

Baum-Welch 학습법을 이용한 HMM 기반 대역폭 확장법

HMM-Based Bandwidth Extension Using Baum-Welch Re-Estimation Algorithm

송근배*, 김석호*
(Geun-Bae Song*, Austin Kim*)

*삼성전자

(접수일자: 2007년 7월 5일, 수정일자: 2007년 8월 2일; 채택일자: 2007년 8월 16일)

본 논문에서는 HMM 기반 통계적인 대역폭 확장 (Bandwidth Extension, BWE) 방법의 개선에 대해 다룬다. 이를 위해 우선, HMM 모델 학습을 위한 기존의 Jax의 학습법과 일반적인 Baum-Welch 학습법의 관계를 비교 검토하고, Jax의 학습법의 한계점 및 문제점을 검토한다. 그리고 이를 바탕으로 Baum-Welch 학습법을 이용한 새로운 HMM 기반 BWE 방법을 제시한다. 결론적으로, Baum-Welch 학습법은 Jax의 학습법의 일반화된 형태로 볼 수 있으며, 보다 유연하고 적응적인 학습능력을 가진 알고리즘임을 알 수 있다. 따라서 학습 데이터에 대한 보다 정확한 HMM 모델링이 가능하며 아울러, 이와 같이 개선된 HMM 모델을 활용함으로써 BWE 시스템의 성능향상을 가져올 수 있었다. 실험결과에 의하면, 제시된 새로운 방법이 기존의 Jax의 방법에 비해 실험의 모든 경우에서 우수한 성능을 보임을 알 수 있다. 주어진 실험조건하에서 근제곱평균 (root-mean-square, RMS) 로그 스펙트럴 왜곡 (Log Spectral Distortion, LSD) 값이 전체적으로 평균 0.52dB 그리고, 최소 0.31dB에서 최대 0.8dB까지 개선되었다.

핵심용어: 대역폭 확장, 은닉 마코프 모델, 가우스 혼합 모델, Baum-Welch 학습법

투고분야: 음성처리분야 (2,3)

This paper contributes to an improvement of the statistical bandwidth extension (BWE) system based on Hidden Markov Model (HMM). First, the existing HMM training method for BWE, which is suggested originally by Jax, is analyzed in comparison with the general Baum-Welch training method. Next, based on this analysis, a new HMM-based BWE method is suggested which adopts the Baum-Welch re-estimation algorithm instead of the Jax's to train HMM model. Conclusionally speaking, the Baum-Welch re-estimation algorithm is a generalized form of the Jax's training method. It is flexible and adaptive in modeling the statistical characteristic of training data. Therefore, it generates a better model to the training data, which results in an enhanced BWE system. According to experimental results, the new method performs much better than the Jax's BWE system in all cases. Under the given test conditions, the RMS log spectral distortion (LSD) scores were improved ranged from 0.31dB to 0.8dB, and 0.52dB in average.

Key words: Bandwidth extension, Hidden markov model, Gaussian mixture model, Baum-welch

Re-estimation

ASK subject classification: : Speech Signal processing (2,3)

I. 서론

현재 상용화된 유/무선 통신 시스템은 제한된 대역폭 (협대역 즉, 0.3-3.4kHz)으로 인하여 만족스러운 음질을

제공하지 못하고 있다. 가령 자음의 경우 4kHz 이상의 고주파 대역에서도 중요한 정보를 많이 가지고 있는데, 이러한 통신 시스템의 제약으로 인한 자음 정보의 손실은 음의 명료도 (intelligibility)를 떨어뜨리는 주요 원인이 된다. 또한 저주파 대역 (0-0.3kHz)의 손실은 음의 자연스러움 (naturalness) 및 현장감 (presence) 등을 떨어뜨리는 원인이 된다. 이와 같은 음질 문제를 해결하기 위해

책임저자: 송근배 (geunbae.song@samsung.com)
442-742 경기도 수원시 팔달구 매탄3동 삼성전자 통신연구소
통신연구소 HIDS팀
(전화: 031-279-3324)

광대역 (즉, 0-8kHz) 통신시스템을 구축하는 것이 필요하지만 이는 막대한 비용이 소요되므로 당장 실현이 곤란하다. 따라서 그 중간 단계로서, 수신단에서 협대역 (혹은 하위대역 즉, 0-4kHz) 신호의 정보만으로 광대역 (혹은 상위대역 즉, 4-8kHz) 정보를 추정 및 재생하는 대폭 확장법 (Bandwidth Extension, BWE)이 연구되어 왔다 [1]-[7]. 특히 지난 수 년에 걸쳐 은닉 마코프 모델 (Hidden Markov Model, HMM)에 기반한 새로운 통계적인 BWE 방법이 개발되었다 [5]-[7]. 이와 같은 은닉 마코프 모델 접근법은 논문 [1]-[4]와 같은 기억능력이 없는 (memory-less) 추정 방법과 달리 인접한 프레임 (frame)간의 상관관계를 활용한다는 장점을 가지며, 기존의 방법들에 비해 인공잡음 (artefact)을 현저히 감소시킨다는 장점을 가지는 것으로 알려져 있다 [5]-[7].

본 논문에서는 BWE를 위한 Jax로 부터 비롯된 기존의 HMM 방법을 되짚어 보고 이의 개선을 도모한다 [5]-[7]. 특히 BWE를 위한 HMM 모델 학습에 있어서, Jax가 제안한 학습법과 일반적인 Baum-Welch 학습법 [8]과의 관계를 비교 분석하고, Jax 학습법의 한계점 및 문제점을 검토한다. 그리고 이를 바탕으로 Baum-Welch 학습법 이용한 새로운 HMM 기반 BWE 방법을 제안한다. Jax 학습법의 특징은 HMM 모델의 상태확률 및 관찰확률을 개별적으로 모델링 한다는 점 그리고 특히 상태확률 계산에 있어서, 경성 분류 (hard classification) 방법인 벡터 양자화 (Vector Quantization, VQ) 알고리즘을 사용한다는 점 등으로 요약될 수 있다. 반면 Baum-Welch 학습법은 상태확률 및 관찰확률을 모두 E-M (Expectation-Maximization) 알고리즘의 틀속에서 통합적으로 계산하며 특히 상태확률 계산에 있어서, E-M 알고리즘 속에 내포된 연성 분류 (soft classification) 방법을 사용한다. 결론적으로, Baum-Welch 학습법은 Jax의 학습법의 일반화된 형태로 볼 수 있으며, 보다 유연하고 적응적인 학습능력을 가진 알고리즘임을 알 수 있다. 따라서 학습 데이터에 대한 보다 정확한 HMM 모델링이 가능하며 아울러, 이와 같이 개선된 HMM 모델을 활용함으로써 BWE 시스템의 성능 향상을 가져 올 수 있었다.

본 논문은 다음과 같이 구성된다. 2절에서는 본 논문에서 다룰 BWE 문제를 함수형태로 간략히 정리하고자 한다. 3절에서는 Jax에 의해 제시된 기존의 HMM 방법을 서술한다. 4절에서는 HMM의 기본 학습법인 Baum-Welch 학습법의 관점에서 Jax가 제시한 학습법을 해석한다. 5절에서는 4절에서의 분석에 기초하여 보다 완전

한 형태의 HMM 기반 BWE 방법을 제시한다. 6절에서는 실험 및 결과분석을 7절에서는 결론을 내린다.

II. BWE 문제의 정의

지금부터 일반성을 훼손하지 않는 범위내에서 BWE 문제를 파라미터화된 하위대역의 스펙트럴 포락선 (spectral envelope) 정보로부터 파라미터화된 상위대역 스펙트럴 포락선 정보를 추정하는 문제로 단순화하기로 한다. 즉,

$$\mathbf{y}^+ = F\{\mathbf{x}\} \quad (1)$$

여기서 \mathbf{x} 는 관찰된 하위대역 특징 벡터 (feature vector)를 나타내며, \mathbf{y}^+ 는 추정된 상위대역 특징 벡터를 나타내고, F 는 하위대역과 상위 대역을 관계짓는 추정 함수 (estimation function) 혹은 사상함수 (mapping function)를 나타낸다. 특징벡터들은 흔히 LPC 계수, 선 스펙트럴 주파수 (Line Spectral Frequency, LSF)나 캡스트럴 계수 (Cepstral Coefficient, CC) 가 사용된다. 본 논문에서는 하위대역 특징 벡터 \mathbf{x} 로는 10차의 멜 주파수 캡스트럴 계수 (Mel-Frequency Cepstral Coefficient, MFCC)를, 그리고 상위대역 특징 벡터 \mathbf{y}^+ 로는 10차의 LPC 캡스트럴 계수 (LPC-CC)에 이득비 (gain ratio, g)를 더하여 총 11차의 특징벡터를 사용하기로 한다. 상위대역 벡터 \mathbf{y}^+ 로서 MFCC 계수 대신 LPC-CC 계수를 사용하는 이유는 멜 스케일에 의해 변형되지 않은 선형 주파수 영역에서 추정함수의 성능을 평가하기 위함이다. 이득비 g 는 다음과 같이 정의된다.

$$g = \frac{\sigma_{high}}{\sigma_{low}} \quad (2)$$

여기서 σ_{low} 와 σ_{high} 는 각각 하위대역 및 상위대역의 LPC 잔차신호의 RMS (root-mean-square) 값이다. 벡터 사상함수 F 는 크게 선형과 비선형, 기억능력 (memory) 유무에 따라 나뉠 수 있는데 HMM 기반 사상함수는 기억능력을 가진 비선형 사상함수라 할 수 있다. BWE의 성능은 이 사상함수를 어떻게 설계하느냐에 따라 크게 달라진다.

III. Jax가 제안한 HMM 기반 BWE 방법

다음 설명은 Jax의 최신 논문 [7]을 중심으로 전개된다. 여기서 제시된 내용은 논문 [5] 및 [6]의 내용들을 포함하고 있다. 또한 설명에 있어서 가급적 원본의 기호법을 그대로 따르려고 노력하였다. HMM 학습 데이터로는 16kHz 샘플링된 음성신호로부터 추출된 특징벡터들을 사용한다. 임의의 광대역 학습벡터를 \mathbf{z} 라 하고 b 차의 임의의 하위대역 부분 벡터를 $\mathbf{x}=[x_1, x_2, \dots, x_b]^T$, d 차의 임의의 상위대역 부분 벡터를 $\mathbf{y}=[y_1, y_2, \dots, y_d]^T$ 로 나타내기로 하면, 광대역 벡터 \mathbf{z} 는 두 행벡터 \mathbf{x} , \mathbf{y} 에 의해 $\mathbf{z}=[\mathbf{x}^T \mathbf{y}^T]^T$ 와 같이 표현될 수 있다. 논문 [7]에서 제시된 방법에 따르면, HMM 학습에 있어서 상태관련 확률 즉, 초기 상태 확률 및 상태전이 확률의 계산에는 상위대역 부분 벡터 \mathbf{y} 를 사용하고, 관찰 확률의 계산에는 (상위대역 부분 벡터 \mathbf{y} 에 일대일 대응하는) 광대역 벡터 \mathbf{z} 를 사용한다. 즉 HMM 모델을 학습함에 있어 단일한 관찰열 (가령 광대역 벡터 \mathbf{z} 의 관찰열)을 사용하지 않고 모델 파라미터에 따라 각자 다른 관찰열 (observation sequence)을 사용한다 (그림 1 참조).

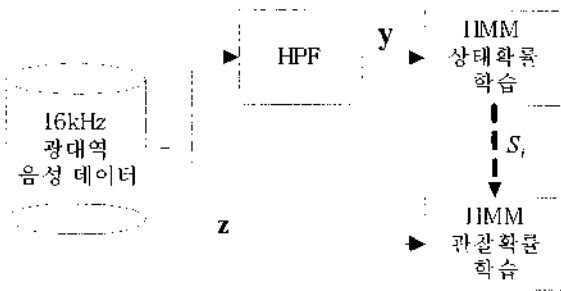


그림 1. Jax의 HMM 모델 학습 방법
Fig. 1. Jax's HMM training method.

3.1. HMM 상태에 대한 정의

HMM의 각 상태 S_i , $1 \leq i \leq N_s$ 는 상위대역 벡터 \mathbf{y} 에 LBG VQ 알고리즘을 적용하여 생성된 코드북 $C = \{\hat{y}_i, 1 \leq i \leq N_s\}$ 의 각 코드벡터 (code-vector) \hat{y}_i 로 정의된다 [9]. 즉 HMM 상태의 수는 코드북의 코드벡터 수와 동일하다.

3.2. 상태 확률

초기 상태 확률을 π_i 즉,

$$\pi_i = P(S_i), \quad 1 \leq i \leq N_s \quad (3)$$

로 나타내기로 하고 상태 전이 확률을 a_{ij} 즉,

$$a_{ij} = P(S_j | S_i), \quad 1 \leq i, j \leq N_s \quad (4)$$

로 나타내기로 하자. 이들 확률 값들은 3.1 절의 상태에 대한 정의로부터 계산될 수 있다. 즉, HMM 학습 데이터 중 m 번째 프레임의 상위대역 벡터를 $\mathbf{y}(m)$ 라 할 때 이 벡터가 속한 상태는 다음 식에 의해 결정된다.

$$i_{opt} = \arg \min_{1 \leq i \leq N_s} \|\mathbf{y}(m) - \hat{\mathbf{y}}_i\|^2 \quad (5)$$

여기서 $\|\cdot\|$ 는 유클리드 거리 (Euclidean distance)를 나타낸다. 식 (5)에 의해 학습 데이터 전체에 대한 상태 열 (state sequence)를 계산할 수 있으며 이 상태열로부터 초기 상태 확률 π_i 및 상태 전이 확률 a_{ij} 을 계산하게 된다.

3.3. 관찰 확률

식 (5)를 이용하여 각 상위대역 부분벡터 \mathbf{y} 및 이에 (일대일) 대응되는 광대역 벡터 \mathbf{z} 가 속한 상태에 대한 분류를 마친 다음, 각 상태 S_i 에 속한 광대역 벡터 \mathbf{z} 를 이용하여 각 상태 S_i 의 관찰 확률을 계산한다. 상태 S_i 의 광대역 벡터 \mathbf{z} 의 관찰 확률은 식 (6)과 같이 L 개의 가우스 혼합 성분 (Gaussian mixture component)들로 구성된 GMM (Gaussian Mixture Model) 모델에 의해 표현된다.

$$P(\mathbf{z} | S_i) = P(\mathbf{x}, \mathbf{y} | S_i) = \sum_{u=1}^L \rho_{x,u} f(\mathbf{z}; \boldsymbol{\mu}_{x,u}, \mathbf{V}_{x,u}) \quad (6)$$

여기서 $\rho_{x,u}$ 는 $0 < \rho_{x,u} \leq 1$ 및 $\sum_{u=1}^L \rho_{x,u} = 1$ 의 성질을 만족하는 가중치를 나타내고, $\boldsymbol{\mu}_{x,u}$ 및 $\mathbf{V}_{x,u}$ 은 가우스 혼합 성분의 평균벡터 (mean vector) 및 공분산 행렬 (covariance matrix)을 나타낸다. 광대역 벡터 \mathbf{z} 의 각 성분들이 비상관관계 (statistically uncorrelated)를 가진다고 가정하면 공분산 행렬 $\mathbf{V}_{x,u}$ 은 대각행렬 (diagonal matrix)이 되며 식 (6)의 가우스 혼합 성분은 하위대역 부분과 상위대역 부분으로 쉽게 나누어진다. 즉,

$$f(\mathbf{z}; \boldsymbol{\mu}_{x,u}, \mathbf{V}_{x,u}) = f(\mathbf{x}; \boldsymbol{\mu}_{x,u}, \mathbf{V}_{x,u}) f(\mathbf{y}; \boldsymbol{\mu}_{y,u}, \mathbf{V}_{y,u}) \quad (7)$$

각 상태 S_i 의 GMM 모델의 학습은 잘 알려진 E-M 학습법을 이용하여 이루어진다 [10]. 편의상, 상태 S_i 에 속하는 광대역 벡터들의 집합을 $Z_i = \{z_i(n), 1 \leq n \leq M_{S_i}\}$ 라 하자. 여기서 M_{S_i} 는 총 M 개의 학습 데이터 중 상태 S_i 에 속하는 광대역 벡터들의 수를 나타내며, n 은 전체 학습 데이터에 대한 프레임 인덱스 m 과 구별하여 상태 S_i 에 속하는 광대역 벡터들의 나열 순서를 가리킨다. 그러면 상태 S_i 의 과거의 GMM 파라미터 $\rho_{i,d}, \mu_{i,d}, V_{i,d}$ 로부터 새로운 GMM 파라미터 $\rho'_{i,d}, \mu'_{i,d}, V'_{i,d}$ 를 계산하는 식은 잘 알려진 바와 같이 식 (8) 및 식 (9)–(11)과 같다.

Expectation 단계:

$$\omega'_{i,d}(n) = P(i | z_i(n)) = \frac{\rho_{i,d} f(z_i(n); \mu_{i,d}, V_{i,d})}{\sum_{i=1}^I \rho_{i,d} f(z_i(n); \mu_{i,d}, V_{i,d})} \quad (8)$$

Maximization 단계:

$$\rho'_{i,d} = \frac{\sum_{n=1}^{M_{S_i}} \omega'_{i,d}(n)}{M_{S_i}} \quad (9)$$

$$\mu'_{i,d} = \frac{\sum_{n=1}^{M_{S_i}} \omega'_{i,d}(n) z_i(n)}{\sum_{n=1}^{M_{S_i}} \omega'_{i,d}(n)} \quad (10)$$

$$\Sigma'_{i,d} = \frac{\sum_{n=1}^{M_{S_i}} \omega'_{i,d}(n) z_i(n) z_i^T(n)}{\sum_{n=1}^{M_{S_i}} \omega'_{i,d}(n)} - \mu'_{i,d} \mu_{i,d}^T \quad (11)$$

여기서 위 첨자 T 는 전치 (transpose) 연산을 가리킨다. 그리고 $\omega'_{i,d}(n)$ 는 상태 S_i 의 임의의 벡터 $z_i(n)$ 가 상태 S_i 의 i 번째 가우스 혼합 성분에서 발생할 확률을 나타낸다.

3.4. MMSE 추정식

앞서 설명된 Jax의 학습법에 의한 HMM 모델이 주어진 조건 하에서, 관찰된 하위대역 벡터 \mathbf{x} 로부터 상위대역 벡터 \mathbf{y} 를 추정하기 위한 MMSE (Minimum Mean Square Error) 추정식 (estimation rule)은 다음과 같다.

$$\mathbf{y}^*(m) = E\{\mathbf{y}(m) | \mathbf{X}(m)\} = \int_{\mathcal{R}^N} \mathbf{y}(m) P(\mathbf{y}(m) | \mathbf{X}(m)) d\mathbf{y}(m) \quad (12)$$

여기서 $\mathbf{y}^*(m)$ 은 m 번째 프레임에서 추정된 상위대역 벡터를 가리키며, $\mathbf{X}(m) = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)\}$ 은 m 번째 프레

임까지 부분 관찰된 하위대역 벡터 관찰열을 가리킨다. 또한 $E\{\cdot\}$ 는 평균 연산을 가리킨다. 식 (12)에서 조건부 확률 $P(\mathbf{y}(m) | \mathbf{X}(m))$ 은 직접 계산하는 것이 어려우므로 다음과 같이 HMM의 상태를 매개변수로 전개한다.

$$P(\mathbf{y}(m) | \mathbf{X}(m)) = \sum_{i=1}^{N_s} P(\mathbf{y}(m), S_i(m) | \mathbf{X}(m)) \quad (13)$$

여기서 $S_i(m)$ 은 m 번째 프레임에서의 상태가 i 라는 것을 가리킨다. 식 (13)의 오른쪽 식은 Bayes rule을 활용하여 다음과 같이 전개될 수 있다.

$$P(\mathbf{y}(m), S_i(m) | \mathbf{X}(m)) = P(\mathbf{y}(m) | S_i(m), \mathbf{X}(m)) P(S_i(m) | \mathbf{X}(m)) = P(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) P(S_i(m) | \mathbf{X}(m)) \quad (14)$$

식 (14)의 두번째 등식은 m 번째 프레임의 출력 벡터 $\mathbf{x}(m)$ 과 $\mathbf{y}(m)$ 은 m 번째 프레임의 상태 $S_i(m)$ 에 의해서만 영향을 받는다는 HMM의 기본 가정에 의해 성립한다. 식 (13)과 (14)를 식 (12)에 대입하여 정리하면 다음 식이 만들어진다.

$$\mathbf{y}^*(m) = \sum_{i=1}^{N_s} P(S_i(m) | \mathbf{X}(m)) \times \int_{\mathcal{R}^N} \mathbf{y}(m) P(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) d\mathbf{y}(m) \quad (15)$$

여기서 오른쪽의 적분식은 $S_i(m)$ 과 $\mathbf{x}(m)$ 에 조건지워진 $\mathbf{y}(m)$ 의 평균을 의미한다. 즉,

$$\int_{\mathcal{R}^N} \mathbf{y}(m) P(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) d\mathbf{y}(m) = E\{\mathbf{y}(m) | S_i(m), \mathbf{x}(m)\} = \mathbf{c}_i(m) \quad (16)$$

식 (16)을 식 (15)에 대입하면 최종적으로 식 (17)과 같이, m 번째 프레임에서 상위대역 벡터 $\mathbf{y}^*(m)$ 을 MMSE 추정하기 위한 계산식이 완성된다.

$$\mathbf{y}^*(m) = \sum_{i=1}^{N_s} P(S_i(m) | \mathbf{X}(m)) \mathbf{c}_i(m) \quad (17)$$

여기서 조건평균벡터 $\mathbf{c}_i(m)$ 는 다음과 같은 유도과정에 의해 계산된다 (편의상 일부 프레임 인덱스 m 생략.) [11].

$$\begin{aligned}
 c_i(m) &= E\{y(m) | S_i(m), \mathbf{x}(m)\} \\
 &= \int_{\mathcal{R}^L} y p_{y,b}(y | S_i, \mathbf{x}) dy \\
 &= \int_{\mathcal{R}^L} \frac{y p_{x,y}(\mathbf{x}, y | S_i)}{p_x(\mathbf{x} | S_i)} dy \\
 &= \frac{1}{p_x(\mathbf{x} | S_i)} \int_{\mathcal{R}^L} y p_x(\mathbf{z} | S_i) dy \\
 &= \frac{1}{p_x(\mathbf{x} | S_i)} \int_{\mathcal{R}^L} y \sum_{i=1}^L \rho_{x,i} f(\mathbf{z}; \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) dy \\
 &= \frac{1}{p_x(\mathbf{x} | S_i)} \int_{\mathcal{R}^L} y \sum_{i=1}^L \rho_{x,i} f(\mathbf{x}; \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) f(y; \boldsymbol{\mu}_{y,i}, \mathbf{V}_{y,i}) dy \\
 &= \frac{1}{p_x(\mathbf{x} | S_i)} \sum_{i=1}^L \rho_{x,i} f(\mathbf{x}; \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) \int_{\mathcal{R}^L} y f(y; \boldsymbol{\mu}_{y,i}, \mathbf{V}_{y,i}) dy \\
 &= \frac{\sum_{i=1}^L \rho_{x,i} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) \boldsymbol{\mu}_{y,i}}{\sum_{i=1}^L \rho_{x,i} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i})}
 \end{aligned} \tag{18}$$

여기서 다섯번째 등식은 광대역 벡터 \mathbf{z} 의 관찰 확률인 식 (6)을 대입한 것이고 여섯번째 등식은 공분산 행렬 $\mathbf{V}_{x,i}$ 이 대각행렬이라는 가정에 의해 성립한다 (식 (7) 참조). 마지막 등식에 있는 평균벡터 $\boldsymbol{\mu}_{y,i}$ 는 i 번째 상태의 l 번째 가우스 혼합 성분의 상위대역 평균벡터를 가리킨다.

식 (17)의 가중치 확률 $P(S_i(m) | \mathbf{X}(m))$ 은 Bayes rule을 활용하여 식 (19)와 같이 전개될 수 있다. 즉,

$$\begin{aligned}
 P(S_i(m) | \mathbf{X}(m)) &= \frac{P(S_i(m), \mathbf{X}(m))}{P(\mathbf{X}(m))} \\
 &= \frac{P(S_i(m), \mathbf{X}(m))}{\sum_{i=1}^{N_s} P(S_i(m), \mathbf{X}(m))}
 \end{aligned} \tag{19}$$

여기서 $P(S_i(m), \mathbf{X}(m))$ 은 잘 알려진 HMM의 전향확률 (forward probability) $\alpha_{x,i}(m)$ 로서 다음의 관계를 가지며 [8],

$$\begin{aligned}
 P(S_i(m), \mathbf{X}(m)) &= \alpha_{x,i}(m) \\
 &= P(S_i(m), \mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m))
 \end{aligned} \tag{20}$$

$\alpha_{x,i}(m)$ 의 회귀적인 특성 즉,

$$\alpha_{x,i}(1) = \pi_i P(\mathbf{x}(1) | S_i) \tag{21}$$

$$\alpha_{x,i}(m+1) = \sum_{k=1}^{N_s} \alpha_{x,k}(m) a_{k,i} P(\mathbf{x}(m) | S_k) \tag{22}$$

을 활용하여 효과적으로 계산된다. 식 (22)에서 하위대역 부분 벡터 $\mathbf{x}(m)$ 의 관찰 확률 $P(\mathbf{x}(m) | S_i)$ 은 식 (6)의 광대역 벡터 $\mathbf{z}(m)$ 의 관찰 확률 $P(\mathbf{z}(m) | S_i)$ 의 주변확률밀도함수 (marginal pdf)로서 식 (7)의 조건으로부터 다음과 같이 직접 구해진다 (편의상 일부 프레임 인덱스 m 생략), 즉,

$$\begin{aligned}
 P(\mathbf{x}(m) | S_i) &= \int_{\mathcal{R}^L} \sum_{i=1}^L \rho_{x,i} f(\mathbf{z}; \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) dy \\
 &= \sum_{i=1}^L \rho_{x,i} f(\mathbf{x}; \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i}) \int_{\mathcal{R}^L} f(y; \boldsymbol{\mu}_{y,i}, \mathbf{V}_{y,i}) dy \\
 &= \sum_{i=1}^L \rho_{x,i} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,i}, \mathbf{V}_{x,i})
 \end{aligned} \tag{23}$$

IV. Baum-Welch 학습법과 Jax 학습법의 비교분석

본 절에서는 Jax의 학습법과 HMM 모델의 일반적인 학습법인 Baum-Welch 학습법을 비교분석하고자 한다. 이를 위해 Jax가 제안한 학습의 틀 즉, HMM 상태에 대한 정의나 그림 1에서와 같이 상태확률과 관찰확률 계산을 위한 관찰열을 상위대역 관찰열 $\mathbf{Y}(M)$ 과 광대역 관찰열 $\mathbf{Z}(M)$ 으로 나누는 등의 틀은 유지하면서 Baum-Welch 학습법을 적용하면서 Jax 학습법과의 관계 및 차이점을 비교 검토한다.

4.1. HMM 상태에 대한 정의

Baum-Welch 학습법을 이용하여 HMM을 생성할 경우에도 상태 S_i 에 대한 정의는 3.1절에서와 마찬가지로 상위대역 벡터 \mathbf{y} 에 대한 코드북의 코드벡터로 대응시킬 수 있다. 그러나 이 경우 각 상위대역 벡터 \mathbf{y} 가 속한 상태 S_i 는 식 (5)에서 처럼 경성 분류되는 것이 아니라 일종의 사후확률 (a posteriori probability) $\gamma_{y,i}(m)$ 에 의해 연성 분류된다. 즉,

$$\begin{aligned}
 \gamma_{y,i}(m) &= P(S_i(m) | \mathbf{Y}(M)) \\
 &= \frac{\alpha_{y,i}(m) \beta_{y,i}(m)}{\sum_{k=1}^{N_s} \alpha_{y,k}(m) \beta_{y,k}(m)}
 \end{aligned} \tag{24}$$

여기서 기호 $\gamma_{y,i}(m)$ 는 HMM 모델과 총 M 개의 상위대역 전체 관찰열 $\mathbf{Y}(M) = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(M)\}$

이 주어진 조건에서 m 번째 프레임 벡터 $y(m)$ 이 상태 S_i 에 속할 확률을 나타낸다. 또한 식 (24)의 두번째 등식의 기호 $\alpha_{y,i}(m)$ 와 $\beta_{y,i}(m)$ 는 각각 상위대역 관찰열 $\mathbf{Y}(M)$ 에 대한 HMM의 전향확률 및 후향확률 (backward probability)을 나타낸다 [8]. 즉,

$$\alpha_{y,i}(m) = P(S_i(m), \mathbf{Y}(m)) \tag{25}$$

$$\beta_{y,i}(m) = P(y(m+1), y(m+2), \dots, y(M) | S_i(m)) \tag{26}$$

4.2. 상태 확률

과거의 HMM 모델로부터 새로운 초기 상태 확률 π'_i 및 상태 천이 확률 a'_{ij} 을 계산하기 위한 Baum-Welch 학습법은 식 (27) 및 (28)과 같이 주어진다.

$$\pi'_i = \gamma_{y,i}(1) \tag{27}$$

$$a'_{ij} = \frac{\sum_{m=1}^M \gamma_{y,i,j}(m)}{\sum_{m=1}^M \gamma_{y,i}(m)} \tag{28}$$

여기서 식 (27)의 $\gamma_{y,i}(1)$ 는 HMM 모델과 총 M 개의 상위대역 전체 관찰열 $\mathbf{Y}(M)$ 이 주어진 조건에서 첫번째 프레임 벡터 $y(1)$ 이 상태 S_i 에 속할 확률을 나타낸다. 또한 식 (28)의 $\gamma_{y,i,j}(m)$ 는 HMM 모델과 전체 관찰열 $\mathbf{Y}(M)$ 이 주어진 조건에서, m 번째 프레임의 상태가 i 이고, $m+1$ 번째 프레임의 상태가 j 일 확률을 나타낸다. 즉,

$$\begin{aligned} \gamma_{y,i,j}(m) &= P(S_i(m), S_j(m+1) | \mathbf{Y}(M)) \\ &= \frac{\alpha_{y,i}(m) a_{ij} P(y(m+1) | S_j) \beta_{y,j}(m+1)}{\sum_{k=1}^{N_s} \alpha_{y,k}(m) \beta_{y,k}(m)} \end{aligned} \tag{29}$$

여기서 식 (29)와 식 (24)는 다음의 관계를 가진다.

$$\gamma_{y,i}(m) = \sum_{j=1}^{N_s} \gamma_{y,i,j}(m) \tag{30}$$

이상에서 알 수 있는 바와 같이, 새로운 초기 상태 확률 π'_i 및 상태 천이 확률 a'_{ij} 을 계산하기 위한 Baum-Welch 학습 과정은 식 (24)를 중심으로한 연성 분류의 기초위에 전개됨을 알 수 있다.

4.3. 관찰 확률

각 상태 S_i 에 대한 광대역 벡터 \mathbf{z} 의 관찰 확률은 3.3절의 식 (6)과 같이 GMM 모델에 의해 정의된다. 과거의 (GMM 파라미터를 포함한) HMM 모델로부터 각 상태 S_i 의 새로운 GMM 파라미터 $\rho'_{z,i}$, $\mu'_{z,i}$, $\mathbf{V}_{z,i}$ 의 계산식은 Baum-Welch 학습법에 의해 식 (31)-(33)과 같이 계산된다.

$$\rho'_{z,i} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m)}{\sum_{m=1}^M \sum_{l=1}^L \gamma'_{z,i,l}(m)} \tag{31}$$

$$\mu'_{z,i} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m) \mathbf{z}(m)}{\sum_{m=1}^M \gamma'_{z,i}(m)} \tag{32}$$

$$\Sigma'_{z,i} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m) \mathbf{z}(m) \mathbf{z}^T(m)}{\sum_{m=1}^M \gamma'_{z,i}(m)} - \mu'_{z,i} \mu_{z,i}^T \tag{33}$$

여기서 $\gamma'_{z,i}(m)$ 는 $\gamma_{z,i}(m)$ 을 일반화한 식으로, 총 M 개의 광대역 전체 관찰열 $\mathbf{Z}(M)$ 이 주어진 조건에서, m 번째 프레임 벡터 $\mathbf{z}(m)$ 이 상태 S_i 의 l 번째 혼합성분에 속할 확률을 나타낸다. 이를 수식으로 나타내면 다음과 같다. 즉,

$$\begin{aligned} \gamma'_{z,i}(m) &= P(S_i(m), l | \mathbf{Z}(M)) \\ &= P(S_i(m) | \mathbf{Z}(M)) P(l | S_i(m), \mathbf{Z}(M)) \\ &= P(S_i(m) | \mathbf{Z}(M)) P(l | S_i(m), \mathbf{z}(m)) \\ &= \gamma_{z,i}(m) \omega'_{z,i,l}(m) \end{aligned} \tag{34}$$

$$= \left[\frac{\alpha_{z,i}(m) \beta_{z,i}(m)}{\sum_{k=1}^{N_s} \alpha_{z,k}(m) \beta_{z,k}(m)} \right] \left[\frac{\rho_{z,i,l} f(\mathbf{z}(m); \mu_{z,i,l}, \mathbf{V}_{z,i,l})}{\sum_{k=1}^L \rho_{z,i,k} f(\mathbf{z}(m); \mu_{z,i,k}, \mathbf{V}_{z,i,k})} \right]$$

여기서 세번째 등식은 HMM의 기본 가정 즉, 현재 프레임의 출력벡터는 현재 프레임의 상태에 의해서만 영향을 받는다는 전제에 의해 성립한다 (식 (14)의 두번째 등식과 비교). 또한 네번째 등식에서 $\gamma_{z,i}(m)$ 기호는 m 번째 프레임 벡터 $\mathbf{z}(m)$ 에 대한 연성 분류와 관련된 사후확률로서, 관찰열 $\mathbf{Z}(M)$ 이 주어진 조건에서, m 번째 프레임 $\mathbf{z}(m)$ 이 상태 S_i 에 속할 확률을 나타낸다. 만약 여기서 이 $\gamma_{z,i}(m)$ 값을 $\gamma_{y,i}(m)$ 로 바꾸면 $\gamma'_{z,i}(m)$ 값에 대한 다음과 같은 근사식 (approximate equation)이 만들어진다. 즉,

$$\begin{aligned} \gamma'_{z,i}(m) &\approx \gamma'_{y,z,i}(m) \\ &= \gamma_{y,i}(m) \omega'_{z,i}(m) \\ &= \left[\frac{\alpha_{y,i}(m) \beta_{y,i}(m)}{\sum_{i=1}^{N_s} \alpha_{y,i}(m) \beta_{y,i}(m)} \right] \left[\frac{\rho_{z,i} f(\mathbf{z}(m); \boldsymbol{\mu}_{z,i}, \mathbf{V}_{z,i})}{\sum_{i=1}^{N_s} \rho_{z,i} f(\mathbf{z}(m); \boldsymbol{\mu}_{z,i}, \mathbf{V}_{z,i})} \right] \\ &= P(S_i(m) | \mathbf{Y}(M)) P(i | S_i(m), \mathbf{z}(m)). \end{aligned} \quad (35)$$

여기서 $\gamma'_{z,i}(m)$ 의 계산을 위해 $\gamma_{z,i}(m)$ 를 쓰느냐 $\gamma_{y,i}(m)$ 를 쓰느냐의 차이는 m 번째 프레임 벡터 $\mathbf{z}(m)$ 이 속한 상태 S_i 를 결정함에 있어 광대역 벡터 관찰열 $\mathbf{Z}(M)$ 을 관찰 조건으로 하느냐 혹은 (이에 대응되는) 상위대역 부분 벡터 관찰열 $\mathbf{Y}(M)$ 을 관찰조건으로 하느냐의 차이이다. 후자는 Jax의 방식 즉, $\mathbf{z}(m)$ 이 속한 상태 S_i 를 결정함에 있어 상위대역 부분벡터 $\mathbf{y}(m)$ 만을 관찰하는 방식에 대응된다. 여기서 한걸음 더 나아가 Jax 학습법의 식 (5)처럼 경성분류기를 도입하기로 하자. 즉, 식 (35)의 $\gamma_{y,i}(m)$ 값 대신 식 (36) 및 (37)의 조건식을 적용한다고 가정하자. 즉,

$$i_{opt} = \arg \max_{1 \leq i \leq N_s} \gamma_{y,i}(m) \quad (36)$$

$$\begin{cases} \gamma_{y,i}(m) = 1, & \text{if } i = i_{opt} \\ \gamma_{y,i}(m) = 0, & \text{if } i \neq i_{opt} \end{cases} \quad (37)$$

그러면 Baum-Welch 학습법에 의한 관찰확률 계산식 (35) 및 (31)–(33)은 Jax의 관찰확률 계산식 (즉, GMM 학습법) (8) 및 식 (9)–(11)과 유사한 형태가 됨을 알 수 있다. 즉,

$$\gamma'_{z,i}(m) \approx \gamma'_{y,z,i}(m) \Big|_{\mathbf{y}(m) \in S_i} \rightarrow \omega'_{z,i}(m) \quad (38)$$

그러나 식 (36)과 (37)을 활용하여 만들어진 식 (38)과 식 (5)를 활용하여 만들어진 식 (8)은 근본적인 차이를 가진다. 왜냐하면, 식 (36)의 $\gamma_{y,i}(m)$ 값은 HMM 모델값에 의해 결정되며 따라서 학습과정을 통하여 HMM 모델이 변화됨에 따라 마찬가지로 변화하게 된다. 이는 매 학습횟수마다 동일한 벡터 $\mathbf{y}(m)$ 에 대해 다른 i_{opt} 결과를 출력한다는 것을 의미한다. 또는 달리 말하면, 상태 S_i 와 관련된 코드북 C 가 매 학습횟수마다 재설계된다는 것을 의미한다. 반면, 식 (5)는 학습초기에 VQ 알고리즘에 의해 결정된 코드북 C 에 의해 계산되므로 학습과정을 통하여

변해가는 GMM 모델에 상관없이 동일한 벡터 $\mathbf{y}(m)$ 에 대해서 항상 동일한 i_{opt} 결과를 출력한다. 정리하자면, 만약 Baum-Welch 학습법에 경성분류기를 도입하면 계산식의 형태는 Jax 학습법과 비슷해지지만 Baum-Welch 학습법의 경우 코드북 C 가 학습과정을 통하여 계속 재설계된다는 점 그리고 이는 다시 모델학습에 영향을 준다는 점에서 Jax 학습법과 다르다.

이상의 검토를 종합하면, Baum-Welch 학습법과 Jax의 학습법의 차이점은 크게 다음 두가지로 요약할 수 있다. 첫째, 각 상태 S_i 의 영역분할에 있어서 연성 분류기를 사용하는냐 경성 분류기를 사용하는냐 이다. 둘째, 각 상태 S_i 와 관련된 코드북 C 를 학습과정을 통하여 재설계 하느냐 고정시키느냐의 차이점이다. Jax 학습법은 VQ 알고리즘을 사용하여 코드북 C 를 생성함으로써 학습벡터들의 시간적 나열순서의 특징들을 코드북 C 의 설계에 반영하지 못하고 있다. 또한 학습과정을 통하여 코드북 C 를 재설계하고 적응시켜나가는 구조가 아니다. 그러므로 Baum-Welch 학습법에 비해 성능이 떨어지는 HMM 모델의 생성을 예상할 수 있다.

4.4. MMSE 추정식

앞서 설명된 Baum-Welch 학습법에 의해 학습된 HMM 모델이 주어졌다고 할 때, 이를 이용한 MMSE 추정식은 3.4절에 제시된 기존의 방법 즉, 식 (17)과 동일하다.

V. 제안하는 HMM 기반 BWE 방법

앞서 4.3절에서 언급한 Baum-Welch 학습법과 Jax의 학습법의 두가지 차이점으로부터 Baum-Welch 학습법이 학습데이터를 모델링하기 위한 보다 유연하고 적응적인 알고리즘임을 알 수 있다. 더불어 Baum-Welch 학습법을 BWE를 위한 학습데이터의 모델링에 적용함에 있어, 그림 1처럼 상태확률 계산 및 관찰확률 계산을 위한 학습데이터를 상위대역 관찰열 $\mathbf{Y}(M)$ 과 광대역 관찰열 $\mathbf{Z}(M)$ 으로 각각 나누는 것은 근사식 (35)가 시사하는 바와 같이 바람직하지 않다. 즉, 이는 E-M 알고리즘이 관찰확률 (파라미터) 추정하는데 있어서 오차를 발생시키며 결국 전체적으로 HMM 학습이 수렴하지 못하고 발진하도록 한다. 따라서 관찰확률뿐만 아니라 상태확률 모델링을

위에서도 광대역 관찰열 $\mathbf{z}(M)$ 을 동일하게 사용하는 것이 바람직하다 (그림 2 참조). 이 경우 상태 확률에 대한 Baum-Welch 학습법은 식 (27) 및 (28)로부터 식 (39) 및 (40)과 같이 수정된다.

$$\pi'_i = \gamma_{z,i}(1) \tag{39}$$

$$\alpha'_{ij} = \frac{\sum_{m=1}^{T-1} \gamma_{z(i,j)}(m)}{\sum_{m=1}^{T-1} \gamma_{z,i}(m)} \tag{40}$$

여기서 아래 첨자 \mathbf{z} 는 광대역 관찰열 $\mathbf{z}(M)$ 을 관찰조건으로 하여 상태확률이 학습됨을 가리킨다. 관찰 확률에 대한 식은 4.3절의 식 (31)-(33)과 동일하다. 또한 HMM 모델이 주어졌다고 할 때 이를 이용한 MMSE 추정식 역시 4.4절 (혹은 3.4절)과 동일하다.

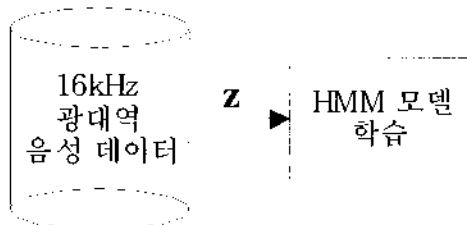


그림 2. Baum-Welch 학습법을 이용한 HMM 모델 학습 방법
Fig. 2. HMM training method using Baum-Welch re-estimation algorithm.

VI. 성능 비교 평가

학습 및 평가를 위한 데이터로는 16kHz 샘플링된 TIMIT 데이터베이스를 이용하였다 [12]. 학습에는 전체 TIMIT 학습 데이터 (화자 총 462명, 발성음 총 4620개)가 이용되었으며 테스트를 위해서는 핵심 테스트 셋 (Core Test Set 화자 총 24명, 발성음 총 192개)만 사용하였다. 각 프레임별 특징벡터로는 앞서 2절에서 언급한 바와 같이, 하위대역 신호에 대해서는 10차의 MFCC 계수를, 상위대역 신호에 대해서는 10차의 LPC-CC계수 및 이득버를 더하여 총 11차의 계수를 추출하여 사용하였다. 따라서 광대역 프레임 벡터는 총 21차의 계수들로 이루어진 벡터가 된다. 각 프레임별 음질왜곡의 정도를 평가하기 위한 척도로는 식 (41)과 같이 정의된 로그 스펙트럴 왜곡 (Log Spectral Distortion, LSD)이 사용되었다.

$$D_{LSD}^2 = \frac{1}{2\pi} \int_{\omega} \left(20 \log_{10} \frac{g}{|A_{kb}(e^{j\omega})|} - 20 \log_{10} \frac{g'}{|A_{kb}(e^{j\omega})|} \right)^2 d\Omega \quad [dB] \tag{41}$$

여기서 적분의 범위는 상위대역 (4k-kHz)에만 대상으로 한다. 이득 g 와 g' 는 식 (2)에서 정의된 바와 같이 하위대역과 상위대역의 에너지 비를 나타내며, g 는 실제값을 g' 는 추정된 값을 나타낸다. A_{kb} 와 $[A_{kb}']^{-1}$ 는 각각 상위대역의 실제 LPC 모델과 추정된 LPC 모델을 나타낸다. 추정된 LPC 모델 $[A_{kb}']^{-1}$ 의 LPC 계수는 BWE 함수를 통해 예측된 10차의 상위대역 LPC-CC 계수 y^+ 로부터 계산된다. 이때 안정된 LPC 모델을 얻기 위해서는 LPC-CC로부터 직접 LPC 계수로의 변환을 시도하기 보다는 먼저, 푸리에 변환을 이용하여 LPC-CC로부터 전력 스펙트럴 밀도 (Power Spectral Density)를 구하고 이를 다시 푸리에 역변환하여 자기상관계수 (auto-correlation coefficient)를 구한 뒤 최종적으로 Levinson-Durbin 알고리즘을 이용하여 안정된 LPC 계수를 구하는 최적적 방법을 사용한다. (보다 자세한 설명은 논문 [6] 참조)

상태 수 N_s 및 가우스 혼합성분 수 L 에 대한 다양한 조합을 만들어 두 학습법의 성능을 평가하였다. 그림 3 및 4는 각각 가우스 혼합성분 수 L 를 8, 16으로 고정시켰을 때 상태 수 N_s 를 16, 32, 64로 변화시킨데 따른 두 학습법의 성능차이를 나타낸다. 반대로, 그림 5 및 6은 N_s 를 32, 64로 고정시켰을 때 가우스 혼합성분 수 L 를 4, 8, 16으로 변화시킨데 따른 두 학습법의 성능차이를 나타낸다. 그림에서 보듯, 두 학습법 모두 상태 수 혹은 혼합 성분 수가 증가함에 따라 성능이 개선됨을 알 수 있다. 또한 예상된 바와 같이, 모든 경우에 대해 Baum-Welch 학습법이 Jax의 학습법에 비해 우수한 성능을 보이고 있다. 전체적으로 평균 0.52dB (RMS LSD 수치) 그리고, 최소 0.31dB에서 최대 0.8dB까지 개선되었다. 이는 Baum-Welch 학습법이 BWE의 MMSE 추정을 위한 보다 향상된 HMM 모델을 생성함을 보여준다. 특히 그림 5 및 6에서 주목할만한 점은, Jax 방법은 Baum-Welch 방법에 비해 상대적으로 상태 수 N_s 보다는 가우스 혼합 성분 수 L 의 증가에 따라 빠르게 성능이 개선된다는 점이다. 이는 Jax 학습법의 특징에 비롯된 결과로, 가우스 혼합 성분 수 L 의 증가에 따라서 E-M 알고리즘에 의한 관찰확률 모델링이 보다 정교해지고 이것이 VQ 알고리즘에 의한 상태확률의 허술한 모델링을 점차로 보완해주는 데 따른 결과이다. 반면, Baum-Welch 학습법은 두 파라미터 모두에 대해 일정한 속도로 완만한 성능개선을 보여주고 있다. 이는 Baum-Welch 방법은

관찰확률뿐만 아니라 상태확률도 B-M 알고리즘의 틀속에서 모델링함으로써 모델 구조 변경에 상관없이 안정된 HMM 모델을 생성하기 때문이다.

HMM 모델을 이용한 BWE 방법의 개발에 있어서 이상과 같은 개선에도 불구하고 여전히 RMS LSD 절대값이 크다는 사실을 알 수 있다. 이 점은 일반적으로 음성부호화기의 성능평가에서 2dB 이상을 이상치 (outlier)로 간주한다는 점을 고려할 때 더욱 분명해진다. 비록 추가적인 최적화 작업 (가령, 특징벡터 선정 등)을 진행한다고 하더라도 현재의 수준에서 볼 때 궁극적으로 이상치의 범위를 벗어나기는 어렵다고 보여진다. 논문 [13]~[15]에서도 지적인 바와 같이 적어도 음성의 경우, 하위대역 정보만을 이용하여 상위대역 정보를 추정하고자 하는 시도는 근본적인 한계를 가진다. 즉, 음성의 하위대역 성분과 상위대역 성분 사이에는 기대만큼의 상관성을 가지고 있지 않기 때문에 단지 추정방법 (혹은 특징벡터)의 개선을 통해서 도달할 수 있는 BWE 성능향상에는 한계가 있다. 따라서 만족할만한 결과를 얻기 위해서는 본 논문에서와 같

이, 손실대역의 정보를 정확히 추정하고자 하는 객관적인 관점에서의 접근과 더불어 심리음향학적 (psycho-acoustic) 관점에서의 접근 즉, 원래의 음성과 상관없이 최종적으로 듣기 좋은 음성을 생성하고자 하는 관점에서의 연구가 함께 요구된다.

VII. 결론

본 논문에서는 새로운 HMM 기반 BWE 방법을 제시하였다. 제시된 방법은 HMM 모델 학습을 위해 기존의 Jax가 제안한 학습법 대신 일반적인 HMM 모델 학습법인 Baum-Welch 학습법을 사용하였다. 두 학습법에 대한 분석 결과에 의하면 Baum-Welch 학습법은 Jax의 학습법의 일반화된 형태로 볼 수 있으며, 보다 유연하고 적응적인 학습능력을 가진 알고리즘임을 알 수 있었다. 따라서 학습 데이터에 대한 보다 정확한 HMM 모델링이 가능

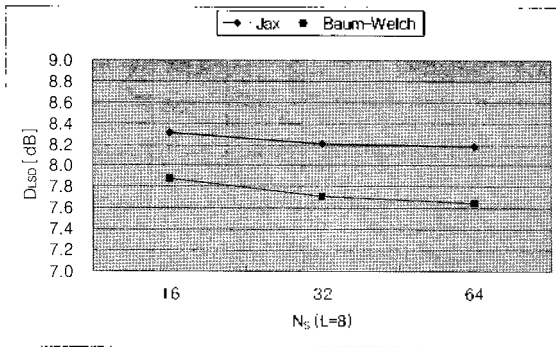


그림 3. 가우스 혼합성분 수 $L=8$ 일 때 $N_s=16, 32, 64$ 에 따른 BWE에 있어서의 Jax 및 Baum-Welch 두 학습법의 성능차이
Fig. 3. Performance comparison of the Jax's and Baum-Welch training methods for BWE when the Gaussian mixture component $L=8$ and the state number $N_s=16, 32, 64$.

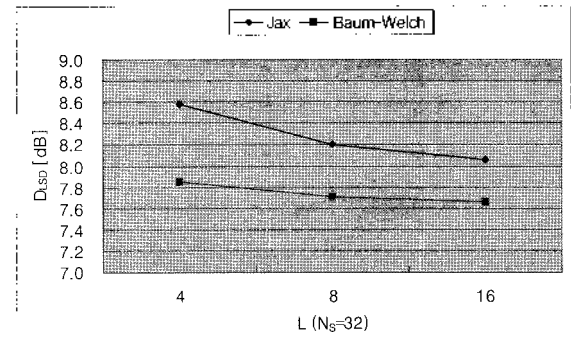


그림 5. 가우스 혼합성분 수 $N_s=32$ 일 때 $L=4, 8, 16$ 에 따른 BWE에 있어서의 Jax 및 Baum-Welch 두 학습법의 성능차이
Fig. 5. Performance comparison of the Jax's and Baum-Welch training methods for BWE when the state number $N_s=32$ and the Gaussian mixture component $L=4, 8, 16$.

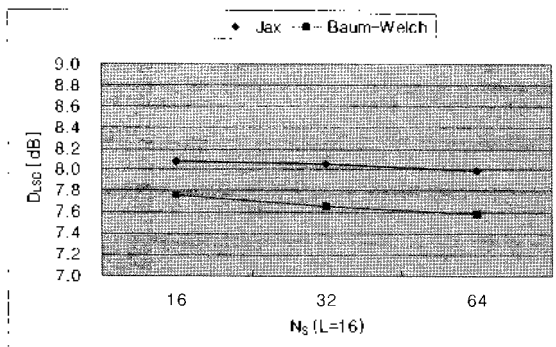


그림 4. 가우스 혼합성분 수 $L=16$ 일 때 $N_s=16, 32, 64$ 에 따른 BWE에 있어서의 Jax 및 Baum-Welch 두 학습법의 성능차이
Fig. 4. Performance comparison of the Jax's and Baum-Welch training methods for BWE when the Gaussian mixture component $L=16$ and the state number $N_s=16, 32, 64$.

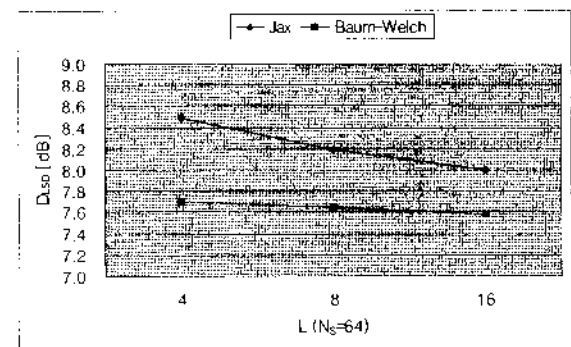


그림 6. 가우스 혼합성분 수 $N_s=64$ 일 때 $L=4, 8, 16$ 에 따른 BWE에 있어서의 Jax 및 Baum-Welch 두 학습법의 성능차이
Fig. 6. Performance comparison of the Jax's and Baum-Welch training methods for BWE when the state number $N_s=64$ and the Gaussian mixture component $L=4, 8, 16$.

하며 아울러, 이와 같이 개선된 HMM 모델을 활용함으로써 BWE 시스템의 성능향상을 가져 올 수 있었다.

한편, 이상과 같은 개선에도 불구하고 여전히 RMS LSD의 절대값이 크다는 사실을 알 수 있다. 적어도 음성 신호에 있어서는 상위대역과 하위대역 사이의 상관성이 충분하지 않다고 알려져 있다. 따라서 만족할만한 BWE 결과를 얻기 위해서는 본 논문에서와 같이 손실 대역의 원래 정보를 정확히 추정하고자 하는 연구와 더불어, 원래의 음성과 상관없이 최종적으로 듣기 좋은 음성을 생성하고자 하는 심리음향학적 관점에서의 접근이 함께 필요하다.

참고 문헌

1. N. Enbom and W. B. Kleijn, "Bandwidth expansion of speech based on vector quantization of the Mel frequency cepstral coefficients," IEEE Workshop on Speech Coding, 171-173, June 1999.
2. K. -Y Park and H. S. Kim, "Narrowband to Wideband Conversion of Speech Using GMM Based Transformation," ICASSP 3, 1843-1846, June 2000.
3. S. Chennoukh, A. Gerrits, and R. Sluijter, "Speech Enhancement via Frequency Bandwidth Extension Using Line Spectral Frequencies," ICASSP 1, 665-668, May 2001.
4. S. Jaisimha and I. Y. Soon, "Bandwidth Extension of Narrow Band Speech Using Cepstral Linear Prediction," Joint Conference of the Fourth International Conference on Multimedia 3, 1404-1407, Dec. 2003.
5. P. Jax and P. Vary, "Wideband extension of telephone speech using a hidden Markov model," IEEE Workshop on Speech Coding, 133-135, Sept. 2000.
6. P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," Signal Processing 83 (8), 1707-1719, Aug. 2003.
7. P. Jax and P. Vary, "Artificial Bandwidth Extension of Speech Signals Using MMSE Estimation Based on a Hidden Markov Model," ICASSP 1, 680-683, April 2003.
8. L. R. Rabiner, "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE 77 (2), 257-286, Feb. 1989.
9. Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun. 28 (1), 84-95, 1980.
10. T. K. Moon, "The expectation-maximization algorithm," IEEE Signal Process. Mag 13 (6), 47-60, Nov. 1996.
11. Wei-shou Hsu, Robust bandwidth Extension of narrowband speech, M.A. thesis, McGill Univ., Dept. of Electrical & Computer Engineering, 26-29, Nov. 2004.
12. J. S. Garofolo, L. F. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA-TIMIT: Acoustic-Phonetic Continuous Speech Corpus," 1990.
13. M. Nilsson, S. V. Andersen, and W. B. Kleijn, "On the mutual information between frequency bands in speech," ICASSP 3, 1327-1330, June 2000.

14. M. Nilsson, H. Gustafsson, S. V. Andersen, and W. B. Kleijn, "Gaussian mixture model based mutual information estimation between frequency bands in speech," ICASSP 1, 525-528, June 2002.
15. Y. Agiomvriannakis and Y. Stylianou, "Combined estimation/coding of highband spectral envelopes for speech spectrum expansion," ICASSP 1, 496-472, May 2004.

저자 약력

• 송근배 (Geun-Bae Song)



1992년 2월 : 아주대학교 전자공학 (학사)
 1995년 2월 : 아주대학교 전자공학 (석사)
 2001년 8월 : 아주대학교 전자공학 (박사)
 1996년 1월~1996년 3월 : 현대전자 S/W 연구소 연구원
 2001년 8월~현재 : 삼성전자 통신연구소 책임연구원
 * 관심분야 : 신호처리, 패턴인식, 기계학습

• 김석호 (Austin Kim)



1986년 5월 : University of Illinois-Urbana-Champaign, 전자공학 (학사)
 1988년 5월 : University of Illinois-Urbana-Champaign, 전자공학 (석사)
 2001년 12월 : Illinois Institute of Technology, 컴퓨터공학 (박사)
 1988년 7월 ~ 2002년 12월 : Lucent Technologies, Bell Labs 근무

2003년 2월~현재 : 삼성전자 통신연구소 상무
 * 주 관심분야 : 신호처리, 멀티미디어, 시스템 simulation, ASIC/FPGA 개발, RF 분석