ISSN 2586-1816 / e-ISSN 2586-4440
융합정보논문지 제7권 제6호 pp. 159-166, 2017 DOI : https://doi.org/10.22156/CS4SMB.2017.7.6.159

SNR 기반 가중 KL 거리를 활용한 화자 변화 검증에 관한 연구

조준범[1], 이지은[2], 이경록[3*]
[1]남부대학교 간호학과, [2]전남과학대학교 생활체육과, [3]남부대학교 IT·디자인학과

The Study on Speaker Change Verification Using SNR based weighted KL distance

Joon-Beom Cho[1], Ji-eun Lee[2], Kyong-Rok Lee[3*]
[1]Department of Nursing, Nambu University
[2]Department of Living physical Training Special Study, Chunnam Techno University
[3]Department of IT & Design, Nambu University

요 약 본 논문에서는 방송 뉴스에서 화자 변화 검증 성능 향상을 위해서 입력소음음성 향상과 SNR(Signal to Noise Ratio)기반 가중 함수 w_m를 적용한 KL 거리 D_s를 실험하였다. GMM-UBM(Gaussian Mixture Model-Universal Background Model) 기반 KL(Kullback Leibler) 거리 D를 이용한 화자 변화 검증 시스템(실험 0)을 기본 시스템으로 한다. 실험 1은 실험 0의 입력소음음성 향상을 위해 MMSE Log-STSA(Minimum Mean Square Error Log-Spectral Amplitude Estimator)를 적용하였다. 실험 2는 실험 1의 기존 KL거리 D 대신에 D_s를 적용하였다. 실험 데이터베이스는 다양한 소음을 반영하기 위해 스포츠 뉴스와 실외 인터뷰를 중심으로 구축하였다. 실험은 화자 변화 정보의 누락을 막기 위해 MDR(Missed Detection Rate) 0%를 기준으로 하였다. 실험 0은 FAR(False Alarm Rate) 71.5%의 성능을 보였다. 실험 1은 FAR 67.3%로 실험0에 비해 4.2% 향상되었고, 실험 2는 FAR 60.7%로 10.8% 향상되었다.

키워드 : 화자 변화검출, Kullback Leibler 거리, 음성향상, Minimum Mean Square Error Log-Spectral Amplitude Estimator, 신호대 잡음 비

Abstract In this paper, we have experimented to improve the verification performance of speaker change detection on broadcast news. It is to enhance the input noisy speech and to apply the KL distance D_s using the SNR-based weighting function w_m. The basic experimental system is the verification system of speaker change using GMM-UBM based KL distance D(Experiment 0). Experiment 1 applies the input noisy speech enhancement using MMSE Log-STSA. Experiment 2 applies the new KL distance D_s to the system of Experiment 1. Experiments were conducted under the condition of 0% MDR in order to prevent missing information of speaker change. The FAR of Experiment 0 was 71.5%. The FAR of Experiment 1 was 67.3%, which was 4.2% higher than that of Experiment 0. The FAR of experiment 2 was 60.7%, which was 10.8% higher than that of experiment 0.

Key Words : Speaker Change Detection, Kullback Leibler distance, Speech Enhancement, Minimum Mean Square Error Log-Spectral Amplitude Estimator, Signal to Noise Ratio

Received 2017-11-08 Revised 2017-11-29 Accepted 2017-12-20 Published 2017-12-31
*Corresponding author : Kyong-Rok Lee (krlee@nambu.ac.kr)

1. 서론

화자 변화검출은 입력음향의 질에 따라서 큰 영향을 받는다. 입력음향의 질은 배경소음과 목적하는 음성의 대비에 의해서 결정된다. 일반적인 미팅 환경에서의 화자 변화 검출은 배경잡음이 단순하며 일관성이 있는 반면에 방송 뉴스 환경의 경우, 실내에서 녹음된 데이터와 실외에서 녹음된 데이터의 배경잡음의 성질의 차이가 크다.

특히, 스포츠 뉴스나 실외 인터뷰의 경우 SNR(Signal to Noise Ratio)이 극히 낮아 사람의 귀로 들어도 제대로 인식하기 어려운 경우도 있다. 이런 고소음 환경의 화자 변화 검출은 전체 성능을 저하시키는 큰 요인이 된다.

본 논문에서는 기존의 GMM-UBM(Gaussian Mixture Model-Universal Background Model) 기반 KL(Kullback Leibler) 거리를 활용한 화자 변화검증 시스템의 문제점이 고소음 환경에서의 성능 저하로 분석하고, 소음에 대한 선 처리와 소음정보를 반영한 새로운 KL 거리를 적용하여 이를 개선하고자 한다.

2. 소음음성의 향상

배경잡음이 존재하는 음성만을 활용할 수 있을 때, 비상관(Uncorrelated) 가산 잡음에 의해 왜곡된 음성을 향상하는 것이 요즘 큰 관심을 받고 있다[1].

본 논문에서는 고소음 환경에서의 화자 변화 검증 성능을 향상하기 위해서 최근까지 다양한 음성향상에 적용되고 있는 MMSE SATA(Minimum Mean Square Error Spectral Amplitude Estimator) 방식을 이용하여 소음음성을 향상하고자 하였다[2-4], 그 후 추정된 클린 음성을 사용하여 프레임별 SNR을 계산하였다.

2.1 MMSE SATA

MMSE는 미지의 변수에 대해 최적의 추정치를 얻기 위해 사용되는 방법 중 하나이다. 수학적으로 취급하기 쉽고 계산이 용이성을 갖춘 평균제곱오차(Mean Square Error, MSE)를 추정오차 최소화의 정량적 판단 기준으로 사용한다. 평균제곱오차를 최소화하는 것이 MMSE 추정방법의 핵심이다[5-7].

Ephraim는 소음배경 음성만을 활용 가능할 때, 비상관(Uncorrelated) 가산 잡음에 의해 저하 된 음성을 향상시키기 위해 단시간 스펙트럼 진폭 (Short Time Spectral Amplitude, STSA)에 대한 최적의 MMSE 추정을 제안했다. 최대 우도(Maximum Likelihood, ML) 접근법과 결정 지향 접근법을 제안하고 성능을 개선하였다. Ephraim가 제안한 알고리즘은 음성 신호의 STSA를 중심으로 MMSE STSA 추정기를 사용하여 잡음이 있는 음성을 향상시킨다[2,3,8].

Ephraim의 연구에서 사용 된 스펙트럼의 평균 제곱 오차의 왜곡 측정은 수학적으로 다루기 쉽고 좋은 결과를 도출한다. Gray의 연구에서 로그 - 스펙트럼의 평균 제곱 오차에 기초한 왜곡 측정이 음성 처리에 더 적합하다는 것이 알려졌다[3,9]. 이에 본 논문에서는 로그-스펙트럼의 평균 제곱 오차를 사용하였다.

2.2 MMSE Log-STSA 추정기

이 절은 Ephraim의 연구에서 제안된 MMSE Log-STSA 추정기에 대한 내용이다[2,3]. STSA 추정 문제는 주어진 소음음성 $\{y(t), 0 \le t \le T\}$에서 음성신호$\{x(t), 0 \le t \le T\}$의 각 푸리에 확장 계수의 진폭을 추정하는 것으로 공식화할 수 있다. 음성 프로세스와 소음 프로세스의 푸리에 확장 계수는 통계적으로 독립적인 가우시안 확률 변수로 모델링된다.

$X_k = A_k e^{j\alpha k}$, D_k, $Y_k = R_k e^{j\theta k}$ 를 구간 [0, T]에서의 음성신호, 소음처리, 관측소음음성의 k번째 푸리에 확장 계수라고 하자. 소음음성을 향상시키기 위해서, 주어진 관측소음음성 $\{y(t),\ 0 \le t \le T\}$에서 왜곡 측정 수식 (1)을 최소화하는 추정기 $\hat{A}_k$ 수식 (2)를 찾는다.

$$E\left\{\left(\log A_k - \log \hat{A}_k\right)^2\right\} \tag{1}$$

$$\hat{A}_k = \exp\left\{E\left[\ln A_k | y(t)\right], 0 \le t \le T\right\} \tag{2}$$

Ephraim의 연구에 따르면 가정된 통계 모델 하에서, $\{y(t), 0 \le t \le T\}$에서 A_k의 기댓 값은 Y_k만 주어진 A_k의 기대 값과 동일하다. 이것은 A_k를 $\ln A_k$로 대체해도 동일하므로 수식 (2)는 수식 (3)과 같다.

$$\hat{A}_k = \exp\left\{E\left[\ln A_k | Y_k\right]\right\} \tag{3}$$

가우시안 모델에 대한 $E\left[\ln A_k | Y_k\right]$는 주어진 Y_k에서

$\ln A_k$의 모멘트(Moment) 생성 함수를 이용한다. $Z_k = \ln A_k$라 하면 주어진 Y_k에서 Z_k의 모멘트 생성 함수 $\Phi_{Z_k|Y_k}(\mu)$는 수식 (4)와 같다. $E[\ln A_k|Y_k]$를 수식 (4)를 이용하여 나타내면 수식 (5)와 같다.

$$\begin{aligned}\Phi_{Z_k|Y_k}(\mu) &= E\{\exp(\mu Z_k)|Y_k\} \\ &= E\{A_k^{\mu}|Y_k\}\end{aligned} \tag{4}$$

$$E[\ln A_k|Y_k] = \frac{d}{d\mu}\Phi_{Z_k|Y_k}(\mu)|_{\mu=0} \tag{5}$$

Ephraim의 연구에 의하면 수식 (4)를 가우시안 모델 기반으로 해석하면 수식 (6)와 같다.

$$\begin{aligned}\Phi_{Z_k|Y_k}(\mu) &= E\{A_k^{\mu}|Y_k\} \\ &= \frac{\int_0^{\infty}\int_0^{2\pi} a_k^{\mu} p(Y_k|a_k,\alpha_k)p(a_k,\alpha_k)d\alpha_k da_k}{\int_0^{\infty}\int_0^{2\pi} p(Y_k|a_k,\alpha_k)p(a_k,\alpha_k)d\alpha_k da_k}\end{aligned} \tag{6}$$

여기서 $p(Y_k|a_k)$는 수식 (7)로 정의되었고, $p(a_k,\alpha_k)$은 수식 (8)로 정의되었다.

$$p(Y_k|a_k) = \frac{1}{\pi\lambda_d(k)} exp\left\{\frac{-1}{\lambda_d(k)}\left|Y_k - a_k e^{j\alpha k}\right|^2\right\} \tag{7}$$

$$p(a_k,\alpha_k) = \frac{a_k}{\pi\lambda_x(k)} exp\left\{-\frac{a_k^2}{\lambda_x(k)}\right\} \tag{8}$$

$\lambda_d(k) \approx E\{|D_k|^2\}$ 와 $\lambda_x(k) \approx E\{|X_k|^2\}$ 는 잡음음성과 음성의 k번째 스펙트럼 성분의 분산이다. 수식 (6)에 수식 (7)과 수식 (8)을 대입하고, 0차 $I_0(\ )$의 수정된 베셀(Bessel) 함수의 적분 표현을 사용하면 수식 (9)와 같다[10].

$$\Phi_{Z_k|Y_k}(\mu) = \frac{\int_0^{\infty} a_k^{\mu+1}\exp(-a_k^2/\lambda_k) I_0\left(2a_k\sqrt{\nu_k/\lambda k}\right)da_k}{\int_0^{\infty} a_k\exp(-a_k^2/\lambda_k) I_0\left(2a_k\sqrt{\nu_k/\lambda k}\right)da_k} \tag{9}$$

$$\frac{1}{\lambda_k} = \frac{1}{\lambda_x(k)} + \frac{1}{\lambda_d(k)} \tag{10}$$

$$v_k \approx \frac{\xi_k}{1+\xi_k}\gamma_k,\ \xi_k \approx \frac{\lambda_x(k)}{\lambda_d(k)},\ \gamma_k \approx \frac{R_k^2}{\lambda_d(k)} \tag{11}$$

ξ_k, γ_k는 각각 사전, 사후 SNR이다. 수식 (9)의 적분을 계산하면 수식 (12)와 같다[10].

$$\Phi_{Z_k|Y_k}(\mu) = \lambda_k^{\mu/2}\Gamma(\mu/2+1)M(-\mu/2;1;-\nu_k) \tag{12}$$

여기서, $\Gamma(\ )$는 감마 함수이고, $M(a;c;x)$는 융합 초기하(Hypergeometric) 함수이다[10]. $\Phi_{Z_k|Y_k}(\mu)$는 라이시안(Rician) 확률 변수의 μ번째 모멘트 공식이다. 수식 (5)를 적용될 μ에 대한 $\Phi_{Z_k|Y_k}(\mu)$의 도함수는 다음과 같다.

먼저, $M(a:c:x)$는 수식 (13)과 같이 정의된다[10].

$$M(a;c;x) = \sum_{r=0}^{\infty}\frac{(a)_r}{(c)_r}\frac{x^r}{r!} \tag{13}$$

여기에서 $(a)_r \approx 1 \bullet a \bullet (a+1),\cdots,(a+r-1)$이고, $(a)_0 \approx 1$이다. Ephraim의 연구에 의하면 $\mu = 0$ 일 때 수식 (12)의 $M(-\mu/2;1;-\nu_k)$ 의 도함수는 수식 (14)와 같다.

$$\begin{aligned}&\frac{\partial}{\partial\mu}M(-\mu/2;1;-\nu_k)|_{\mu=0} \\ &= -\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-\nu)^r}{r!}\frac{1}{r}\end{aligned} \tag{14}$$

$\Gamma(\mu/2+1)$의 도함수는 수식 (15)와 같고, $\ln\Gamma(\mu/2+1)$의 도함수를 사용하여 수식 (16)과 같이 정의할 수 있다[10. eq. 8.342.1].

$$\frac{\partial}{\partial\mu}\Gamma\left(\frac{\mu}{2}+1\right) = \Gamma\left(\frac{\mu}{2}+1\right)\frac{d}{d\mu}ln\Gamma\left(\frac{\mu}{2}+1\right) \tag{15}$$

$$\begin{aligned}&\ln\Gamma(\mu/2+1) \\ &= -c\frac{\mu}{2} + \sum_{r=2}^{\infty}\frac{(-\mu)^r}{2^r r}\alpha_r \quad |\mu| < 2\end{aligned} \tag{16}$$

여기서, $\alpha_r \approx \sum_{n=1}^{\infty} \frac{1}{n^r}$이고 $c = 0.57721566490$는 오일러 상수이다. 수식 (16)을 미분하여 수식 (15)에 대입하면 수식 (17)을 구할 수 있다.

$$\frac{d}{d\mu} I\left(\frac{\mu}{2}+1\right)|_{\mu=0} = -c/2 \tag{17}$$

수식 (12)에 수식 (14)와 수식 (17)을 대입하면 수식 (18)과 같다.

$$\begin{aligned} &\frac{d}{d\mu}\Phi_{Z_k|Y_k}(\mu)|_{\mu=0} \\ &= \frac{1}{2} ln\lambda_k - \frac{1}{2}\left(c + \sum_{r=1}^{\infty} \frac{(-\nu_k)^r}{r!}\frac{1}{r}\right) \\ &= \frac{1}{2} ln\lambda_k + \frac{1}{2}\left(\ln\nu_k + \int_{\nu_k}^{\infty} \frac{e^{-t}}{t} dt\right) \end{aligned} \tag{18}$$

수식 (18)을 수식 (5)에 대입하고 수식 (11)과 수식 (3)를 사용하면 식 (19)를 구할 수 있다. 식 (19)를 사용하여 원하는 진폭 추정기를 얻을 수 있다.

$$\hat{A} = \frac{\xi_k}{1+\xi_k} exp\left\{\frac{1}{2}\int_{\nu_k}^{\infty} \frac{e^{-t}}{t} dt\right\} R_k \tag{19}$$

2.3 결정지향 접근법을 활용한 $\xi_k(n)$ 계산

결정 지향 접근법은 ML(Maximum Likelihood) 추정치의 가중 평균과 향상된 음성으로부터 결정된 이전 프레임의 SNR 추정치를 취함으로써 SNR 추정치를 계산한다. 사용된 가중치는 0.98이다. 이 때, 평균 잡음 파워 스펙트럼은 미리 알고 있다고 가정한다[8].

다음은 Ephraim, Soni의 연구에서 제안된 결정지향 접근법을 나타낸 것이다[2,8].

사전 SNR $\xi_k(n)$는 n번째 분석 프레임의 k 번째 스펙트럼에서의 실제 SNR으로 수식 (20)과 같다.

$$\xi_k(n) = \frac{E\{A_k^2(n)\}}{\lambda_d(k,n)} \tag{20}$$

사후 SNR $\gamma(n)$는 관찰된 잡음음성의 크기의 제곱과 잡음 파워의 비율에 의해 주어지는 잡음이 추가 된 후 n번째 분석 프레임에서 관측되고 측정 된 SNR로 간주될 수 있다. $\xi_k(n)$의 제안된 추정기 $\hat{\xi}_k(n)$는 수식 (21)과 같고, 이 때 $P[x]$는 수식 (22)와 같다.

$$\begin{aligned} \hat{\xi}_k(n) &= \alpha \frac{\hat{A}_k^2(n-1)}{\lambda_d(k,n-1)} \\ &\quad + (1-\alpha) P[\gamma_k(n)-1], \ 0 \le \alpha < 1 \end{aligned} \tag{21}$$

$$P[x] = \begin{cases} x & \text{if } x \ge 0 \\ 0 & otherwise \end{cases} \tag{22}$$

다수의 실험을 통해서 결정된 초기 조건은 수식 (23)과 같다.

$$\hat{\xi}_k(0) = \alpha + (1-\alpha) P[\gamma_k(0)-1] \tag{23}$$

Ephraim의 연구에서는 재귀적 추정기 $\xi_k(n)$의 비선형적인 특성을 반영하여 시뮬레이션을 통해서 α의 최적의 값을 0.98로 결정하였다. 이는 본 논문에서도 그대로 적용되었다[2].

3. W_m을 반영한 KL 거리 D_s

다음은 Wu, Campbell, Lu의 연구에서 활용되고, Cho의 연구에서 실험한 KL 거리 D는 수식 (24)와 같다[11-14]. 수식 (24)는 두 분석 윈도우 X, Y에서 훈련된 GMM의 공분산을 활용한 각 모델간 거리를 계산한다.

$$D = \frac{1}{2} tr\left[\left(\Sigma_x - \Sigma_y\right)\left(\Sigma_y^{-1} - \Sigma_x^{-1}\right)\right] \tag{24}$$

Σ_x, Σ_y는 분석 윈도우 X, Y의 GMM의 공분산이다. D가 작으면 두 GMM간의 거리가 가깝다는 뜻이므로 두 분석 윈도우가 동일 화자에게서 발성되었다는 것이다. 실험에서는 문턱치를 조절하면서 문턱치보다 낮은 D값을 가진 화자 변화 지점은 화자 변화가 일어나지 않은 것으로 처리하였다.

GMM을 활용한 거리 기반 분할방식 알고리즘들의 일반적인 문제점인 GMM의 훈련 데이터 부족으로 인한 왜곡을 보상하기 위해서 기존의 수식 (24)에 GMM-UBM을 적용하고, 이를 수식 (25)로 정의하였다.

$$D=\frac{1}{2}tr\left[\left(\Sigma_{ux}-\Sigma_{uy}\right)\left(\Sigma_{uy}^{-1}-\Sigma_{ux}^{-1}\right)\right] \quad (25)$$

Σ_{ux}, Σ_{uy}는 분석 윈도우 X, Y의 GMM-UBM의 공분산이다. 실험에서는 KL 거리 방법과 동일하게 문턱치를 조절하면서 문턱치보다 낮은 D값을 가진 화자 변화 지점을 거절하였다.

기존의 KL 거리 D를 적용한 화자 변화 지점 검증의 문제점은 동일화자에 의해 발성된 구간에서 양쪽 분석 윈도우의 소음환경이 다를 경우에 그 성능 저하가 크다는 점이다. 본 논문에서는 이를 극복하기 위해서 분석윈도우의 음성 프레임 중 상대적으로 소음에 의한 왜곡이 적은 부분만을 분석에 활용하여 성능을 향상하고자 하였다. 뉴스 데이터의 특성 상 소음환경의 음성만이 획득 가능한 상황을 고려하여 2절의 방법을 사용하여 추정한 소음향상 음성과 원본 음성간의 SNR을 계산한 다음, SNR 기반 음성 프레임 가중 함수를 사용하여 분석 윈도우 모델링을 실시하였다.

Alam1의 연구에서는 GMM 기반 음성존재확률을 사용하여 노이즈 스펙트럼을 계산할 때 시그모이드 가중(Sigmoid weighting)을 사용하여 GMM 모델링에 대한 소음 가중치를 적용하였다[15]. 본 논문에서는 음성 신호의 배경잡음의 강도에 따른 분석 윈도우 모델링 가중치를 주기 위해 다음과 같이 SNR 기반 가중 함수 W_m을 수식 (26)과 같이 제안한다. 이러한 W_m을 KL 거리 D에 사용되는 GMM 모델링에 적용하였다.

$$W_m=\begin{cases}1 & \text{if } S_m \geq \beta \bullet S_{mean} \\ 0 & else\end{cases} \quad (26)$$

S_m은 2절에 의해 추정된 m번째 프레임의 SNR, S_{mean}은 2절에 의해 추정된 양쪽 분석 윈도우 전체의 SNR 평균, β는 0~1 사이의 값이다. 이를 통해서 문턱치보다 낮은 SNR을 가진 프레임을 분석윈도우의 KL 거리 연산을 위한 GMM 모델링에서 제외하였다.

SNR 기반 가중 함수 W_m을 적용한 KL 거리는 D_s 는 수식 (27)과 같다. $\Sigma_{ux_s}, \Sigma_{uy_s}$ 는 분석 윈도우 X, Y에서 W_m을 적용한 GMM-UBM의 공분산이다.

$$D_s=\frac{1}{2}tr\left[\left(\Sigma_{ux_s}-\Sigma_{uy_s}\right)\left(\Sigma_{uy_s}^{-1}-\Sigma_{ux_s}^{-1}\right)\right] \quad (27)$$

4. 실험결과

4.1 실험 데이터베이스

화자 변화 검출을 위하여 실제 뉴스 데이터 3회분을 대상으로 잡음환경, 화자정보(성별) 등의 특징을 고려하여 훈련, 테스트 데이터를 구축하였다. 특히, 잡음환경에서의 화자 변화를 실험하기 위해서 스포츠 뉴스와 실외 인터뷰 등을 중심으로 데이터를 수집하였다. 실험을 위한 데이터베이스 구성은 Table 1과 같다.

Table 1. Description of database

Aspects of Noise	Number of speaker change	Pattern of speaker change
Low SNR	25	Male ↔ Male
	16	Male ↔ Female
High SNR	6	Male ↔ Male
	4	Male ↔ Female

4.2 실험 시스템

본 논문은 Cho가 연구한 GMM-UBM 기반 KL 거리를 이용한 화자 변화 검증 시스템을 기본 시스템으로 사용하였다[14]. 4.3은 2절의 내용을 적용하여 소음음성을 향상시키고 이를 화자 변화 검증 시스템을 이용하여 실험한다. 4.4는 4.3에서 도출된 SNR을 기반으로 한 W_m을 적용한 KL 거리 D_s를 사용하여 화자 변화 검출을 검증한다.

실험결과 분석을 위하여 MDR, FAR을 수식 (28)과 같이 정의하였다.

$$MDR=\frac{N_{scd}-N_{cd}}{N_{scd}}\times 100 \quad (28)$$

$$FAR=\frac{N_{all}-N_{cd}}{N_{all}}\times 100$$

N_{scd}는 실제 화자 변화 지점의 수, N_{cd}는 실제 화자 변화 지점 중 검출된 수, N_{all}는 검출된 전체 화자 변화 지점의 수이다.

Fig. 1은 실험 2를 간략하게 나타낸 것으로 기존의 KL 거리 D를 활용했던 화자 변화 지점 검증을 SNR 기반

SNR 기반 가중 함수 W_m을 적용한 KL 거리는 D_s 로 개선한 것을 나타낸 것이다.

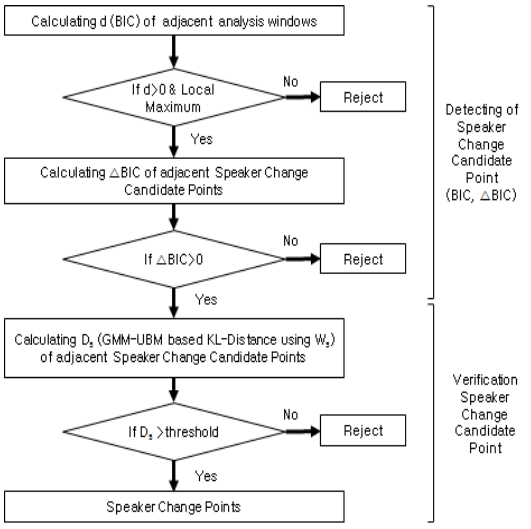

Fig. 1. A brief block diagram of the verification of Speaker Change Candidate Points using KL-Distance D_s

4.3 실험 1. MMSE Log-STSA 기반 음성 향상

소음음성만이 사용가능한 상황에서 고소음음성 향상이 화자 변화 검출에 미치는 영향을 실험하기 위해서 MMSE Log-STSA 추정기를 활용하여 소음음성을 향상시켰다.

실험은 차후 화자인식을 통한 화자 클러스터링을 위해서 MDR이 0인 경우로 한정하여 화자 변화 정보의 누락을 방지하였다. 실험결과 Table 2에서 보는바와 같이 실험 0에 비해 FAR이 4.2% 향상되었다.

FA(False Alarm) 패턴을 분석한 결과, 고소음 환경에서 발행한 FA 지점의 양쪽 분석 윈도우에 대한 음성향상에도 불구하고 소음이 잔존하여 화자 변화가 발생했다고 오인식하는 경우가 많았다.

Table 2. Experiment result of the KL distance D based verification of SCP using Speech Enhancement by MMSE log-STSA.

Speech Enhancement	Threshold	MDR	FAR
Before	0.028	0	71.5
After	0.024	0	67.3

4.4 실험 2. W_m을 반영한 KL 거리 D_s 적용

실험 1에서는 고소음환경 음성 향상의 한계로 인하여 성능개선이 제한적이었다. 이를 극복하기 위하여 4.3절의 실험에 각 분석 윈도우의 GMM을 모델링할 때 SNR 가중치 함수 W_m을 반영한 KL 거리 D_s를 적용하였다. 실험결과 Table 3에서 보는바와 같이 β가 0.42일 때 실험 0에 비해서 10.8% 향상되었다.

실험 1과 실험 2의 FA 패턴을 비교분석하면, 실험1에서 빈발한 고소음 환경에서의 FA가 해당 지점의 양쪽 분석 윈도우의 음성 프레임 중 SNR 기반 문턱치보다 낮은, 즉 소음에 의한 왜곡이 심한 음성 프레임이 모델링에서 제외되어 분석 윈도우 모델의 변별력이 강화되었음을 확인할 수 있었다.

Table 3. Experiment result of the KL distance D_s based verification of SCP using Speech Enhancement by MMSE log-STSA.

Applying D_s	Threshold	β	MDR	FAR
Before	0.024	-	0	67.3
After	0.024	0.42	0	60.7

5. 결론

본 논문에서는 다양한 소음이 존재하는 방송뉴스의 화자 변화 검출 검증 성능 향상을 위해서 입력음성 향상과 SNR 기반 가중 함수 W_m을 적용한 KL 거리 D_s를 실험하였다. 실험시스템은 Cho의 연구에서 제안된 시스템을 기반으로 실험 1은 MMSE Log-STSA를 이용한 입력 소음음성 향상을 적용했을 경우, 실험 2는 실험 1에 기존의 KL 거리 D 대신에 KL 거리 D_s를 적용했을 경우로 구성되었다[14].

실험 데이터베이스는 다양한 소음환경을 반영하기 위해서 스포츠 뉴스와 실외 인터뷰 등을 중심으로 구축하였다. 총 51개의 화자 변화 지점은 저 SNR 41개, 고 SNR 10개로 구성되었다.

실험은 화자 변화 정보 누락을 방지하기 위해서 MDR 0%를 기본조건으로 한다. Cho의 시스템을 적용한 경우(실험 0), FAR 71.5%였다[14]. 입력 소음음성 향상을 적용했을 경우(실험 1), FAR 67.3%로 실험 0에 비해 4.2%

향상되었다. SNR 가중치 함수 W_m를 반영한 KL 거리 D_s를 추가 적용했을 경우(실험 2), FAR 60.7%로 실험 0에 비해 10.8% 향상되었다.

ACKNOWLEDGMENTS

This study was supported by research funds from Nambu University, 2016.

REFERENCES

[1] J. S. Lim & A. V. Oppenheim. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67(12),* 1586-1604. USA : IEEE.
DOI : 10.21236/ada073139

[2] Y. Ephraim & D. Malah. (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(6),* 1109-1121.
DOI : 10.1109/tassp.1984.1164453

[3] Y. Ephraim & D. Malah. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 33(2), 443-445.*
DOI : 10.1109/icmcs.2014.6911142

[4] K. Paliwal, B. Schwerin & K. Wójcicki. (2012). Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. *Speech Communication, 54(2),* 282-305.
DOI : 10.1016/j.specom.2011.09.003

[5] J. B. Cha. (2017). *Minimum Mean Square Error, Glossary of ICT.* Ktword. www.ktword.co.kr

[6] P. C. Loizou. (2013). *Speech enhancement : theory and practice.* USA : CRC press.

[7] V. O. Alan & C. Ve. George. (2010). *CHAPTER 8 Estimation with Minimum Mean Square Error.* MIT Open Course Ware. https://ocw.mit.edu

[8] B. A. Soni & K. Vaghela. (2017). Spectral Subtraction and MMSE : A Hybrid Approach For Speech Enhancement. *International Reaserch Journal of Engineering and Technology, 4(4),* 2340-2343.

[9] R. Gray, A. Buzo, A. Gray & Y. Matsuyama. (1980). Distortion measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4),* 367-376.
DOI : 10.1109/TASSP.1980.1163421

[10] I. S. Gradshteyn & Z. M. Ryzhik. (1980). *Table of integrals, series, and products.* New York : Academic Press.

[11] T. Y. Wu, L. Lu, K. Chen & H. Zhang. (2003). Universal Background Models for Real-time Speaker Change Detection. *In MMM* (pp. 135-149). Russia : MMM.

[12] J. P. Campbell. (1997). Speaker recognition : A tutorial. *Proceedings of the IEEE, 85(9),* 1437-1462. USA : IEEE.
DOI : 10.1109/5.628714

[13] L. Lu & H. J. Zhang. (2002). Speaker change detection and tracking in real-time news broadcasting analysis. *In Proceedings of the tenth ACM international conference on Multimedia* (pp. 602-610). USA : ACM.
DOI : 10.1145/641007.641127

[14] J. B. Cho, J. E. Lee & K. R. Lee. (2016). The Study on the Verification of Speaker Change using GMM-UBM based KL distance. *Journal of Convergence for Information Technology, 6(1),* 71-77.
DOI : 10.22156/cs4smb.2016.6.4.071

[15] M. J. Alam1, P. Kenny1, P. Dumouchel & D. O'Shaughnessy. (2014). Noise Spectrum Estimation using Gaussian Mixture Model-based Speech Presence Probability for Robust Speech Recognition. *INTERSPEECH 2014,* 2759-2763. Singapore : INTERSPEECH.

저 자 소 개

조 준 범(Joon-Beom Cho) [정회원]

- 1989년 2월 : 원광대학교 무역학과 학사
- 1995년 2월 : 원광대학교 영어영문학과 석사
- 2005년 8월 : 원광대학교 영어영문학과 박사
- 1999년 3월 ~ 현재 : 남부대학교 교수

<관심분야> : 멀티미디어 콘텐츠 인덱싱, 언어학, 통사론

이 지 은(Ji-Eun Lee) [정회원]

- 2001년 2월 : 조선대학교 무용과 학사
- 2003년 2월 : 조선대학교 체육학 석사
- 2006년 8월 : 전남대학교 체육학 박사
- 2015년 9월 ~ 현재 : 전남과학대학교 부교수

<관심분야> : 스포츠 콘텐츠 인덱싱, 스포츠융합, 통계

이 경 록(Kynog-Rok Lee) [정회원]

- 1997년 2월 : 호남대학교 전자공학과 공학사
- 2001년 8월 : 전남대학교 정보통신협동과정 공학석사
- 2006년 2월 : 전남대학교 컴퓨터전자공학부 공학박사
- 2008년 4월 ~ 현재 : 남부대학교 IT·디자인학과 부교수

<관심분야> : 멀티미디어 콘텐츠 인덱싱, 화자인식