

시간 연속성을 갖는 비음수 행렬 분해를 이용한 음질 개선

Speech Enhancement Using Nonnegative Matrix Factorization with Temporal Continuity

남승현

(Seung-Hyon Nam)

배재대학교 전자공학과

(Received January 27, 2015; revised February 25, 2015; accepted March 27, 2015)

초 록: 본 논문은 시간 연속성을 갖는 비음수 행렬 분해(Nonnegative Matrix Factorization, NMF)를 이용하여 잡음에 열화된 음성 신호의 음질을 개선하는 문제를 다룬다. 음성과 잡음 신호는 포아송 분포로 모델되며, NMF의 기본 벡터와 이득 벡터는 감마 분포로 모델된다. 이득 벡터의 시간 연속성은 음질 개선에 중요한 영향을 미치는 것으로 알려져 있다. 본 논문에서 시간의 연속성은 이득 벡터를 감마-마르코프 연쇄(Gamma-Markov chain, GMC) 사전 분포로 모델함으로써 이루어진다. 실험 결과는 제안된 알고리즘이 잡음 신호의 시간 연속성을 효과적으로 모델하는 것을 보여준다. **핵심용어:** 음질 개선, 비음수행렬분해, 변분-베이지 추정, 감마-마르코프 연쇄, 시간 연속성

ABSTRACT: In this paper, speech enhancement using nonnegative matrix factorization with temporal continuity has been addressed. Speech and noise signals are modeled as Poisson distributions, and basis vectors and gain vectors of NMF are modeled as Gamma distributions. Temporal continuity of the gain vector is known to be critical to the quality of enhanced speech signals. In this paper, temporal continuity is implemented by adopting Gamma-Markov chain priors for noise gain vectors during the separation phase. Simulation results show that the Gamma-Markov chain models temporal continuity of noise signals and track changes in noise effectively.

Keywords: Speech enhancement, Nonnegative matrix factorization, Variational Bayesian inference, Gamma-Markov chain, Temporal continuity

PACS numbers: 43.60.Cg

1. 서 론

잡음에 열화된 음성 신호로부터 잡음을 제거하고 음질을 개선하는 연구는 수 십년 간 많은 진보를 이루었지만, 여전히 다양한 잡음환경에서의 음질 개선 연구는 주요 대상 중 하나이다.^[1] 최근에는 NMF를 이용하여 잡음을 제거하고 음질을 개선하는 연구들이 발표되고 있다.^[2]

NMF는 입력 행렬을 두 행렬의 곱으로 분해하는 것으로 행렬의 원소들이 모두 비음수 값을 갖도록 제한된다.^[3] 비음수 조건 덕분에 입력 신호는 기본 벡

터들의 선형조합으로 표현된다. NMF의 이러한 특징은 음성과 잡음 신호의 분리를 통한 음질 개선을 가능하게 한다.

NMF가 처음 제안되었을 때에는 확률통계적 관점이 명확히 설명되지 않았지만, 이후 입력 신호가 포아송(Poisson) 분포를 갖는다는 가정을 하면, NMF는 최대우도(maximum likelihood) 추정을 위한 기댓값 최대화(Expectation-Maximization, EM) 알고리즘과 동일하다는 사실이 확인되었다.^[4] 이는 NMF 알고리즘이 계층적 사전(prior)분포 구조를 갖도록 확장될 수 있음을 의미한다. 나아가 Cemgil은 포아송 분포와 켈레 관계를 갖는 감마 사전분포를 적용한 Variational Bayesian NMF(VB-NMF)를 제안하였다.^[4] VB-NMF 기법에서는 사후분포들이 서로 독립적인 것으로 근

†Corresponding author: Seung-Hyon Nam (shnam@pcu.ac.kr)
Department of Electronic Engineering, Paichai University, 155-40,
Baejae-Ro(Doma-dong), Seo-Gu, Daejeon 303-735, Republic
of Korea
(Tel: 82-42-520-5607, Fax: 822-70-4362-6311)

사화된다.^[5] 변수들이 서로 독립적인 관계를 갖는다는 것은 일견 현실적이지 않지만, 많은 경우 매우 효율적인 알고리즘 수단을 제공한다.

감마 분포는 두 개의 초모수(hyperparameter)를 갖으며, 이들 모수들은 신호의 다양한 시간-주파수 특성을 포착하는데 매우 유용하다. VB-NMF 기법은 잡음에 열화된 음성의 음질 개선에 적용된 바 있으며 우수한 잡음 제거 성능을 제공하는 것으로 확인되었다.^[6] Mohammadiha *et al.*은 Reference [6]에서 초모수들의 적절한 선택이 음질 개선 성능에 중요한 영향을 미치는 것으로 보고하였다. 뿐만 아니라, 동적 잡음 신호의 시간 연속성을 모델하기 위해 매 프레임마다 이전 프레임으로부터 얻은 사후 평균을 이용하여 사전분포의 초모수를 초기화하였다. 또한 Reference [7]에서는 GMC를 적용하여 음악 신호에 내재된 시간 연속성을 모델한 최대사후(maximum a posteriori) 방식 알고리즘을 도출하였다.

본 논문에서는 VB-NMF에 GMC를 적용하여 시간 연속성을 모델한다. 이 논문은 GMC를 VB-NMF에 적용하였다는 점에서 MAP에 적용한 Reference [7]과 다르며, Signal-to-Noise Ratio(SNR) 계산없이 단지 GMC의 모양 초모수의 비를 이용하여 시간 연속성을 모델한다는 점에서 Reference [6]과 다르다.

이 논문의 구성은 다음과 같다. II장에서는 NMF의 확률통계적 측면과 이를 확장한 VB-NMF를 간단히 소개한다. III장에서는 VB-NMF에 적용되는 GMC 사전 분포 모델의 특성을 살펴보고, 이를 적용한 VB-NMF 알고리즘을 기술한다. 마지막으로 IV장에서는 잡음에 열화된 음성 신호의 음질 개선 실험을 통해 제안된 알고리즘의 성능을 비교 확인한다.

II. 통계적 관점의 NMF

통계적 관점의 NMF는 Reference [4]에 상세히 기술되어 있지만, 여기서는 새로 제안되는 알고리즘을 소개하기 위한 범위 안에서 간략하게 기술되며, 편의상 표기는 Reference [4]와 일치시킨다.

2.1 NMF

NMF는 비음수 원소를 갖는 행렬 $X = \{x_{\nu\tau}\}$, $\nu =$

1.. W , $\tau = 1..K$ 을 양의 원소를 갖는 행렬 T 와 V 의 곱으로 다음과 같이 분해한다.^[3]

$$x_{\nu\tau} \approx [TV]_{\nu\tau} = \sum_{i=1}^I t_{\nu i} v_{i\tau}, \quad (1)$$

여기서 T 는 I 개의 기본 벡터 열로 구성된 $W \times I$ 행렬이고, V 는 I 개의 이득 벡터 행으로 구성된 $I \times K$ 행렬이다. Eq.(1)의 해는 Kullback-Leibler 발산을 최소화함으로써 구해질 수 있지만, 통계적 관점에서는 최대 우도 추정 기법을 통해서도 구해질 수 있다.^[4]

최대 우도 추정 기법을 이용한 NMF에서 입력 신호는 $x_{\nu\tau} = \sum_i s_{\nu i\tau}$ 와 같이 잠재변수의 합으로 가정되고, 각각의 잠재변수 $s_{\nu i\tau}$ 는 포아송 분포 $PO(s_{\nu i\tau}; t_{\nu i} v_{i\tau})$ 를 갖는다고 가정된다. 여기서 포아송 분포는 다음과 같이 정의된다.

$$PO(s; \lambda) = \exp[s \log \lambda - \lambda - \log \Gamma(s + 1)].$$

이제 Eq.(1)을 만족하는 해(T, V)는 로그 우도(log-likelihood) 함수 $L_x(T, V) = \log p(X|T, V)$ 를 최대화함으로써 구해진다.

$$L_x(T, V) = \log \left[\sum_S p(X|S) p(S|T, V) \right] \geq \sum_q q(S) \log \frac{p(X, S|T, V)}{q(S)}. \quad (2)$$

그런데 로그 함수는 비볼록 특성을 가지므로 대신 볼록 특성을 갖는 로그 함수의 하계(lower bound)를 다룬다.

EM 알고리즘은 Eq.(2)의 하계를 반복적으로 최대화하는 과정으로 다음과 같다. 먼저 E 단계에서는 (T, V)를 고정하고 하계를 최대화 만드는 잠재변수 S 의 사후분포

$$q(S)^{(n)} = p[S|X, T^{(n-1)}, V^{(n-1)}] \quad (3)$$

를 추정한다. 다음 M 단계에서는 추정된 사후분포 q 를 고정시키고 완전 데이터 결합 분포의 기댓값을 최대화하는 (T, V)를 구한다.

$$[T^{(n)}, V^{(n)}] = \operatorname{argmax}_{T, V} \langle \log p(X, S | T, V) \rangle_{q(S)^{(n)}}, \quad (4)$$

여기서 $\langle \cdot \rangle_q$ 는 q 에 대한 기댓값을 의미한다. 이 E와 M 단계는 하계가 수렴할 때 까지 반복된다.

2.2 VB-NMF

2.1절에서 기술한 최대우도 NMF에서는 T 와 V 를 직접 추정하였지만, Bayesian 접근 방식에서는 (T, V)의 사후분포를 추정한다. 이를 위해 T 와 V 가 잠재 변수 S 의 포아송 분포와 켈레 관계를 갖는 감마 분포로 가정된다. 감마 분포는 다음과 같이 정의된다.

$$G(v; a, b) = \exp\left[(a-1)\log v - \frac{v}{b} - \log \Gamma(a) - a \log b\right]. \quad (5)$$

감마 분포에서 a 와 b 는 각각 모양과 크기를 제어하는 초모수이다. 감마 분포의 기댓값은 b 이다.

$$\begin{aligned} (T, V) \text{의 사후분포는 주변 우도} \\ p(X|\theta) = \\ \int dT dV \sum_S p(X|S) p(S|T, V) p(T, V|\theta) \end{aligned} \quad (6)$$

를 최대로 하는 것으로 추정된다. 여기서 θ 는 감마 분포의 초모수들을 표현한다. ML-NMF의 경우와 동일하게 주변 우도는 직접 다루는 대신 주변 우도의 하계를 다룬다.

$$\log p(X|\theta) \geq \sum_s \int d(T, V) q \log \frac{p(X, S, T, V|\theta)}{q}. \quad (7)$$

Variational Bayesian 접근 방식은 사후분포 $q(S, T, V)$ 의 꼴을 독립적인 사후분포의 항들로 단순화하는 것이다.^[5]

$$q(S, T, V) = q(S)q(T)q(V). \quad (8)$$

Eq.(4)를 이용하여 하계를 최대로 만드는 사후분포는 EM 알고리즘을 이용하여 다음과 같은 꼴로 구해진다.

$$q_\alpha^{(n+1)} \propto \exp[\langle \log p(X, S, T, V|\theta) \rangle_{q_\alpha^{(n)}}], \quad (9)$$

여기서 $q_\alpha^{(n)}$ 은 q 에서 q_α 항을 제외한 나머지를 의미한다. Eq.(9)를 반복적으로 적용하면 우도함수의 하계 Eq.(7)을 단조적으로 증가시킨다. 알고리즘이 수렴한 이후 잠재변수는 다항분포를 가지며 조건기댓값을 이용하여 다음과 같이 구해진다.^[4]

$$\begin{aligned} \langle s_{i\tau} \rangle &= x_{i\tau} p_{i\tau} \\ &= x_{i\tau} \frac{\exp(\langle \log t_{i\tau} \rangle) \exp(\langle \log v_{i\tau} \rangle)}{\sum_i \exp(\langle \log t_{i\tau} \rangle) \exp(\langle \log v_{i\tau} \rangle)}. \end{aligned} \quad (10)$$

VB-NMF 알고리즘은 Reference [4]에 상세하게 기술되어 있다.

2.3 VB-NMF를 이용한 음성 신호의 음질 개선

VB-NMF 알고리즘을 이용한 음질 개선 절차는 다음과 같다.^[6,8] 먼저 음성 신호와 잡음신호의 기본 벡터를 미리 각각의 신호로부터 훈련을 통해 구한다. 이때 VB-NMF에서 $\langle t_{i\tau} \rangle$ 와 $\langle v_{i\tau} \rangle$ 는 2.1절에 기술된 KL-NMF를 통해 초기화된다. 잡음에 열화된 신호에서 잡음을 제거하고 음성 신호를 복원할 때에는 훈련 과정에서 얻은 기본 벡터를 고정시키고, VB-NMF 알고리즘을 이용하여 음성 신호와 잡음 신호의 이득 벡터들을 구한다. 마지막으로 Eq.(10)을 이용하여 음성신호에 해당하는 잠재변수의 조건기댓값을 계산함으로써 음성신호를 추정한다.

III. GMC를 이용한 음질 개선 방안

일반적으로 NMF에 신호의 연속성을 활용하면 음질개선 효과를 증진된다.^[2] VB-NMF에서 신호의 연속성을 보장하는 한가지 방법은 이전 프레임의 이득의 사후분포를 이용하여 현재 프레임 τ 의 이득 $v_{i\tau}$ 의 사전분포를 다음과 같이 초기화하는 방법이다.^[6]

$$b_{i\tau}^v = \alpha b_{i\tau-1}^v + (1-\alpha) \langle b_{i\tau-1}^v \rangle, \quad (11)$$

여기서 평활 파라미터 α 는 실험에 의해 결정된 값으

로 SNR의 함수로 주어진다. 본 논문에서는 별도의 SNR 추정없이 GMC를 이용하여 프레임 간 연속성을 구현한다.

3.1 GMC 모델

본 논문에서는 잡음 신호의 이득 벡터에 Fig. 1과 같은 GMC 모델^[7]을 적용한 VB-NMF를 고려한다. 여기서 $z_{i\tau}$ 는 역 감마분포를 갖으며 $v_{i\tau}$ 는 감마분포를 갖는다고 가정된다. 즉

$$\begin{aligned} z_{i1} &\sim IG(z_1; a_1^z, b_1 a_1^z), \\ v_{i\tau} | z_{i\tau} &\sim G(v_{i\tau}; a_{i\tau}^v, z_{i\tau} / a_{i\tau}^v), \\ z_{i\tau} | v_{i\tau} &\sim IG(z_{i\tau}; a_{i\tau}^z, v_{i\tau-1} a_{i\tau}^z). \end{aligned} \quad (12)$$

신호의 결합 특성은 천이 커널(transition kernel) $p(v_{i\tau} | v_{i\tau-1})$ 로 설명된다. 천이 커널은 초모수 $a_{i\tau}^z$ 와 $a_{i\tau}^v$ 의 함수로 나타나는데, Fig. 2는 다양한 초모수에 대한 천이 형태를 보여준다. $a_{i\tau}^v < a_{i\tau}^z$ 이면 음의 방향으로 $a_{i\tau}^v > a_{i\tau}^z$ 양의 방향으로 천이된다.

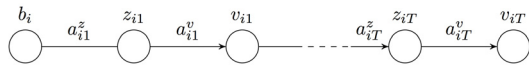


Fig. 1. A GMC model of a gain vector (Z,V).

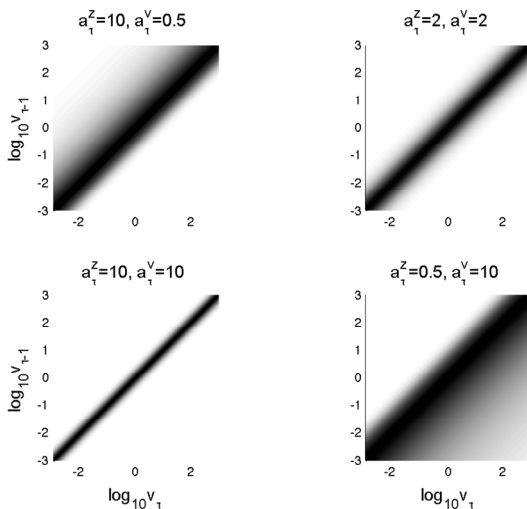


Fig. 2. Transition kernels of the GMC model used.

3.2 GMC 모델을 이용한 VB-NMF

3.1절에서 언급한 GMC 모델을 적용하여 VB-NMF를 재설계할 수 있다. 여기서 주변우도함수는 다음과 같이 표현된다.

$$p(X|\theta) = \int dT dV dZ \sum_S p(X|S) \cdot p(S|T, V, Z) p(T, V, Z|\theta). \quad (13)$$

이제 문제는 2.2절과 마찬가지로 주변우도 함수의 하계를 최대로 하는 변수들의 사후분포 $q(S, T, V, Z)$ 를 구하는 것으로 귀착된다. Variational Bayesian 접근 방법에 따라 사후분포를 다음과 같은 결합으로 분해한다.

$$q(S, T, V, Z) = q(S)q(T)q(V)q(Z). \quad (14)$$

이 근사식을 적용하여 Reference [6]과 같은 과정을 이용하여 Z 의 사후분포 $IG(z_{i\tau}; \alpha_{i\tau}^z, \beta_{i\tau}^z)$ 의 초모수 $\alpha_{i\tau}^z, \beta_{i\tau}^z$ 를 유도하면 각각 다음과 같다.

$$\begin{aligned} \alpha_{i\tau}^z &= a_{i\tau}^z + a_{i\tau}^v, \\ \beta_{i\tau}^z &= a_{i\tau}^z \langle v_{i\tau-1} \rangle + a_{i\tau}^v \langle v_{i\tau} \rangle, \end{aligned} \quad (15)$$

여기서 $\langle v_{i0} \rangle = b_i$ 이다. 이로부터

$$\begin{aligned} \left\langle \frac{1}{z_{i\tau}} \right\rangle &= \frac{\alpha_{i\tau}^z}{\beta_{i\tau}^z}, \\ &= \frac{a_{i\tau}^z + a_{i\tau}^v}{a_{i\tau}^z \langle v_{i\tau-1} \rangle + a_{i\tau}^v \langle v_{i\tau} \rangle}, \end{aligned} \quad (16)$$

$$\begin{aligned} \langle z_{i\tau} \rangle &= \frac{\beta_{i\tau}^z}{\alpha_{i\tau}^z - 1}, \\ &= \frac{a_{i\tau}^z \langle v_{i\tau-1} \rangle + a_{i\tau}^v \langle v_{i\tau} \rangle}{a_{i\tau}^z + a_{i\tau}^v - 1}. \end{aligned} \quad (17)$$

마찬가지로 V 의 사후분포 $G(v_{i\tau}; \alpha_{i\tau}^v, \beta_{i\tau}^v)$ 의 초모수 $\alpha_{i\tau}^v, \beta_{i\tau}^v$ 는 다음과 같이 유도된다.

$$\alpha_{i\tau}^v = a_{i\tau}^v + a_{i\tau+1}^z + \sum_v \langle s_{vi\tau} \rangle, \\ \beta_{i\tau}^v = \left(a_{i\tau}^v \left\langle \frac{1}{z_{i\tau}} \right\rangle + a_{i\tau+1}^z \left\langle \frac{1}{z_{i\tau+1}} \right\rangle + \sum_v \langle t_{vi} \rangle \right)^{-1}, \quad (18)$$

이로부터

$$\langle v_{i\tau} \rangle = \alpha_{i\tau}^v * \beta_{i\tau}^v. \quad (19)$$

Variational Bayesian에서 적용된 근사식(14)는 (Z, V) 사이에 존재하는 결합을 무시하는 것이므로 일견 GMC 모델의 의미를 훼손하는 것처럼 보이지만, 유도된 사후분포는 여전히 (Z, V) 사이의 결합을 보여 준다. 프레임 간 $v_{i\tau}$ 의 결합력은 $a_{i\tau}^z$ 와 $a_{i\tau}^v$ 에 의해 결정되는데, 이 초모수들은 Newton-Raphson 방식에 의해 계산될 수 있다.^[5]

3.3 GMC-VB-NMF를 이용한 음질 개선

앞에서 유도한 알고리즘을 프레임 별로 음질 개선에 적용하기 위해서는 적절한 변형이 필요하다. 프레임 별 처리에서는 미래 프레임이 가용하지 않으므로 GMC 모델은 Fig 3과 같이 표현될 수 있다. 여기서 초모수들은 i 에 대하여 동일하다고 가정된다. 또한 이전 프레임의 변수는 b_p, z_p, v_p 로, 현재 프레임의 변수는 b, z, v 로 표시하였다. Eqs.(17)과 (18)에서 $v_{i\tau-1}$ 과 $v_{i\tau}$ 는 각각 이전과 현재 프레임의 변수를 가르킨다. $v_{i\tau+1}$ 은 존재하지 않으므로 현재 프레임의 값으로 대체된다.

음성 부분에 대한 감마 분포의 초모수는 Reference [6]과 [7]과 동일하게 설정하며 잡음 부분에만 GMC를 적용한다. GMC 모델의 특성 상 b 는 초모수라기 보다는 잠재 변수로 해석될 수 있다. 즉 이전 프레임의 v_p 는 현재 프레임의 b 로 해석될 수 있다. 따라서 Reference [6]에서 Eq.(11)과 같이 v_p 의 사후 분포를 이용하여 b 를 초기화하는 것은 자연스러운 일이다.

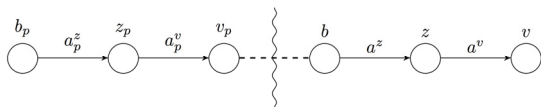


Fig. 3. The modified GMC model for frame-by-frame processing.

Reference [6]에서는 α 값을 실험적으로 구한 SNR의 함수로 표현하였지만, 여기서는 Eq.(17)의 사후 평균으로부터 힌트를 얻어 다음과 같이 정한다.

$$\alpha = \frac{a_{smoothed}^v}{a^v + a^z - 1}, \quad (20)$$

여기서 a^z 는 고정시키고 $a_{smoothed}^v$ 는 프레임 별로 계산된 a^v 의 변동성을 줄이기 위해 다음과 같이 평활된 것이다.

$$a_{smoothed}^v \leftarrow \gamma a_{smoothed}^v + (1-\gamma) a^v, \quad (21)$$

여기서 $\gamma = 0.9$ 를 적용하였다. 실험을 통해 잡음 신호의 사전분포 초모수인 a^z 와 a^v 가 거의 유사하게 1보다 큰 값을 갖는다는 사실을 관찰할 수 있었다. 그 결과 Eq.(20)의 α 는 0.54 ~ 0.6 사이의 변동성을 갖으며 이때 천이 커널은 Fig. 2의 왼쪽아래 패널 유형에 해당된다. 이는 충격성 잡음을 제외한 일반적인 잡음 신호에서는 사전 분포의 시간 결합성이 높다는 것으로 이해될 수 있다. Eq.(20)에서 산출된 α 값은 Reference [6]에서와 같이 0 ~ 1 사이의 값을 갖도록 다음과 같이 변환된다.

$$\alpha = \max [0.01, \min \{20(\alpha - 0.54), 0.99\}].$$

IV. 실험 결과

제안된 알고리즘의 성능 평가는 다음과 같이 이루어졌다. 음성 신호는 TIMIT 데이터를 사용하였다. 훈련에는 남자 여자 각각 4명으로부터 10개씩 총 80개의 음성 발화를 이용하였고, 시험용에는 남녀 각각 2명으로부터 총 40개의 음성 발화를 사용하였다. 잡음 신호는 NOISEX-92 데이터 베이스에서 destroyerengine과 leopard 신호를 이용하였다. 훈련에 사용된 잡음 신호의 길이는 90s이다. 음성과 잡음 신호들은 16 kHz로 변환된 것이며 훈련과 시험에 사용된 신호들은 서로 중복되지 않도록 선택되었다. 훈련으로부터 얻은 기본 벡터의 수는 음성과 잡음 신호 각각 60개와 100개이다.

실험을 통한 관찰 결과 초모수 값들은 Fig. 2의 왼쪽 아래의 경우와 유사한 값들로 수렴한다는 것을 확인하였다. 따라서 초모수 α^z 는 항상 9로 고정하고, α^v 는 10으로 초기화하고 Eqs.(11), (20)과 (21)을 이용하여 프레임 마다 적응적으로 계산하였다.

프레임 크기는 512 샘플에 50% 중첩을 사용하였고 sinebell 윈도우를 취한 다음 512 길이의 FFT를 통해 변환되었다. 복원된 음성 신호의 위상은 열화된 입력 신호의 위상을 사용하였다. KL-NMF의 랜덤 초기화에 의한 영향을 줄이기 위해, 모든 실험 결과는 10번의 독립적인 실험의 평균을 산출하였다.

성능 비교를 위해 0, 5, 10 dB 입력 조건에서 본 논문에서 제안된 GMC 방식과 Reference [6]에서 제안된 SNR 방식의 입출력 segSNR(segmental SNR),^[10] SDR (Signal Distortion Ratio)^[11]의 차이인 segSNRg, SDRg를 비교하였다. Table 1에서 확인할 수 있듯, destroyer-

engine의 경우 두 방식은 거의 유사한 성능을 보인다. 반면에 leopard의 경우 제안된 GMC 방식이 SNR 방식보다 약간 우수하다.

SNR의 변화에 따른 성능을 확인하기 위해 destroyerengine 잡음 신호를 구간 별로 0, 10, 0, 10 dB로 변하게 하고 α 의 변화를 비교하였다. Fig. 4에서 보는 바와 같이, SNR 방식에서 α 는 10 dB에서 0 dB로 변하는 경계에서는 빠르게 적응하지만, 반대로 0 dB에서 10 dB로 변하는 경계에서는 매우 느리게 반응한다. 이에 반해 GMC의 α 는 어느 경우어나 빠르게 반응한다. 결과적으로 Table 2에서 확인할 수 있듯, 0 dB에서 10 dB로 변하는 경계 구간에서 제안된 GMC 방식이 SNR 방식보다 약간 나은 segSNR과 SDR 이득을 제공한다.

V. 결론

본 논문은 VB-NMF에서 프레임 간 이득 벡터의 시간 연속성을 이용한 음질 개선에 대하여 기술하였다. 잡음 신호의 프레임 간 시간 연속성은 GMC를 이용하여 효과적으로 모델될 수 있음을 확인하였다. GMC를 이용한 모델은 기존의 SNR을 이용한 방식과 유사한 성능을 보이며, 잡음 수준이 변하는 환경에서 더 나은 적응력을 보여준다.

Table 1. Comparison of segSNRg/SDRg at various input SNRs.

Noise	Methods	Input SNR (dB)		
		0	5	10
destroyer-engine	SNR	5.48/8.64	5.12/7.01	4.14/4.95
	GMC	5.33/8.52	5.16/7.10	4.13/4.94
leopard	SNR	7.03/12.74	6.17/10.41	4.70/7.32
	GMC	7.52/12.75	6.71/10.60	4.83/7.38

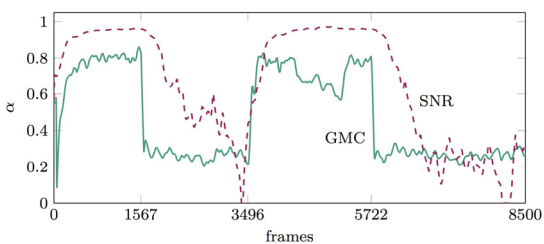


Fig. 4. Evolutions of α in the presence of SNR changes.

Table 2. Comparison of segSNRg and SDRg when SNR level changes.

Measures	Methods	Input SNR (intervals)			
		0	10	0	10
SegSNRg	SNR	5.40	3.09	6.13	3.86
	GMC	5.28	3.74	5.93	4.72
SDRg	SNR	8.21	3.72	9.68	4.15
	GMC	8.12	4.66	9.47	4.72

References

1. P. Loizou, *Speech Enhancement: Theory and Practice*, 2nd Ed., (CRC Press, Inc. Boca Raton, FL, 2013), pp. 1-6.
2. P. Smaragdis, C. Fevotte, G. J. Mysore, N. Mohammadiha, and M. Hoffman, "Static and dynamic source separation using nonnegative factorization," *IEEE Sigal Processing Magazine*, 66-75 (2014).
3. D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, **401**, 788-791 (1999).
4. A. T. Cemgil, "Bayesian inference in non-negative matrix factorisation models," *Computational Intelligence and Neuroscience*, **2009**, Article ID 785152, 1-17 (2009).
5. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, NewYork, 2006), pp. 462-466.
6. N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using NMF," *IEEE*

- Trans. Audio, Speech, Lang. Processing, **21**, 2140-2151, (2013).
7. T. Virtanen, A.T. Cemgil, and S. Godsill, "Bayesian extension to non-negative matrix factorization for audio signal modeling," IEEE Int. Conf. on Acousts, Speech and Signal Process. 1825-1828, (2008).
 8. <http://www.uni-oldenburg.de/en/medipysics-acoustics/sigproc/staff/nasser-mohammadiha/matlab-codes/>
 9. A. T. Cemgil, and O. Dikmen, "Conjugate Gamma Markov random fields for modeling nonstationary sources," 7th Int. Conf. on Independent Component Analysis and Signal Separation, 697-705, (2007)
 10. S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality* (Prentice Hall, New Jersey, 1988), pp. 45.
 11. E. Vincent, R. Grivonval, and C. Févotte, "Performance measurement in blind audio source separation." IEEE Trans. Audio, Speech, and Language Process. **14**, 1462-1469 (2006).

저자 약력

▶ 남 승 현 (Seung-Hyon Nam)



1980년 2월: 서강대학교 전자공학과 학사
 1987년 12월: 앨러바마 헨츠빌 컴퓨터 및 전기공학과 석사
 1992년 12월: 텍사스A&M 전기공학과 박사
 1980 ~ 1985년: 국방과학연구소 연구원
 1993년 3월 ~ 현재: 배재대학교 전자공학과