

# 음성신호의 대역폭 확장을 위한 GMM 방법 및 HMM 방법의 성능평가

## Performance Comparison of GMM and HMM Approaches for Bandwidth Extension of Speech Signals

송 근 배\*, 김 석 호\*  
(Geun-Bae Song\*, Austin Kim\*)

\*삼성전자

(접수일자: 2007년 9월 5일; 수정일자: 2007년 12월 10일; 채택일자: 2008년 2월 28일)

본 논문에서는 대역폭 확장 (Bandwidth Extension, BWE)을 위한 대표적인 통계적 방법인 가우스 혼합 모델 (Gaussian Mixture Model, GMM) 방법과 은닉마코프 모델 (Hidden Markov Model, HMM) 방법의 관계를 분석하고 성능을 비교한다. HMM 방법은 GMM 방법과 달리 기억능력을 가진 시스템으로서 인접한 음성 프레임간의 상관성을 모델링하고 이를 BWE 시스템에 활용한다는 장점을 가진다. 따라서 원래 신호의 프레임간 스펙트럼 변화특성을 보다 잘 추정할 수 있으리라 예상할 수 있다. 이 점을 확인하기 위해 정적 측도 외에 음성 스펙트럼의 일차 도 함수와 관련된 동적 측도를 적용하였다. 성능 평가 결과, 정적 측도 관점에서는 두 방법은 대등한 성능을 보였지만 동적 측도 관점에서는 HMM 방법이 우수한 성능을 보였다. 또한 이러한 차이는 HMM 모델의 상태 수에 비례하여 증가함을 확인할 수 있었다. 이와 같은 실험결과는 HMM 방법이 적어도 'blind BWE' 문제에 있어서 적절한 해법임을 시사한다. 한편, 동적 측도의 관점에서는 비록 열세로 나타났지만 GMM 방법은 상대적으로 단순하다는 장점을 가지고 있으며 특히, 정적 측도에 있어서 HMM 방법과 대등하다는 사실은 응용분야에 따라서는 HMM 방법의 효과적인 대안이 될 수 있음을 시사한다.

**핵심용어:** 대역폭 확장법, 가우스 혼합모델, 은닉 마코프 모델, Baum-Welch 학습법

**투고분야:** 음성처리분야 (2.3)

This paper analyzes the relationship between two representative statistical methods for bandwidth extension (BWE): Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) ones, and compares their performances. The HMM method is a memory-based system which was developed to take advantage of the inter-frame dependency of speech signals. Therefore, it could be expected to estimate better the transitional information of the original spectra from frame to frame. To verify it, a dynamic measure that is an approximation of the 1st-order derivative of spectral function over time was introduced in addition to a static measure. The comparison result shows that the two methods are similar in the static measure, while, in the dynamic measure, the HMM method outperforms explicitly the GMM one. Moreover, this difference increases in proportion to the number of states of HMM model. This indicates that the HMM method would be more appropriate at least for the 'blind BWE' problem. On the other hand, nevertheless, the GMM method could be treated as a preferable alternative of the HMM one in some applications where the static performance and algorithm complexity are critical.

**Keywords:** Bandwidth extension; Gaussian mixture model; Hidden Markov Model; Baum-Welch Re-estimation algorithm

**ASK subject classification:** Speech Signal processing (2.3)

## I. 서론

대역폭 확장 (Bandwidth Extension, BWE) 문제는 협대역 (narrow-band 즉, 0.3-3.4 kHz) 신호 (혹은 하위대역 (low-band 즉, 0-4 kHz) 신호)의 정보를 활용하여 광대역 (wide-band 즉, 0-8 kHz) 신호 (혹은 상위대역 (high-band 즉, 4-8 kHz) 신호)를 추정하는 문제로 요약될 수 있다. 이를 함수형태로 간단히 나타내면 다음 식과 같다. 즉,

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

여기서  $\mathbf{x}$ 는 관찰된 하위대역 특징 벡터를 나타내고,  $\mathbf{y}^+$ 는 추정된 상위대역 특징 벡터를 나타내고,  $F$ 는 하위대역과 상위대역을 관계를 나타내는 추정함수 (estimation function) 혹은 사상함수 (mapping function)이다. BWE 시스템의 성능은 적절한 특징벡터의 선택과 더불어 적절한 추정함수의 선택에 의해 좌우된다. 특징 벡터로는 흔히 LPC 계수, 선 스펙트럴 주파수 (Line Spectral Frequency, LSF), 켈프스트럴 계수 (Cepstral Coefficient, CC) 등을 사용하며, 추정함수로는 그동안 선형, 비선형 형태의 다양한 방법이 개발되었다 [1]-[8]. 특히 추정함수와 관련하여 최근 수년간, 가우스 혼합 모델 (Gaussian Mixture Model, GMM) [4] 및 은닉마코프 모델 (Hidden Markov Model, HMM)에 기반한 통계적 방법이 개발되어 기존의 방법에 비해 향상된 성능을 보여주었다 [5]-[8]. HMM 방법은 설계조건에 따라 GMM 방법을 포함한 기존의 모든 방법들을 포괄할 수 있는 가장 일반적인 알고리즘이다 [9]. 또한, 기억능력을 가지 시스템으로서 인접한 프레임간의 상관성 (inter-frame dependency) 즉, 동적 특성 (dynamic characteristic)을 모델링하고 이를 BWE 추정식 (estimation rule)에 활용한다는 점에서 장점을 가진다. 본 논문에서는 이러한 HMM의 장점이 BWE 시스템의 성능향상에 얼마나 개선된 효과를 주는지를 GMM 방법과의 비교를 통하여 알아보려고 한다.

HMM 모델의 관찰 확률 (observation probability)로서 GMM 모델을 사용한다고 가정하면, HMM 모델의 상태 수 (state number)를 1로 설정할 경우 HMM 모델은 단순히 GMM 모델이 된다 [10][11]. 마찬가지로 HMM 기반 BWE 추정법도 HMM 모델의 상태 수를 1로 설정하면 GMM 모델 기반 BWE 추정법이 됨을 쉽게 알 수 있다 [9]. 그러나 비록 두 방법이 알고리즘 형태상으로는 유사하더라도 여기에서 상태 수를 1로 하느냐 혹은 그 이상으로

하느냐는 BWE 시스템이 기억능력을 가지느냐 안 가지느냐의 차이를 의미한다. 아울러, 상태 수가 1에 가까울수록 HMM 기반 BWE 시스템의 기억능력은 약해지고 그 반대의 경우 강해진다. 따라서 다양한 HMM 모델 구조에 따른 BWE 시스템의 성능 변화를 관찰함으로써 BWE 시스템의 특성을 보다 깊이 이해할 수 있으며 또한, BWE 시스템 설계에 있어서 최적의 HMM 모델 구조를 결정할 수 있을 것이다.

성능 평가를 위한 측도로는 정적 측도 (static measure)와 동적 측도 (dynamic measure) 두 가지를 이용하였다 (본문 5절 참고). 정적 측도는 각 프레임 단위로 추정법의 성능을 평가하기 위한 목적으로 사용한다. 동적 측도는 스펙트럼상에서 원래 신호와 추정된 신호의 시간적 변화 특성이 얼마나 비슷한지를 비교하는 측도이다. 기본적으로 (1차 마코프) HMM 모델은 프레임간 상태 천이에 있어서 가능성이 낮은 상태로의 천이를 억제함으로써 결과적으로 원래 신호와 추정된 신호의 스펙트럼상의 궤적을 비슷하게 맞춰주는 효과를 줄 수 있다. 동적 측도는 이러한 효과를 평가하기 위한 목적으로 사용한다.

HMM 모델 학습법과 관련하여, Jax는 논문 [5]-[7]에서 다소 ad hoc한 학습법을 제안하였다. 이 문제에 대해 본 논문의 선행 논문인 [12]에서는 Jax 학습법과 Baum-Welch 학습법의 차이점을 분석하였고 Baum-Welch 학습법의 우수성을 실험을 통하여 보였다. 따라서 BWE를 위한 HMM 모델의 학습법으로서 Jax 학습법 대신 일반적인 HMM 학습법인 Baum-Welch 알고리즘을 이용하기로 한다. 또한, HMM 모델을 이용한 MMSE (Minimum Mean Square Error) 추정법과 관련하여, Jax는 [9]에서 크게 'soft classification' 추정법과 'cascaded estimation' 추정법 두 가지를 분류하고 있다. 본 논문에서는 두 추정법의 차이점을 분석하였으며 결론적으로, soft classification은 cascaded estimation의 근사적 형태로 볼 수 있고 또한 성능에 있어서도 열세이므로 따라서, 본 논문에서는 cascaded estimation 추정법만을 다루기로 한다.

본 논문은 다음과 같이 구성된다. 2절에서는 BWE를 위한 Baum-Welch HMM 모델 학습법을 요약한다. 3절에서는 HMM 모델을 이용하여 상위대역 벡터를 MMSE 추정하는 soft classification 및 cascaded estimation, 두 추정법에 대해 살펴본다. 4절에서는 HMM 기반 BWE 방법과 GMM 기반 BWE 방법의 관계에 대해 살펴본다. 5절에서는 두 방법의 성능을 평가하고, 6절에서는 결론을 내린다.

## II. HMM 모델 학습

광대역 학습데이터의 통계적 특성에 대한 정확한 모델링은 HMM 기반 BWE 시스템의 성능향상을 위해 중요하다. 본 절에서는 논문 [12]에서 설명된 Baum-Welch 학습법에 대한 내용을 간략히 요약한다. Baum-Welch 학습법은 가장 일반적인 HMM 모델 학습법으로서 그 밖에 보다 향상된 학습법 가령, [13]의 방법을 이용한다면 비슷하거나 혹은 향상된 HMM 모델 및 BWE 성능을 기대할 수 있을 것이다.

광대역 16 kHz 음성 데이터에 대해, 프레임별로 추출된 임의의 광대역 학습벡터를  $\mathbf{z}$ 라 하자. 그리고  $b$ 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,  $(b+d)$  차의 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. HMM 모델은 다음과 같이 초기 상태 확률 (initial state probability)과 상태전이 확률 (state transition probability) 즉,

$$\pi_i = P(S_i), \quad 1 \leq i \leq N_s \quad (2)$$

$$a_{ij} = P(S_j | S_i), \quad 1 \leq i, j \leq N_s, \quad (3)$$

그리고 확률 밀도 함수 (probability density function)로 표현된 관찰 확률 즉,

$$p(\mathbf{z} | S_i) = p(\mathbf{x}, \mathbf{y} | S_i) = \sum_{l=1}^L \rho_{z,il} f(\mathbf{z}; \boldsymbol{\mu}_{z,il}, \mathbf{V}_{z,il}) \quad (4)$$

로 구성된다. 여기서 식 (3)은 HMM의 형태가 'ergodic' 즉, 모든 상태들간에 상호전이 가능한 형태임을 함축하고 있다. 식 (4)의 두 번째 등식은 관찰 확률이  $L$ 개의 가우스 혼합 성분 (Gaussian mixture component)으로 구성된 GMM 모델로 모델링됨을 의미한다. 여기서  $\rho_{z,il}$  및  $\boldsymbol{\mu}_{z,il}$ ,  $\mathbf{V}_{z,il}$ 는 상태  $S_i$ 의  $l$ 번째 가우스 성분  $f(\cdot)$ 의 가중치, 평균 벡터, 공분산 (co-variance) 행렬을 각각 나타낸다.

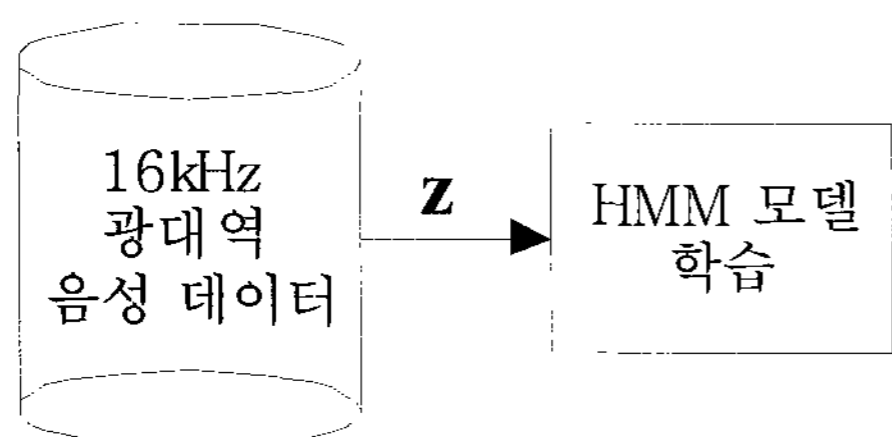


그림 1. Baum-Welch 학습법을 이용한 HMM 모델 학습 방법  
Fig. 1. HMM model training with Baum-Welch re-estimation algorithm.

HMM 모델의 초기화를 위해 먼저, 벡터 양자화 (Vector Quantization, VQ) 알고리즘 (가령, LBG [14])을 이용하여 총  $M$ 개의 광대역 관찰열  $Z(M) = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$ 을  $N_s$ 개의 코드벡터 (code-vector)  $\hat{\mathbf{z}}_i$ 로 이루어진 코드북  $C = \{\hat{\mathbf{z}}_i, 1 \leq i \leq N_s\}$ 로 분류한다. 즉, HMM의 상태  $S_i$ 는 코드벡터 (code-vector)  $\hat{\mathbf{z}}_i$ 로 표현된 대표적인 특징벡터 혹은 그 영역으로 정의된다. 코드북  $C$ 를 광대역 관찰열  $Z(M)$ 에 적용하면 (초기) 상태열을 구할 수 있다. 이로부터 상태 확률  $\pi_i$  및  $a_{ij}$ 의 초기치의 직접적인 계산이 가능하다. 다음으로, 가우스 성분 수  $L$ 만큼 각 상태  $S_i$ 의 벡터들을 VQ 알고리즘을 이용하여 분류한다. 이와 같이 분류된  $L$ 개의 하위그룹들은 각 상태  $S_i$ 의 관찰 확률 파라미터  $\rho_{z,il}$ ,  $\boldsymbol{\mu}_{z,il}$ ,  $\mathbf{V}_{z,il}$ 의 초기치를 계산하는데 이용된다. 가령, 평균 벡터  $\mathbf{V}_{z,il}$ 는 각 하위 그룹의 중심벡터 (centroid)로 초기화 된다.

HMM 모델의 초기치가 결정되면 다음으로 Baum-Welch 알고리즘이 반복적으로 적용된다. 상태 확률에 대한 Baum-Welch 계산식은 다음과 같다. 즉,

$$\pi'_i = \gamma_{z,i}(1) \quad (5)$$

$$a'_{ij} = \frac{\sum_{m=1}^M \gamma_{z,(i,j)}(m)}{\sum_{m=1}^M \gamma_{z,i}(m)} \quad (6)$$

여기서  $\pi'_i$ ,  $a'_{ij}$ 의 프라임 기호는 과거값  $\pi_i$ ,  $a_{ij}$ 으로부터 갱신된 새로운 값을 의미한다. 관찰확률 즉, GMM 모델의 파라미터의 계산식은 다음과 같다. 즉,

$$\rho'_{z,il} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m)}{\sum_{m=1}^M \sum_{l=1}^L \gamma'_{z,i}(m)} \quad (7)$$

$$\boldsymbol{\mu}'_{z,il} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m) \mathbf{z}(m)}{\sum_{m=1}^M \gamma'_{z,i}(m)} \quad (8)$$

$$\mathbf{V}'_{z,il} = \frac{\sum_{m=1}^M \gamma'_{z,i}(m) \mathbf{z}(m) \mathbf{z}^T(m)}{\sum_{m=1}^M \gamma'_{z,i}(m)} - \boldsymbol{\mu}'_{z,il} \boldsymbol{\mu}'_{z,il}{}^T \quad (9)$$

여기서도 마찬가지로 프라임 기호는 갱신된 파라미터를 의미하며, 위첨자  $T$ 는 전치 (transpose) 연산을 가리킨다. 기타,  $\gamma_{z,i}(m)$  및  $\gamma_{z,(i,j)}(m)$ ,  $\gamma'_{z,i}(m)$  등은 다음과 같이 정의된 전향확률 (forward probability)  $\alpha_{z,i}(m)$  및 후향확률 (backward probability)  $\beta_{z,i}(m)$  과 관련된 식으로서 일반적으로 널리 알려진 내용 이므로 본 논문에서 자세한 설명은 생략하기로 한다 (논문 [10], [11] 및 [12] 등 참조). 즉,

$$\alpha_{z,i}(m) = p(S_i(m), \mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(m)) \quad (10)$$

$$\beta_{z,i}(m) = p(\mathbf{z}(m+1), \mathbf{z}(m+2), \dots, \mathbf{z}(M) | S_i(m)) \quad (11)$$

여기서  $S_i(m)$  은  $m$  번째 프레임의 상태가  $i$  라는 것을 의미한다.

### III. MMSE 추정식

본 절에서는 광대역 학습데이터에 대한 HMM 모델이 주어진 조건하에서, 관찰된 하위대역 벡터  $\mathbf{x}$ 로부터 상위대역 벡터  $\mathbf{y}$ 를 MMSE 추정하는 방법에 대해 설명한다 [9].

#### 3.1. Cascaded estimation 추정식

제  $m$  번째 프레임까지 관찰된 하위대역 관찰열을  $\mathbf{X}(m) = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)\}$  이라 할 때, 상위대역 벡터  $\mathbf{y}$  에 대한 MMSE 추정식은 다음과 같다.

$$\begin{aligned} \mathbf{y}^+(m)_{MMSE} &= E\{\mathbf{y}(m) | \mathbf{X}(m)\} \\ &= \int_{\mathbb{R}^V} \mathbf{y}(m) p(\mathbf{y}(m) | \mathbf{X}(m)) d\mathbf{y}(m) \end{aligned} \quad (12)$$

여기서  $E\{\}$  는 평균 연산을 가리킨다. 이 식의 두 번째 등식에서 조건 확률  $p(\mathbf{y}(m) | \mathbf{X}(m))$  을 직접 계산하는 것은 어렵다. 따라서 다음과 같이 HMM의 상태  $S_i$  를 매개로 전개한다.

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

이 식의 오른쪽은 Bayes' rule을 이용하여 다음과 같이 전개된다.

$$\begin{aligned} p(\mathbf{y}(m), S_i(m) | \mathbf{X}(m)) &= p(\mathbf{y}(m) | S_i(m), \mathbf{X}(m)) P(S_i(m) | \mathbf{X}(m)) \\ &= P(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) P(S_i(m) | \mathbf{X}(m)) \end{aligned} \quad (14)$$

여기서 두 번째 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. 즉,

$$\mathbf{y}^+(m)_{MMSE} = \sum_{i=1}^{N_s} P(S_i(m) | \mathbf{X}(m)) \quad (15)$$

$$\times \int_{\mathbb{R}^V} \mathbf{y}(m) p(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) d\mathbf{y}(m)$$

여기서 오른쪽의 적분식은 상태  $S_i(m)$  과 하위대역 벡터  $\mathbf{x}(m)$  에 조건 지워진 상위대역 벡터  $\mathbf{y}(m)$  의 평균을 의미한다. 즉,

$$\begin{aligned} \int_{\mathbb{R}^V} \mathbf{y}(m) p(\mathbf{y}(m) | S_i(m), \mathbf{x}(m)) d\mathbf{y}(m) \\ = E\{\mathbf{y}(m) | S_i(m), \mathbf{x}(m)\} \end{aligned} \quad (16)$$

여기서  $E\{\mathbf{y}(m) | S_i(m), \mathbf{x}(m)\} \equiv \mathbf{c}_i(m)$  라 하면, 식 (15)의 MMSE 추정식은 다음과 같이 정리된다. 즉,

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

여기서 알 수 있는 바와 같이, MMSE 추정식은 각 상태  $S_i$  의 대표 벡터인  $\mathbf{c}_i(m)$  와 각 상태  $S_i$  의 발생확률  $P(S_i(m) | \mathbf{X}(m))$  로 이루어진다. 여기서, 확률  $P(S_i(m) | \mathbf{X}(m))$  은 대표 벡터  $\mathbf{c}_i(m)$  의 가중치의 역할을 하고 있다.

가중치 확률  $P(S_i(m) | \mathbf{X}(m))$  은 Bayes' rule을 활용하여 다음 식과 같이 전개될 수 있다. 즉,

$$\begin{aligned} P(S_i(m) | \mathbf{X}(m)) &= \frac{p(S_i(m), \mathbf{X}(m))}{p(\mathbf{X}(m))} \\ &= \frac{p(S_i(m), \mathbf{X}(m))}{\sum_{i=1}^{N_s} p(S_i(m), \mathbf{X}(m))} \end{aligned} \quad (18)$$

여기서 확률  $P(S_i(m) | \mathbf{X}(m))$  은 다음 식과 같이 하위대역 벡터 관찰열  $\mathbf{x}(m)$  에 대한 전향확률  $\alpha_{x,i}(m)$  을 의미한다 (식 (10) 참조). 즉,

$$p(S_i(m), \mathbf{X}(m)) = \alpha_{x,i}(m) \quad (19)$$

식 (18) 및 (19)를 식 (17)에 대입하면 MMSE 추정식은 최종적으로 다음과 같이 정리된다. 즉,

$$\mathbf{y}^+(m)_{MMSE} = \frac{\sum_{i=1}^{N_s} \alpha_{x,i}(m) \mathbf{c}_i(m)}{\sum_{i=1}^{N_s} \alpha_{x,i}(m)} \quad (20)$$

여기서 전향확률  $\alpha_{x,i}(m)$  은 잘 알려진 바와 같이 다음의 회귀적인 특성을 가진다. 즉,

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

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

여기서 하위대역 벡터  $\mathbf{x}(m)$  의 관찰 확률  $p(\mathbf{x}(m) | S_i)$  은

식 (4)의 광대역 벡터  $\mathbf{z}(m)$ 의 관찰 확률  $p(\mathbf{z}(m)|S_i)$ 의 주변 확률밀도 함수 (marginal pdf)로서  $p(\mathbf{z}(m)|S_i)$ 의 파라미터들로부터 직접 계산될 수 있다.

한편, 식 (16)에 정의된 대표벡터  $\mathbf{c}_i(m)$ 은 다음과 같이 계산된다 (편의상 프레임 번호  $(m)$  생략) [15].

$$\begin{aligned}
 \mathbf{c}_i(m) &= E\{\mathbf{y}(m)|S_i, \mathbf{x}(m)\} \\
 &= \int_{\mathcal{R}^y} \mathbf{y} p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|S_i, \mathbf{x}) d\mathbf{y} \\
 &= \int_{\mathcal{R}^y} \frac{\mathbf{y} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}|S_i)}{p_{\mathbf{x}}(\mathbf{x}|S_i)} d\mathbf{y} \\
 &= \frac{1}{p_{\mathbf{x}}(\mathbf{x}|S_i)} \int_{\mathcal{R}^y} \mathbf{y} p_{\mathbf{z}}(\mathbf{z}|S_i) d\mathbf{y} \\
 &= \frac{1}{p_{\mathbf{x}}(\mathbf{x}|S_i)} \int_{\mathcal{R}^y} \mathbf{y} \sum_{l=1}^L \rho_{z,il} f(\mathbf{z}; \boldsymbol{\mu}_{z,il}, \mathbf{V}_{z,il}) d\mathbf{y} \\
 &= \frac{1}{p_{\mathbf{x}}(\mathbf{x}|S_i)} \int_{\mathcal{R}^y} \mathbf{y} \sum_{l=1}^L \rho_{z,il} f(\mathbf{x}; \boldsymbol{\mu}_{x,il}, \mathbf{V}_{x,il}) f(\mathbf{y}; \boldsymbol{\mu}_{y,il}, \mathbf{V}_{y,il}) d\mathbf{y} \\
 &= \frac{1}{p_{\mathbf{x}}(\mathbf{x}|S_i)} \sum_{l=1}^L \rho_{z,il} f(\mathbf{x}; \boldsymbol{\mu}_{x,il}, \mathbf{V}_{x,il}) \int_{\mathcal{R}^y} \mathbf{y} f(\mathbf{y}; \boldsymbol{\mu}_{y,il}, \mathbf{V}_{y,il}) d\mathbf{y} \\
 &= \frac{\sum_{l=1}^L \rho_{z,il} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,il}, \mathbf{V}_{x,il}) \boldsymbol{\mu}_{y,il}}{\sum_{l=1}^L \rho_{z,il} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,il}, \mathbf{V}_{x,il})}
 \end{aligned} \tag{23}$$

여기서 다섯 번째 등식은 광대역 벡터  $\mathbf{z}$ 의 관찰확률인 식 (4)를 대입한 것이고, 여섯 번째 등식은 공분산 행렬  $\mathbf{V}_{z,il}$ 이 대각 행렬 (diagonal matrix) 즉, 벡터  $\mathbf{z}$ 의 성분들이 상호 비 상관관계라는 가정에 의해 성립한다. 마지막 등식에 있는 기호  $\boldsymbol{\mu}_{y,il} = \int_{\mathcal{R}^y} \mathbf{y} f(\mathbf{y}; \boldsymbol{\mu}_{y,il}, \mathbf{V}_{y,il})$ 는  $i$ 번째 상태의  $l$ 번째 가우스 성분의 상위대역 평균벡터를 의미한다. 결론적으로, 'cascaded estimation'란 명칭은 이 MMSE 추정식이 먼저 1단계로 식 (23)의 대표 벡터  $\mathbf{c}_i(m)$ 를 계산하고 2단계로 식 (18)의 가중치 확률  $P(S_i(m)|\mathbf{X}(m))$ 을 계산하여 적용하는 데서 비롯된다.

### 3.2. Soft classification 추정식

Jax에 의해 soft classification으로 명명된, 논문 [5], [6] 및 [8] 등에서 적용된 MMSE 추정식은 앞서 설명한 cascaded estimation의 근사적 형태로 간주할 수 있다. 즉,

$$\begin{aligned}
 \mathbf{y}^+(m)_{MMSE} &= \sum_{i=1}^{N_s} P(S_i(m)|\mathbf{X}(m)) \\
 &\quad \times \int_{\mathcal{R}^y} \mathbf{y}(m) p(\mathbf{y}(m)|S_i, \mathbf{x}(m)) d\mathbf{y} \\
 &= \sum_{i=1}^{N_s} P(S_i(m)|\mathbf{X}(m)) \mathbf{c}_i
 \end{aligned}$$

$$\begin{aligned}
 &\approx \sum_{i=1}^{N_s} P(S_i(m)|\mathbf{X}(m)) \\
 &\quad \times \int_{\mathcal{R}^y} \mathbf{y}(m) p(\mathbf{y}(m)|S_i) d\mathbf{y} \\
 &= \sum_{i=1}^{N_s} P(S_i(m)|\mathbf{X}(m)) \hat{\mathbf{y}}_i \\
 &= \mathbf{y}^+(m)_{MMSE\_SC}
 \end{aligned} \tag{24}$$

여기서 마지막 등식의 아래첨자 즉, 'MMSE\_SC'는 cascaded estimation 추정식과의 구분을 위해 사용하였다. 세 번째 등식 (즉, 근사식)에서 알 수 있는 바와 같이 soft classification 식은 첫 번째 등식의 피 적분 확률식  $p(\mathbf{y}|S_i, \mathbf{x})$ 를  $p(\mathbf{y}|S_i)$ 로 근사시켰음을 뜻한다. 즉, 두 랜덤벡터  $\mathbf{x}$ 와  $\mathbf{y}$ 가 서로 독립이라고 가정하였음을 뜻한다. 결과적으로, 변형된 적분식  $\int_{\mathcal{R}^y} \mathbf{y} p(\mathbf{y}|S_i) d\mathbf{y}$ 은 상태  $S_i$ 에 속한 상위대역  $\mathbf{y}$ 의 평균  $\hat{\mathbf{y}}_i$ 에 해당한다 (식 (23)의 평균벡터  $\boldsymbol{\mu}_{z,il}$ 와 비교). 각 상태  $S_i$ 의 평균  $\hat{\mathbf{y}}_i$ 는 논문 [5] 및 [6]에서처럼 HMM 학습 초기에 VQ 알고리즘을 이용하여 계산할 수도 있고 혹은, 논문 [8]에서처럼 소위 HMM 모델의 '그림자 코드북 (shadow codebook)'의 형태로 HMM 모델의 학습과정과 병렬 형태로 계산할 수도 있다.

한편, 여기서 두 랜덤벡터  $\mathbf{x}$ 와  $\mathbf{y}$ 가 서로독립이라는 가정은 사실과 맞지 않으며 또한, 두 벡터가 독립이 아니라는 가정에서 출발한 BWE 시스템의 전제와도 모순된다. 적절치 않은 가정을 도입한 만큼 성능실험 결과에서도 cascaded estimation 추정식에 비해 열세로 나타났다. 이 차이는 결국, 대표 벡터를  $\mathbf{c}_i(m)$ 에서  $\hat{\mathbf{y}}_i$ 로 단순화한 데서 비롯된다. 따라서 본 논문에서는 보다 일반적인 cascaded estimation 추정법만을 다루기로 한다.

## IV. HMM 방법과 GMM 방법의 관계

본 절에서는 [9]에서 언급된 내용을 수식을 이용하여 보다 자세히 다룬다.

식 (17)에서 보는 바와 같이, HMM 기반 MMSE 추정식은 각 상태  $S_i$ 의 대표 벡터  $\mathbf{c}_i(m)$ 와 가중치 확률  $P(S_i(m)|\mathbf{X}(m))$ 로 구성된다. 특히 가중치 확률  $P(S_i(m)|\mathbf{X}(m))$ 은  $m$  번째 프레임까지의 관찰열  $\mathbf{X}(m) = \{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(m)\}$ 을 관찰조건으로 하는 통계량으로서 식 (18)–(22)에서 보는 바와 같이, 과거 프레임의 영향을 누적하여 반영한다. 즉, HMM 기반 BWE 시스템이 기억능력을 가지도록 하는 통계량이다. 여기서 만약 HMM 모델의 상태 수를  $N_S=1$ 로 설정하면, 가중치 확률은  $P(S_1|\mathbf{X}(m))=1$ 이 된다. 즉,

$$\begin{aligned}
\mathbf{y}^+(m)_{MMSE} |_{N_s=1} &= P(S_1(m) | \mathbf{X}(m)) \mathbf{c}_1(m) \\
&= \mathbf{c}_1(m) \\
&= \frac{\sum_{l=1}^L \rho_{z,l} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,l}, \mathbf{V}_{x,l}) \boldsymbol{\mu}_{y,l}}{\sum_{l=1}^L \rho_{z,l} f(\mathbf{x}(m); \boldsymbol{\mu}_{x,l}, \mathbf{V}_{x,l})} \\
&= \sum_{l=1}^L P(l | \mathbf{x}(m)) \boldsymbol{\mu}_{y,l}
\end{aligned} \tag{25}$$

여기서 세 번째 등식은 식 (23)을 참조하여 전개하였다. 네 번째 등식의 확률  $P(l | \mathbf{x}(m))$  은  $\mathbf{x}(m)$  을 관찰조건으로  $l$  번째 가우스 성분의 발생확률을 의미한다. 또한  $\boldsymbol{\mu}_{z,l}$  는 앞서 살펴본 바와 같이,  $l$  번째 가우스 성분의 상위대역 평균벡터를 의미한다. 따라서 식 (25)는  $\boldsymbol{\mu}_{z,l}$  를  $l$  번째 가우스 성분의 대표 벡터로 하고  $P(l | \mathbf{x}(m))$  를 그 가우스 성분의 가중치 확률로 하는 GMM 기반 MMSE 추정식임을 알 수 있다 [4][15]. 식 (17)과 (25)를 비교해 보면, 대표 벡터로는  $\mathbf{c}_i(m)$  와  $\boldsymbol{\mu}_{z,i}$  가 대응되고 가중치 확률로는  $P(S_i(m) | \mathbf{X}(m))$  와  $P(l | \mathbf{x}(m))$  가 각각 대응됨을 알 수 있다. 두 식은 다만  $P(l | \mathbf{x}(m))$  는  $m$  번째 프레임의 벡터  $\mathbf{x}(m)$  만을 관찰조건으로 하는 반면,  $P(S_i(m) | \mathbf{X}(m))$   $m$  번째 프레임까지의 전체관찰열  $\mathbf{x}(m)$  을 관찰조건으로 한다는 점에서 다르다.

## V. 성능 비교 평가

본 절에서는 특징 벡터 및 성능평가 측도에 대한 설명 그리고, 실험결과 분석 등을 다룬다. 학습 및 평가를 위한 데이터로는 16 kHz 샘플링 된 TIMIT 데이터베이스를 이용하였다 [16]. 학습에는 전체 TIMIT 학습 데이터 (화자 총 462명, 음성파일 총 4620개)가 사용되었으며 테스트를 위해서는 핵심 테스트 셋 (Core Test Set; 화자 총 24명, 음성파일 총 192개)만을 사용하였다. 광대역 신호는 저역 및 고역필터 처리되어 하위대역과 상위대역 신호로 나뉜 뒤, 각각 20ms 프레임 단위로 처리되었다. 이때 하위대역 신호에 대해 특별히 협대역 부호화 처리를 하지는 않았다. (논문 [12]에서는 상업 환경과 비슷하게 EVRC 부호화된 하위대역 신호를 이용하였다. 따라서 동일한 실험조건에 대해 '절대적 수치'에 있어서는 본 논문의 결과와 다르다.)

### 5.1. 특징 벡터 및 평가 측도

각 프레임별 특징벡터로는, 하위대역 신호에 대해서는

10차의 MFCC 계수를, 상위대역 신호에 대해서는 10차의 LPC-CC계수 및 로그 이득 비  $\ln g$  를 더하여 총 11차의 계수를 추출하여 사용하였다. 여기서 이득 비  $g$  는 다음과 같이 정의된다.

$$g = \frac{g_{high}}{g_{low}} \tag{26}$$

여기서  $g_{low}$  와  $g_{high}$  는 각각 하위대역 및 상위대역의 LPC 잔차 신호 (residual signal)의 RMS (Root-Mean-Square) 값이다. 따라서 광대역 프레임 벡터는 총 21차의 계수들로 이루어진 벡터가 된다. 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. 이와 같은 CC 계수는 LPC 관련 다른 계수들 가령, LSF나 자기 상관 계수 (auto-correlation coefficient) 등에 비해 벡터 성분들이 근사적으로 비 상관관계를 가지고 따라서, 관찰 확률의 공분산 행렬  $\mathbf{V}_{z,l}$  은 대각행렬로 근사할 수 있다는 장점을 가진다. 또한 성능에 있어서도 완전 공분산 (full covariance) 행렬을 이용한 다른 계수들에 비해 적어도 본 논문의 실험조건에서는 대등함을 확인할 수 있었다.

성능 평가를 위한 측도로는 정적 측도와 동적 측도 두 가지를 이용하였다. 정적 측도로 흔히 쓰이는 로그 스펙트럴 왜곡 (Log Spectral Distortion, LSD)는 음성 코덱분야에서 먼저 널리 이용되었으며 주관적 음질과 잘 부합하는 측도로 알려져 있다. 이는 프레임 단위로 BWE 시스템의 성능을 평가한다. 임의의  $m$  번째 프레임에 대한 LSD 정의 식은 다음과 같다. 즉,

$$\begin{aligned}
D_{LSD}^2 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [20 \log_{10} \frac{g}{|\mathbf{A}_{hb}(e^{j\Omega})|} \\
&\quad - 20 \log_{10} \frac{g^+}{|\mathbf{A}_{hb}^+(e^{j\Omega})|}]^2 d\Omega
\end{aligned} \tag{27}$$

여기서 적분의 범위는 상위대역 (4 k-8 kHz)만을 대상으로 한다. 이득 비  $g$  와  $g^+$  는 각각 실제 값 및 추정된 값을 나타낸다. 또한  $\mathbf{A}_{hb}$  와  $[\mathbf{A}_{hb}^+]^{-1}$  는 각각 (상위대역) 원래의 LPC 모델과 추정된 LPC 모델을 나타낸다. 추정된 LPC 모델  $[\mathbf{A}_{hb}^+]^{-1}$  의 LPC 계수는 BWE 함수를 통해 예측된 11차의 상위대역 벡터  $\mathbf{y}^+$  중 이득 비  $g^+$  를 제외한 나머지 10차의 상위대역 LPC-CC 계수로부터 계산된다. 이때 안정된 LPC 모델을 얻기 위해서는 LPC-CC로부터 직접 LPC 계수로 변환을 시도하기 보다는 먼저, 푸리에 변환

을 이용하여 LPC-CC로부터 전력 스펙트럴 밀도 (power spectral density)를 구하고, 이를 다시 푸리에 역 변환하여 자기 상관계수를 구한 뒤, 최종적으로 Levinson-Durbin 알고리즘을 이용하여 LPC 계수를 구하는 최적 방법을 사용한다 (자세한 설명은 논문 [6] 참조.)

동적 측도 역시 그동안 음성 코덱 분야에서 많은 연구가 있었다 [17]–[19]. 앞선 여러 연구 결과들에 의하면 프레임 단위의 음의 왜곡 특성 즉, 정적 왜곡 특성뿐만 아니라 왜곡의 시간적 변화 특성 역시 음질에 중요함을 알 수 있다. 가령, 논문 [18]에서는 동일한 평균 LSD 값을 가지는 음에 대해 LSD의 프레임간 변화율이 작은 음 즉, LSD 값이 시간적으로 부드럽게 변화하는 음을 청취자들이 선호함을 보여주고 있다. 여기서 LSD의 변화율을 측정하기 위한 다양한 동적 측도가 고려될 수 있지만, 본 논문에서는 논문 [19]에서 제시된 차분 로그 스펙트럴 왜곡 (Differenced Log Spectral Distortion, DLSD)을 이용하기로 한다. 임의의  $m$  번째 프레임에 대한 DLSD의 정의식은 다음과 같다. 즉,

$$dD_{LSD}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\Delta(\Omega, m) - \Delta^+(\Omega, m)]^2 d\Omega \quad (28)$$

여기서  $\Delta(\Omega, m)$  및  $\Delta^+(\Omega, m)$ 은 각각 다음과 같이 정의된다. 즉,

$$\Delta(\Omega, m) = 20 \log_{10} \frac{g(m)}{|\mathbf{A}_{hb}(e^{j\Omega(m)})|} - 20 \log_{10} \frac{g(m-1)}{|\mathbf{A}_{hb}(e^{j\Omega(m-1)})|} \quad (29)$$

$$\Delta^+(\Omega, m) = 20 \log_{10} \frac{g^+(m)}{|\mathbf{A}_{hb}^+(e^{j\Omega(m)})|} - 20 \log_{10} \frac{g^+(m-1)}{|\mathbf{A}_{hb}^+(e^{j\Omega(m-1)})|}. \quad (30)$$

즉,  $dD_{LSD}$ 는 원래 신호의 LPC 모델 (1차) 차분 스펙트럼과 추정된 신호의 LPC 모델 (1차) 차분 스펙트럼의 유클리드 측도 (Euclidean measure)에 해당한다. 만약 두 LPC 모델의 시간적 변화율이 동일하다면  $dD_{LSD} = 0$ 이 됨을 알 수 있다. 다른 관점에서 보면, 이 측도는 원래 신호의 차분 LPC-CC 벡터  $\Delta \mathbf{y} = \mathbf{y}(m) - \mathbf{y}(m-1)$ 와 추정된 신호의 차분 LPC-CC 벡터  $\Delta \mathbf{y}^+ = \mathbf{y}^+(m) - \mathbf{y}^+(m-1)$ 에 대한 유클리드 측도와 동등하다 [20]. 즉, 원래 신호의 차분 벡터  $\Delta \mathbf{y}$ 를 추정하는 BWE 시스템의 능력을 평가하는 측도로도 볼 수도 있다.

## 5.2. GMM 방법과 HMM 방법의 성능비교

앞서 4절에서 살펴 본 바와 같이, HMM 기반 BWE 방법은 HMM 모델 상태수를 줄어나감에 따라 점점 GMM 방법으로 수렴해간다. 그리고 최종적으로 상태수를  $N_S = 1$ 로 하면 두 방법은 동일하게 된다.

본 논문에서는 GMM 방법과 HMM 방법의 성능을 비교하기 위한 첫 번째 실험으로 HMM 모델의 복잡도를  $N_S \times L = 128$ 로 일정하게 유지하면서 상태 수  $N_S$ 를 변화시켰을 때의 성능변화 추이를 관찰하였다. 이 경우  $1 \times 128$  조합이 가우스 성분 128개의 GMM 방법에 해당한다. 그림 2의 (a) 및 (b)는 각각 다양한  $N_S \times L$  조합에 대한 평균  $D_{LSD}$  및 평균  $dD_{LSD}$ 의 그래프를 보여준다. 또한 그림 3의 (a) 및 (b)는 각각  $D_{LSD}$  및  $dD_{LSD}$ 의 분산  $\sigma$ , 약 이상치 (mild outliers, MO), 강 이상치 (extreme outliers, EO)의 변화 추이를 보여준다.  $D_{LSD}$ 의 MO와 EO의 경계 값 즉, 이상치 안쪽 경계와 바깥쪽 경계는 GMM 방법에서 산출한  $D_{LSD}$  값의 분포를 기준으로 계산하였다. 계산 결과 MO의 경계값은 대략 15 dB, EO의 경계 값은 대략 21dB로 설정되었다. 따라서 그림 3 (a)의 그래프상의 MO [%]는  $15 \text{ dB} \leq D_{LSD} < 21 \text{ dB}$ 인 프레임의 비율을 나타내고, EO [%]는  $21 \text{ dB} \leq D_{LSD}$ 인 프레임의 비율을 나타낸다.  $dD_{LSD}$ 의 MO와 EO의 경계 값 역시 GMM의  $dD_{LSD}$  값 분포를 기준으로 산출하였다.  $dD_{LSD}$ 의 MO와 EO의 경계 값은 각각 0.000231 dB, 0.000358 dB이다.

그림 2 (a)에서 보는 바와 같이, 평균  $D_{LSD}$  값은  $N_S$  값이 클 경우 약간 개선됨을 알 수 있다. 양 극단의 HMM 모델 즉, (GMM)  $1 \times 128$  조합과  $128 \times 1$  조합의 차이는 약 0.1dB로 나타났다. 그 밖에  $D_{LSD}$ 와 관련된 다른 값 즉, 분산 및 이상치 비율은 그림 3 (a)에서 보는 바와 같이 HMM 모델 구조에 따라 거의 차이가 없다. 즉, 평균  $D_{LSD}$ 에서만 약간 차이를 보일뿐 기타 수치에서는 의미있는 차이를 보이지 않고 있다. 한편,  $dD_{LSD}$  측도 관점에서는 평균  $dD_{LSD}$  뿐만 아니라  $dD_{LSD}$ 의 분산 및 이상치 등 모든 수치에 있어서 HMM 방법이 우수한 성능을 보임을 알 수 있다. 또한 이는  $N_S$ 가 클수록 증가하고 있다. 평균  $dD_{LSD}$ 의 값은 그림 2 (b)에서 보는 바와 같이,  $1 \times 128$  조합과  $128 \times 1$  조합에 대해 약 0.2 dB의 차이를 보였다. 이는  $dD_{LSD}$ 가 차분 값에 대한 측도라는 점을 고려할 때 평균  $D_{LSD}$ 에 비해 상대적으로 큰 차이임을 알 수 있다. 아울러,  $dD_{LSD}$ 의 분산 및 이상치 역시 뚜렷한 차이를 보이고 있다. 따라서, 두 BWE 방법은  $D_{LSD}$  관점에서는 큰 차이가 없지만  $dD_{LSD}$  관점에서는 뚜렷한 차이를 보임을 알 수 있다.

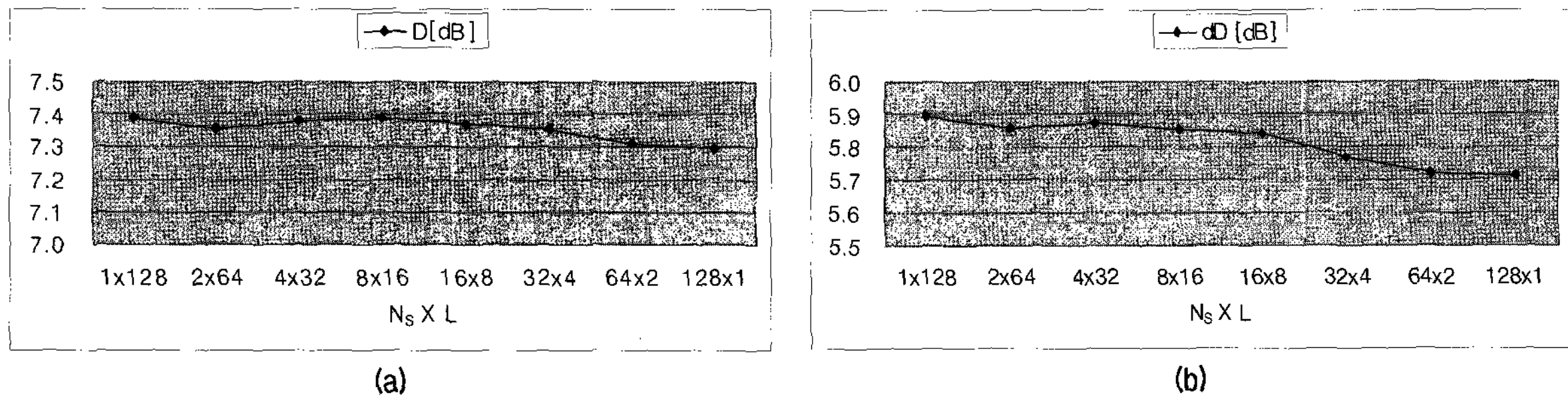


그림 2. 다양한  $N_S \times L$  조합에 대한 HMM 기반 BWE 시스템의 성능실험결과 (a) 평균  $D_{LSD}$  (b) 평균  $dD_{LSD}$   
 Fig.2. Test results of HMM-based BWE system for various  $N_S \times L$  combinations (a) averaged  $D_{LSD}$  (b) averaged  $dD_{LSD}$ .

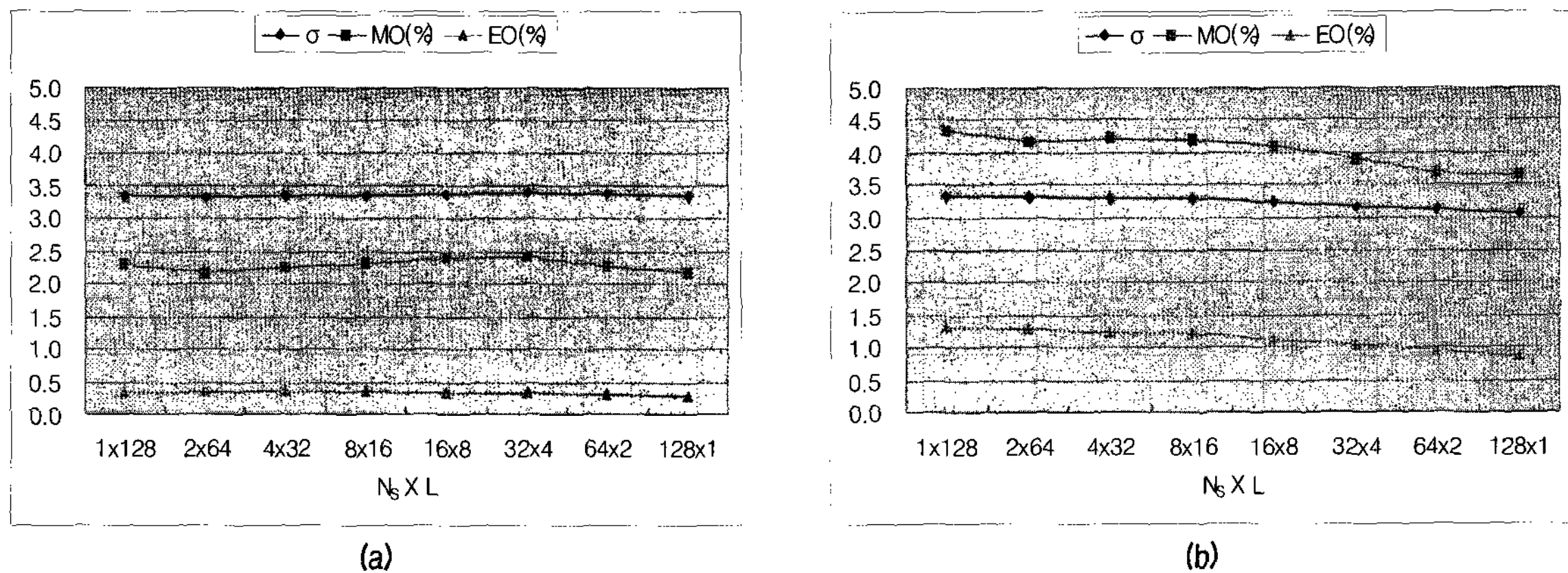


그림 3. 다양한  $N_S \times L$  조합에 대한 HMM 기반 BWE 시스템의 성능실험결과  
 (a)  $D_{LSD}$ 의 분산, 약 이상치, 강 이상치 (b)  $dD_{LSD}$ 의 분산 (주의, 가운데 그래프), 약 이상치, 강 이상치  
 Fig. 3. Test results of HMM-based BWE system for various  $N_S \times L$  combinations  
 (a)  $\sigma$ , MO, and EO of  $D_{LSD}$  (b)  $\sigma$ , MO, and EO of  $dD_{LSD}$ .

동적 측도  $dD_{LSD}$ 에 있어서의 차이를 만드는 주요 원인이  $L$ 보다는  $N_S$ 와 관련이 있다는 점을 보다 확실히 하기 위해,  $N_S = const. = 1$  일 때  $L = 64, 128, 256$ 에 대한  $D_{LSD}$  및  $dD_{LSD}$ 의 변화 추이 그리고 반대로,  $L = const. = 1$  일 때  $N_S = 16, 32, 64$ 에 대한  $D_{LSD}$  및  $dD_{LSD}$ 의 변화 추이를 살펴 보았다. 그림 4 (a) 및 (b)의 비교를 통해, 평균  $D_{LSD}$ 의 변화율은  $N_S = const.$  및  $L = const.$  두 경우 모두 비슷함을 알 수 있다. 반면 평균  $dD_{LSD}$ 의 변화율은 그림 4 (c) 및 (d)에서 보는 바와 같이,  $N_S = const.$ 인 경우에는 미미한 수준이지만  $L = const.$ 인 경우에는 크다는 사실을, 알 수 있다. 즉, 평균  $dD_{LSD}$ 은  $N_S$ 의 변화에 더 민감하게 반응한다. 이와 같은 현상은 표 1 및 2에 기록된  $D_{LSD}$  및  $dD_{LSD}$ 의 분산과 이상치의 비교에서도 마찬가지로이다. (여기서 사용된 MO와 EO의 경계 값은 그림 3에서와 동일하다.) 결론적으로, GMM 및 HMM 방법은 정적 측도  $D_{LSD}$ 의 관점에서는 비슷한 성능을 보이지만 동적 측도  $dD_{LSD}$  관점에서는 뚜렷한 차이를 보이며 이 차이는 HMM 모델의  $N_S$ 의 크기와 주로 관련이 있음을 알 수 있다.

동적 측도 관점에서 HMM 방법이 우수하다는 실험결

과는 앞서 설명한 바와 마찬가지로 직접적인 의미로는 HMM 방법이 상위대역 신호의 차분 스펙트럼을 추정하는데 우수하다는 것을 의미하며 따라서, 시간축을 따라 원래 신호의 변화 특성과 유사한 추정된 스펙트럼을 재생함을 의미한다. 이는 또한 LSD 값이 시간축을 따라 상대적으로 일정하다는 것을 의미하며 따라서, 주관적 음질에도 긍정적 효과로 작용할 수 있다. 즉, 이상과 같은 실험결과는 HMM 모델링의 효과를 명시적으로 보여주는 것으로 적어도 'blind BWE' 문제에 있어서는 HMM 방법이 적절한 해법임을 의미한다.

한편, 정적 측도 관점에서 두 방법이 대등하다는 사실은 여전히 GMM 방법이 응용분야에 따라서는 HMM 방법의 효과적인 대안이 될 수 있음을 의미한다. 가령 입력신호를 하위대역과 상위대역으로 나누어 부호화하는 대역분할 부호화 (split-band coding) 방식에 있어서, BWE 추정법은 두 대역간의 예측 부호화 (predictive coding)를 위한 예측기로 이용될 수 있다 [21]. 즉, 상위대역 정보 중에서 하위대역으로부터 예측 가능한 부분을 제거하고 나머지 정보만을 부호화하기 위한 예측기가 그것이다.



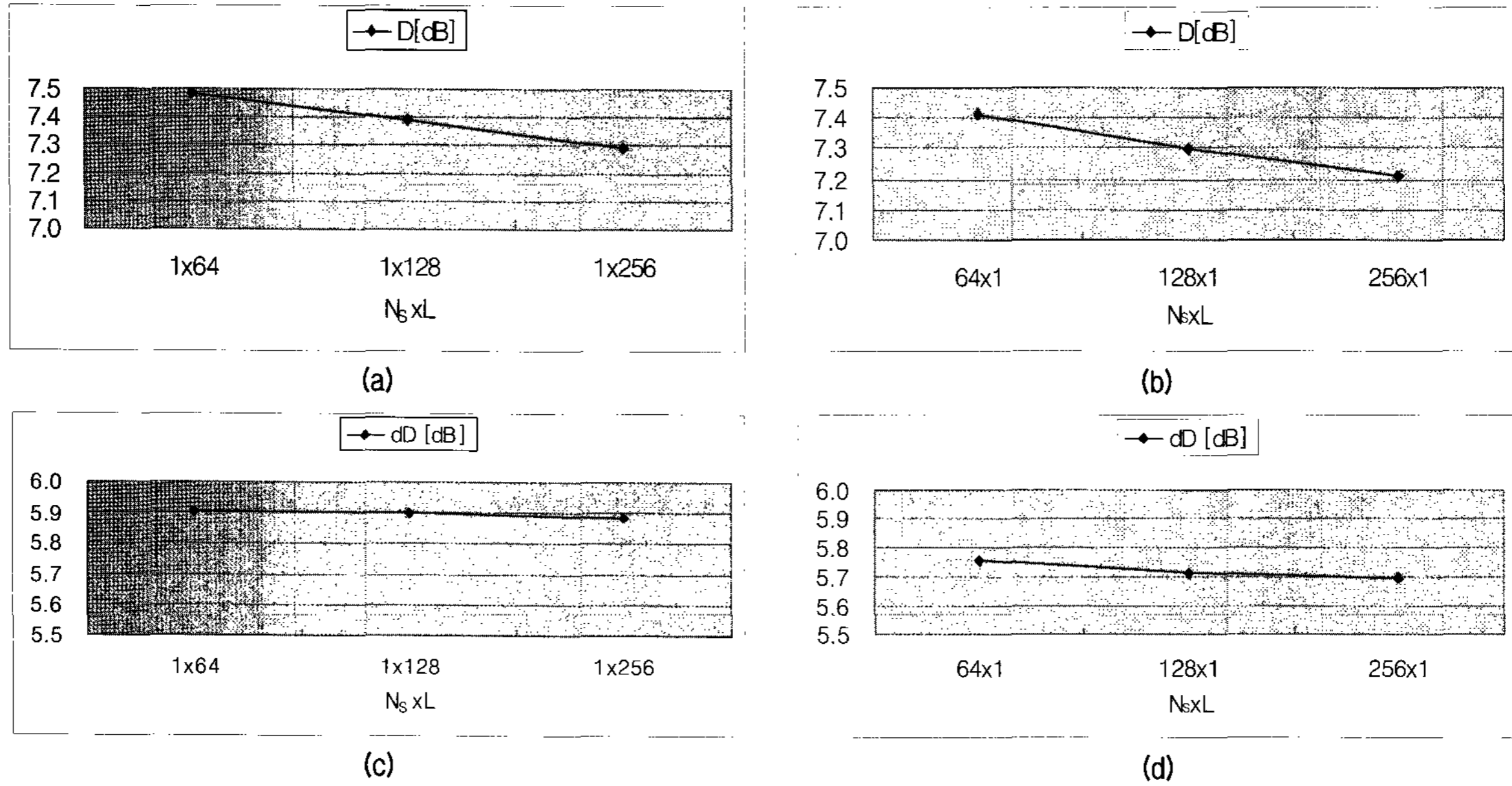


그림 4. (a) 몇몇  $1 \times L$  조합에 대한 평균  $D_{LSD}$  (b) 몇몇  $N_S \times 1$  조합에 대한 평균  $D_{LSD}$   
 (c) 몇몇  $1 \times L$  조합에 대한 평균  $dD_{LSD}$  (d) 몇몇  $N_S \times 1$  조합에 대한 평균  $dD_{LSD}$   
 Fig. 4. (a) Averaged  $D_{LSD}$  for some  $1 \times L$  combinations (b) Averaged  $D_{LSD}$  for some  $N_S \times 1$  combinations  
 (c) Averaged  $dD_{LSD}$  for some  $1 \times L$  combinations (d) Averaged  $dD_{LSD}$  for some  $N_S \times 1$  combinations.

이 경우 부호화 효율은 예측잔차의 분산 즉,  $D_{LSD}$  의 분산  $\sigma$ 과 관련이 되므로 알고리즘의 복잡도 및 크기 등을 고려할 때 GMM 추정법이 더 바람직한 해법이라 할 수 있다. 더욱이 전송 선로상에서는 프레임 손실 (frame error)이 빈번히 발생하므로 기억능력이 없는 GMM 방법이 기억능력을 가진 HMM 방법에 비해 강건한 특성을 보일 것으로 예상된다. 프레임 손실에 따른 HMM 방법의 성능저하 문제는 향후 연구과제 중 하나이다.

### 6. 결론

본 논문에서는 대표적인 두 가지 통계적 BWE 방법인 GMM 방법과 HMM 방법의 관계 및 성능을 분석하였다. HMM 모델의 상태 수를 1로 할 경우 HMM 방법은 GMM 방법이 된다. 따라서 GMM 방법은 HMM 방법의 특수한 형태로 볼 수 있다. 성능평가와 관련하여, 정적 특성의 파악을 위해서는 ‘로그 스펙트럴 왜곡’ 측도를 이용하였고 동적 특성을 파악하기 위해서는 ‘차분 로그 스펙트럴 왜곡’ 측도를 이용하였다. 성능평가 결과 정적 측도 관점에서 두 방법은 거의 대등한 성능을 보였지만 동적 측도 관점에서는 HMM 방법이 우수한 성능을 보였다. 그리고 이러한 차이는 HMM 모델의 상태 수에 비례함을 확인할 수 있었다. 이와 같은 실험결과는 적어도 blind BWE 문제에

표 1. 다양한  $1 \times L$  혹은  $N_S \times 1$  조합에 대한 평균  $D_{LSD}$ , 분산, 약 이상치, 강 이상치  
 Table 1. Averaged  $D_{LSD}$ ,  $\sigma$ , MO and EO for some  $1 \times L$  or  $N_S \times 1$  combinations.

CASE	1x64	1x128	1x256	64x1	128x1	256x1
$\sigma$	3.359	3.337	3.297	3.384	3.338	3.310
MO [%]	2.421	2.286	2.127	2.376	2.179	2.138
EO [%]	0.352	0.348	0.307	0.353	0.295	0.267

표 2. 다양한  $1 \times L$  혹은  $N_S \times 1$  조합에 대한 평균  $dD_{LSD}$ , 분산, 약 이상치, 강 이상치  
 Table 2. Averaged  $dD_{LSD}$ ,  $\sigma$ , MO, and EO for some  $1 \times L$  or  $N_S \times 1$  combinations.

CASE	1x64	1x128	1x256	64x1	128x1	256x1
$\sigma$	3.364	3.332	3.296	3.124	3.056	3.026
MO [%]	4.369	4.335	4.145	3.808	3.663	3.465
EO [%]	1.367	1.313	1.255	0.953	0.849	0.857

있어서는 HMM 방법이 적절한 해법임을 의미한다. 한편, 동적 측도의 관점에서는 비록 열세로 나타났지만 GMM 방법은 그 나름의 장점을 가지고 있으며 특히, 정적 측도에 있어서 HMM 방법과 대등하다는 사실은 응용분야에 따라서는 HMM 방법의 효과적인 대안이 될 수 있음을 시사한다.

전체적으로는 이상의 분석 및 실험을 통하여, 현재까지 알려진 BWE 문제의 가장 일반적인 해법인 HMM 방법의 동작특성 파악 및 응용분야에 따른 적절한 설계지침을 위한 기초지식을 마련할 수 있었다.

## 참고 문헌

1. Y. Nakatoh, M. Tsushima, and T. Norimatsu, "Generation of broadband speech from narrowband speech using piecewise linear mapping," Proc. European Conf. Speech Commun., Technol., 1643-1646, Sept. 1997.
2. S. Chennoukh, A. Gerrits, and R. Sluijter, "Speech enhancement via frequency bandwidth extension using line spectral frequencies," ICASSP 1, 665-668, May 2001.
3. 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.
4. K. -Y. Park and H. S. Kim, "Narrowband to wideband conversion of speech using GMM based transformation," ICASSP 3, 1843-1846, June 2000.
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. G. Chen and V. Parsa, "HMM-Based frequency bandwidth extension for speech enhancement using line spectra frequencies," ICASSP 1, 17-21, May 2004.
9. E. Larson and R. M. Aarts, *Audio Bandwidth Extension* (John Wiley & Sons, Ltd., 2004), Chap. 6, 226-235.
10. J. A. Bilmes, "A gentle tutorial of the EM Algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models," U. C. Berkely, TR-97-021, April, 1998.
11. 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.
12. 송근배, 김석호, "Baum-Welch 학습법을 이용한 HMM 기반 대역폭 확장법", 한국음향학회지, 26(6), 207-213, 2007.
13. A. Rao and K. Rose, "Deterministically annealed design of Hidden Markov Model speech recognizers," IEEE Trans. Speech, Audio Processing 9(6), 111-126, Feb. 2001.
14. Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun. 28(1), 84-95, 1980.
15. Wei-shou Hsu, "Robust bandwidth extension of narrowband speech," M.A. thesis, McGill Univ., Dept. of Electrical & Computer Engineering, 26-29, Nov. 2004.
16. J. S. Garofolo, L. F. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, *DARPA-TIMIT: Acoustic-Phonetic Continuous Speech Corpus*, (1990).
17. H. P. Knagenhjelm and W. B. Kleijn, "Spectral dynamics is more important than spectral distortion," ICASSP 1, 665-668, May 1995.
18. F. Norden and T. Eriksson, "A speech spectral distortion measure with interframe memory," ICASSP 1, 665-668, May 2001.
19. F. Norden and T. Eriksson, "Time evolution in LPC spectrum coding," IEEE Trans. Speech, Audio Processing 12(3), 290-301, May 2004.
20. F. K. Soong and A. E. Rosenberg, "On the use of ins-

tantaneous and transitional spectral information in speaker recognition," IEEE Trans. Acoust., Speech, Signal Processing 36(6), 871-879, June 1998.

21. B. Geiser and P. Vary, "Backwards compatible wideband telephony in mobile networks: CELP watermarking and bandwidth extension," ICASSP 4, 533-536, April 2007.

## 저자 약력

### • 송근배 (Geun-Bae Song)



1992년 2월: 아주대학교 전자공학 (학사)  
 1995년 2월: 아주대학교 전자공학 (석사)  
 2001년 8월: 아주대학교 전자공학 (박사)  
 1995년 1월~1996년 3월: 현대전자 S/W 연구소 연구원  
 2001년 8월~현재: 삼성전자 책임연구원  
 ※주관심분야: 음성/오디오 신호처리, 패턴인식, 기계 학습

### • 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  
 ※주관심분야: 멀티미디어 신호처리, 통신 시스템/단말 simulation, ASIC/FPGA 개발, RF 분석, 선행 antenna 개발