

# 화자식별을 위한 히스토그램 등화 기법의 개선방법

김 명 재\*, 양 일 호\*, 소 병 민\*, 김 민 석\*\*, 유 하 진\*

\*서울시립대학교 컴퓨터과학부

\*\*LG전자 전자기술원

## Improvement of Histogram Equalization for Speaker Identification

Myung-Jae Kim\*, IL-Ho Yang\*, Byung-Min So\*, Min-Seok Kim\*\*, Ha-Jin Yu\*

\*School of Computer Science, University of Seoul

\*\*Advanced Research Institute LG Electronics

mj@uos.ac.kr, heisco@hanmail.net, sbm1210@naver.com,

minseok3.kim@lge.com, hjyu@uos.ac.kr

### Abstract

In this paper, we propose an improvement method of histogram equalization. The method is based on order-statistics histogram equalization. Ranks of speech features of a test utterance are calculated on the UBM and the test utterance. Our system reduced the relative error rate 14.1% from average baseline systems.

### 1. 서론

일반적으로 히스토그램 등화 기법 (HEQ, Histogram Equalization)은 충분한 음성 발화가 필요하며, 음성 발화가 충분하지 않으면 다른 채널 보상 방법에 비해 낮은 화자 식별 성능을 보인다. 이러한 단점을 개선하기 위해 UBM (universal background model) 학습 데이터를 이용하여, 짧은 음성 발화의 화자 식별 성능을 개선할 수 있다. 개선된 결과를 다른 채널 보상 기법인 CMN (cepstral mean normalization)[1], MVN (mean and variance normalization)[2], 히스토그램 등화 기법과 비교하였다. 본 논문은 2장에서 제안한 히스토그램 등화 기법에 대해 설명하고, 3장에서 실험 설계 및 실험 결과를 보이고, 5장에서 결론을 맺는다.

### 2. 제안한 히스토그램 등화 기법

본 논문은 순서 기반의 히스토그램 추정 방식을 기반으로 한 히스토그램 등화 기법을 제안한다. 제안한 히스토그램 등화 기법은 UBM 학습 데이터에 대한 음성 데이터의 오름차순 서열과 음성 데이터의 오름차순 서열의 합을 새로운 서열로 하여 히스토그램을 추정하

며, 추정과정은 다음과 같다.

1. UBM 학습 데이터 전체에서 수집한 특정한 차원의 특징계수 열을  $U$ 라 정의한다.

$$U = \{u_1, u_2, \dots, u_m, \dots, u_M\} \quad (1)$$

2. 실험에 사용할 발화 음성 데이터에서 추출한 특정 차원의 특징계수 열을  $S$ 라 정의한다.

$$S = \{s_1, s_2, \dots, s_n, \dots, s_N\} \quad (2)$$

3.  $S$ 의 특징계수의 크기에 따른 오름차순 서열과  $U$ 에 대한  $S$ 의 오름차순 서열을 계산하고, 각각  $R_s$ 와  $R_u$ 로 정의한다.

4. 3에서 정의한 두 서열을 가지고 다음과 같이 새로운 서열  $R_{\text{new}}$ 를 정의한다.

$$R_{\text{new}}(s_n) = R_s(s_n) + R_u(s_n) \quad (3)$$

5. 새롭게 정의한 서열을 이용하여 히스토그램 누적 분포를 추정한다.

$$\Phi_n = \frac{R_{\text{new}}(s_n) - 0.5}{K}, \quad K = N + M \quad (4)$$

6. 추정된 누적분포  $\Phi_n$ 과 표준 정규 분포의 누적 분포 역함수  $C_{\text{ref}}^{-1}$ 를 이용하여 변환된 특징계수 열  $T(S)$ 를 얻을 수 있다

$$T(s_n) = C_{ref}^{-1}(\Phi_n) \quad (5)$$

### 3. 실험 설계 및 결과

본 실험은 ETRI에서 배포한 한국어 중가 마이크 화자인식용 음성 데이터베이스를 사용하였다. UBM을 위하여 월차 100명, 3개월차 50명의 첫 번째 시차 발생 중에서 첫 번째 회차의 10 ~ 19번 문장 발성을 사용하였다. 화자 식별 실험을 위해 주차 100명의 첫 번째 시차 발생 중에서 첫 번째 회차의 10 ~ 19번 문장 발성으로 화자 적용하였으며, 화자 적용 데이터 녹음 시점으로부터 1주차, 2주차, 3주차에 녹음된 첫 번째 회차의 10 ~ 19번 문장 발성으로 화자 식별 실험을 진행하였다. 데이터베이스의 표본화 주파수는 16kHz이다.

본 실험은 음성특징으로 18차 MFCC를 사용하였다. Hamming 윈도우 크기는 25ms, 이동주기는 10ms이다.

화자 모델 학습에는 GMM-UBM[3] 방법을 사용하였다. UBM은 혼합 수 128개의 GMM을 사용하였고, UBM으로부터 MAP 적용 ( $\tau = 1$ )을 통하여 화자모델을 구성하였다.

표 1은 주차별 화자 식별 오류율이며, 표 2는 다른 채널 보상 방법 대비 제안한 방법의 상대적 오류 감소율이다. MFCC는 채널 보상을 하지 않은 방법이다. CHEQ[4]는 누적 기반의 히스토그램 등화기법이며, OHEQ[4]는 순서 기반의 히스토그램 등화기법이다. 또, CHEQ를 위한 bin의 개수는 1,000개로 하였다. 1주차, 2주차, 3주차 데이터의 화자 식별 결과를 각각 WEEK 1, WEEK 2, WEEK 3으로 표시하였다.

표 1. 주차별 화자 식별 오류율

보상방법	오류율 (%)		
	WEEK 1	WEEK 2	WEEK 3
baseline	2.1	3.2	3.9
CMN	1.8	4.2	4.4
MVN	1.7	4.1	4.3
CHEQ	2.4	4.3	5.5
OHEQ	1.9	3.7	4.9
Proposed	<b>1.6</b>	<b>3.1</b>	<b>3.3</b>

표 2. 제안한 방법의 상대적 오류 감소율

보상방법	오류 감소율 (%)		
	WEEK 1	WEEK 2	WEEK 3
baseline	23.8	3.1	15.3
CMN	11.1	26.2	25
MVN	5.9	24.3	23.3
CHEQ	33.3	27.9	40
OHEQ	15.7	16.2	32.7

실험결과, 제안한 방법은 baseline보다 평균적으로 14.1%의 상대적 오류 감소율을 보였다. 특히 CHEQ와 OHEQ는 다른 채널 보상 방법보다 성능이 떨어지는 것을 확인할 수 있다. 학습과 테스트 환경이 같으므로 채널 보상 방법을 적용하면, baseline보다 성능이 떨어지는 현상을 볼 수 있다. 하지만 제안한 방법을 이용하였을 때 baseline보다 성능이 향상되는 것을 볼 수 있다. 실제 환경에서 제안한 방법을 사용하였을 때 더 높은 화자 식별 성능을 기대할 수 있다.

### 4. 결론

본 논문에서는 UBM 데이터를 이용한 개선된 히스토그램 등화 기법을 제안하였다. 이 방법은 음성 특징의 서열과 UBM에 대한 음성 특징의 서열을 이용하여 음성 특징을 비선형 변환하였다. 기존의 히스토그램 등화 기법은 발화 길이가 짧은 음성에서는 다른 채널 보상 기법보다 성능이 떨어지나, 제안한 방법의 히스토그램 등화 기법은 다른 채널 보상 기법보다 화자 식별 성능이 향상되었다.

### 감사의 글

본 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국 연구재단의 기초연구사업 지원을 받아 수행되었습니다. (2010-0024047)

### 참고문헌

- [1] Atal. B. S. (1974). "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *Journal of the Acoustical Society of America*, vol. 55, No. 6, pp. 1304-1312
- [2] Viikki. O. and Laurila. K. (1998). "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Communication*, Vol 25, pp. 133-147
- [3] Reynolds. D. A., Quatieri. T. F. and Dunn. R. B. (2000). "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, Vol. 10, pp. 19-41
- [4] Skosan. M. and Mashao. D. (2006). "Modified segmental histogram equalization for robust speaker verification," *Pattern Recognition Letters*, Vol 27, No. 5, pp. 479-486