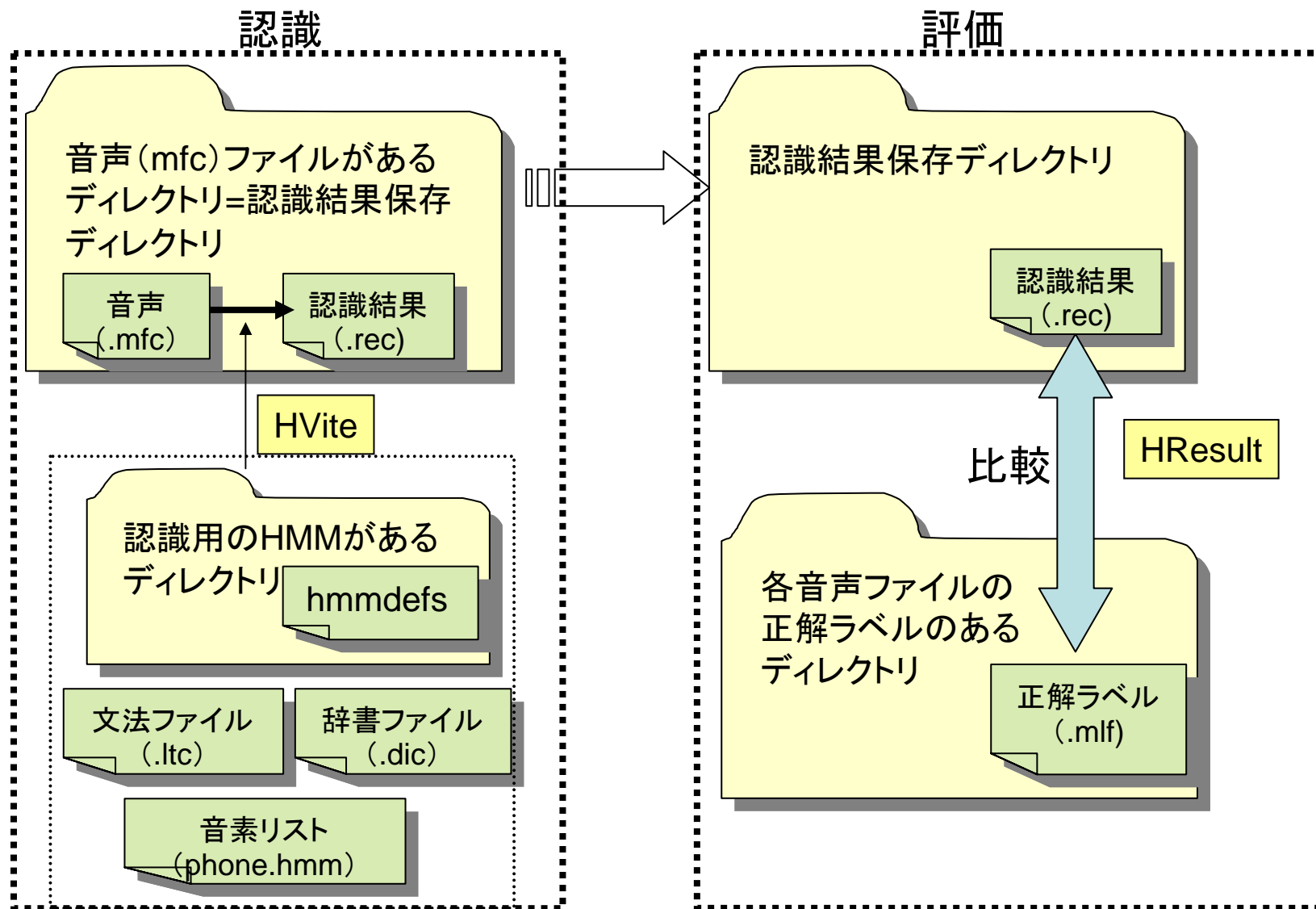
1-1. HMM学習の流れ

既存モデルの利用
学習→混合数増加→学習

初期モデルの作成から学習まで



1-1. 音素HMMを用いた認識と評価の流れ



1-2. 音素HMMの学習と認識

初期モデルの作成(1)

【フラットスタート】

モデルの平均および分散の初期値を推定(全モデルで初期値を等しくする)
与えられた音声(音素バランス文50文)で推定

```
% HCompV -T 1  
-C ../config/config.train  
-m  
-v 0.01  
-M ../model/phone/mono-mix1-0  
proto_5states  
../mfcc/balance.xxxx/*.mfc
```

女性はfemale
男性はmale

proto_5states

読み上げ50文
で平均と分散
を推定

mono-mix1-0

proto_5states

1-2. 音素HMMの学習と認識

初期モデルの作成(2)

確認

```
% less proto_5states          : 学習前
% cd ../model/phone/mono-mix1-0/
% less proto_5states          : 学習後
```

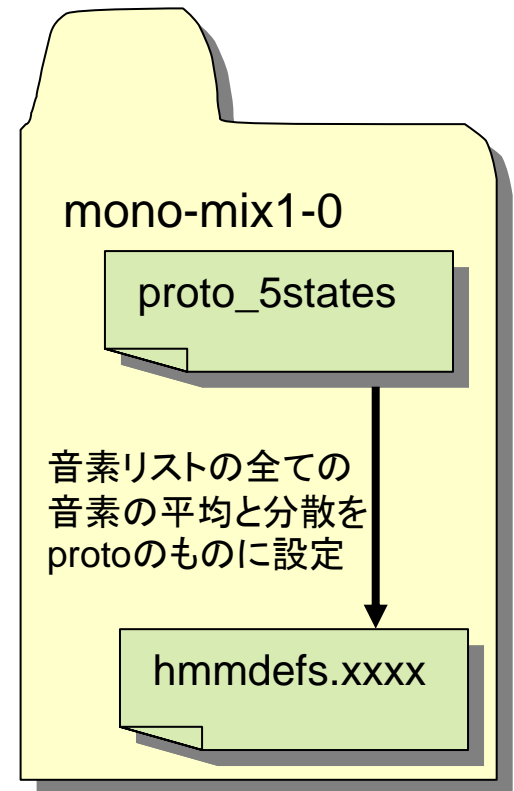
平均と分散を全ての音素モデルに割り当て

```
% ./mkhmmdefs.pl proto_5states
../.../work/monophones > hmmdefs.xxxx
```

女性はfemale
男性はmale

確認

```
% less hmmdefs.xxxx
% cd ../.../work
```



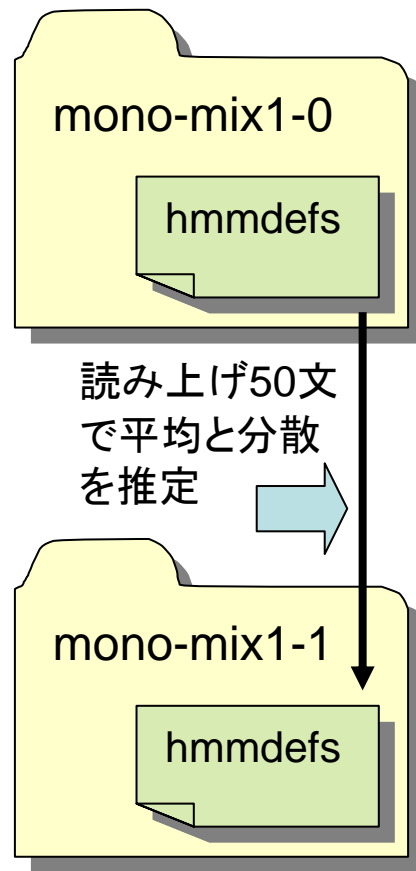
1-2. 音素HMMの学習と認識

学習

【連結学習】 音素バランス文50文で連結学習

```
% HERest -T 1
          -C ../config/config.train
          -v 0.01
          -l ../lab/train.mlf
          -H ../model/phone/mono-mix1-0/hmmdefs.xxxx
          -M ../model/phone/mono-mix1-1
            phone.hmm
            ../mfcc/balance.xxxx/*.mfc
```

女性はfemale
男性はmale



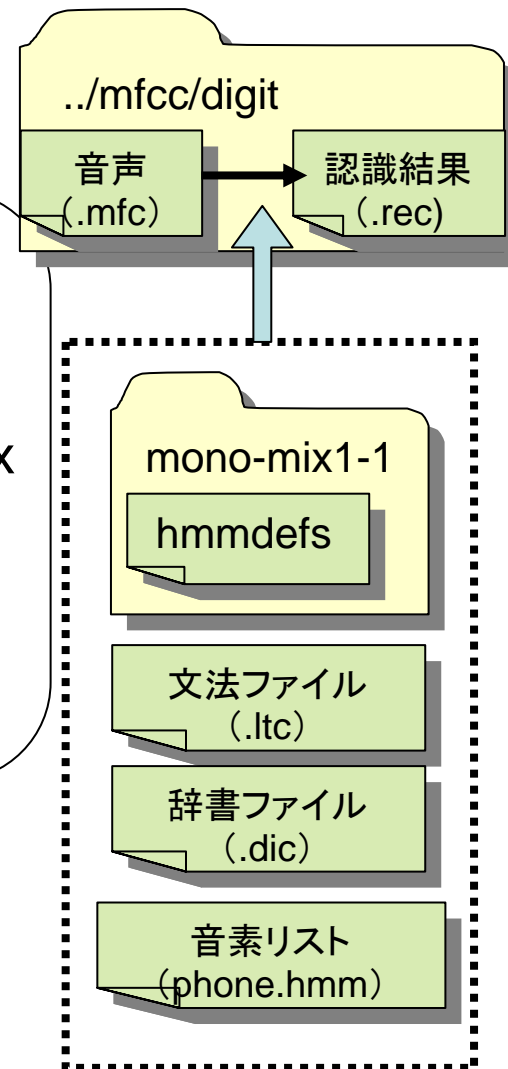
1-2. 音素HMMの学習と認識

認識(1)

【認識】数字発声50発話を認識

```
% HVite -T 1
-C ../config/config.rec
-w digit2.phone.ltc
-H ../model/phone/mono-mix1-1/hmmdefs.xxxx
digit2.phone.dic
phone.hmm
../mfcc/digit/d*.mfc
```

女性はfemale
男性はmale

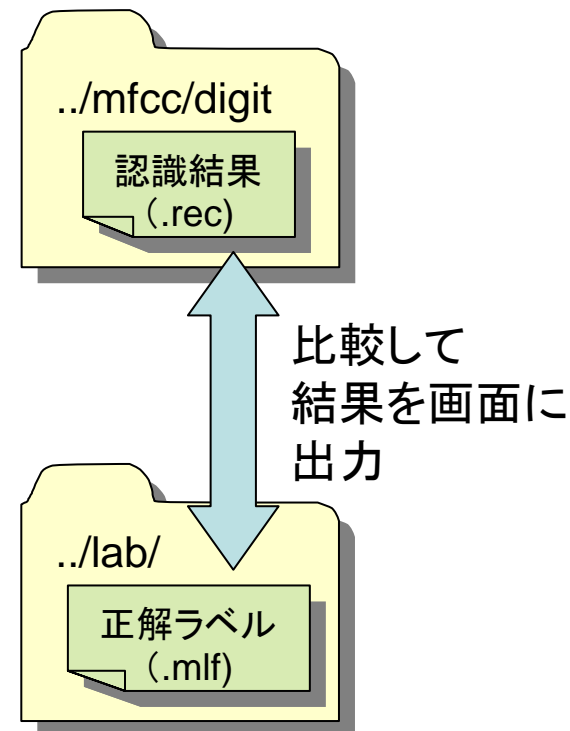


1-2. 音素HMMの学習と認識

認識(2)

【認識率の計算】

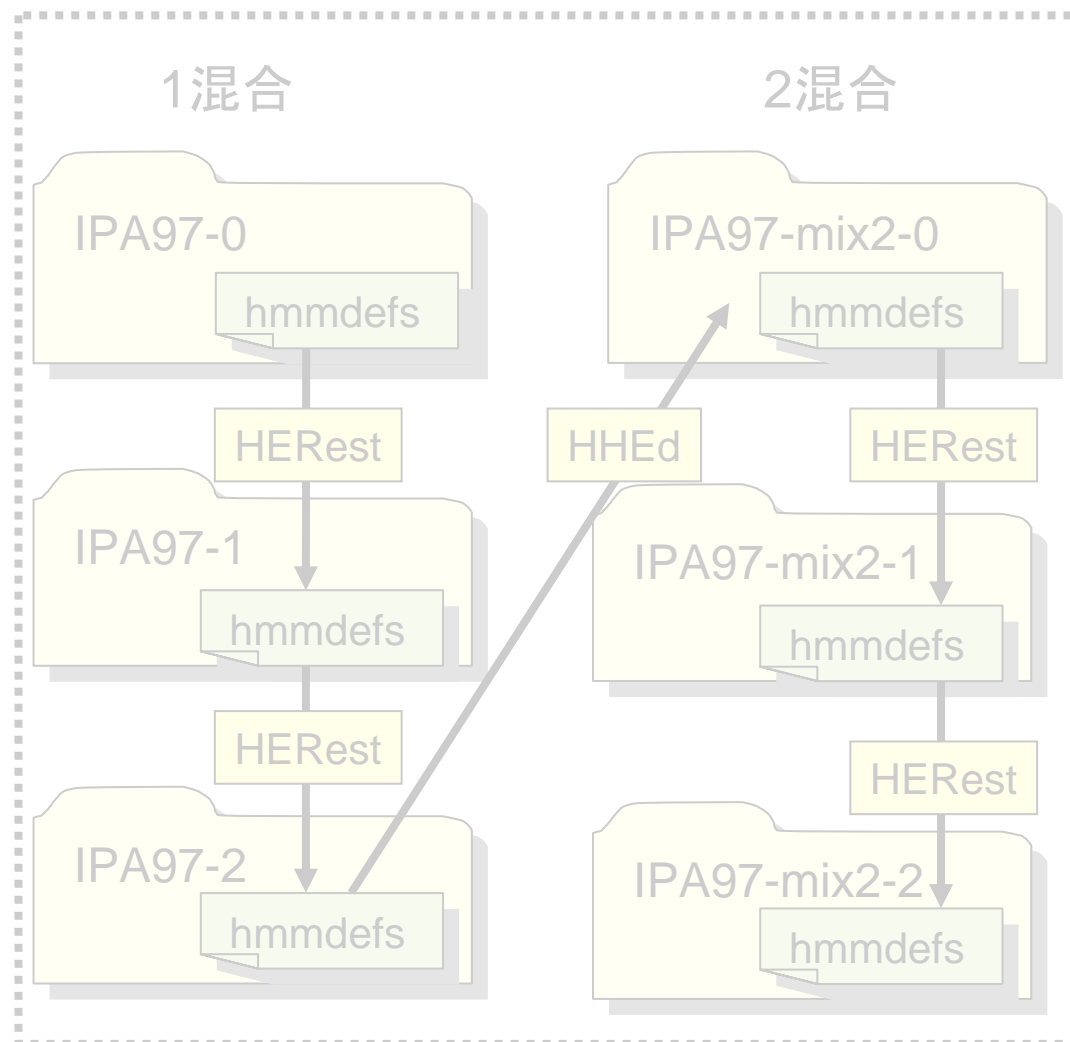
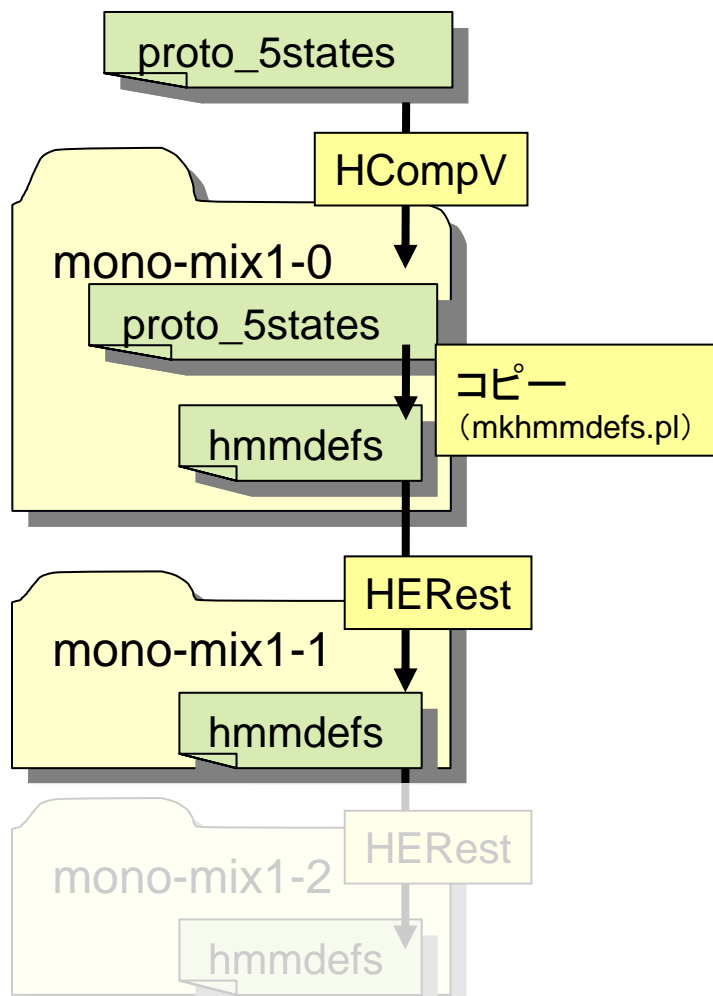
```
% HResults -T 1  
            -l ../lab/digit2.mlf  
            phone.hmm  
            ../mfcc/digit/*.rec
```



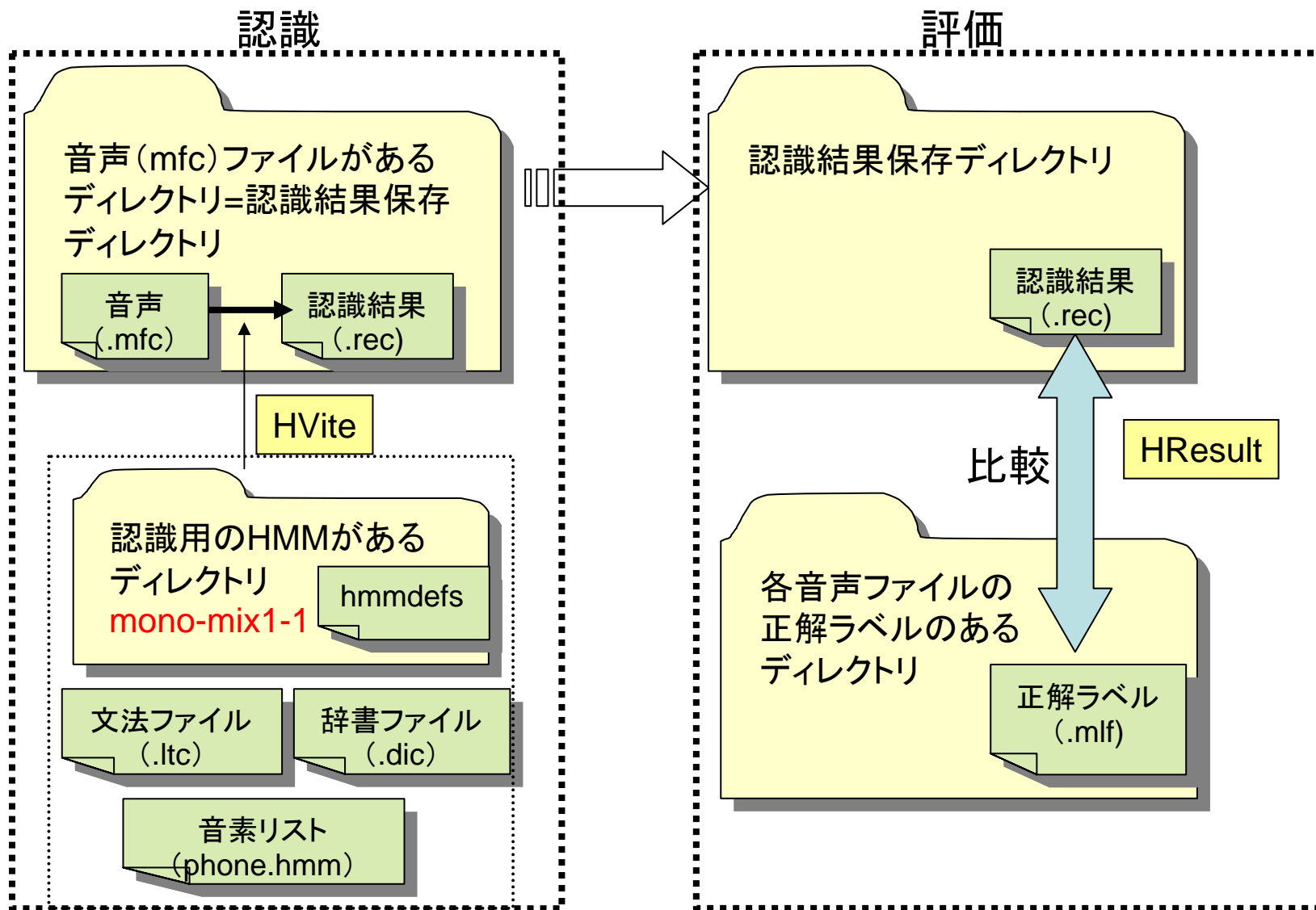
1-2. 行った作業

既存モデルの利用
学習→混合数増加→学習

初期モデルの作成から学習まで



1-2. 行った作業



1-3. 音素HMMの学習と認識

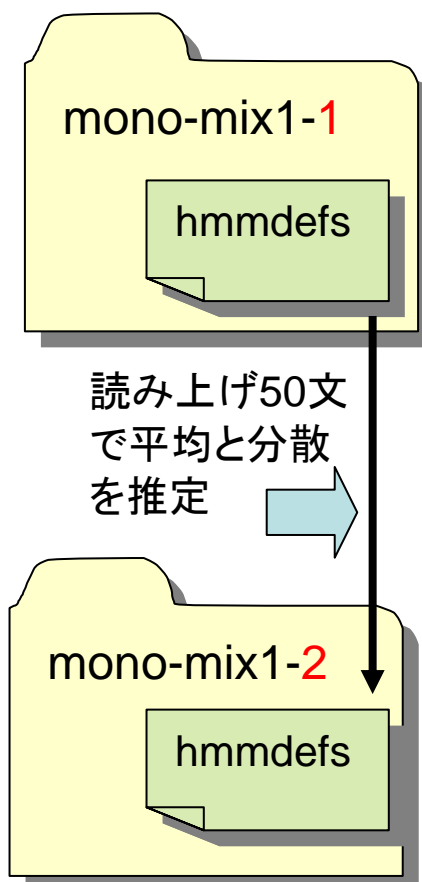
繰り返し学習

【連結学習(2回目)】

音素バランス文50文でさらに連結学習

```
% HERest -T 1
          -C ../config/config.train
          -v 0.01
          -I ../lab/train.mlf
          -H ../model/phone/mono-mix1-1/hmmdefs.xxxx
          -M ../model/phone/mono-mix1-2
phone.hmm
../mfcc/balance.xxxx/*.mfc
```

女性はfemale
男性はmale



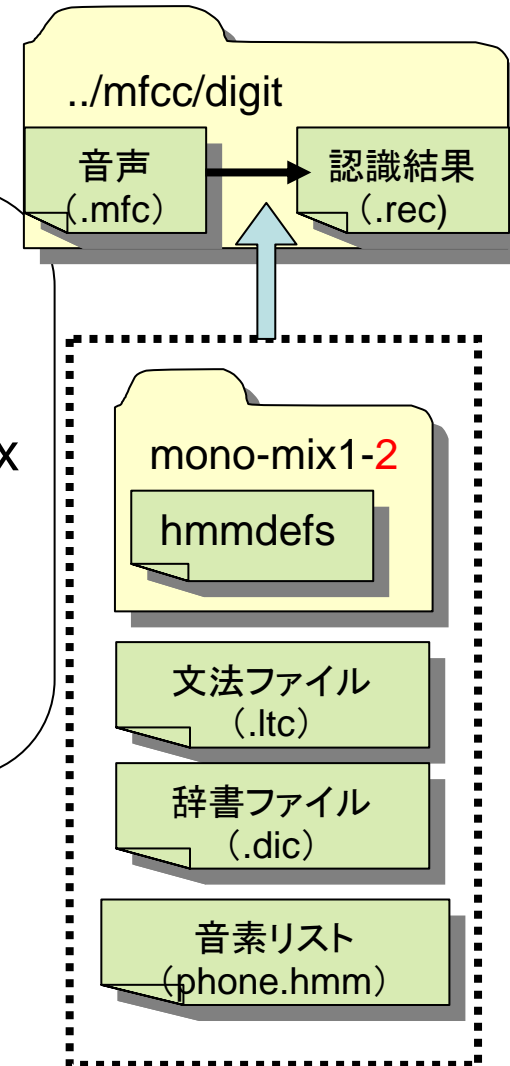
1-3. 音素HMMの学習と認識

繰り返し学習

【認識】 数字発声50発話を認識

```
% HVite -T 1  
-C ../config/config.rec  
-w digit2.phone.ltc  
-H ../model/phone/mono-mix1-2/hmmdefs.xxxx  
digit2.phone.dic  
phone.hmm  
../mfcc/digit/d*.mfc
```

女性はfemale
男性はmale



1-3. 音素HMMの学習と認識

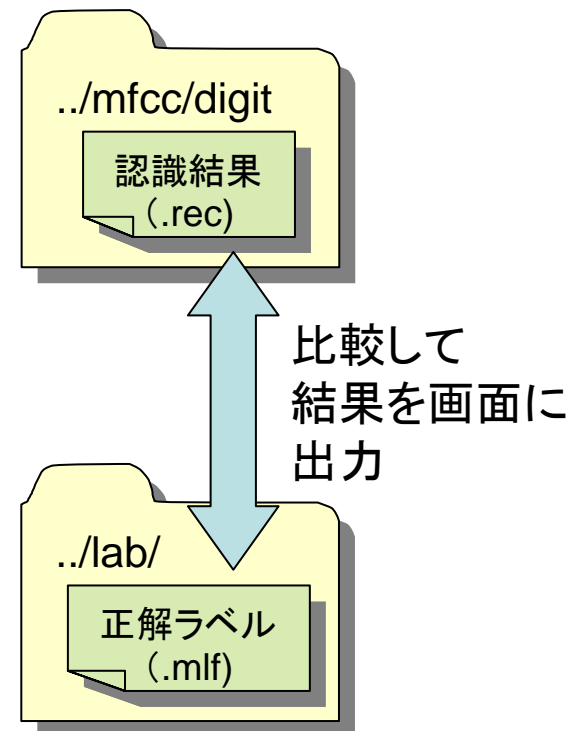
繰り返し学習

【認識率の計算】

```
% HResults -T 1  
            -l ../lab/digit2.mlf  
            phone.hmm  
            ../mfcc/digit/*.rec
```

【演習】

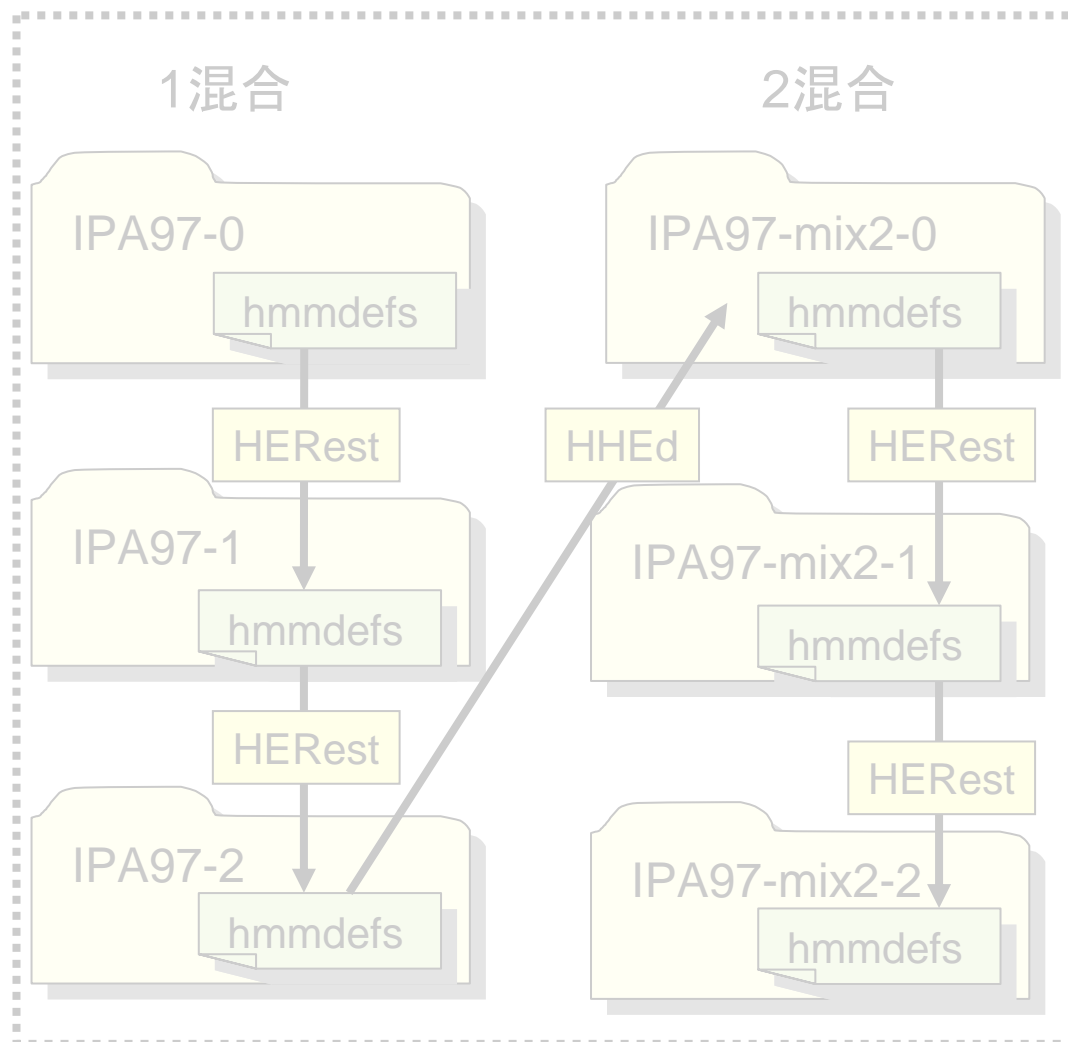
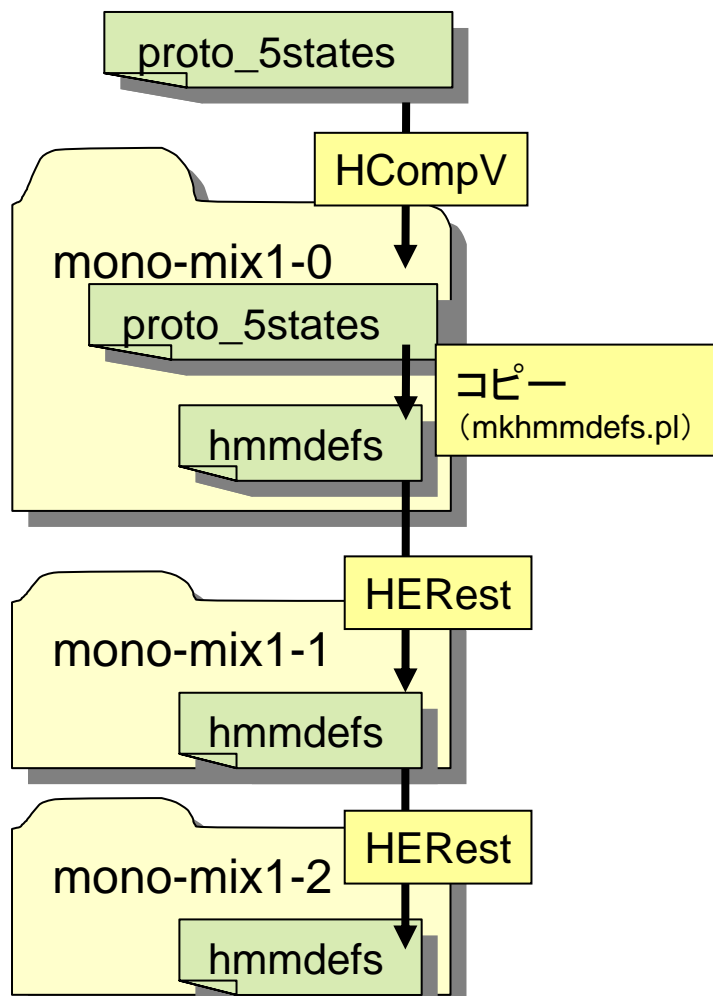
連結学習1回目の結果と比較
してみよ. (認識率, ゆう度など)



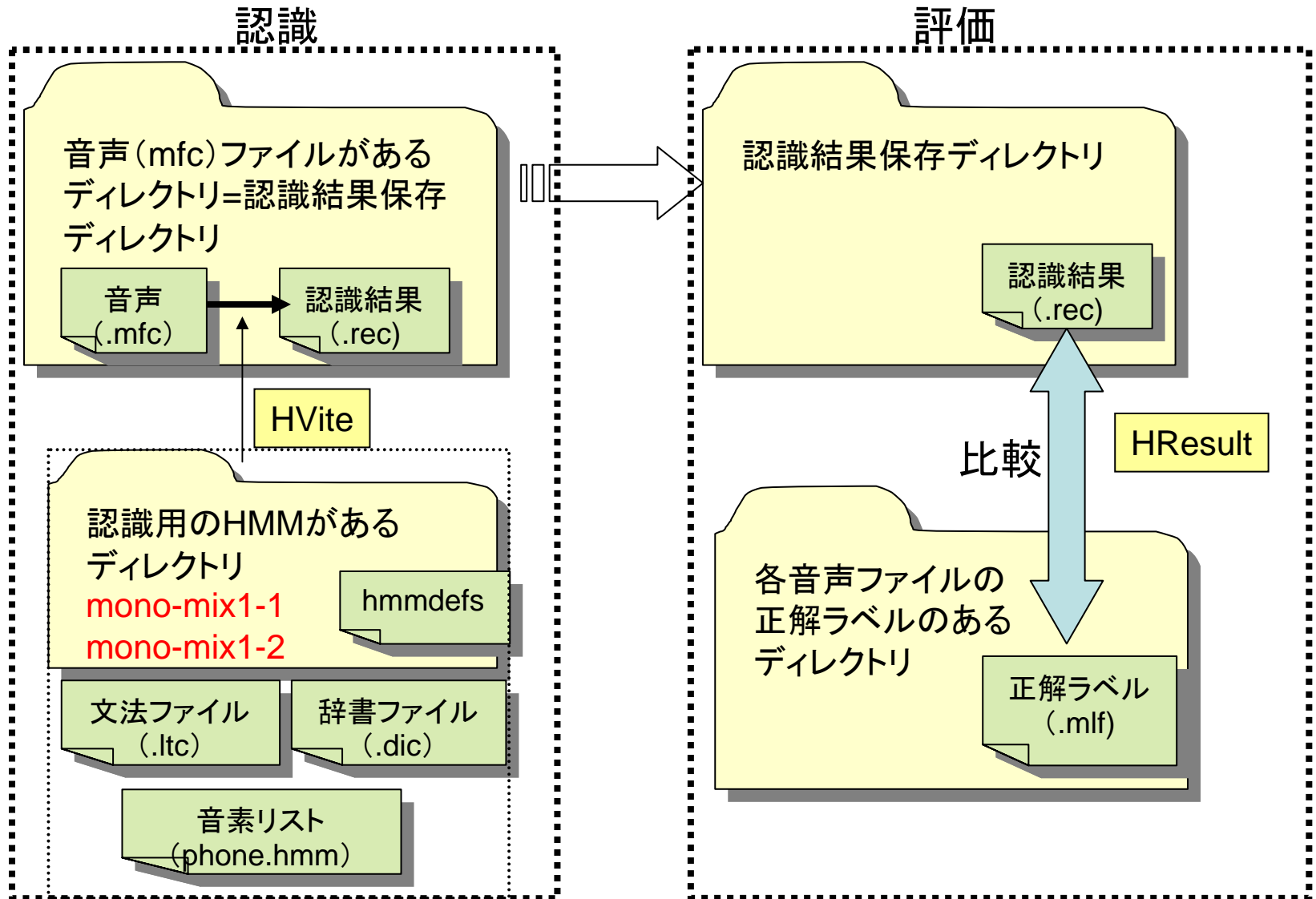
1-2～1-3. 行った作業

既存モデルの利用
学習→混合数増加→学習

初期モデルの作成から学習まで



1-2～1-3. 行った作業



1-3. [比較]既存音素HMMを用いた認識

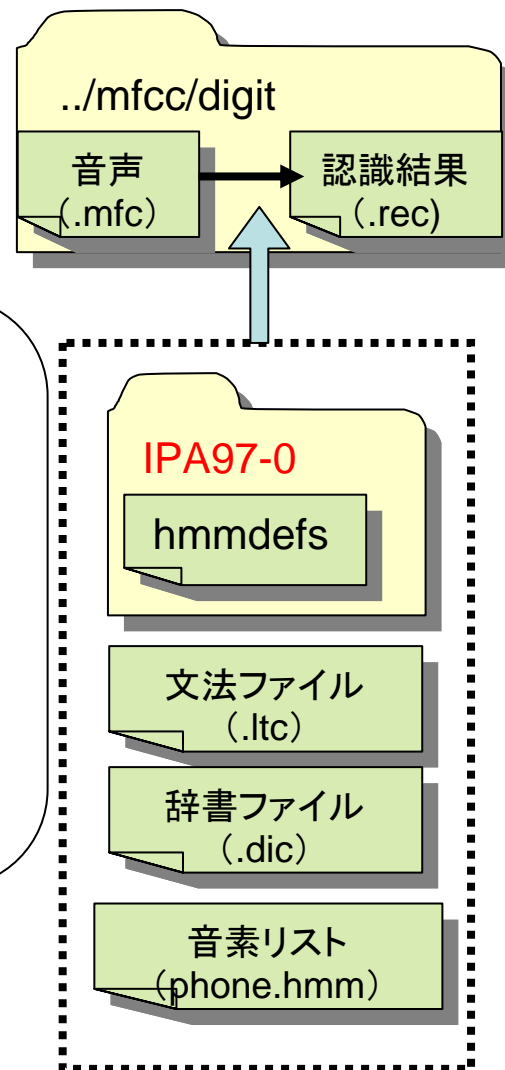
【認識】 数字発声50発話を認識

大規模な学習データで学習した既存の
音素HMM(IPAモデル)を用いる

```
% HVite -T 1
-C ../config/config.rec
-w digit2.phone.ltc
-H ../model/phone/IPA97-0/hmmdefs.xxxx
digit2.phone.dic
phone.hmm
../mfcc/digit/d*.mfc
```

女性はfemale
男性はmale

【演習】 自分で作成したモデルを用いた
場合の結果と比較せよ。(認識率, ゆう度など)



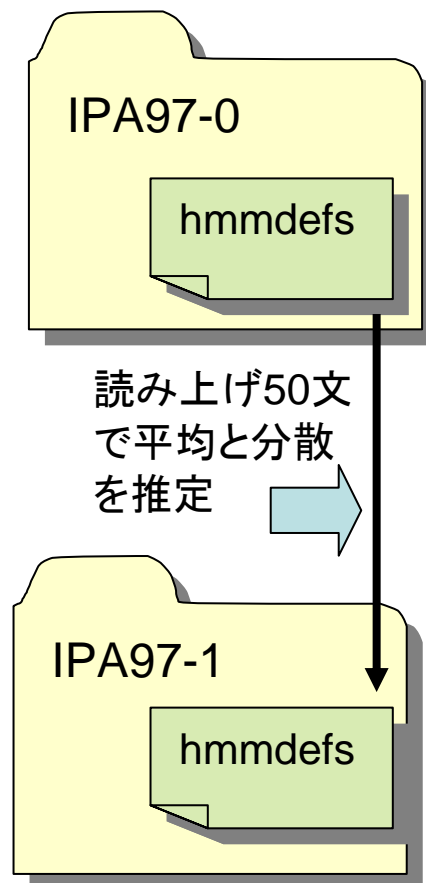
1-4. 既存音素HMMを初期モデルとした 音響モデルの学習と認識

【連結学習(1回目)】

予め用意した50文の発話と自分で発声した収録発話20文で連結学習

```
% HERest -T 1
-C ../config/config.train
-v 0.01
-l ../lab/train.mlf
-H ../model/phone/IPA97-0/hmmdefs.xxxx
-M ../model/phone/IPA97-1
phone.hmm
../mfcc/balance.xxxx/*.mfc
../mfcc/balance/*.mfc
```

女性はfemale
男性はmale



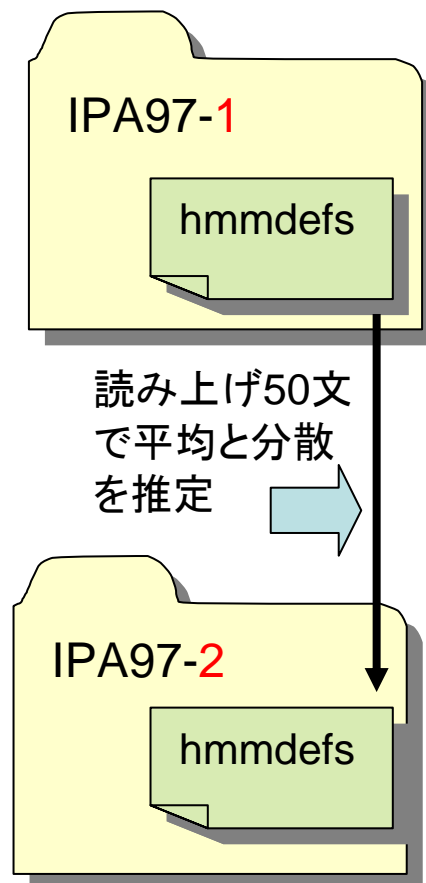
1-4. 既存音素HMMを初期モデルとした 音響モデルの学習と認識

【連結学習(2回目)】

予め用意した50文の発話と自分で発声した収録発話20文でさらに連結学習

```
% HERest -T 1
-C ../config/config.train
-v 0.01
-l ../lab/train.mlf
-H ../model/phone/IPA97-1/hmmdefs.xxxx
-M ../model/phone/IPA97-2
phone.hmm
../mfcc/balance.xxxx/*.mfc
../mfcc/balance/*.mfc
```

女性はfemale
男性はmale

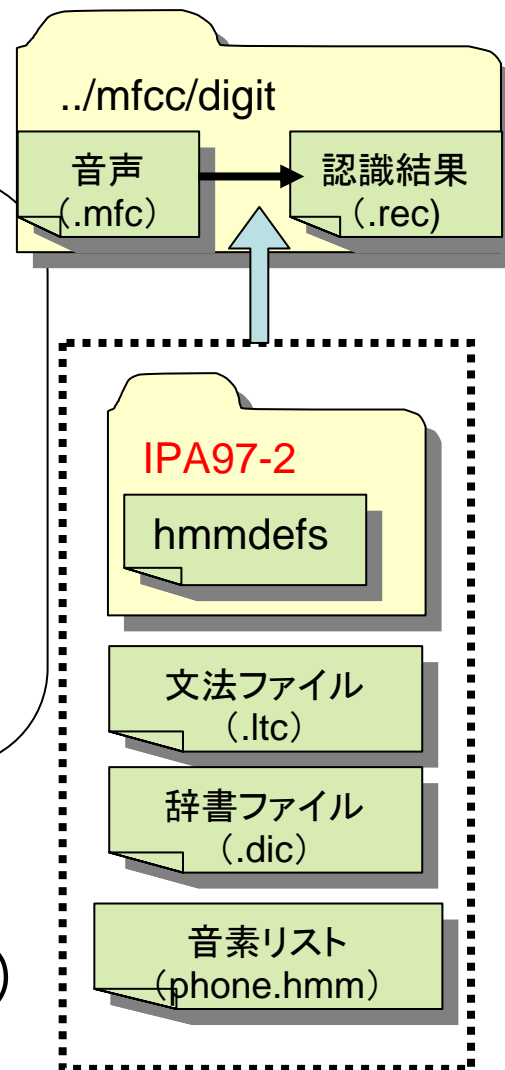


1-4. 既存音素HMMを初期モデルとした 音響モデルの学習と認識

【認識】 数字発声50発話を認識

```
% HVite -T 1  
-C ../config/config.rec  
-w digit2.phone.ltc  
-H ../model/phone/IPA97-2/hmmdefs.xxxx  
digit2.phone.dic  
phone.hmm  
../mfcc/digit/d*.mfc
```

女性はfemale
男性はmale

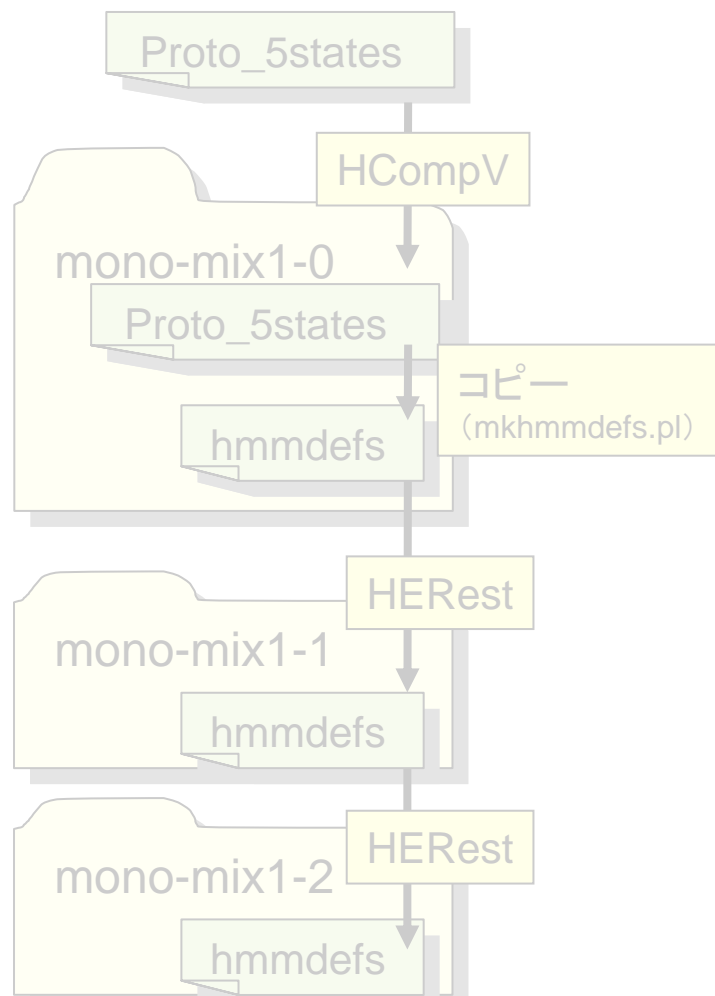


【演習】 学習前のモデルを用いた場合の
認識結果と比較せよ。（認識率, ゆう度など）

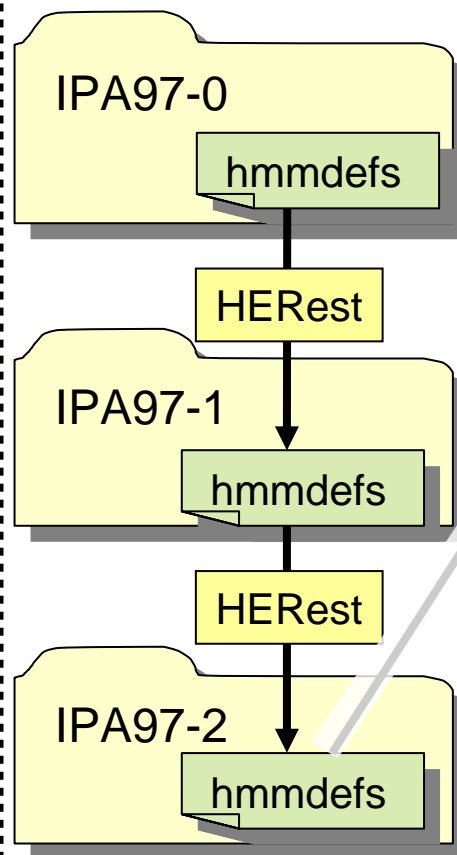
1-4. 行った作業

既存モデルの利用
学習→混合数増加→学習

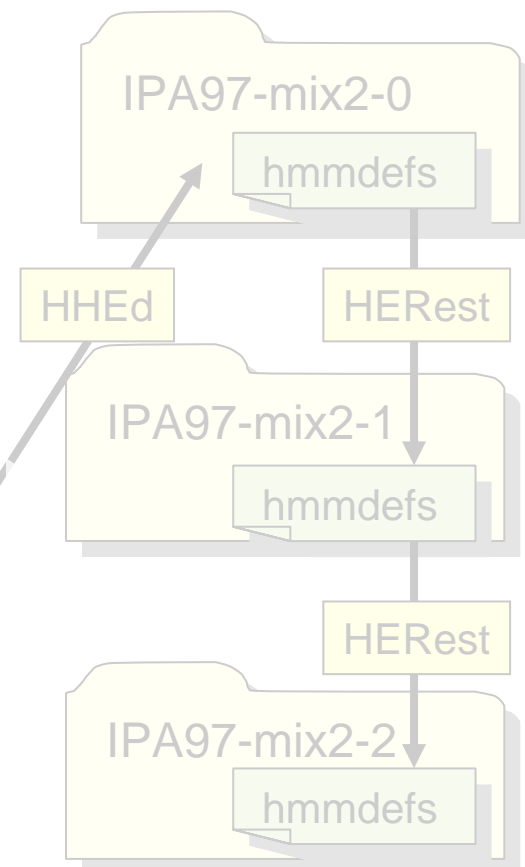
初期モデルの作成から学習まで



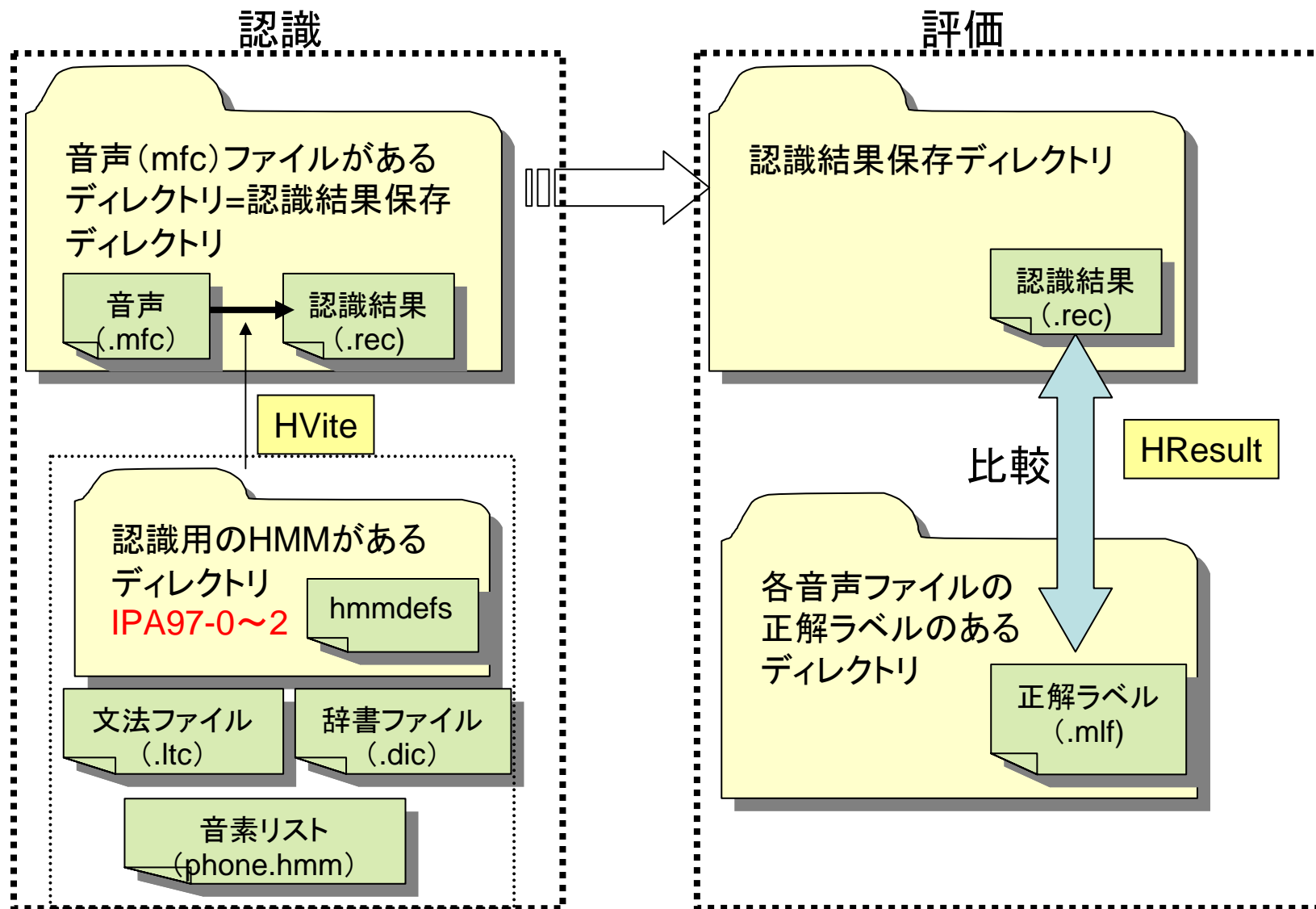
1混合



2混合



1-4. 行った作業



1-5. 音素HMMの学習と認識 (混合数の増加)

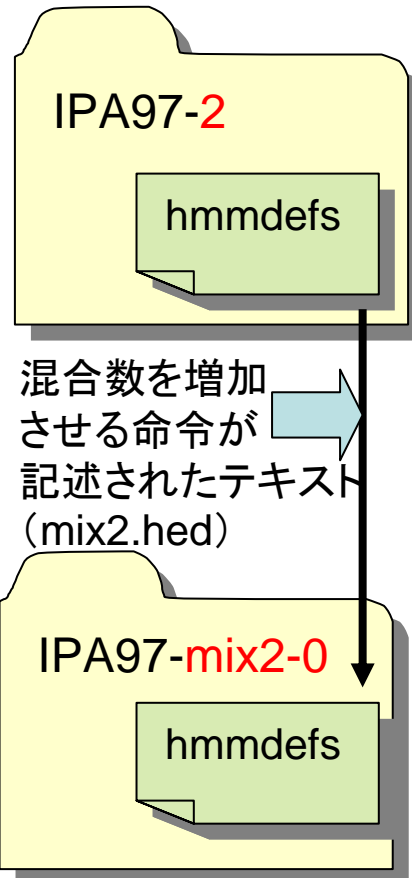
【混合数の増加(1→2)】

```
% HHEd -T 1  
-H ../model/phone/IPA97-2/hmmdefs.xxxx  
-M ../model/phone/IPA97-mix2-0  
mix2.hed  
phone.hmm
```

女性はfemale
男性はmale

確認

```
% less ../model/phone/IPA97-mix2-0/hmmdefs.xxxx
```

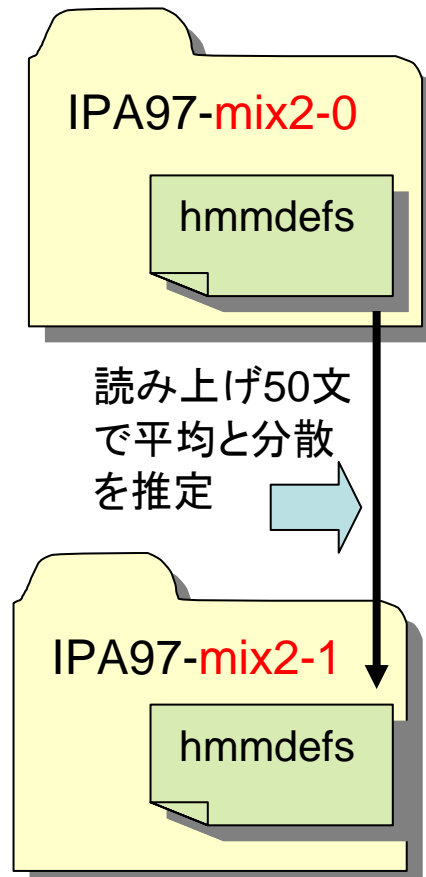


1-5. 音素HMMの学習と認識 (混合数の増加)

【連結学習】

```
% HERest -T 1
-C ../config/config.train
-v 0.01
-I ../lab/train.mlf
-H ../model/phone/IPA97-mix2-0/hmmdefs.xxxx
-M ../model/phone/IPA97-mix2-1
phone.hmm
../mfcc/balance.xxxx/*.mfc
../mfcc/balance/*.mfc
```

女性 ← 女性はfemale
男性 ← 男性はmale



【演習】2度目の連結学習を行え.

保存先のディレクトリは「IPA97-mix2-2」とせよ

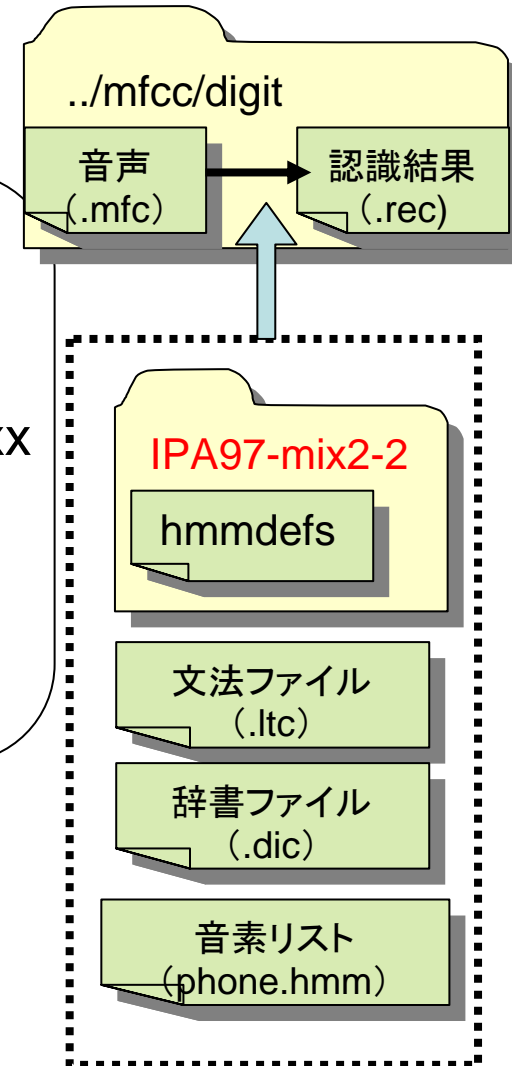
1-5. 音素HMMの学習と認識 (混合数の増加)

【認識】 数字発声50発話を認識

```
% HVite -T 1
-C ../config/config.rec
-w digit2.phone.ltc
-H ../model/phone/IPA97-mix2-2/hmmdefs.xxxx
digit2.phone.dic
phone.hmm
../mfcc/digit/d*.mfc
```

女性はfemale
男性はmale

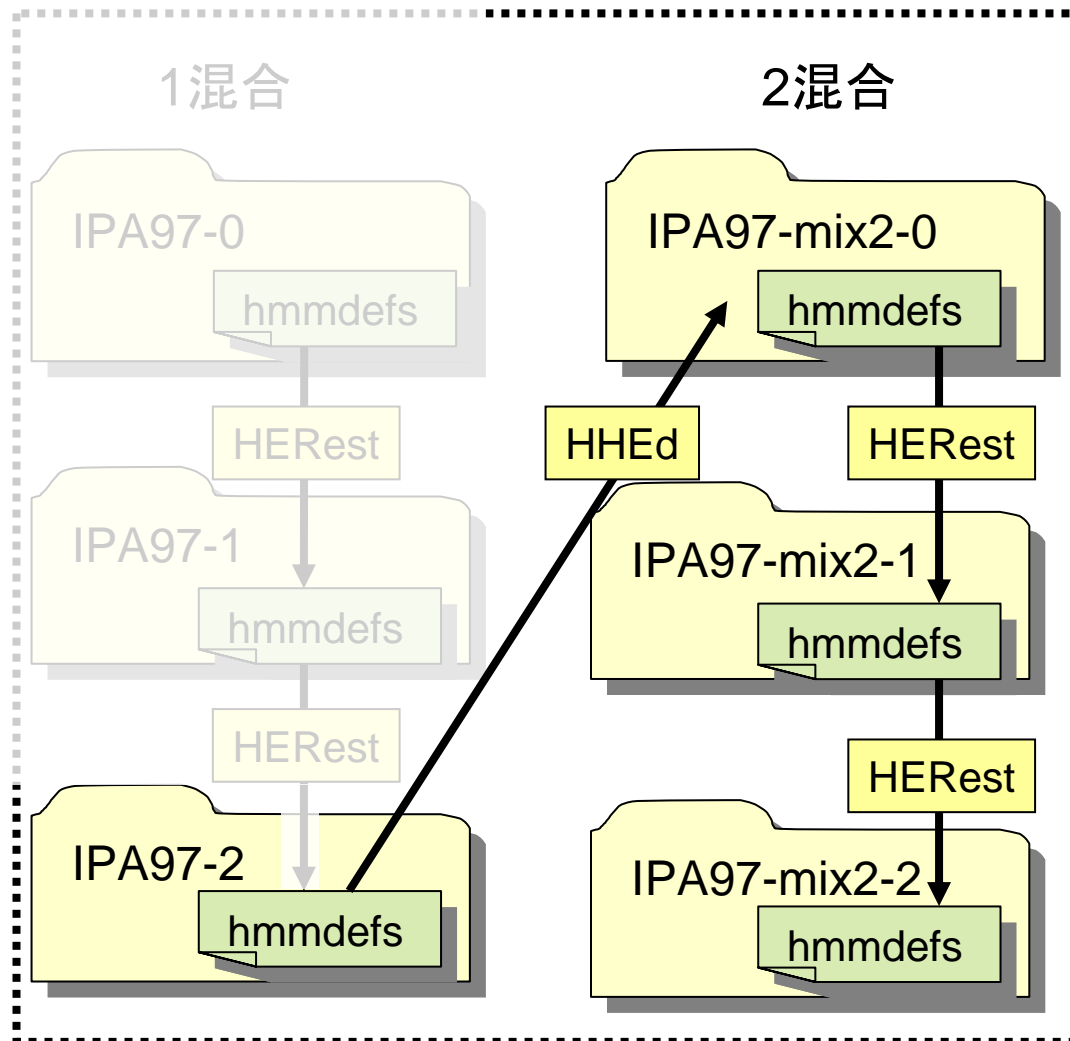
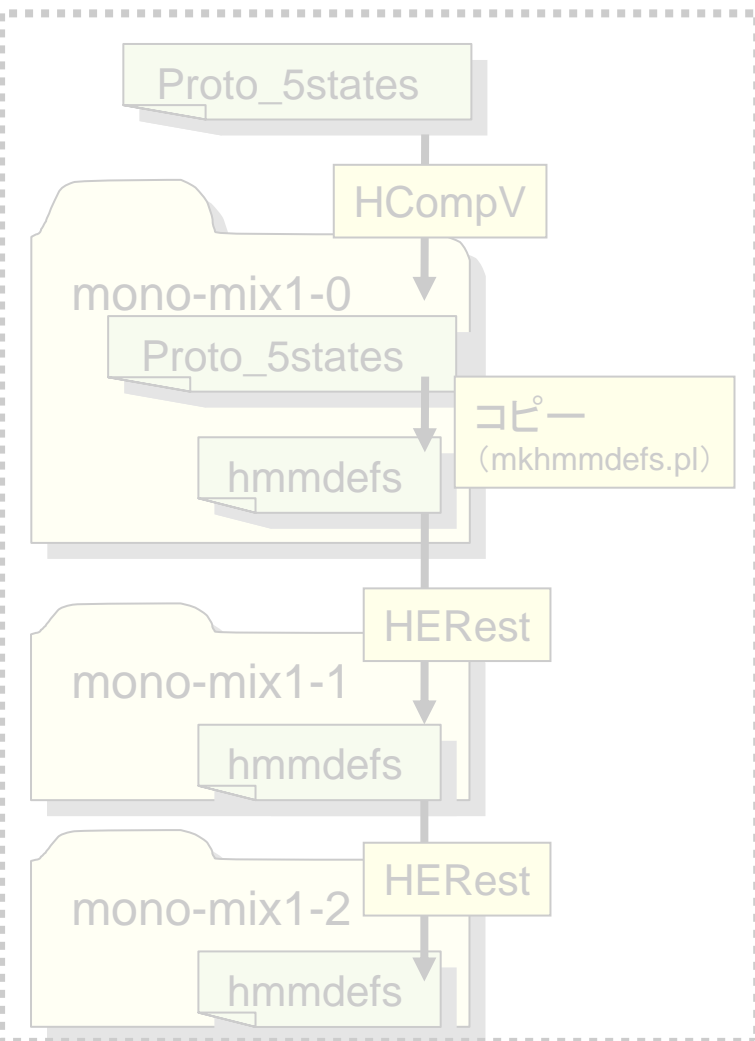
【演習】 学習前のモデルを用いた場合の
認識結果と比較せよ。(認識率, ゆう度など)



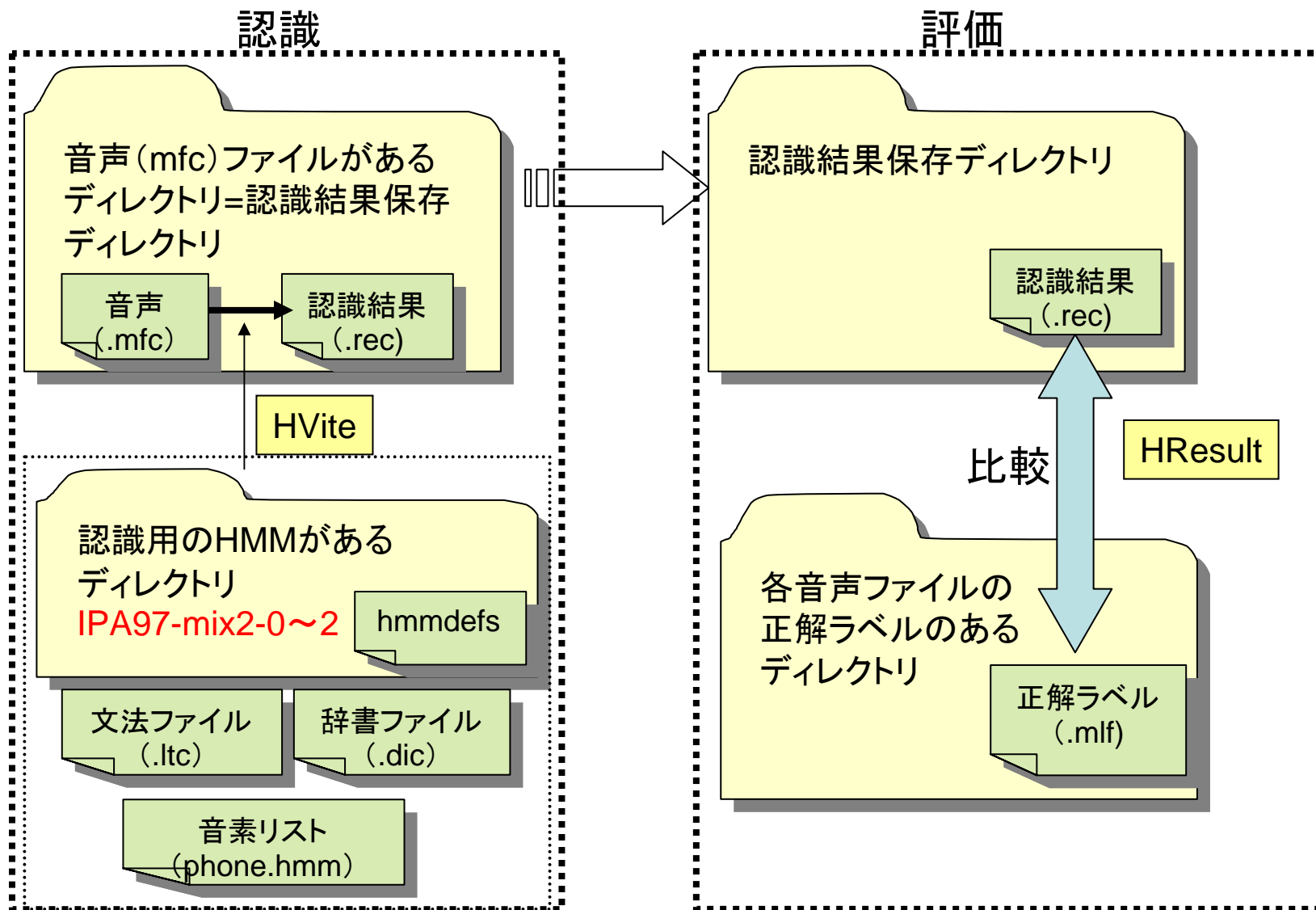
1-5. 行った作業

既存モデルの利用
学習→混合数増加→学習

初期モデルの作成から学習まで



1-5. 行った作業



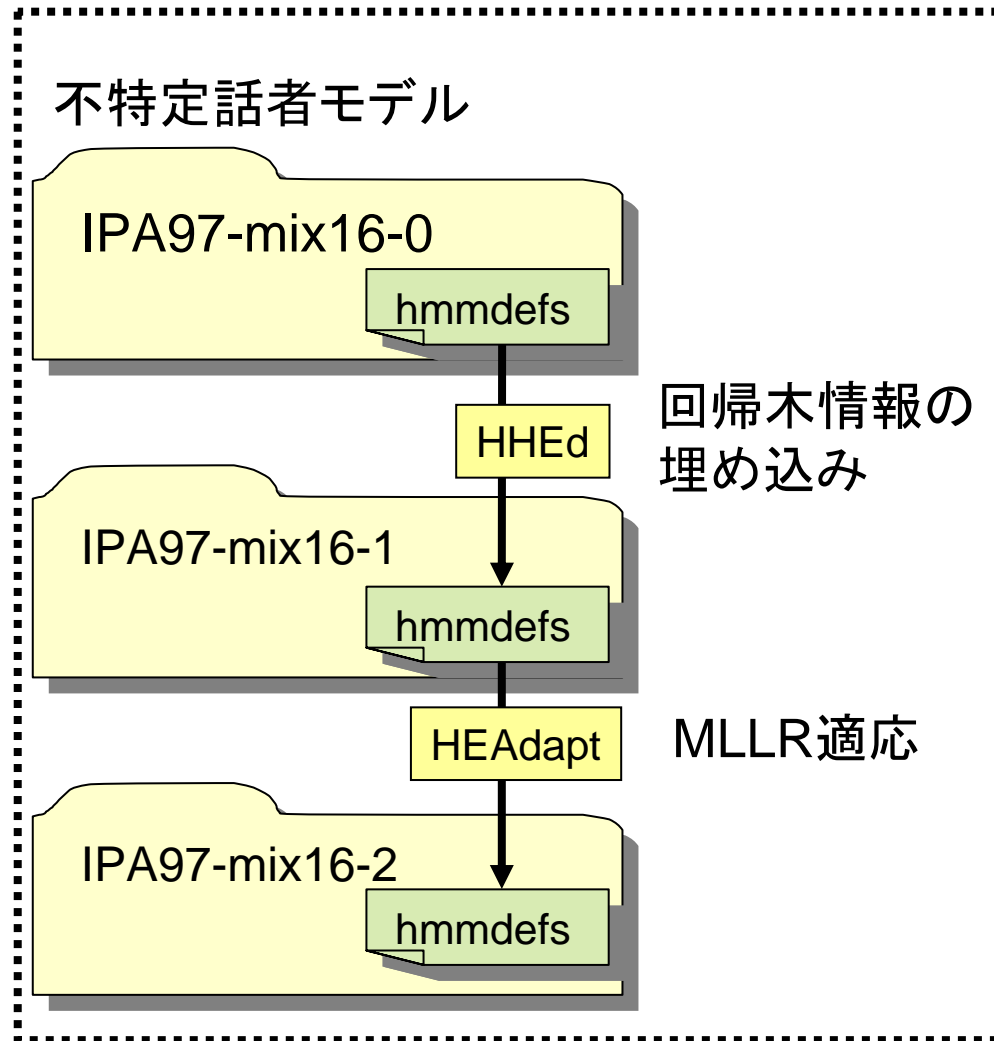
2. 話者適応(MLLR)

講義2のプリントの終盤部に対応

概要

1. 回帰木情報の埋め込み(クラスタリング) [HHEd]
2. 話者適応(MLLR) [HEAdapt]
3. 認識 [Hvite]
4. Julius用音響モデルの作成 [HEAdapt,mktmix.pl]

2. MLLR適応の流れ



2-1. 話者適応(MLLR)

回帰木情報の埋め込み(クラスタリング)

各分布の平均・分散を線形移動
(類似する分布は同じ方向に移動)

【クラスタリング】 類似する分布をまとめる

```
% HHEd -T 1  
-H ../model/phone/IPA97-mix16-0/hmmdefs  
-M ../model/phone/IPA97-mix16-1  
mlr_monof-mix16.hed  
phone.hmm
```

回帰木情報の確認

```
% less ../model/phone/IPA97-mix16-1/hmmdefs
```

2-1. 話者適応(MLLR) 適応

【適応】収録した20文で適応

```
% HEAdapt -T 1  
-C ../config/config.train  
-I ../lab/train.mlf  
-H ../model/phone/IPA97-mix16-1/hmmdefs  
-M ../model/phone/IPA97-mix16-2  
phone.hmm  
../mfcc/balance/*.mfc
```

※最新版のHTK(バージョン3.3)では, HEAdapt は HERest に統合されている
3.3を利用する際は, 注意が必要

2-1. 話者適応(MLLR)

適応後のモデルを用いた認識

【演習】話者適応の前後のモデルを用いて認識を実行し、尤度と認識率を比べてみること。

認識

```
% HVite -T 1
-C ../config/config.rec
-w digit2.phone.ltc
-H ../model/phone/IPA97-mix16-X/hmmdefs
digit2.phone.dic
phone.hmm
../mfcc/digit/d*.mfc
```

↑
1 or 2

評価 HResultを利用

2-2. 話者適応(MLLR)

Julius用 PTMモデルの作成(適応)

JuliusでPTMモデルを利用した認識を行うには, HMMファイルの記述形式を変換しておくこと(tmix化)が重要

PTMモデルの適応(MLLR)

```
% HEAdapt
-T 1
-C ../config/config.train
-I ../lab/train-tri.mlf
-H ../model/phone/ptm-3000x64-0/hmmdefs,rclass
-M ../model/phone/ptm-3000x64-1
logicalTri
../mfcc/balance/*.mfc
```

2-2. 話者適応(MLLR)

Julius用PTMモデルの作成(形式変換)

変換コマンド

```
% mktmix.pl  
< ../model/phone/ptm-3000x64-1/hmmdefs,rclass  
> ../model/phone/ptm-3000x64-1/hmmdefs,rclass,tmix
```

カンマ



このモデル(ファイル名 “**hmmdefs,rclass,tmix**”)は
ディクテーションの実習(実習5)で使用するので必ず作成すること！

3. 単語HMMの学習と認識

講習会テキスト演習課題2に相当

小語彙音声認識には単語認識が有効

-- 音素間の渡りの部分(調音結合)のモデル化が可能

本演習では、数字を認識するための音響モデルを学習する

※数字認識

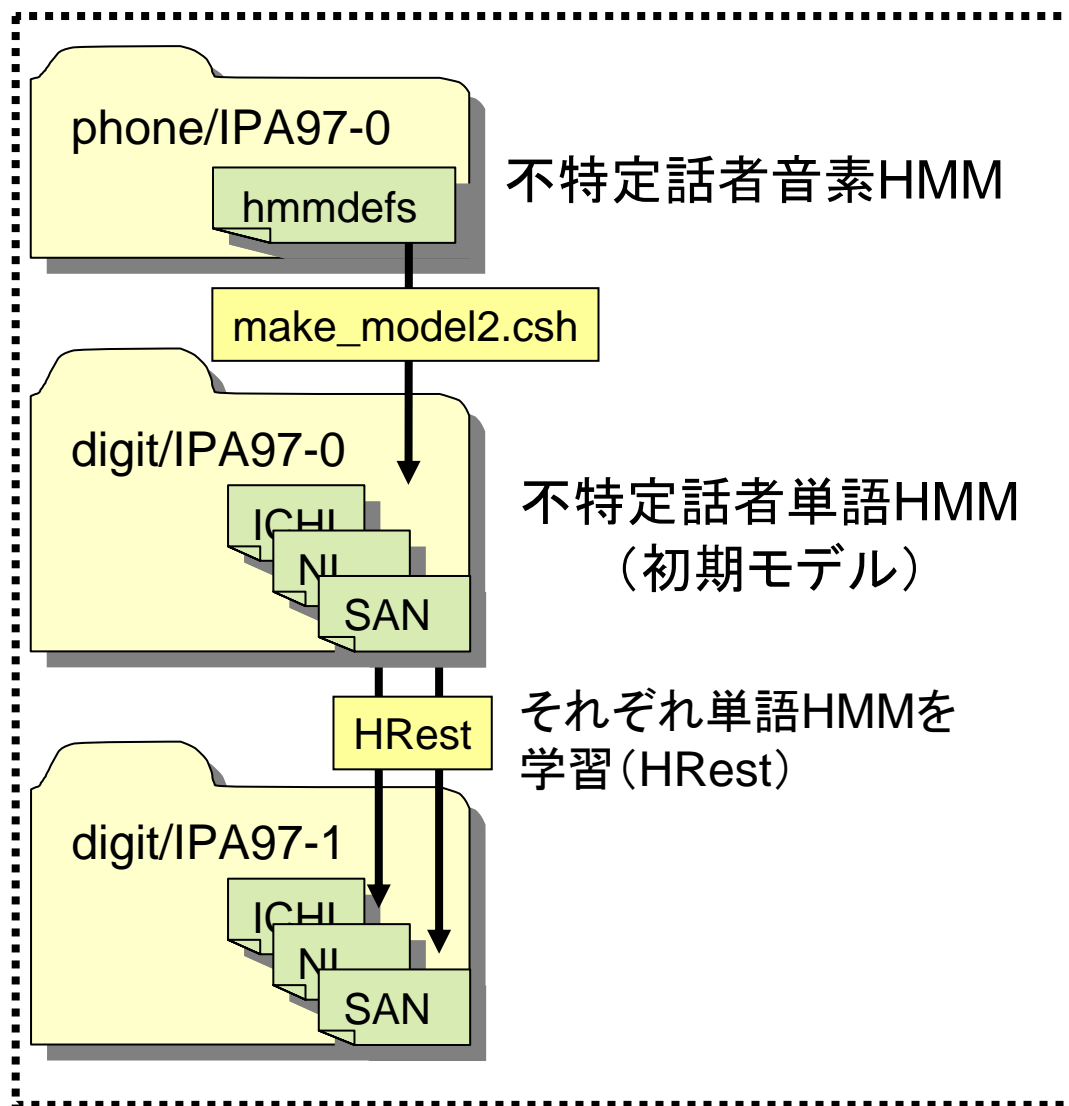
-- 音声認識の基本タスク

-- 実際に、AURORA2J(雑音下音声認識評価環境)で評価タスクとして用いられている

概要

1. 初期モデルの作成
2. 学習 (HRest)
3. 認識 (HVite)

3-1. 単語HMM学習の流れ



3-1. 単語HMMの学習と認識

音素HMMからの単語HMMの作成

既存の音素モデルをつないで各単語HMMの初期モデルを作成

```
% GENDER=male  
% GENDER=female
```

 } 自分の性別に合わせて
どちらかを実行

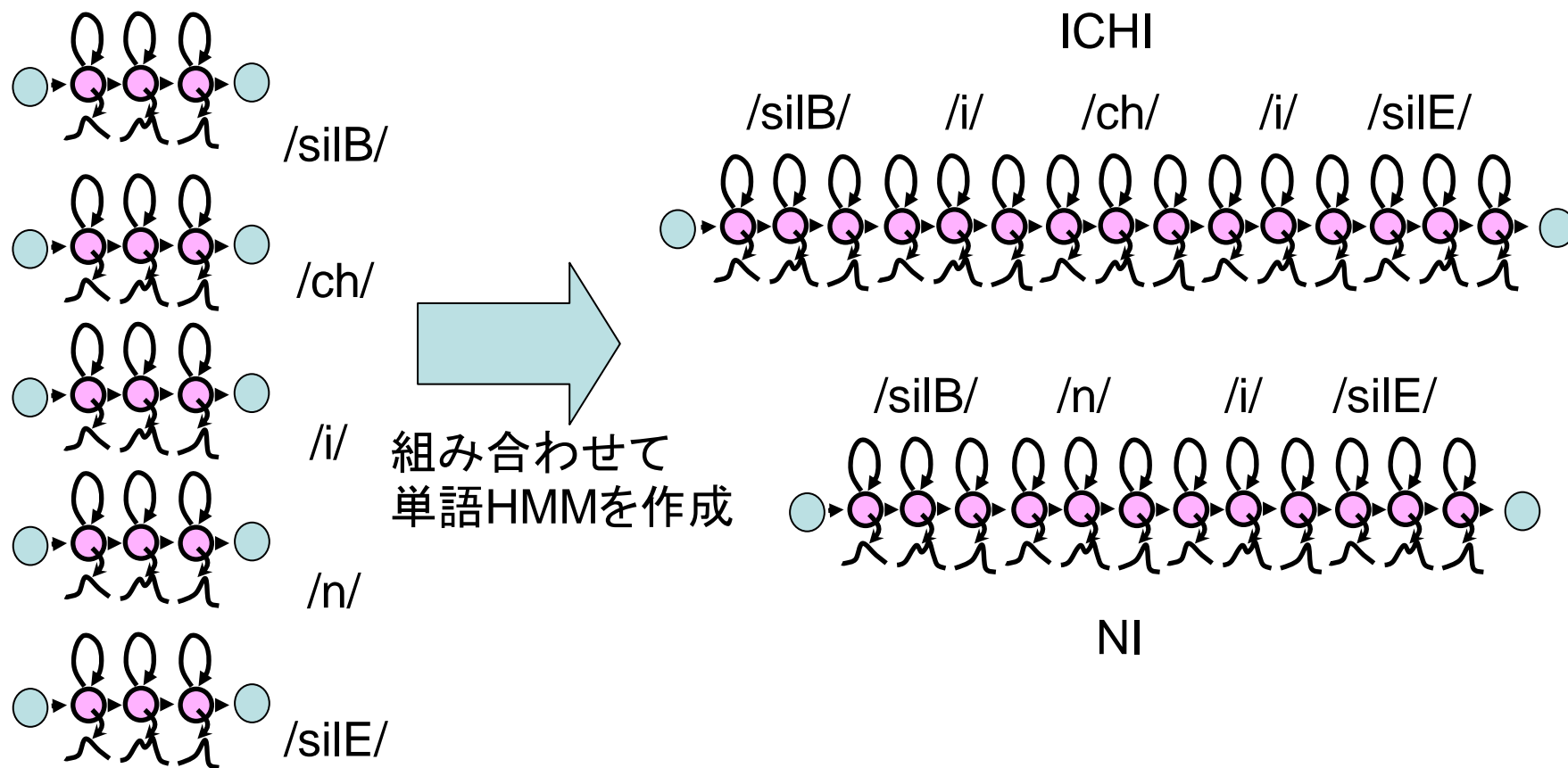
```
% export GENDER  
% cd ../model/digit/IPA97-0/  
% ./make_models2.csh  
% cd ../../../../work/
```

確認

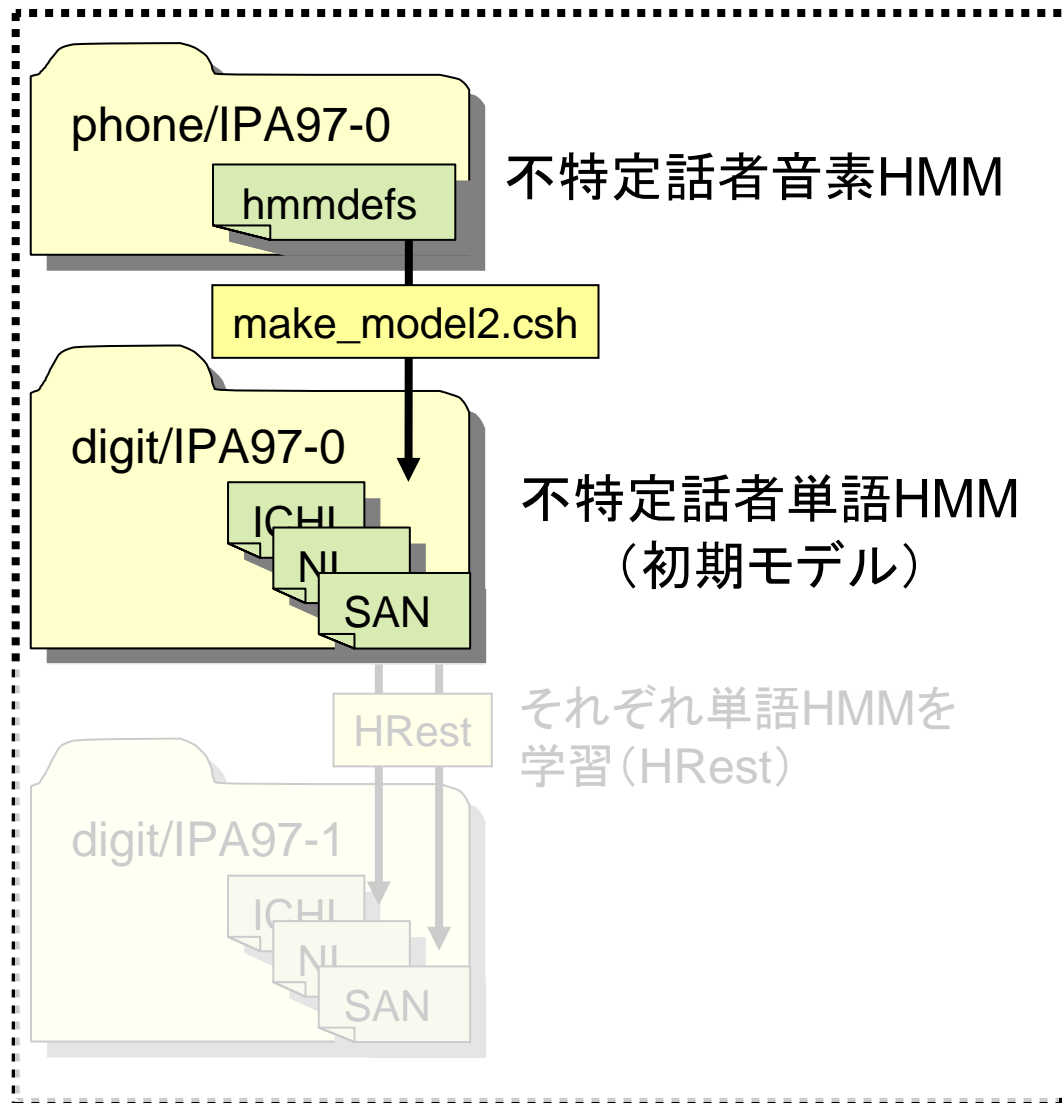
```
% less ../model/digit/IPA97-0/ZERO      (zeroのモデル)  
% less ../model/digit/IPA97-0/ICHI     (ichiのモデル)
```

3-1. 単語HMMの学習と認識

音素HMMからの単語HMMの作成



3-1. 音素HMMからの単語HMMの作成



3-2. 単語HMMの学習と認識 学習

【学習】 収録した数字単語発声のうち、最初の4回分を学習データにして単語HMMを学習

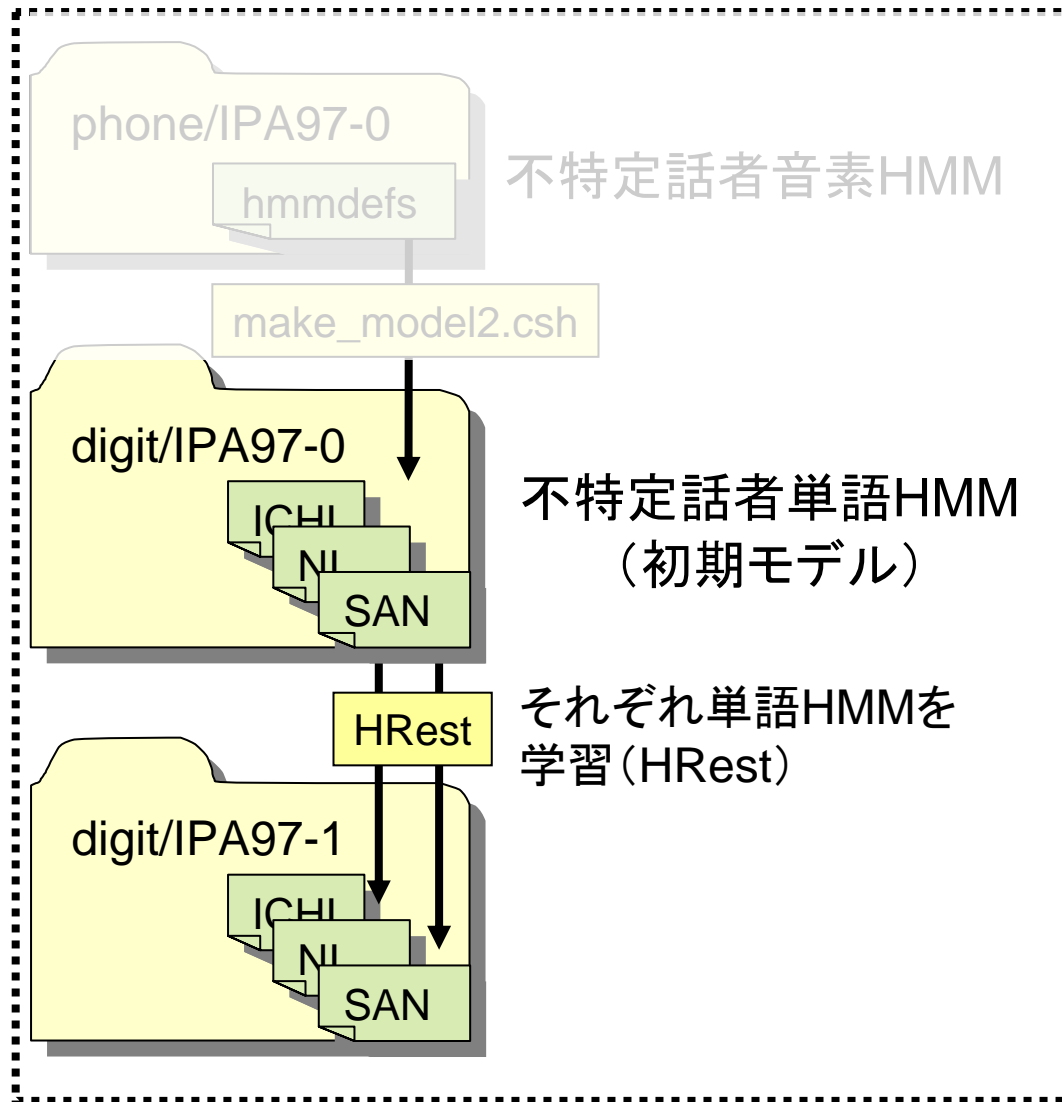
ZEROのモデルの学習

```
% HRest -T 1
        -i 40
        -C ../config/config.train
        -M ../model/digit/IPA97-1
        -v 0.01
        ../model/digit/IPA97-0/ZERO
        ../mfcc/digit/d0[1-4].mfc
```

【演習】

「**ICHI**」(d**1**[1-4].mfc) ~ 「**KYU:**」(d**9**[1-4].mfc)
についても同様に学習すること.

3-2. 単語HMM学習



3-2. 単語HMMの学習と認識

認識

【認識】 5回目に発声した音声(数字)を認識

```
% HVite -T 1  
-C ../config/config.rec  
-w digit2.ltc  
-d ../model/digit/IPA97-1  
digit2.dic  
digit2.hmm  
../mfcc/digit/d?5.mfc
```

【演習】

学習前のモデル「IPA97-0」を用いて認識し、認識時の尤度を比べてみること。

3-3. 単語HMMの学習と認識(16混合)

余力のある人のみ

【演習】

「digit/IPA97-mix16-0」を用いて学習と認識を実行し、認識時の尤度を比較すること(モデルの保存場所は、「digit/IPA97-mix16-1」)

3-1から3-2まで、適宜読みかえて実行すること.

```
% GENDER=xxxx  
% export GENDER  
% cd ../model/digit/IPA97-mix16-0/  
% ./make_models2.csh  
% cd ../../../../work/  
.....
```