

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

---

# **XỬ LÝ NGÔN NGỮ NÓI**

## **BÀI TẬP 2: NHẬN DẠNG TIẾNG NÓI RỜI RẠC**

Giáo viên: **PGS.TS. VŨ HẢI QUÂN**

Họ và tên học viên: **Nguyễn Bá Công**

Mã số học viên: **0209 480 107**

Ngành: **Khoa học máy tính – K19**

## I. Tổng quan

**Nhận dạng tiếng nói** là một quá trình nhận dạng mẫu, với mục đích là phân lớp (classify) thông tin đầu vào là tín hiệu tiếng nói thành một dãy tuần tự các mẫu đã được học trước đó và lưu trữ trong bộ nhớ. Các mẫu là các đơn vị nhận dạng, chúng có thể là các từ, hoặc các âm vị. Nếu các mẫu này là bất biến và không thay đổi thì công việc nhận dạng tiếng nói trở nên đơn giản bằng cách so sánh dữ liệu tiếng nói cần nhận dạng với các mẫu đã được học và lưu trữ trong bộ nhớ. Khó khăn cơ bản của nhận dạng tiếng nói đó là tiếng nói luôn biến thiên theo thời gian và có sự khác biệt lớn giữa tiếng nói của những người nói khác nhau, tốc độ nói, ngữ cảnh và môi trường âm học khác nhau. Xác định những thông tin biến thiên nào của tiếng nói là có ích và những thông tin nào là không có ích đối với nhận dạng tiếng nói là rất quan trọng. Đây là một nhiệm vụ rất khó khăn mà ngay cả với các kỹ thuật xác suất thống kê mạnh cũng khó khăn trong việc tổng quát hoá từ các mẫu tiếng nói những biến thiên quan trọng cần thiết trong nhận dạng tiếng nói.

Các nghiên cứu về nhận dạng tiếng nói dựa trên ba nguyên tắc cơ bản:

- Tín hiệu tiếng nói được biểu diễn chính xác bởi các giá trị phổ trong một khung thời gian ngắn (short-term amplitude spectrum). Nhờ vậy ta có thể trích ra các đặc điểm tiếng nói từ những khoảng thời gian ngắn và dùng các đặc điểm này làm dữ liệu để nhận dạng tiếng nói.
- Nội dung của tiếng nói được biểu diễn dưới dạng chữ viết, là một dãy các ký hiệu ngữ âm. Do đó ý nghĩa của một phát âm được bảo toàn khi chúng ta phiên âm phát âm thành dãy các ký hiệu ngữ âm.
- Nhận dạng tiếng nói là một quá trình nhận thức. Thông tin về ngữ nghĩa (semantics) và suy đoán (pragmatics) có giá trị trong quá trình nhận dạng tiếng nói, nhất là khi thông tin về âm học là không rõ ràng.

Các cách tiếp cận nhận dạng tiếng nói bằng thống kê bao gồm: sử dụng mô hình Markov ẩn, mạng nơ-ron, ...

**Phát biểu bài toán:** Sử dụng công cụ/thư viện HTK (Hidden Markov Model Toolkit) để xây dựng công cụ nhận dạng tiếng nói (Tiếng Việt) rời rạc từng từ một.

**Hướng giải quyết:** Trên thế giới đã có rất nhiều hệ thống nhận dạng tiếng nói đã và đang được ứng dụng rất hiệu quả như LVia Voice của IBM, Spoken Toolkit của CSLU (Central of Spoken Language Understanding). Tuy nhiên ở bài trình bày này; trình bày về phương pháp sử dụng công cụ HTK (Hidden Markov Model Toolkit), bằng việc sử dụng đặc trưng MFCC (Mel-Frequency Cepstrums Coefficients) và mô hình Markov ẩn (HMM).

## II. Cơ sở lý thuyết

### 1. Trích chọn đặc trưng

Trích chọn đặc trưng trong xử lý ngôn ngữ nói là biến đổi từ giọng nói thô thành những đặc trưng của người nói. Đặc trưng được sử dụng trong bài này là đặc trưng MFCC (Mel Frequency Cepstral Coefficients), đặc trưng MFCC thể hiện hình thái của cuống họng, ... khi ta nói.

Các bước trích chọn đặc trưng MFCC: gồm 5 bước cơ bản

- Phân khung (frame blocking): Chia tín hiệu đầu vào thành các đoạn nhỏ khoảng 20ms-30ms. Phân khung tín hiệu mỗi khung N mẫu, hai khung kề nhau lệch M mẫu.
- Lấy cửa sổ (Windowing): Lấy cửa sổ nhằm giảm sự gián đoạn của tín hiệu ở đầu và cuối mỗi khung vừa được chia (ví dụ dùng cửa sổ Hamming, ...).
- Biến đổi FFT (Fast Fourier Transform): Chuyển đổi mỗi khung với N mẫu từ miền thời gian sang miền tần số.
- Biến đổi sang thang đo Mel (Mel-frequency Wrapping)
- Hệ số Cepstrum (Cepstrum Coefficients)

### 2. Mô hình Markov ẩn (HMM)

#### a. Mô hình Markov

##### i. Các khái niệm cơ bản

Mô hình Markov là mô hình dựa trên thống kê để mô hình hóa các tín hiệu dạng chuỗi (theo thời gian) được sử dụng rất tốt trong các lĩnh vực nhận dạng.

Các tín hiệu dạng chuỗi:

- Chữ viết tay
- Tín hiệu âm thanh, ánh sáng
- Chuỗi DNA
- ...

Phân loại mô hình chuỗi trạng thái

- **Mô hình chuỗi đơn định**

Là mô hình các trạng thái trong đó liên kết chuyển từ trạng thái này sang trạng thái khác là hoàn toàn xác định. Ví dụ: mô hình chuyển trạng thái của đèn giao thông (đỏ-xanh-vàng)

- **Mô hình chuỗi xác suất**

Là mô hình các trạng thái trong đó liên kết chuyển từ trạng thái này sang trạng thái khác là giá trị xác suất, sao cho tổng xác suất đầu ra từ 1 trạng thái tới tất cả các trạng thái khác bao gồm chính nó luôn là 100%.

Như vậy theo nguyên lý xác suất Bayes, xác suất mô hình ở trạng thái  $q_i$  phụ thuộc vào xác suất tồn tại của tất cả trạng thái  $q_{j, j < i}$  trước nó.

## ii. Giả thuyết Markov

- Xác suất tại 1 trạng thái trong mô hình Markov chỉ phụ thuộc vào trạng thái trước nó

$$P(q_t | q_{t-1}, q_{t-2}, \dots, q_1) = P(q_t | q_{t-1})$$

- Dựa trên lý thuyết xác suất trạng thái đồng thời

$$P(A, B) = P(A|B)P(B)$$

- Xác suất chuỗi trạng thái được tính

$$\begin{aligned} &P(q_1, q_2, \dots, q_{t-1}, q_t) \\ &= P(q_t | q_{t-1}, q_{t-2}, \dots, q_1) P(q_{t-1}, q_{t-2}, \dots, q_1) \\ &= \dots \\ &= P(q_t | q_{t-1}) P(q_{t-1} | q_{t-2}) \dots P(q_2 | q_1) P(q_1) \end{aligned}$$

## iii. Mô hình Markov

Mô hình Markov được định nghĩa là 1 mô hình bao gồm các thành phần:

- Các trạng thái trong mô hình

$$1, 2, \dots, N$$

- Chuỗi trạng thái quan sát được

$$q_1, q_2, \dots, q_t, \dots, q_T$$

- Ma trận chuyển vị trạng thái A thể hiện xác suất chuyển vị từ trạng thái này sang trạng thái khác

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

trong đó

$$\begin{aligned} a_{ij} &= P(q_j | q_i) \quad 1 \leq i, j \leq N \\ a_{ij} &> 0 \quad \forall i, j \\ \sum_{j=1}^N a_{ij} &= 1 \quad \forall i \end{aligned}$$

- Xác suất ban đầu  $\pi$

$$\pi_1, \pi_2, \dots, \pi_N$$

trong đó

$$\begin{aligned} \pi_i &= P(q_i) \quad 1 \leq i \leq N \\ \sum_{i=1}^N \pi_i &= 1 \quad \forall i \end{aligned}$$

Tóm gọn lại mô hình Markov có thể được gọi là 1 bộ  $\lambda(A, \pi)$

#### iv. Ví dụ mô hình Markov

Mô hình nghiên cứu thời tiết

- Các trạng thái thời tiết:

$$(Sun, Cloudy, Rain), N = 3$$

- Xác suất ban đầu  $\pi$ :

$$(\pi_{Sun}, \pi_{Cloudy}, \pi_{Rain}) = (1.0, 0.0, 0.0)$$

- Ma trận chuyển vị trạng thái A:

$$\begin{array}{l} \text{sun} \\ \text{cloud} \\ \text{rain} \end{array} \begin{bmatrix} \text{sun} & \text{cloud} & \text{rain} \\ 0.50 & 0.375 & 0.125 \\ 0.25 & 0.125 & 0.625 \\ 0.25 & 0.375 & 0.375 \end{bmatrix}$$

#### b. Mô hình Markov ẩn

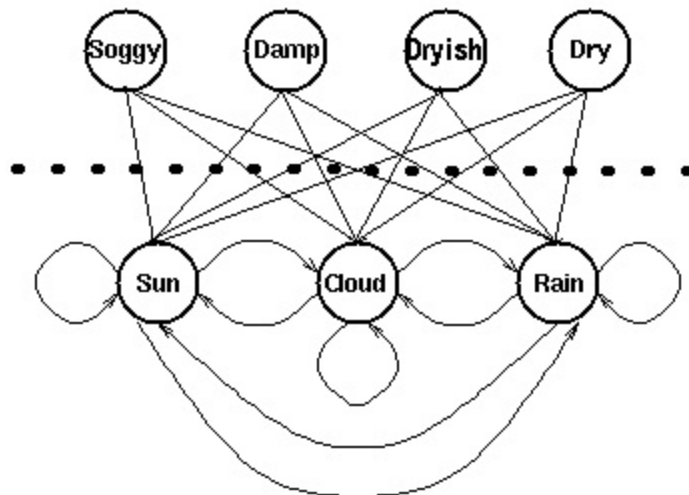
##### ■ Các trạng thái ẩn

Mô hình Markov được trình bày ở phần trước tồn tại 1 số khuyết điểm

trong trường hợp tổng quát, khi các trạng thái quan sát được không đồng thời là các trạng thái mà mô hình đang thể hiện. Với ví dụ mô hình nghiên cứu thời tiết đã nêu. Giả sử trong trường hợp các trạng thái chúng ta quan sát được không phải là các trạng thái thời tiết trực tiếp mà là các trạng thái của tảo biển bao gồm [*Soggy*, *Damp*, *Dryish*, *Dry*] (với các thực nghiệm cho thấy các trạng thái của tảo biển có thể hiện 1 phần trạng thái thời tiết theo 1 xác suất nghiên cứu biết trước), chúng ta phải nâng cấp mô hình Markov để có thể chứa thêm thông tin này.

	Dry	Dryish	Damp	Soggy
Sun	0.60	0.20	0.15	0.05
Cloud	0.25	0.25	0.25	0.25
Rain	0.05	0.10	0.35	0.50

Như vậy, mô hình Markov ẩn có thể được hình dung thông qua sơ đồ sau



Bên cạnh các giá trị xác suất khởi tạo và xác suất chuyển vị trạng thái thời tiết, mô hình cần phải lưu thêm các xác suất tương ứng của trạng thái tảo biển quan sát được tương ứng với trạng thái thời tiết cần nghiên cứu.

#### ■ Mô hình Markov ẩn được định nghĩa bao gồm

- Các trạng thái nghiên cứu trong mô hình  
1, 2, ...,  $N$

- Các trạng thái quan sát được

$$1, 2, \dots, M$$

- Chuỗi trạng thái quan sát được

$$o_1, o_2, \dots, o_t, \dots, o_T$$

- Ma trận chuyển vị trạng thái A thể hiện xác suất chuyển vị từ trạng thái này sang trạng thái khác

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{pmatrix}$$

trong đó

$$a_{ij} = P(q_j | q_i) \quad 1 \leq i, j \leq N$$

$$a_{ij} > 0 \quad \forall i, j$$

$$\sum_{j=1}^N a_{ij} = 1 \quad \forall i$$

- Xác suất ban đầu  $\pi$

$$\pi_1, \pi_2, \dots, \pi_N$$

trong đó

$$\pi_i = P(q_i) \quad 1 \leq i \leq N$$

$$\sum_{i=1}^N \pi_i = 1 \quad \forall i$$

- Ma trận chuyển vị trạng thái B thể hiện xác suất chuyển vị từ trạng thái quan sát được sang trạng thái ẩn

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \dots & \dots & \dots & \dots \\ b_{N1} & b_{N2} & \dots & b_{NM} \end{pmatrix}$$

trong đó

$$b_{ij} = P(o_j | q_i) \quad 1 \leq i \leq N, 1 \leq j \leq M$$

$$b_{ij} > 0 \quad \forall i, j$$

$$\sum_{j=1}^M b_{ij} = 1 \quad \forall i$$

Tóm gọn lại mô hình Markov ẩn có thể được gọi là 1 bộ  $\lambda(A, B, \pi)$ .



### c. 3 Bài toán Mô hình Markov ẩn

Với mô hình Markov ẩn được nêu, áp dụng vào các vấn đề thực tế làm nảy sinh ra 3 bài toán chính sau:

#### Bài toán 1: Tính điểm (Evaluation)

Cho trước chuỗi quan sát  $O = \{o_1, o_2, \dots, o_T\}$  tương ứng với mô hình  $\lambda(A, B, \pi)$ , xác định xác suất tương ứng có điều kiện  $P(O|\lambda)$ .

Ví dụ không hình thức:

- Cho trước 2 mô hình thời tiết  $\lambda_{Summer}$  và  $\lambda_{Winter}$  cùng với 1 chuỗi quan sát trạng thái tảo biển  $O$  trong thời gian 1 tuần. Chúng ta cần xác định xem tuần được quan sát thuộc về mô hình nào tương ứng với mùa nào trong năm bằng cách tính điểm chuỗi  $O$  cho cả 2 mô hình và chọn mô hình có điểm cao nhất.

#### Bài toán 2: So khớp (Matching)

Cho trước chuỗi quan sát  $O = \{o_1, o_2, \dots, o_T\}$  tương ứng với mô hình  $\lambda(A, B, \pi)$ , xác định chuỗi trạng thái nghiên cứu tương ứng  $Q = \{q_1, q_2, \dots, q_T\}$  sao cho xác suất tương  $P(O|Q, \lambda)$  đạt cực đại.

Ví dụ không hình thức:

- Cho trước 2 mô hình thời tiết  $\lambda$  cùng với 1 chuỗi quan sát trạng thái tảo biển  $O$  trong thời gian 1 tuần. Chúng ta cần xác định xem tuần được quan sát thuộc có thời tiết cụ thể như thế nào.

#### Bài toán 3: Huấn luyện (Training)

Cho trước tập chuỗi quan sát  $\{O_1, O_2, \dots, O_k\}$  xác định các tham số phù hợp cho mô hình  $\lambda(A, B, \pi)$  sao cho xác suất tương ứng có điều kiện  $P(O_{i, i \in [1..k]}|\lambda)$  là cực đại.

Ví dụ không hình thức:

- Cho trước tập 1 chuỗi quan sát trạng thái tảo biển  $O$  mỗi tuần trong vòng nhiều tháng. Chúng ta cần xác định 1 mô hình thời tiết tương ứng tại khu vực đó sao cho phù hợp nhất với chuỗi quan sát này.

## III. Thực nghiệm

## 1. Các chức năng được sử dụng trong công cụ HTK

HTK cung cấp các công cụ cho việc xây dựng một mô hình Markov ẩn (HMMs). HMMs có thể dùng cho các bài toán với dữ liệu là một chuỗi theo thời gian, và HTK cũng có thể sử dụng cho các bài toán dạng này. Tuy nhiên, HTK được thiết kế chính là phục vụ cho các bài toán về tiếng nói. Dưới đây là một vài công cụ cơ bản thường được dùng:

- **HCopy**  
Copy và thực hiện việc rút trích đặc trưng tùy thuộc vào file cấu hình.
- **HCompV**  
Công cụ khởi tạo tham số cho các hàm phân bố xác suất có trong từng trạng thái của mô hình Markov ẩn tương ứng với dữ liệu huấn luyện.
- **HRest**  
Công cụ huấn luyện mô hình bằng việc tính toán, ước lượng lại các tham số của mô hình.
- **HParse**  
Công cụ chuyển đổi (parse) từ dữ liệu đầu vào thành dữ liệu theo format HTK.
- **HVite**  
Công cụ so khớp/nhận dạng với đầu vào là dữ liệu cần nhận dạng, và trả về mô hình tốt nhất/khớp nhất mà nó thuộc vào.
- **HResults**  
Công cụ thống kê, đánh giá kết quả nhận dạng.

## 2. Chuẩn bị dữ liệu

Dữ liệu thực nghiệm được lấy từ dữ liệu thu âm trên lab:

- K22\_SV34 (250 files tương ứng 250 câu)

### a. Tập dữ liệu huấn luyện (train)

200 files

### b. Tập dữ liệu test

Gồm 327 files, là các từ được tách ra từ 50 câu.

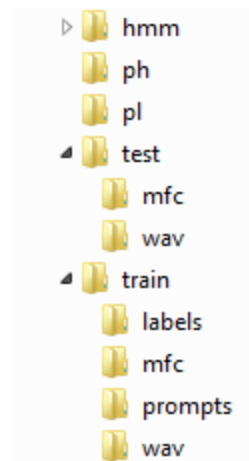
Chúng ta cũng cần chuẩn bị các file sau cho việc cài đặt mô hình:

- Các file prompts tương ứng với các file âm thanh (wav),
- File từ điển bao gồm tất cả các từ trong bộ dữ liệu thực nghiệm,
- Chuẩn bị các file cho việc huấn luyện mô hình dùng công cụ HTK:
  - proto (prototype model)
  - config
  - HCompV.cfg
  - HCopy.cfg
  - HERest.cfg
  - HVite.cfg
  - mktri.led
  - sil.hed
  - monophones0
  - monophones1
  - Các file perl-scripts

### 3. Cài đặt

Chương trình được xây dựng dựa vào công cụ HTK và các “Active Perl” scripts. Dưới đây là mô tả về cách chương trình được cài đặt cụ thể.

#### Cấu trúc chương trình



#### Mô hình

- Sử dụng HTK toolkit để xây dựng mô hình gồm:
  - Module **Huấn luyện**
  - Module **Nhận dạng**
- Đặc trưng: 39 đặc trưng MFCC\_0\_D\_A, trong đó
  - 0: 12 MFCC features + (1) total energy in the frame,
  - D: 13 “Delta coefficients” (đạo hàm bậc 1 của MFCC),

- A: 13 “Acceleration coefficients” (đạo hàm bậc 2 của MFCC)
- Mỗi một âm tiết là một mô hình Markov ẩn
- Số trạng thái: 5

#### a. Huấn luyện

- Tạo file "words.mlf" từ file prompts.txt

```
Perl pl\prompts2mlf.pl words.mlf prompts.txt
```

- Tạo 2 file "phones0.mlf" và "phones1.mlf" : phones0 chỉ có "sil" nhưng phones1 có thêm "sp"

```
HLEd -l * -d dict.txt -i phones0.mlf mkphones0.led words.mlf
```

```
HLEd -l * -d dict.txt -i phones1.mlf mkphones1.led words.mlf
```

- Rút trích đặc trưng MFCC 0 D A

```
perl pl\listwavmfc.pl train\wav train\mfc.scf
```

```
Hcopy -C config -S train\mfc.scf
```

- Tạo file liệt kê dữ liệu huấn luyện

```
perl pl\mkTrainFile.pl train\mfc train\train.scf
```

- Bước 2. Khởi tạo mô hình

```
HCompV -A -D -T 1 -C HCompV.cfg -f 0.01 -m -S train\train.scf -M hmm\hmm0 proto
```

- Bước 3. Tạo mô hình chuẩn

```
perl pl\mkMacrosFile.pl hmm\hmm0\ Floors hmm\hmm0\macros
```

```
perl pl\mkHmmddefsFile.pl hmm\hmm0\proto monophones0 hmm\hmm0\hmmddefs
```

- Bước 4. Huấn luyện 3 vòng -> hmm3

```
HERest -A -D -T 1 -C HERest.cfg -l phones0.mlf -t 250.0 150.0 1000.0 -S train\train.scf  
-H hmm\hmm0\macros -H hmm\hmm0\hmmddefs -M hmm\hmm1 monophones0
```

```
HERest -A -D -T 1 -C HERest.cfg -l phones0.mlf -t 250.0 150.0 1000.0 -S train\train.scf  
-H hmm\hmm1\macros -H hmm\hmm1\hmmddefs -M hmm\hmm2 monophones0
```

```
HERest -A -D -T 1 -C HERest.cfg -l phones0.mlf -t 250.0 150.0 1000.0 -S train\train.scf  
-H hmm\hmm2\macros -H hmm\hmm2\hmmddefs -M hmm\hmm3 monophones0
```

- Bước 5. Thêm “sp” vào danh sách mô hình

```
perl pl\makesp.pl hmm\hmm3\hmmdefs hmm\hmm4\hmmdefs hmm\hmm3\macros
hmm\hmm4\macros
```

■ Bước 6. Liên kết **sil-sp**

```
HHEd -H hmm\hmm4\macros -H hmm\hmm4\hmmdefs -M hmm\hmm5 sil.hed
monophones1
```

■ Bước 7. Huấn luyện 2 vòng **hmm5 -> hmm7**

```
HERest -A -D -T 1 -C HERest.cfg -I phones1.mlf -t 250.0 150.0 1000.0 -S train\train.scp
-H hmm\hmm5\macros -H hmm\hmm5\hmmdefs -M hmm\hmm6 monophones1
```

```
HERest -A -D -T 1 -C HERest.cfg -I phones1.mlf -t 250.0 150.0 1000.0 -S train\train.scp
-H hmm\hmm6\macros -H hmm\hmm6\hmmdefs -M hmm\hmm7 monophones1
```

■ Bước 8. Clone/copy **phones1.mlf -> aligned.mlf** (xác định cách đọc chính cho từng từ)

```
copy phones1.mlf aligned.mlf
```

■ Bước 9. Huấn luyện 2 vòng -> **hmm9**

```
HERest -A -D -T 1 -C HERest.cfg -I aligned.mlf -t 250.0 150.0 1000.0 -S train\train.scp -H
hmm\hmm7\macros -H hmm\hmm7\hmmdefs -M hmm\hmm8 monophones1
```

```
HERest -A -D -T 1 -C HERest.cfg -I aligned.mlf -t 250.0 150.0 1000.0 -S train\train.scp -H
hmm\hmm8\macros -H hmm\hmm8\hmmdefs -M hmm\hmm9 monophones1
```

■ Bước 10. Tạo triphones

```
HLEd -n ph\triphones1 -I * -i wintri.mlf mktri.led aligned.mlf
perl pl\mkTriHed.pl monophones1 ph\triphones1 mktri.hed
HHEd -B -H hmm\hmm9\macros -H hmm\hmm9\hmmdefs -M hmm\hmm10 mktri.hed
monophones1
```

Huấn luyện 2 vòng -> **hmm12**

```
HERest -B -C HERest.cfg -I wintri.mlf -s stats -S train\train.scp -H hmm\hmm10\macros
-H hmm\hmm10\hmmdefs -M hmm\hmm11 ph\triphones1
```

```
HERest -B -C HERest.cfg -I wintri.mlf -s stats -S train\train.scp -H hmm\hmm11\macros
-H hmm\hmm11\hmmdefs -M hmm\hmm12 ph\triphones1
```

■ Bước 11. Liên phần **share/chung**

```
perl pl\mkFullList.pl monophones0 fulllist
perl pl\mkTree.pl 350 monophones0 tree.hed
```

```
HHEd -B -H hmm\hmm12\macros -H hmm\hmm12\hmmdefs -M hmm\hmm13 tree.hed
ph\triphones1 > log
```

- Bước 12. Huấn luyện 2 vòng -> **hmm15** (mô hình cuối cùng)

```
HERest -B -C HERest.cfg -I wintri.mlf -s stats -S train\train.scp -H hmm\hmm13\macros
-H hmm\hmm13\hmmdefs -M hmm\hmm14 tiedlist
```

```
HERest -B -C HERest.cfg -I wintri.mlf -s stats -S train\train.scp -H hmm\hmm14\macros
-H hmm\hmm14\hmmdefs -M hmm\hmm15 tiedlist
```

## b. Nhận dạng

- Chuẩn bị dữ liệu kiểm thử đúng/truth file

```
Perl pl\prompts2mlf.pl test\truth.mlf test\prompts.txt
```

- Chuẩn bị dữ liệu nhận dạng

```
perl pl\listwavmfc.pl test\wav test\mfc.scp
```

```
HCOPY -T 1 -C config -S test\mfc.scp
```

```
perl pl\mkTrainFile.pl test\mfc test\test.scp
```

- Parse gram thành wdnets

```
HParse gram.txt wdnets
```

- Thực hiện nhận dạng

```
HVite -T 1 -C HVite.cfg -H hmm\hmm15\macros -H hmm\hmm15\hmmdefs -S
test\test.scp -i test\recout.mlf -w wdnets -p 0.0 -s 5.0 gram_dict.txt tiedlist
```

- Thực hiện thống kê kết quả nhận dạng

```
HResults -I test\truth.mlf tiedlist test\recout.mlf > test\result.txt
```

## 4. Kết quả

Kết quả được thống kê dựa trên tập dữ liệu test:

- K22\_SV34 (327 mẫu là các từ được tách ra từ 50 câu)

```
===== HTK Results Analysis =====
Date: Sat Jun 01 13:23:12 2013
Ref : test\truth.mlf
Rec : test\recout.mlf
----- Overall Results -----
SENT: %Correct=35.38 [H=115, S=210, N=325]
WORD: %Corr=35.38, Acc=31.38 [H=115, D=0, S=210, I=13, N=325]
=====
```

## IV. Kết luận

Việc xây dựng một mô hình để nhận dạng tiếng nói (Tiếng Việt) rời rạc sử dụng công cụ

HTK được thực hiện khá sơ sài. Tuy nhiên kết quả thực nghiệm lại tương đối thấp ( $Corr=35.38$ ,  $Acc=31.38$ ). Nguyên nhân có thể do:

- Dữ liệu huấn luyện ít, không tổng quát được dữ liệu kiểm thử;
- Dữ liệu kiểm thử (test) được tách ra từ một câu cho chất lượng không tốt;
- Dữ liệu dùng để huấn luyện cũng như kiểm thử do tác giả đọc nghe không rõ, có các từ được phát âm gần như nhau;
- Mô hình chưa đủ mạnh.

Hướng tiếp cận trong thời gian tới:

- Thu thập thêm nhiều dữ liệu;
- Nâng cấp mô hình (Adaptive Model).