

Eigenvoice 기반 화자가중치 거리측정 방식을 이용한 화자 분할 시스템

Speaker Segmentation System Using Eigenvoice-based Speaker Weight Distance Method

최무열 · 김형순

(Mu Yeol Choi and Hyung Soon Kim)

부산대학교 전자공학과

(접수일자: 2012년 3월 6일; 채택일자: 2012년 4월 23일)

초 록: 화자 분할 기술은 오디오 데이터로부터 자동적으로 화자 경계 구간을 검출하는 것이다. 화자 분할 방식은 화자에 대한 선행 지식 사용 여부에 따라 거리기반 방식과 모델기반 방식으로 나누어진다. 본 논문에서는 eigenvoice 기반의 화자가중치 거리를 이용한 화자 분할 방식을 도입하고, 이 방식을 대표적인 거리 기반 방식들과 비교한다. 또한, 화자가중치의 거리 측정 함수로 유클리드 거리와 cosine 유사도를 사용하여 화자 분할 성능을 비교하고, eigenvoice 방식에 의해 화자 적응된 모델들 사이의 직접적인 거리를 이용한 화자 분할 방식과의 비교를 통해 화자가중치 거리를 이용한 방식이 계산량면에서 효율적인 점을 검증한다.

핵심용어: 화자 분할, 화자경계 검출, Eigenvoice, 화자적응

투고분야: 음성처리 분야(2.5)

ABSTRACT: Speaker segmentation is a process of automatically detecting the speaker boundary points in the audio data. Speaker segmentation methods are divided into two categories depending on whether they use a prior knowledge or not: One is the model-based segmentation and the other is the metric-based segmentation. In this paper, we introduce the eigenvoice-based speaker weight distance method and compare it with the representative metric-based methods. Also, we employ and compare the Euclidean and cosine similarity functions to calculate the distance between speaker weight vectors. And we verify that the speaker weight distance method is computationally very efficient compared with the method directly using the distance between the speaker adapted models constructed by the eigenvoice technique.

Key words: Speaker segmentation, Speaker change detection, Eigenvoice, Speaker adaptation

ASK subject classification: Speech Signal Processing (2.5)

1. 서 론

화자 분할 시스템의 목적은 방대한 양의 축적된 오디오 데이터베이스의 검색과 분류를 위해 스트림 형태의 음성신호 내에 존재하는 화자변화 구간을 검출하는 것이다. 화자 분할 기술은 주로 두 가지 방식으로 연구되고 있는데 하나는 거리기반 분할 방식이

고, 다른 하나는 모델 기반 분할 방식이다^[1]. 거리 기반 분할 방식은 음성 신호를 따라서 이동하는 인접한 두 분석 윈도우 간의 유사도를 측정하는데 Bayesian Information Criterion (BIC)^[2]과 Kullback Leibler (KL)^[3]와 같은 유사도 함수를 사용한다. 이 방식은 입력된 음성 신호만을 사용하기 때문에 화자에 대한 선행 지식을 필요로 하지 않는다는 장점이 있으나 비교적 긴 분석 윈도우가 필요하다는 단점이 있다. 반면 모델 기반 방식은 데이터의 선행지식을 이용하여 화자 모델을 미리 구축하기 때문에 짧은 구간의 화자경계

*Corresponding author: 김형순 (kimhs@pusan.ac.kr)
609-735 부산시 금정구 장전동 산 30 부산대학교 전자공학과
(전화: 051-510-2452; 팩스: 051-515-5190)

점 검출 성능이 비교적 높다는 장점이 있다. 화자 모델을 구축하기 위해서 Universal Background Model (UBM), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), 그리고 eigenvoice^[4] 방식등이 사용되며 특히 eigenvoice 방식은 효과적인 화자모델링 방식으로 화자 분할 시스템에 성공적으로 적용되어 비교적 높은 화자경계점 검출 성능을 나타낸다.

Eigenvoice를 화자 분할 기술에 적용한 기존 방식의 경우^[5] 추정된 화자가중치를 blind segmentation에 사용한 뒤 이를 통해 구축된 화자 모델간의 거리를 이용하여 화자 모델에 대한 가설을 검증한다. 화자 가설은 일정 길이의 데이터 내에 포함된 화자의 수를 추정하고 그에 따라 화자 모델을 만드는 것이다. 이 경우 화자의 수와 데이터의 길이 사이에 trade-off가 발생한다. 즉, 일정 길이의 데이터 내에 2-3명의 화자가 존재하는 경우 비교적 높은 성능을 나타내지만 다수의 화자가 존재할 경우 화자 모델링 성능의 저하로 인해 화자경계점 검출 성능도 현저히 떨어지는 결과를 나타낸다. 다양한 방송 뉴스 환경에서는 서로 다른 화자의 짧고 빈번한 발화 경계가 나타날 가능성이 크기 때문에 화자수를 제한하기는 어렵다. 따라서 eigenvoice 방식을 사용하는 다중 화자 분할 기술에서는 입력 데이터 기반의 화자모델링 방식을 사용하기보다 화자특성을 표현하는 화자가중치 사이의 거리를 사용하는 것이 효과적이다.

따라서 본 논문에서는 eigenvoice 기반의 화자가중치 거리를 이용한 화자 분할 방식을 도입하고, 이 방식을 대표적인 거리 기반 방식인 BIC와 KL 방식과 비교한다. 또한, 화자가중치의 거리 측정 함수로 유클리드 거리와 cosine 유사도를 사용하여 화자 분할 성능을 비교하고, eigenvoice 적응 방식에 의해 화자 적응된 모델들 사이의 직접적인 거리를 이용한 화자 분할 방식과의 비교를 통해 화자가중치 거리를 이용한 방식이 계산량면에서 효율적인 점을 검증한다.

II. 화자가중치 거리를 이용한 eigenvoice 기반의 화자분할 방식

2.1 Eigenvoice 화자 가중치 추정

Eigenvoice 화자 적응 방식에서는 각 화자들 간의

변동을 가장 잘 대표하는 부공간 (Subspace)의 기저 벡터 (Basis vector)를 설정하고 새로운 화자에 대하여 기저벡터 성분의 가중치를 추정한다. 그림 1은 eigenvoice 방식의 전체적인 구조를 나타낸 것이다.

우선, 그림 1의 off-line 단계에서는 T 개의 잘 훈련된 화자중속 모델을 구성한 뒤 모델의 평균 벡터를 연결하여 D 차원의 수퍼벡터 (Supervector)를 만든다. 이 수퍼벡터를 이용하여 구한 분산행렬에 Principal Component Analysis (PCA)를 적용하여 D 차원을 가지는 T 개의 eigenvector를 구하면 그것이 바로 “eigenvoice”가 된다. T 개의 eigenvector를 eigenvalue가 큰 순서대로 정렬하여 그 중 K 개의 eigenvoice를 선택하면, $e(1)...e(K)$ 만으로 전체 화자 모델의 변동을 대표할 수 있게 된다($K < T \ll D$). 이와 같이 선택된 K 개의 eigenvoice는 “eigenspace”를 생성한다. 이것을 벡터-행렬식으로 표현하면 다음과 같다.

$$\mu_s = \mu_{SI} + E w \tag{1}$$

여기서 E 와 w 는 각각 K 개의 eigenvoice로 이루어진 행렬과 추정된 화자가중치(Speaker weight) 벡터이고, μ_{SI} 와 μ_s 는 각각 훈련 화자들로 구축된 SI 모델의 수퍼벡터와 eigenvoice 방법으로 적용된 새로운 화자의 수퍼벡터이다.

그림 2의 on-line 단계는 새로운 화자의 적응 데이터를 이용하여 화자가중치를 구하는 과정으로 Maximum Likelihood Eigen-Decomposition (MLE^[6])을 이용한다. MLED의 결과식을 벡터-행렬 형태로

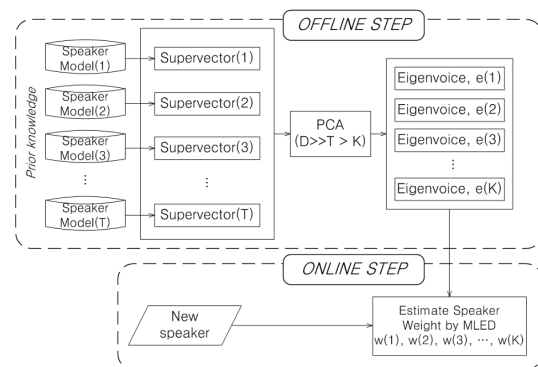


그림 1. Eigenvoice 화자적응 시스템
Fig. 1. Eigenvoice speaker adaptation system.

나타내면 다음과 같다.

$$\mathbf{w} = \mathbf{A}^{-1} \mathbf{y} \quad (2)$$

여기서 행렬 \mathbf{A} 와 \mathbf{y} 의 원소는 다음과 같다.

$$y_k = \sum_{m=1}^M \sum_{t=1}^N \gamma_m(\mathbf{o}_t) (\mathbf{e}_m(k))^T \mathbf{C}_m^{-1} \mathbf{o}_t \quad (3)$$

$$a_{j,k} = \sum_{m=1}^M \sum_{t=1}^N \gamma_m(\mathbf{o}_t) (\mathbf{e}_m(j))^T \mathbf{C}_m^{-1} \mathbf{e}_m(k) \quad (4)$$

여기서 M 은 Gaussian mixture 개수 이고, N 은 적응 데이터의 프레임 수이다. 그리고 $\mathbf{e}_m(k)$ 는 mixture m 에 해당하는 eigenvoice k 의 부벡터이고, \mathbf{C}_m 은 공분산 행렬이며, $\gamma_m(\mathbf{o}_t)$ 는 적응데이터 \mathbf{o}_t 가 주어졌을 때 시간 t 에서의 사후확률이다.

2.2 화자가중치 거리를 이용한 eigenvoice 기반의 화자 분할

서로 다른 두 음성 신호 구간에서 추정된 화자가중치는 eigenspace 상에서 두 좌표로 표시되는데 그 좌표들의 거리가 가까우면 두 음성 신호는 동일 화자의 것으로 보고 거리가 멀면 서로 다른 화자로 볼 수 있다. 이와 같은 특성을 이용하면 화자가중치 사이의 거리측정 만으로도 화자경계점 검출이 가능하다. 화자가중치 거리를 이용한 eigenvoice 방식의 화자분할 시스템을 그림 2에 나타냈다. 먼저, 훈련용 화자 데이터를 이용하여 화자모델을 구축하고 이를 통해 eigenvoice를 구한다. 복수 화자의 발화들이 임의의 순서대로 연이어 나타나는 스트림 형태로 구성된 테스트 데이터는 L 만큼의 길이를 갖는 분석 윈도우에 할당된다. 두 개의 인접한 분석 윈도우는 일정한 길이의 shift rate을 갖는 슬라이딩 윈도우 방식에 의해 이동되며 식 (2)에 의해 화자가중치 추정에 사용된다. 마지막으로, 추정된 화자가중치를 사용하여 두 분석 윈도우 사이의 유사도를 측정한다.

본 논문에서는 화자가중치를 이용한 유사도 측정 방식으로 다음 두 가지 거리측정 방식을 사용한다.

$$d(\mathbf{w}_{t-L}, \mathbf{w}_t) = \sqrt{\sum_{k=1}^K (w_{t-L}(k) - w_t(k))^2} \quad (5)$$

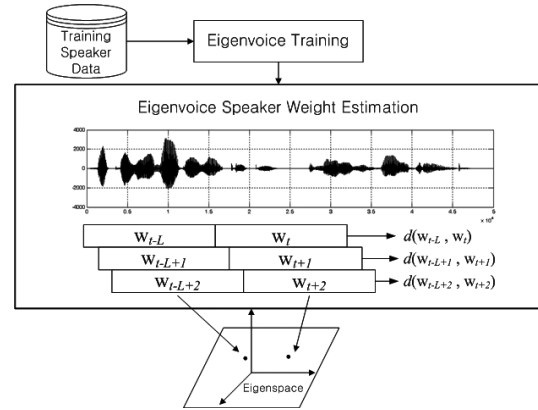


그림 2. 화자가중치 거리를 이용한 eigenvoice 기반의 화자분할

Fig. 2. Eigenvoice-based speaker segmentation using speaker weight distance.

$$s(\mathbf{w}_{t-L}, \mathbf{w}_t) = \frac{\mathbf{w}_{t-L}^T \mathbf{w}_t}{\|\mathbf{w}_{t-L}\| \|\mathbf{w}_t\|} \quad (6)$$

식 (5)는 L 만큼의 길이를 갖는 인접한 두 분석 윈도우로부터 추정된 화자가중치 사이의 유클리드 거리이고, 식 (6)은 코사인 (Cosine) 유사도이다.

또한, 식 (5)에 의한 유사도 측정방식이 화자 경계점 검출에 효과적인 방법임을 검증하기 위해 화자 적응된 모델들 사이의 직접적인 거리 측정 방법으로 적응 모델의 수퍼벡터 사이의 유클리드 거리를 다음 식과 같이 구한다.

$$d(\boldsymbol{\mu}_{t-L}, \boldsymbol{\mu}_t) = \sqrt{\sum_{i=1}^D (\mu_{t-L}(i) - \mu_t(i))^2} \quad (7)$$

여기서 $\boldsymbol{\mu}_t$ 는 화자가중치 \mathbf{w}_t 에 의해 식 (1)을 이용하여 화자 적응된 모델의 수퍼벡터이고 D 는 수퍼벡터의 차원이다.

2.3 화자 분할 성능 평가

서로 다른 두 음성 구간의 경계점인 시간 t 에서 화자 변화의 유무를 검출하기 위해 시간에 따른 유사도의 변화를 나타내는 거리곡선의 극대값을 이용한다. 화자가중치를 이용한 유사도 측정 결과로 나타난 거리곡선을 이용하여 화자경계점을 찾는 과정을 그림 3에 나타냈다. 그림의 세로 점선은 실제 화자의

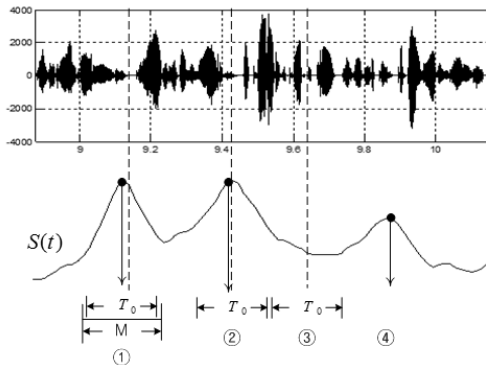


그림 3. 거리곡선에서의 극대값 검출과 화자 분할 성능 평가방법의 예
 Fig. 3. Peak picking method using the distance contour and the evaluation example of speaker segmentation.

경계점을 나타낸 것이다.

그림 3에서 보는 바와 같이 거리곡선 $S(t)$ 는 식 (5), (6), 또는 (7)로부터 구하며 식 (8)과 같은 조건을 만족시키는 극대값의 시간 τ 를 구하면 이것이 추정된 화자변화 경계점이 된다.

$$\tau = \underset{\{\tau - M/2 < t < \tau + M/2\}}{\operatorname{argmax}} S(t) \quad (8)$$

그림 3 및 식 (8)에서 M 은 거리곡선 $S(t)$ 의 극대값을 찾는 범위를 나타낸다. 식 (8)을 만족하는 극대값의 시간축 인덱스인 τ 가 결정되면 실제 화자 경계점을 기준으로 T_0 크기의 tolerance 구간 사이에서 τ 가 존재하지 않는 경우 missed detection이 발생하고, 반대로 τ 를 기준으로 T_0 크기의 tolerance 구간 사이에서 실제 화자 경계점이 발견되지 않으면 false alarm이 발생한다고 판단한다. 그림에서 보는 바와 같이 검출된 극대값으로 나타낸 흑점이 T_0 크기의 tolerance 구간 내에 있는 ①과 ②는 정확한 화자경계점을 검출한 예가 되며, ③은 missed detection이 발생한 경우이고, ④는 false alarm이 발생한 경우가 된다.

III. 실험 및 결과

본 논문에서는 화자분할 성능 평가를 위해 DARPA TIMIT^[7] 코퍼스를 사용하였다. TIMIT은 총

6300문장으로 구성되어 있으며 630명의 화자가 10개의 문장을 발성한 것으로 훈련용 데이터와 테스트용 데이터로 나누어져 있다. 본 논문에서는 462명의 훈련용 화자로부터 무작위로 200명의 화자를 선택하여 훈련에 사용하였다. 선택된 200명의 화자 데이터를 이용하여 GMM 기반의 화자독립모델을 만들었으며, Maximum A Posteriori (MAP) 기반 적응 기술을 이용하여 200명의 화자종속모델을 만들었다.

Eigenvoice 훈련시 화자가중치의 차원은 3~30차원의 범위를 설정해 실험한 후 최적의 차원을 선택하였다. 음성 특징 파라미터는 12차 static MFCC와 프레임 에너지를 포함하여 1차 미분한 26차 MFCC 파라미터를 사용하였으며 화자모델의 GMM mixture 개수는 64개를 사용하였다.

테스트 데이터는 총 168명의 훈련에 사용되지 않은 테스트용 화자가 발성한 1344문장으로 구성되어 있으며 테스트 데이터 내에 동일 화자의 문장이 연속해서 나오지 않도록 고려하면서 총 1343개수의 화자경계점이 나타나도록 구성하였다. 발성 문장의 길이로 살펴본 통계를 표 1에 나타냈다.

화자분할 성능 비교를 위한 거리 기반의 BIC 방식과 KL 방식은 다음과 같이 구하였다. BIC 방식의 1단계에는 3초 길이의 분석 윈도우를 사용하여 윈도우 간격을 1초씩 이동하면서 거리를 구하였고 1단계에서 화자경계가 검출되었다고 판단되면, 2단계에서는 2초 길이의 분석 윈도우를 사용하여 1단계에서 검출된 경계점을 기준으로 양쪽 0.5초 구간을 100ms 간격으로 이동하면서 총 1초 구간 동안 BIC 거리를 계산하였다. 또한 BIC 2단계의 분석 윈도우 길이를 3초, 2초, 1초로 변화시켜 그 성능 변화를 비교하였다. KL 방식에서는 3초, 2초, 1초 길이의 분석 윈도우를 사용하여 먼저 Linear Discriminant Analysis (LDA)^[8]를 수행한 다음 KL 거리를 계산하였다.

화자분할 성능 평가를 위해 Receiver Operating

표 1. 발성 길이에 따른 테스트용 데이터베이스의 문장 개수
 Table 1. The number of utterances by their length in test database.

Duration(sec)	1~2	2~3	3~4	4~5	over 5
# of sentence	181	654	344	131	34

Characteristic (ROC) 곡선을 사용하였으며 ROC 곡선은 missed detection rate (MDR)과 false alarm rate (FAR)을 통해 나타낸다. FAR과 MDR 값을 결정하는 파라미터는 그림 3에 나타낸 극대값을 찾는 구간 M 으로 이 값이 작으면 높은 FAR을 나타내고 값이 크면 낮은 FAR를 나타낸다. 사용한 M 값은 최소 0.6초에서부터 0.2초 간격으로 최대 4초의 구간을 갖도록 변화시키면서 FAR과 MDR을 계산하였다. 또한 화자경계점을 기준으로 양쪽에 각각 0.5초의 구간을 갖는 총 1초 길이의 tolerance 를 허용하였다^[9].

분석 윈도우의 길이를 3초, 2초, 그리고 1초로 줄여가며 BIC, KL 방식과 화자가중치의 유클리드 거리를 이용한 eigenvoice 방식의 화자분할 성능을 그림 4에 나타냈다. 거리 기반 방식에서는 분석 윈도우 길이 2초인 경우가 3초와 1초인 경우보다 높은 성능을 보이는 것을 확인할 수 있다. 이것은 분석 윈도우 길이가 길수록 거리 기반 방식에서는 화자 특징을 충분히 얻을 수 있으나 분석 윈도우 구간 내에 화자경계가 포함될 가능성도 동시에 커지기 때문에 항상 높은 성능을 보이지 않는다는 사실을 보여준다. 이것은 eigenvoice 방식의 분석 윈도우 3초의 결과에서도 동일하게 나타났다. 따라서 거리 기반 방식은 적절한 분석 윈도우 길이를 설정할 경우 상대적으로 높은 성능을 나타낸다는 것을 확인할 수 있다. 또한

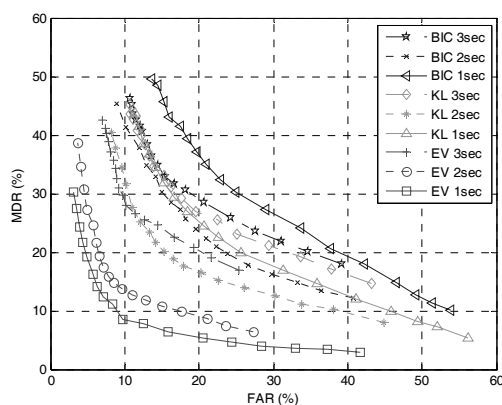


그림 4. BIC, KL, 그리고 유클리드 거리를 이용한 eigenvoice 기반의 분석 윈도우 길이에 따른 화자분할 성능 ROC 곡선

Fig. 4. ROC curves for speaker segmentation according to the analysis window of three different lengths by BIC, KL, and eigenvoice-based Euclidean distance method.

본 논문에서는 분석 윈도우 길이별 비교에서 KL 방식이 BIC 방식보다 높은 화자경계 검출률을 나타냈다. 한편, 모든 분석 윈도우 길이에 대해 eigenvoice 방식이 거리 기반 분할 방식보다 우수한 성능을 나타냈는데 특히 분석 윈도우가 짧은 2초, 그리고 1초 길이를 갖는 경우 매우 높은 성능을 나타냈다. 이는 화자특성 모델링을 통해 추정할 파라미터를 줄여 적응 데이터가 적은 경우에도 비교적 안정적인 파라미터 추정이 가능하다는 장점 때문으로 분석 된다.

Eigenvoice 방식의 화자분할 성능에 영향을 미치는 eigenvoice 개수에 대한 성능 비교를 그림 5에 나타냈다. 1초 길이의 분석 윈도우를 사용하여 eigenvoice 개수를 5, 10, 20, 30으로 변화시킨 결과 eigenvoice 개수가 10인 경우 가장 높은 화자분할 성능을 보였다. 본 논문의 eigenvoice 방식의 실험은 eigenvoice 개수가 10인 경우의 결과를 나타낸 것이다.

Eigenvoice 방식의 화자가중치를 이용한 거리측정 함수에 따른 성능을 비교하기 위하여 거리 기반 방식과의 비교 실험에서 사용한 최소 분석 윈도우 1초 외에 추가적으로 0.5초의 분석 윈도우를 사용하여 실험하였다. 거리 측정 함수로는 식(5), (6)과 같이 유클리드 거리와 코사인 유사도를 사용하였으며, 분석 윈도우 1초와 0.5초에 대한 화자 분할 결과를 그림 6과 7에 각각 나타냈다.

그림 6에서 보는 바와 같이 1초의 분석 윈도우에

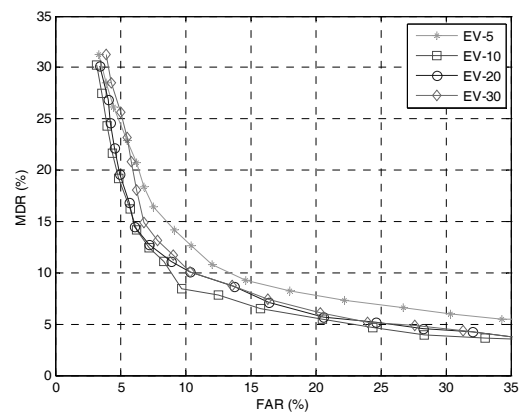


그림 5. Eigenvoice 개수에 따른 eigenvoice 기반의 화자분할 성능 비교

Fig. 5. Performance comparison of eigenvoice-based speaker segmentation according to the number of eigenvoices.

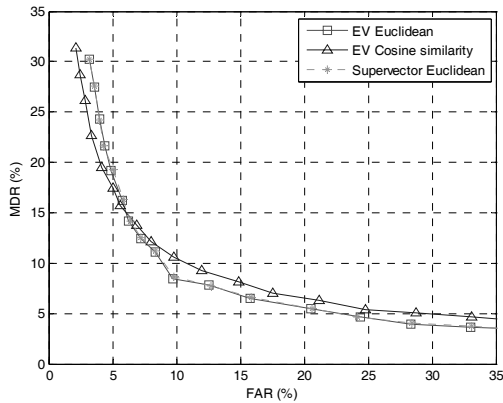


그림 6. 거리 측정 방식에 따른 1초 분석 윈도우를 사용한 eigenvoice 기반의 화자 분할 성능 비교

Fig. 6. Performance comparison of eigenvoice-based speaker segmentation using the analysis window of 1 second according to different distance measurements.

서 추정된 화자가중치의 유클리드 거리와 코사인 유사도의 결과는 equal error rate (EER) 관점에서 각각 9.3%, 10.3%를 나타냈으며, 그림 7의 0.5초 분석 윈도우에서는 각각 11.5%, 12.6%를 나타냈다. EER의 error rate reduction (ERR)은 유클리드 거리 방식이 1초와 0.5초 분석 윈도우에서 각각 9.7%, 8.7%를 얻었다. 이상의 결과를 종합해보면 eigenspace 상에 표현된 화자가중치는 두 벡터의 방향성보다는 단순한 거리를 이용하는 것이 화자 분할 성능 면에서 우수하다는 것을 확인할 수 있다.

또한 화자가중치의 거리를 이용하는 방식의 효율성을 평가하기 위해 식 (7)과 같이 eigenvoice에 의해 화자 적응된 모델들 사이의 직접적인 거리를 이용한 화자 분할 결과를 그림 6과 7에 함께 나타냈다. 그림에서 보는 바와 같이 성능면에서 eigenvoice 유클리드 거리의 결과와 거의 유사한 것을 확인할 수 있다. 그런데 화자적응된 모델들 사이의 직접적인 거리를 구하기 위해서는 화자가중치 추정 이후 식 (1)을 이용한 화자모델 적응과 식 (4)와 같이 고차원 수퍼벡터 (본 논문에서 $D=1664$)들 사이의 거리 계산이 필요하다. 따라서 저차원 ($K=10$) 화자가중치를 이용한 유클리드 거리 기반의 화자 분할 방식은 거의 유사한 성능을 나타내는 화자적응 모델들 사이의 직접적인 거리를 이용한 화자 분할 방식보다 계산량 면에

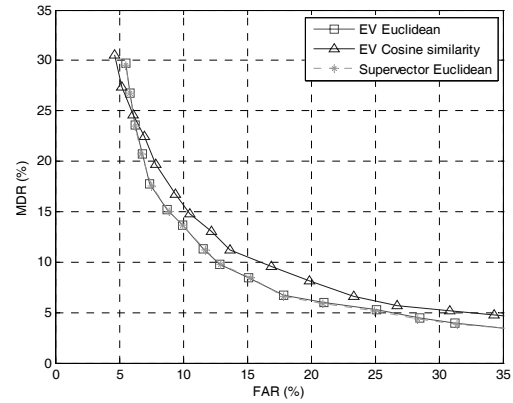


그림 7. 거리 측정 방식에 따른 0.5초 분석윈도우를 사용한 eigenvoice 기반의 화자 분할 성능 비교

Fig. 7. Performance comparison of eigenvoice-based speaker segmentation using the analysis window of 0.5 seconds according to different distance measurements.

서 매우 효율적인 것을 확인할 수 있다.

IV. 결 론

본 논문에서는 eigenvoice 기반의 화자가중치 거리를 이용한 화자 분할 방식을 도입하고, 이 방식을 대표적인 거리 기반 방식인 BIC와 KL 방식과 비교하였다. 그리고 화자가중치의 거리 측정 함수로 유클리드 거리와 cosine 유사도를 사용하여 화자 분할 성능을 비교하였다. 그 결과 eigenvoice에 의해 추정된 화자가중치의 유클리드 거리를 사용한 경우 기존의 거리 기반 방식 보다 월등한 성능 향상을 보였으며, 1초와 0.5초의 분석 윈도우를 사용한 경우 코사인 유사도를 사용한 결과보다도 우수한 성능을 보이는 것을 확인하였다. 또한, eigenvoice 적응 방식에 의해 화자 적응된 모델들 사이의 직접적인 거리를 이용한 화자 분할 방식과의 비교를 통해 화자가중치 거리를 이용한 방식이 계산량면에서 효율적인 것을 확인하였다.

감사의 글

이 논문은 부산대학교 자유과제 학술연구비 (2년)에 의하여 연구되었으며, 논문의 일부는 타 학술대회에서 발표된 바 있다^[10].

참고문헌

1. S. E. Tranter and D. A. Reynolds, "An overview of automatic speaker diarization systems," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1557-1565, 2006.
2. S. S. Chen, and P. S. Gopalakrishnan, "Speaker, environment and channel change detection and clustering via the Bayesian information criterion," in *Proc. DARPA Broadcast News Transcription and Understanding Workshop*, pp. 127-132, 1998.
3. M. A. Siegler, U. Jain, B. Raj, and R. M. Stern, "Automatic segmentation, classification and clustering of broadcast news," in *Proc. DARPA Speech Recognition Workshop*, pp. 97-99, Feb., 1997.
4. R. Kuhn, P. Nguyen, J. C. Jungua, L. Goldwasser, N. Niedzielski, S. Finche, K. Field and M. Contolini, "Eigenvoices for speaker adaptation," in *Proc. of Int. Conf. on Spoken Language Processing*, vol. 5, pp. 1771-1774, Nov. 1998.
5. F. Castaldo, D. Colibro, E. Dalmasso, P. Laface, C. Vair, "Stream-based speaker segmentation using speaker factors and eigenvoices." in *Proc. of Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 4133-4136, 2008.
6. R. Kuhn, J. C. Junqua, P. Nguyen, N. Neidzielski, "Rapid speaker adaptation in eigenvoice space," *IEEE Trans. Speech and Audio Proc.*, vol. 8, no. 6, pp. 695-707, 2000.
7. J. S. Garofalo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, "The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus," CDROM, NIST, 1990.
8. P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
9. M. Kotti, E. Benetos, C. Kotropoulos, "Computationally efficient and robust BIC-based speaker segmentation," *IEEE Trans. Audio, Speech, and Language Proc.*, vol. 16, no. 5, pp. 920-933, 2008.
10. 최무열, 김형순, "Eigenvoice를 이용한 화자분할 시스템의 성능 비교", *한국음성학회 가을 학술대회 발표논문집*, 87-88쪽, 2011.

저자 약력

▶ 최 무 열 (Mu Yeol Choi)



1997년 2월: 동의대학교 전자통신공학과 학사
 1999년 8월: 부산대학교 전자공학과 석사
 2001년 12월 ~ 2003년 5월: ㈜보이스웨어 연구원
 2011년 8월: 부산대학교 전자공학과 박사
 2012년 5월 ~ : ETRI 선임연구원

▶ 김 형 순 (Hyung Soon Kim)



1983년 2월: 서울대학교 전자공학과 (학사)
 1984년 2월: 한국과학기술원 전기및전자공학과 (박사과정 조기진학)
 1989년 2월: 한국과학기술원 전기및전자공학과 (박사)
 1987년 ~ 1992년: 디지털정보통신연구소 선임연구원
 1992년 ~ 현재: 부산대학교 전자공학과 교수
 (관심분야) 음성인식, 음성합성, 음성신호처리