

# 다수 투표 기반의 화자 식별을 위한 배경 화자 데이터의 퍼지 C-Means 중심을 이용한 히스토그램 등화기법

## Histogram Equalization Using Centroids of Fuzzy C-Means of Background Speakers' Utterances for Majority Voting Based Speaker Identification

김명재, 양일호, 유하진<sup>†</sup>

(Myung-Jae Kim, IL-Ho Yang, and Ha-Jin Yu<sup>†</sup>)

서울시립대학교 컴퓨터과학부

(접수일자: 2013년 10월 23일; 채택일자: 2013년 11월 22일)

**초록:** 이전 연구에서 퍼지 C-Means의 중심 데이터로 이루어진 보조 데이터를 이용한 히스토그램 등화기법을 제안하였다. 보조 데이터를 이용한 히스토그램 등화기법은 사용하는 참조 집합의 크기에 따라 화자 식별 성능에 영향을 받는다. 그러나 인식 시점에서 최적의 파라미터를 찾기는 어렵다. 이 문제를 해결하기 위해 본 논문에서는 화자 식별을 위한 다수 투표 방식에 기반을 둔 보조 데이터를 이용한 히스토그램 등화기법을 제안한다. 다수 투표 기반의 제안한 방법은 여러 종류의 보조 데이터를 이용한 히스토그램 등화기법으로 입력 음성을 분류한다. 본 연구에서 제안한 방법을 CMN(Cepstral Mean Normalization), MVN(Mean and Variance Normalization), HEQ(Histogram Equalization)와 같은 기존의 특징 정규화 방법 및 보조 데이터를 이용한 히스토그램 등화기법과 비교한다.

**핵심용어:** 화자인식, 화자식별, 히스토그램 등화기법, 다수 투표

**ABSTRACT:** In a previous work, we proposed a novel approach of histogram equalization using a supplement set which is composed of centroids of Fuzzy C-Means of the background utterances. The performance of the proposed method is affected by the size of the supplement set, but it is difficult to find the best size at the point of recognition. In this paper, we propose a histogram equalization using a supplement set for majority voting based speaker identification. The proposed method identifies test utterances using a majority voting on the histogram equalization methods with various sizes of supplement sets. The proposed method is compared with the conventional feature normalization methods such as CMN(Cepstral Mean Normalization), MVN(Mean and Variance Normalization), and HEQ(Histogram Equalization) and the histogram equalization method using a supplement set.

**Keywords:** Speaker recognition, Speaker identification, Histogram equalization, Majority voting

**PACS numbers:** 43.72.Fx

### 1. 서론

채널 불일치에 의한 화자 인식 시스템의 성능 저하를 방지하기 위해, 일반적으로 특징 정규화 방법을 사용한다. 기존의 특징 정규화 방법으로 캡스트

럼 평균 정규화(Cepstral Mean Normalization)<sup>[1]</sup> 평균-분산 정규화(Mean and Variance Normalization)<sup>[2]</sup> 등이 있다. 기존의 특징 정규화 방법은 간단한 선형 변환을 사용하여 특징을 정규화 한다. 선형 변환은 음성 특징에 선형적으로 작용하는 채널 잡음을 제거하는데 효과적이다. 그러나 비선형적으로 작용하는 잡음에는 효과적이지 못하다. 이를 해결하려는 노력으로 비선형적인 잡음 특성을 선형으로 근사하여 처리하

<sup>†</sup>Corresponding author: Ha-Jin Yu (hjyu@uos.ac.kr)  
School of Computer Science, University of Seoul, 163 Siripdaero,  
Dongdaemun-Gu, Seoul 139-743, Republic of Korea  
(Tel: 82-2-6490-2448, Fax: 82-2-6490-2444)

는 VTS(Vector Taylor Series)<sup>[3]</sup>와 SLA(Static Linear Approximation)<sup>[4]</sup> 등의 연구가 진행되었다. 이러한 접근 방법과 다른 접근 방법으로 히스토그램 등화기법(Histogram Equalization)<sup>[5]</sup>이 제안되었다. 히스토그램 등화기법은 디지털 이미지 프로세싱에서 디지털 이미지의 밝기와 대비를 조절하기 위해 사용되었다.<sup>[5]</sup> 히스토그램 등화기법은 이미지 프로세싱 뿐만 아니라, 음성 인식 분야<sup>[6-7]</sup>와 화자 인식 분야<sup>[10]</sup>에서 특징 정규화 방법으로 사용되었다. 또한, 히스토그램 등화기법의 응용방법으로 feature warping,<sup>[8]</sup> modified segmental HEQ<sup>[9]</sup>가 제안되었다. 히스토그램 등화기법은 누적 분포 함수(Cumulative Distribution Function)를 기반으로 입력 샘플의 비선형 변환을 가능하게 한다. 그러나 히스토그램 등화기법은 누적 분포 함수 추정을 위해 충분한 길이의 입력 샘플을 필요로 하므로, 입력 샘플의 길이가 짧은 경우 기존의 특징 정규화 방법보다 화자 식별 성능이 떨어질 수 있다. 이런 짧은 입력 샘플에 대한 누적 분포 함수의 강인한 추정을 위해, 이전 연구에서는 배경 화자 모델 학습 데이터에 퍼지 C-Means를 수행한 결과물인 클러스터 중심들을 보조 데이터로 이용한다.<sup>[13]</sup> 하지만 이 방법은 인식 시점에서 최적의 보조 데이터 수를 추정하기 어려운 문제점이 있었다. 이러한 문제점을 해결하기 위해, 본 논문에서는 다양한 길이의 보조 데이터를 이용하여 각기 히스토그램 등화기법을 수행하고, 다수 투표 기반으로 앙상블 결합함으로써 이전의 제안한 방법을 개선하고자 한다.

본 논문은 2장에서 히스토그램 등화기법에 대해 설명하고, 3장에서는 보조 데이터를 이용한 히스토그램 등화기법에 대해 설명한다. 4장에서 실험 설계 및 실험 결과를 보이고, 마지막으로 결론을 맺는다.

## II. 히스토그램 등화기법

히스토그램 등화기법의 목적은 주어진 확률 분포를 기준이 되는 확률 분포로 변환하는 것이다. 주어진 확률 분포를 변환하기 위해 누적 분포 함수값을 기준 분포의 동일한 누적 분포 함수 값으로 변환한다. 히스토그램 등화기법의 변환식은 다음과 같다.<sup>[9]</sup>

$$y = C_{ref}^{-1}(C_x(x)). \quad (1)$$

$x$ 는 입력된 샘플이며,  $y$ 는 변환된 샘플이다.  $C_x$ 는  $x$ 의 누적 분포 함수이며,  $C_{ref}^{-1}$ 는 기준 누적 분포 함수의 역함수이다.

히스토그램 등화기법의 구현 방법으로는 누적 히스토그램 기반과 순서 기반의 히스토그램 등화기법이 있다. 누적 히스토그램 등화기법의 누적 분포 함수 추정 과정은 다음과 같다.<sup>[9]</sup>

$d$  번째 차원, 길이  $N$ 으로 이루어진 입력 시퀀스  $X$ 를 다음과 같이 정의한다.

$$X = \{x_1, x_2, \dots, x_n, \dots, x_N\}. \quad (2)$$

입력된 시퀀스  $X$ 에서 최대값  $x_{max}$ 와 최소값  $x_{min}$ 을 찾는다.  $[x_{min}, x_{max}]$ 의 범위를 동일한 크기로  $M$  등분한다. 등분한 범위를  $B$ 라 하고, 다음과 같이 정의한다.

$$x_{min} = b_1 < b_2 < \dots < b_{M+1} = x_{max}. \quad (3)$$

$B_i$ 는  $[b_i, b_{i+1})$ 의 범위를 갖는다.  $B_i$ 를 빈(bin)이라 한다. 이렇게 얻어진 빈들을 가지고 히스토그램을 측정한다. 각 빈에 속하는 입력 시퀀스  $X$ 의 원소 개수를 세어 히스토그램을 측정한다. 다음과 같이 측정한 히스토그램을 정규화한다.

$$p_x(x \in B_i) = \frac{n_i}{N}. \quad (4)$$

$n_i$ 는 빈  $B_i$ 에 속하는 원소 개수이며,  $N$ 은 입력된 시퀀스의 전체 원소 개수이다.

정규화한 히스토그램을 이용하여 다음과 같이 누적 히스토그램을 계산한다.

$$C_x(x \in B_i) = \sum_{j=1}^i \frac{n_j}{N}. \quad (5)$$

누적 히스토그램을 이용한 입력 시퀀스의 변환은 식(1)을 이용한다.

순서 기반 히스토그램 등화기법의 누적 분포 함수 추정 과정은 다음과 같다.

$d$  번째 차원, 길이  $N$ 으로 이루어진 입력 시퀀스  $X$ 를 식(2)와 같이 정의한다. 입력 시퀀스  $X$ 의 오름차순 서열을 구한다.

$$x_{r(1)} \leq x_{r(2)} \leq \dots \leq x_{r(n)} \leq \dots \leq x_{r(N)}. \quad (6)$$

이때,  $r(n)$ 은  $n$ 번째 서열을 갖는 원소의 인덱스이다. 측정된 서열은 입력 시퀀스  $X$ 의 누적 분포 함수를 측정하는데 사용한다.

$$C_n = \frac{R_x(x_n) - 0.5}{N}. \quad (7)$$

$R_x(x_n)$ 은  $n$ 번째 원소의 오름차순 서열을 의미한다. 추정된 누적 분포 함수를 식(1)에 적용하여 입력 시퀀스를 변환한다.

본 논문에서는 기준 분포 함수로 표준 정규 분포를 사용하였다. 표준 정규 분포를 이용한 누적 분포 함수의 역변환은 누적 분포 함수 표를 사용하면 쉽게 구할 수 있다. 표준 정규 분포의 확률 밀도 함수는 다음과 같다.

$$p_x(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (8)$$

### III. 보조 데이터를 이용한 히스토그램 등화기법

보조 데이터를 이용한 히스토그램 등화기법<sup>[13]</sup>은 순서 기반 히스토그램 등화기법의 응용 방법이다. 보조 데이터를 만들기 위해 배경 화자 모델 학습을 위한 데이터에서 화자별로 퍼지 C-Means<sup>[11]</sup>를 수행한다. 각 화자의 퍼지 C-Means 중심을 모으고, 오름차순으로 정렬하여 보조 데이터를 구성한다. 여기서 소속 함수의 퍼지성을 결정하는 변수  $m$ 은 2로 설정하였다. 보조 데이터를 이용한 히스토그램 등화기법의 자세한 방법은 다음과 같다. 먼저, 배경 화자 모델 학습 데이터 중  $K$ 명 화자의 음성 데이터를 무작위로 선택한 집합을  $U$ 라 하자.

$$U = \{u(1), u(2), \dots, u(k), \dots, u(K)\}. \quad (9)$$

이때,  $u(k)$ 는 선택된  $k$ 번째 화자의 음성 데이터이며,  $D$  차원 특징 벡터의 관측 시퀀스이다.

각 화자의 음성 데이터  $u(k)$ 에 퍼지 C-Means를 수행하여 중심 벡터 집합  $F$ 를 얻는다.

$$F = \{f(1), f(2), \dots, f(k), \dots, f(K)\}. \quad (10)$$

이때,  $f(k)$ 는  $C$ 개의 퍼지 C-Means 중심 벡터의 집합으로 구성된다.  $F$ 의  $d$  번째 차원 원소를 모두 연결한 길이  $T(C \times K)$ 의 보조 데이터  $S$ 를 다음과 같이 구성한다.

$$S = \{s_1, s_2, \dots, s_p, \dots, s_T\}. \quad (11)$$

변환하고자 하는  $d$  번째 차원, 길이  $N$ 으로 이루어진 입력 시퀀스  $X$ 를 식(2)와 같이 정의한다.

입력 시퀀스  $X$ 의 오름차순 서열은 다음과 같이 구한다.

$$r_n = r_n^x + r_n^s. \quad (12)$$

이때,  $r_n^x$ 는 시퀀스  $X$ 를 기준으로 한 입력 시퀀스  $X$ 에 속하는  $n$ 번째 원소의 오름차순 서열이다.  $r_n^s$ 는 시퀀스  $S$ 를 기준으로 한 입력 시퀀스  $X$ 에 속하는  $n$ 번째 원소의 오름차순 서열이다. 즉,  $S$  내에서  $X$ 의  $n$ 번째 원소가 오름차순으로 정렬되었을 때 몇 번째인지를 나타낸다. 새로운 서열은 두 오름차순 서열의 합으로 정하며, 새로운 서열을 이용하여 누적 분포 함수를 추정한다.

$$C_n = \frac{r_n - 0.5}{T + N}. \quad (13)$$

새롭게 추정된 누적 분포 함수를 식(1)에 적용하여 입력 시퀀스를 변환한다.

## IV. 실험 설계 및 결과 분석

### 4.1 다수 투표

본 논문에서는 분류기 앙상블 결합 방법으로 다수 투표<sup>[14]</sup>를 사용하였다.  $M$ 명의 화자 모델이 존재할 때, 부류 표지 벡터  $L$ 은 다음과 같이 정의할 수 있다.

$$L = (l_1, l_2, \dots, l_m, \dots, l_M). \quad (14)$$

이때,  $l_m \in \{0, 1\}$ 은  $m$  번째 화자의 선택 유무이다. 또,  $E$ 개의 분류기가 존재한다고 가정하면,  $e$  번째 분류기의 출력은 다음과 같다.

$$L_e = (l_{e1}, l_{e2}, \dots, l_{em}, \dots, l_{eM}). \quad (15)$$

$l_{em}$ 은  $e$  번째 분류기에서  $m$  번째 화자의 선택 유무를 나타낸다. 이 부류 표지 벡터를 이용한 다수 투표의 분류 방법은 다음과 같다.

$$q = \arg \max_j \sum_{e=1}^E l_{ej} \quad (j = 1 \dots M), \quad (16)$$

### 4.2 데이터베이스

제안한 방법의 성능 평가를 위해 ETRI(Electronics and Telecommunications Research Institute)에서 배포한 한국어 중가 마이크 화자인식용 음성 데이터베이스를 이용하였다. ETRI 데이터베이스는 조용한 사무실 환경에서 수집되었으며, 주차 100명, 월차 100명 3개월차 50명, 총 250명의 화자로 구성되어 있다. 각 화자는 2연 숫자, 4연 숫자, 문장 발성을 수행하였으며, 기간별로 4회 시차 발성과 5회 반복발성을 수행하였다. ETRI 데이터베이스는 16 kHz 표본화 주파수로 녹음되었으며, 본 연구에서는 VoIP 코덱 적용을 위해 8 kHz 표본화 주파수로 변환하였다.

배경 화자 학습을 위해 월차 100명과 3개월차 50명의 1시차 1회차 10~19 번 문장 발성을 이용하였다. 배경 화자 학습에 사용한 총 발화 수는 1500개(150명 × 10문장)이다.

화자 모델 적응을 위해 주차 100명의 1시차 1회차

10~19 번 문장 발성을 이용하였다. 화자 모델 적응을 위해 사용한 총 발화 수는 1000개(100명 × 10문장)이다.

시스템 성능을 평가하기 위해, 주차 100명의 2, 3, 4 시차 1회차 10~19 번 문장 발성을 사용하였다. 평가를 위해 사용한 총 발화 수는 3000개(3 시차 × 100 명 × 10 문장)이다. 테스트 데이터는 VoIP 코덱을 이용하여 시뮬레이션 되었다.

### 4.3 특징 추출 및 화자 모델 학습

본 연구에서는 18차 MFCC(Mel Frequency Cepstral Coefficients)를 사용하였다. MFCC 특징 추출을 위해 pre-emphasis filter( $\alpha=0.97$ )를 수행하고, 25 ms 크기의 윈도우로 10 ms 마다 분할하였다. 각 프레임에는 해밍 윈도우가 적용되었고, 26개의 mel 필터를 거쳐 이산 코사인 변환을 수행하였다.

GMM-UBM<sup>[12]</sup> 방법으로 화자 모델을 학습하였다. 화자 모델은 128개 혼합 수의 배경 화자 모델로부터 MAP(Maximum A Posteriori)<sup>[12]</sup> 적응 방법을 사용하여 화자 모델을 구성하였다. ( $\tau=1$ )

### 4.5 사용한 VoIP 코덱

화자 식별 실험을 위해 사용한 테스트 데이터는 VoIP 코덱으로 시뮬레이션 되었다. 사용한 VoIP 코덱은 G.729,<sup>[15]</sup> SILK,<sup>[16]</sup> Speex<sup>[17]</sup>이다. Table 1은 사용한 코덱의 상세정보를 보여준다.

### 4.4 실험 결과

제안한 방법(proposed)의 성능을 평가하기 위해 캡스트럼 평균 정규화(CMN), 평균-분산 정규화(MVN), 누적 히스토그램 기반 히스토그램 등화기법(CHEQ, Cumulative histogram based HEQ), 순서 기반 히스토그램 등화기법(OHEQ, Order statistics based HEQ), 보조 데이터를 이용한 히스토그램 등화기법(SHEQ(FCM),

Table 1. Specification of the codecs used in the experiments.

Codec	Organization	Sampling rate	Version
G.729	ITU-T	8 kHz	3.3
SILK	Skype	8, 12, 16, 24 kHz	1.0.7
Speex	Xiph.org	8, 16, 32 kHz	1.2rc1

Supplement set based HEQ(Fuzzy C-Means)]과 비교하였다. 누적 히스토그램기반 히스토그램 등화기법을 위해 1000개의 빈을 사용하였고, 보조 데이터를 이용한 히스토그램 등화기법을 위해 배경 화자 모델 학습 데이터에서 10명의 화자를 무작위로 선택하였다. 보조 데이터를 이용한 히스토그램 등화기법은 선택한 화자의 음성 데이터로부터 퍼지 C-Means의 중심 수를 늘려가며 실험을 수행하였다.

다수 투표 기반의 제안한 방법은 보조 데이터를 이용한 히스토그램 등화기법에서 사용한 9개 분류기(화자 당 중심개수: 10, 15, 20, 25, 30, 35, 40, 45, 50)를 다수 투표 방식으로 결합하여 화자 식별 오류율을 측정하였다.

Fig. 1은 G.729 환경에서 화자 식별 실험을 수행한 결과를 보여준다. 다수 투표 기반의 제안한 방법은

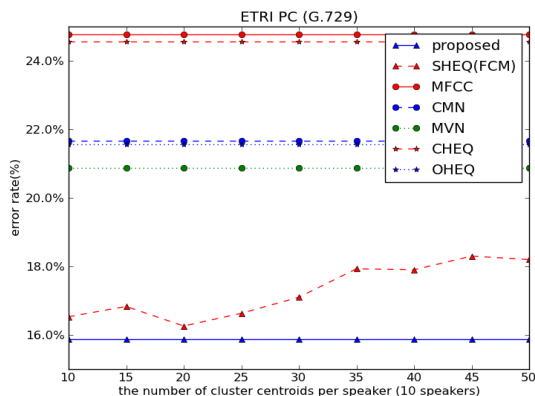


Fig. 1. Speaker identification error rates in G.729 environment.

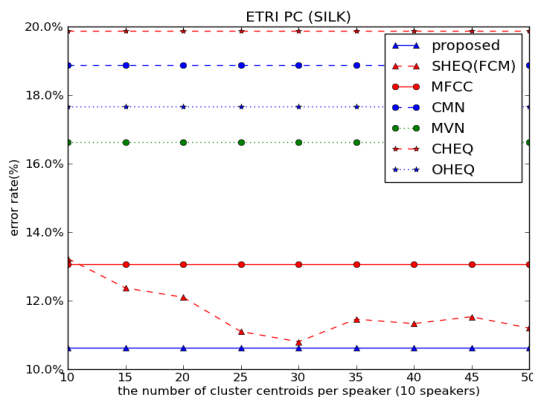


Fig. 2. Speaker identification error rates in SILK environment.

15.9%의 화자 식별 오류율을 보였다. 보조 데이터를 이용한 히스토그램 등화기법은 16.27%~18.2%의 화자 식별 오류율을 보였으며, 평균 화자 식별 오류율은 17.3%였다. 다수 투표기반의 제안한 방법은 기존의 특징 정규화 방법보다 상대 오류가 24%(MVN)~35.9%(MFCC) 감소하였다. 보조 데이터를 이용한 히스토그램 등화기법이 최고 성능을 가질 때보다 2.5%상대오류가 감소하였고, 평균 성능에 비해 8.4%상대 오류가 감소하였다.

Fig. 2는 SILK 환경에서 화자 식별 실험을 수행한 결과를 보여준다. 다수 투표 기반의 제안한 방법은 10.6%의 화자 식별 오류율을 보였다. 보조 데이터를 이용한 히스토그램 등화기법은 10.8%~12.4%의 화자 식별 오류율을 보였고, 평균 성능은 11.7%였다. 제안한 방법은 기존의 특징 정규화 방법에 비해 상대 오류가 18.6%(MFCC)~46.5%(CHEQ) 감소하였다. 보조 데이터를 이용한 히스토그램 등화기법이 가장 좋은 성능을 가질 때, 상대 오류가 1.5%감소하였고, 평균 성능을 가질 때, 상대 오류가 9%감소하였다.

Fig. 3은 Speex 환경에서 화자 식별 실험을 수행한 결과이다. 제안한 방법은 16.4%의 화자 식별 오류율을 보였다. 보조 데이터를 이용한 히스토그램 등화기법은 15.1%~21.1%의 화자 식별 오류율을 나타내며, 평균 성능은 17.9%였다. 제안한 방법은 다른 특징 정규화 방법에 비해 상대 오류가 48.4%(MFCC)~10%(MVN) 감소하였다. 보조 데이터를 이용한 히스

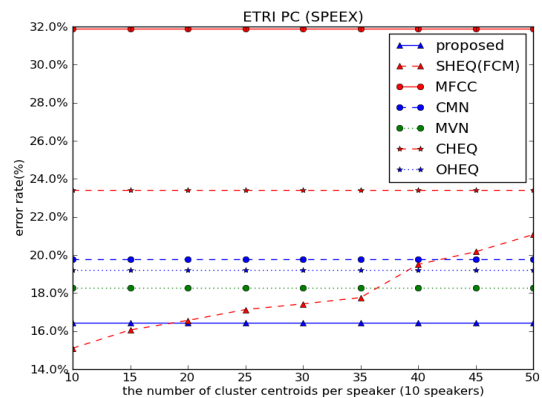


Fig. 3. Speaker identification error rates in Speex environment.

토그램 등화기법이 최고 성능을 가질 때, 제안한 성능의 상대 오류가 8.8% 늘었지만, 평균 성능에 비해 상대 오류가 8% 감소하였다. 실시간 인식의 경우 최고 성능을 갖는 파라미터를 찾기 어렵기 때문에, 다수 투표를 기반으로 한 방법을 적용하여 이러한 문제를 해결할 수 있음을 확인하였다.

## V. 결 론

본 논문에서는 다수 투표 기반의 히스토그램 등화기법을 제안하였다. 다수 투표 시스템을 이루는 분류기는 이전에 제안한 보조 데이터를 이용한 히스토그램 등화기법에서 얻은 화자 모델로 구성하였다. 각 화자 모델은 다양한 크기의 보조 데이터를 사용하여 히스토그램 등화기법을 수행하고, 화자 적응 방법을 이용하여 구성하였다. 채널 불일치 환경을 시뮬레이션 하기 위해 테스트 데이터에 VoIP 코덱을 적용하였다. 따라서 학습 데이터는 클린 환경의 음성 발화이고, 테스트 데이터는 VoIP 환경의 음성 발화이다. 다수 투표 기반의 제안한 방법은 기존의 특징 정규화 방법보다 성능이 향상되었고, 이전에 제안한 보조 데이터를 이용한 히스토그램 등화기법보다도 성능이 향상되었다.

Speex 환경에서는 이전에 제안한 방법이 최고 성능을 가질 때보다 성능이 하락하였지만 평균 성능 보다는 상승하였다. 성능 하락의 원인은 보조 데이터를 이용한 히스토그램 등화기법의 성능 하락 구간에 위치하는 여러 분류기를 포함하여 제안한 방법의 성능 향상 폭이 크지 않았던 것으로 생각된다.

추후, 더 다양한 분류기를 포함하는 투표 시스템을 구성하고, 분류기의 가중치를 고려하는 앙상블 시스템을 구성하면 추가적인 성능 개선이 가능할 것으로 기대된다.

## 감사의 글

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

## References

1. B. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Am.* **55**, 1304 (1974).
2. O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Commun.* **25**, 133-147 (1998).
3. P. J. Moreno, B. Raj, and R. M. Stern, "A vector taylor series approach for environment-independent speech recognition," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.* **2**, 733-736 (1996).
4. N. S. Kim, "Statistical linear approximation for environment compensation," *IEEE Signal Proc. Lett.* **5**, 8-10 (1998).
5. R. C. Gonzalez and P. wintz, *Digital Image Processing* (Addison-Wesley Publishing Company, Reading, 1987), pp. 275-281.
6. J. C. Segura, C. Benitez, A. De La Torre, A. J. Rubio, and J. Ramirez, "Cepstral domain segmental nonlinear feature transformations for robust speech recognition," *IEEE Signal Proc. Lett.* **11**, 517-520 (2004).
7. A. de la Torre, A. M. Peinado, J. C. Segura, J. L. Perez-Cordoba, M. C. Benitez, and A. J. Rubio, "Histogram equalization of speech representation for robust speech recognition," *IEEE Trans. Speech Audio Proc.* **13**, 355-366 (2005).
8. J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. Speaker Odyssey*, 213-218 (2001).
9. M. Skosan and D. Mashao, "Modified segmental histogram equalization for robust speaker verification," *Pattern Recognit. Lett.* **27**, 479-486 (2006).
10. M. Skosan and D. Mashao, "Matching feature distributions for robust speaker verification," in *Proc. PRASA*, 42-47 (2004).
11. R. O. Duda and P. E. Hart, D. G. Stork, *Pattern Classification* (John Wiley & Sons, New York, 2012), pp. 528-530.
12. D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digit. Signal Process.* **10**, 19-41 (2000).
13. M. Kim, I. Yang, and H. Yu, "Histogram Equalization Using Centroids of Fuzzy C-Means of Background Speakers' Utterances for Speaker Identification," in *Proc. Stat. Lang. and Speech Proc.* **1**, 143-151 (2013).
14. I. Oh, *Pattern Recognition*, (in korean, Kyobobook, Seoul, 2008), pp. 438-441.
15. G.729 [online], <http://www.itu.int/rec/T-REC-G.729-200701-S/en>.
16. SILK [online], <https://developer.skype.com/silk>.
17. Speex [online], <http://www.speex.org/>.

## 저자 약력

### ▶ 김 명 재(Myung-Jae Kim)



2010년 8월: 서울시립대학교 컴퓨터과학부  
학사  
2010년 9월 ~ 현재: 서울시립대학교 컴  
퓨터과학부 석·박사통합과정

### ▶ 양 일 호(IL-Ho Yang)



2008년 2월: 서울시립대학교 컴퓨터과학부  
학사  
2010년 2월: 서울시립대학교 컴퓨터과학부  
석사  
2010년 3월 ~ 현재: 서울시립대학교 컴퓨터  
과학부 박사과정

### ▶ 유 하 진(Ha-Jin Yu)



1990년 2월: KAIST 전산학과 학사  
1992년 2월: KAIST 전산학과 석사  
1997년 2월: KAIST 전산학과 박사  
1997년 ~ 2000년: LG 전자 전자기술원 선임  
연구원  
2000년 ~ 2002년: SL2(주) 연구소장  
2002년 ~ 현재: 서울시립대학교 컴퓨터  
과학부 교수