

## 한국어 음소 HMM 모델 결정을 위한 파라미터 비교 연구

권혁재, 신유식, \*유창규, \*임형규, 김종교

전북대학교 전자공학과

\*서남대학교 전자전기공학과

### A Comparative Study on Parameter for Korean Phoneme-based HMM Model Decision

Hyuck-Je Kwon, You-Shik Shin, \*Chang-Kyu Yu, \*Hyung-Kyu Im, Chong-Kyo Kim

Dept. of Electronic Eng., Chonbuk Nat'l Univ.

\*Dept. of Electronic Electric Eng., Seonam Univ.

#### 요 약

본 논문은 음소의 확률적 분포를 이용하는 음소 HMM 모델을 결정하기 위한 여러 가지 거리 측정방법에 대한 연구이다. 음소 HMM 모델 결정을 위해서 LPC 계수를 이용하고, 거리 측정자를 LPC 계수, LPC 스펙트럼, LPC 첵스트림 등의 파라미터를 이용하고, 또한 양자화 과정은 k-means와 LBG 알고리즘을 혼합한 하이브리드 알고리즘을 사용하였다. LPC 코드북을 구성하기 위해 세 가지 파라미터를 유클리디안 거리로 거리측정에 이용하였다. 이렇게 양자화한 파라미터의 평균과 분산을 구하고, 양자화한 파라미터 코드북의 확률값을 비교해 한국어 음소 HMM 모델 결정을 위한 거리 측정 파라미터를 비교하였으며, 그 결과 LPC계수를 주파수 영역으로 변환하여 유클리디안 거리를 이용한 코드북의 분산이 작으므로 상대적으로 높은 확률을 가짐을 보았다.

#### 1. 서론

본 논문은 음소를 이용하여 연속 HMM을 구현하기 위한 여러 가지 거리 측정자를 사용한 모델 파라미터 분석에 관한 연구이다. 즉, HMM 모델 결정에 필요한, 모델 파라미터를 추출하여 실제 시스템 구현단계에서 이용한다. 현재 추세는 monophone,

diphone, triphone을 이용한 음소인식 분야로 연구가 진행되고 있다. 연속 HMM의 모델 파라미터로서는 평균 및 분산, 상태 천이 확률, 가중치 등을 사용하고 있다. 이 경우 모델 파라미터를 추정하는 방식으로 EM(expectation maximization)과 코드북 등을 사용한다. 코드북을 만드는 방식으로 k-means, LBG, genetic, MKM (modified k-means)등이 있다. 본 논문은 각 음소에 대한 모델을 결정하기 위해서 세 가지의 LPC 계열의 파라미터를 거리 측정자로 이용하였다. 파라미터는 LPC 계수, LPC 스펙트럼, LPC 첵스트림 등이다. LPC 계수를 제외한 나머지 두 파라미터는 LPC 계수로부터 직접 구한다. 모델 파라미터인 평균과 분산을 구하는 과정에서 사용하는 양자화 알고리즘은 k-means와 LBG를 각각 사용하였다. 코드북 크기(S)는 5개로 하였으며, 4(2<sup>n</sup>)개까지는 LBG를, 나머지 1개(S-2<sup>n</sup>)는 modified k-means를 사용하였다. 이렇게 구한 평균과 분산으로 각 양자화 알고리즘과 거리측정 파라미터들의 확률값을 비교하였다.

#### 2. 음성 파라미터

본 논문에서 사용하고 있는 데이터는 전화음성 데이터이다. 전화음성 데이터를 이용해 segmental

k-means의 반복 수행과정에서 첫 번째 반복 부분에서 추출한 상태값들이다. 전체 시스템에서 음소구간을 검출하고 이를 인식에 이용하는 과정 중에서 본 논문은 음소구간 검출을 위한 거리 측정 파라미터들을 비교하였다. 구성하는 코드북은 LPC 코드북이며, 12차 LPC 계수, 256차 LPC 스펙트럼과 12차 LPC 캡스트럼을 거리 측정자로 이용하였으며, 사용한 거리 계산은 유클리디안 거리를 사용하였다.

### 3. 양자화

양자화 알고리즘은 k-means와 LBG알고리즘을 이용하여 수행한다. 전체 코드북 크기는 보통 음소 모델링의 경우 mixture model을 5개로 결정하기 때문에, 4개까지는 LBG 알고리즘을 이용하여 구하고, 나머지 1개의 평균벡터는 LBG에서 구한 코드워드들에 대해서 LPC 데이터와 비교하여 가장 거리가 큰 벡터를 또 다른 코드워드로 선정한다. 이렇게 선정된 5개의 코드워드와 다시 중심값을 갱신함으로써 k-means와 동일한 코드북의 크기를 만든다.

#### 3.1 k-means

$$d(v_i, v_j) = d_{ij} \begin{cases} = 0 & \text{if } v_i = v_j \\ > 0 & \text{otherwise} \end{cases} \text{ - Euclidean}$$

- ▷ 초기화 - 임의의 S 벡터를 코드북의 초기 코드워드로 결정한다.
- ▷ 근접 검색 - 각 훈련 벡터에 대해서 가장 적은 거리를 가지고 있는 코드워드로 각 벡터의 인덱스를 할당한다.
- ▷ 중심값 - 각 할당된 벡터에 대해서 코드워드의 새로운 중심값을 찾는다.
- ▷ 반복 - 근접 검색과 중심값에 대한 갱신을 문턱값 이하가 될 때까지 반복 수행한다.

#### 3.2 LBG

$$y_{n^+} = y_n(1 + \epsilon)$$

$$y_{n^-} = y_n(1 - \epsilon)$$

- ▷ 초기화 - 임의의 한 벡터를 코드북의 초기 코

드워드로 결정한다.

- ▷ 분할 - 정해진 한 벡터에 대해 임의의  $\pm\epsilon$ 을 취해 두 개의 코드워드로 결정한다 ( $\epsilon=0.05$ )
- ▷ 근접 검색 - 각 훈련 벡터에 대해서 가장 적은 거리를 가지고 있는 코드워드로 각 벡터의 인덱스를 할당한다.
- ▷ 중심값 - 각 할당된 벡터에 대해서 코드워드의 새로운 중심값을 찾는다.
- ▷ 반복 - 분할 및 근접 검색과 중심값에 대한 갱신을 문턱값 이하가 될 때까지 반복 수행한다.

#### 3.3 하이브리드

LBG알고리즘을 이용하여 2<sup>n</sup>개의 코드워드를 구성하고, 나머지 S-2<sup>n</sup>개의 평균은 이미 구해진 2<sup>n</sup>개의 평균을 이용하여 입력 데이터 중 가장 먼 거리의 벡터로 한다. 전체 코드북 크기에 맞춘 평균을 다시 k-means를 수행해 새로운 중심값을 찾는다.

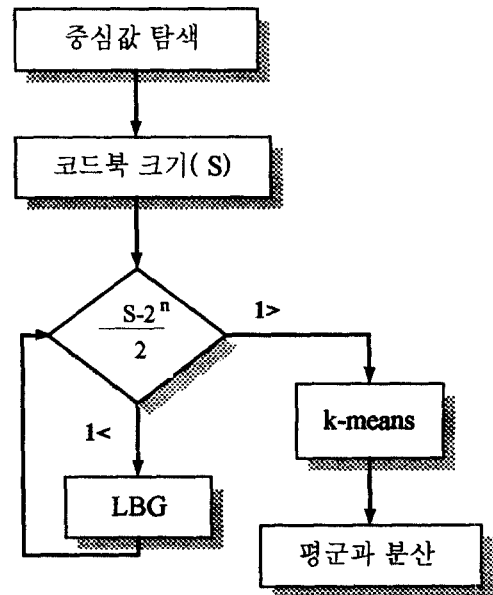


그림 1. 하이브리드 양자화

### 4. 분석 방법

각 양자화 방법과 거리측정 파라미터를 이용하여 각각의 코드북을 구성하고, 코드워드에 대한 평균과 분산을 구한다. 코드북의 구성은 크게 두 가지 방법으로 구별할 수 있다. LPC 코드북은 이미 구해진 거

리측정 파라미터들의 양자화과정을 통해 이에 해당하는 코드워드의 인덱스를 직접 반환해 구성한다. 각 코드북에서 구한 평균과 분산을 이용해 코드북의 입력값의 변화에 따라 확률값을 계산하게 된다. 확률값 계산을 위한 식은

$$F(c) = \frac{1}{(2\pi)^{D/2} \sqrt{|A|}} \text{Exp}\left(-\frac{1}{2}(x-m^c) A^{-1}(x-m^c)\right)$$

이며,  $A$ 는 분산의 대각행렬,  $c$ 는 코드워드,  $D$ 는 각 코드워드의 차수,  $x$ 는 입력벡터,  $m$ 은 평균 벡터이다.

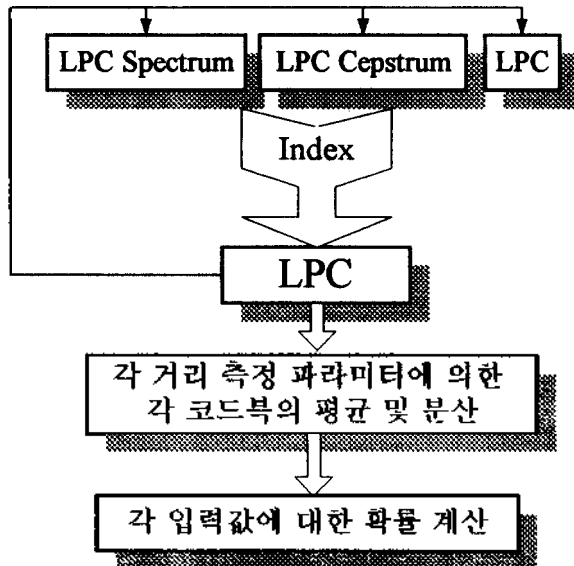


그림 2. 각 거리 측정 파라미터에 의한 코드북의 입력에 대한 확률계산

### 5. 실험 및 결과

총 3개의 LPC 데이터를 이용하였고 k-means와 LBG 알고리즘을 이용하여 실험하였다. 각 양자화 알고리즘에서는 각 거리 측정 파라미터를 이용해 인덱스 직접 반환 방법과 간접 반환방법으로 실험하였다. 모델 파라미터 평균 및 분산을 추정하고, 구성된 코드북에 대한 입력 데이터의 확률값을 계산하였다. 실험결과는 다음 표1, 2와 같다.

<표 1> k-means

거리측정 \ 입력(x)		2-0	2-1	2-2
LPC	2-0	2.183	2.502	1.989
	2-1	1.339	1.732	0.8229
	2-2	2.175	2.559	2.373
Cepstrum	2-0	11.52	12.90	19.96
	2-1	3.445	3.646	3.448
	2-2	1.270	1.382	1.521
Spectrum	2-0	2.606	2.210	1.962
	2-1	3.921	5.077	3.523
	2-2	2.590	1.940	1.952

<표 2> LBG

거리측정 \ 입력(x)		2-0	2-1	2-2
LPC	2-0	2.174	2.510	1.988
	2-1	1.334	1.733	0.8226
	2-2	2.172	2.583	2.389
Cepstrum	2-0	4.069	4.330	4.307
	2-1	3.745	4.215	3.540
	2-2	1.889	2.072	2.469
Spectrum	2-0	13.30	10.37	49.832
	2-1	4.909	6.193	4.432
	2-2	4.790	4.432	3.601

### 6. 결론

본 논문에서 사용하고 있는 데이터는 전화음성 데이터로 segmental k-means의 반복 수행과정에서 추출한 데이터이다. 3개의 LPC 데이터를 이용해 각 데이터의 코드북을 구성하였다. 코드북을 구성하기 위해 사용한 거리측정 파라미터는 LPC 계수와 LPC 첵스트럼, LPC 스펙트럼이며, 유클리디안 거리를 이용하였다. LPC를 제외한 두 거리 측정 파라미터는 데이터로 사용한 LPC에서 직접 추출하였다.

코드북 크기를 5개, 양자화알고리즘은 k-means와

LBG를 사용하였다. 4개까지는 LBG를 사용하고, 나머지 평균벡터에 대해서는 LBG에서 추출한 평균벡터들과 비교해 입력벡터들 중 가장 먼 거리의 벡터를 또 하나의 평균벡터로 정해 새로운 중심값을 정하였다.

각 양자화알고리즘에서 사용하는 거리측정은 세가지 LPC 계열파라미터를 이용하여 LPC 코드북을 구성하였다. 이 과정에서 거리측정 파라미터에서 직접 인덱스를 넘겨주는 방법과 LPC에서 평균 벡터를 구하고 이것을 거리 측정 파라미터와 동일 데이터 형태로 변환후 새로운 중심값을 찾아 인덱스를 넘겨주는 간접 반환방법을 취하였다.

표1, 표2는 각 거리 측정 파라미터로 구성된 LPC 코드북에 대해서 입력 데이터에 대한 확률계산이다. LPC계수를 이용한 코드북의 구성은 나머지 쉼스트림이나 스펙트럼에 의해 구해진 코드북의 입력 데이터에 대한 오류 확률이 더 크다. Euclidean distance를 LPC계수에 직접 사용하여 코드북을 작성한 것보다는 LPC계수를 주파수 영역으로 변환하여 유클리디안 거리를 이용한 코드북의 분산이 작으므로 상대적으로 높은 확률을 가짐을 보았다.

쉼스트림 및 스펙트럼은 두 가지 양자화 알고리즘에 대해서 유사한 결과를 보였다.

이상에서 나타난 결과로 보아 음소 HMM모델인 평균 및 분산, 상태 천이확률, 가중치 중에서 코드북을 사용한 평균, 분산을 추출할 경우에 주파수 영역에서의 거리 측정자를 사용하는 것이 보다 정확한 음소 HMM모델을 결정할 수 있었다.

## 참고문헌

- [1] Sadaoki Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., 1991.
- [2] John R. Deller, Jr., John G. Proakis, John H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.
- [3] Lawrence Rabiner, Biing-Hwang Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Inc., 1993.
- [4] Allen Gersho, Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1991.
- [5] A. M. Kondoz, *Digital Speech, Coding for Low Bit Rate Communication Systems*, John Wiley & Sons, 1994.