클래스 히스토그램 등화 기법에 의한 강인한 음성인식*

서영주(ICU), 김회린(ICU), 이윤근(ETRI)

<차 례>

<Abstract>

Robust Speech Recognition by Utilizing Class Histogram Equalization

Yungjoo Suh, Hoirin Kim, and Yunkeun Lee

This paper proposes class histogram equalization (CHEQ) to compensate noisy acoustic features for robust speech recognition. CHEQ aims to compensate for the acoustic mismatch between training and test speech recognition environments as well as to reduce the limitations of the conventional histogram equalization (HEQ). In contrast to HEQ, CHEQ adopts multiple class-specific distribution functions for training and test environments and equalizes the features by using their class-specific training and test distributions. According to the class-information extraction methods, CHEQ is further classified into two forms such as hard-CHEQ based on vector quantization and soft-CHEQ using the Gaussian mixture model. Experiments on the Aurora 2 database confirmed the effectiveness of CHEQ by producing a relative word error reduction of 61.17% over the baseline mel-cepstral features and that of 19.62% over the conventional HEQ.

* Keywords: Class histogram equalization, Feature compensation, Robust speech recognition.

* 본 연구는 과학기술부 기초과학연구사업 중 지방연구중심대학 육성사업인 헬스케어기술 개발사업단의 지원에 의해 수행되었음.

1. 서 론

음성은 의사소통을 위한 가장 편리한 수단 중의 하나이다. 따라서, 인간이 개발한 기계와의 통신 수단으로서 음성을 사용하려는 노력들이 지난 수십 년간 활발히 이루어져 왔다. 그러나, 현재까지 개발된 음성인식 기술들도 사용자의 요구를 만족할 만한 수준에 이르지 못하고 있는 실정이다. 인간과는 다르게, 현재 개발된 음성인식기들은 훈련 환경과 다른 잡음 환경에서 동작될 경우에 심각한 성능 저하를 나타낸다. 즉, 훈련 환경과 음향 부정합(acoustic mismatch)인 시험 환경에서 음성인식기를 운용할 경우에는 큰 폭의 성능저하를 나타낸다. 이러한 음향 부정합의 원인으로 실제 음성인식 환경에서 발생하는 가산성 잡음과 채널 잡음 등에 의한 음질 저하를 들 수 있다[1]-[4]. 강인한 음성인식은 이러한 음향 부정합 효과를 제거하여, 음향 부정합 환경에서 음성인식기의 성능이 음향학적으로 일치된 환경에서의 성능과 가깝도록 향상시키는데 그 목적이 있다[5][6]. 일반적으로 강인한 음성인식을 위한 대부분의 기술들은 적용 영역에 따라서 다음과 같이 크게 세 가지로 나뉘어진다[5][6]: 신호영역에서의 음성개선, 특징영역에서의 특징보상, 모델영역에서의 모델적응. 이 중에서 특징보상 기법은 음성인식에서 사용되는 음성특징을 보상하여 음향 부정합 현상을 제거하거나 감소시키는 방법이다. 이 기법은 구현 상의 용이함, 계산량, 그리고, 음성인식기의 성능 향상 측면에서의 상대적인 이점으로 인하여 음향 부정합의 감소에 널리 사용되고 있다. 초기 단계에서의 특징보상 기법들로 선형 변환에 기반한 켑스트럼 평균 정규화(CMN: cepstral mean normalization)[7][8] 또는 켑스트럼 평균-분산 정규화(CMVN: cepstral mean and variance normalization)[9] 등과 같은 방법들이 제안되었는데 이들은 뛰어난 계산상의 효율성과 음성인식 성능 향상 효과를 나타내었다. 그러나, 음향 부정합을 야기하는 간섭 인자들은 음성인식의 특징으로 널리 사용되는 로그 에너지나 켑스트럼 영역에서 비선형적으로 작용한다[4]. 따라서, 켑스트럼 평균 정규화나 켑스트럼 평균-분산 정규화와 같은 선형 변환 기반의 특징보상 기법들은 음향 부정합 환경에서 음성인식기의 성능향상을 위한 방법으로서 근본적인 한계를 가지고 있다. 음향 부정합에서 야기되는 비선형적인 특성을 선형으로 근사화하여 보상하는 방법들로 VTS(vector Taylor series)[10]와 SLA(statistical linear approximation)[11]와 같은 불연속적 선형 근사화(piecewise linear approximation) 방법들이 제안되었다. 관련 연구 결과에 따르면, 이러한 특징 보상 방법들은 기존의 선형 변환 방법들에 비해 훨씬 개선된 보상 성능을 나타낸다고 보고 되었다.

이러한 방법들과는 다른 접근법인 특징영역 정규화 방법의 하나로서 히스토그램 등화 기법(HEQ: histogram equalization)이 제안되었다[12]-[17]. 앞에서 언급한 바와 같이, 음성인식에서 훈련 환경과 음향 부정합을 보이는 시험 환경은 음성특징 값을 비선형적으로 변화시켜 훈련 환경에서 구한 특징 값들과의 차이를 야기한다.

따라서, 음향 부정합이 존재하는 훈련 및 시험 환경에서 음성특징을 통계적으로 분석하면 서로 간에 확률분포 상의 차이를 나타낸다. 히스토그램 등화 기법은 시험 음성특징의 확률밀도함수를 훈련(training) 음성특징의 확률밀도함수와 동일하도록 변환시켜 두 환경 간에 존재하는 음향 부정합을 제거하려는 접근방법이다[16]. 이러한 히스토그램 등화 기법은 알고리즘의 구현이나 계산량 측면에서 매우 효율적이면서도 특징보상을 통한 인식성능 개선 효과가 뛰어나 현재 활발히 연구되고 있다[12]-[17].

그러나, 음향 부정합을 효과적으로 제거하기 위해서 기존의 히스토그램 등화 기법은 음향 부정합의 단조 변환(monotonic transformation) 및 훈련과 시험 음향 클래스 분포의 동일성과 같은 몇 가지 근본적인 가정을 전제로 한다[18]-[21]. 그러나, 랜덤 특성을 띤 배경 잡음의 부가나 짧은 시험 음성 발화는 비단조 변환(non-monotonic transformation)과 음향 클래스 분포의 상이함을 각각 야기할 수 있다. 따라서, 이런 조건의 시험 환경에서는, 기존의 히스토그램 등화 기법이 제대로 동작하지 않아 음향 부정합 제거 효과가 감소하는 문제가 발생된다.

본 논문에서는 이와 같은 환경에서 기존의 히스토그램 등화 기법이 가지는 문제점을 해결하기 위하여, 음향 클래스 정보를 이용하는 클래스 히스토그램 등화 기법을 제안하였다. 단일 전역 누적분포함수를 사용하는 기존의 히스토그램 등화 기법과는 달리, 제안된 클래스 히스토그램 등화 기법은 복수의 클래스 기반 훈련 및 시험 누적분포함수들을 이용하여 잡음이 부가된 시험 음성특징을 클래스별로 나누어 보상하는 구조를 취하고 있다. 즉, 본 논문에서는 훈련 환경과 비교하여 음향 부정합이 존재하는 시험 환경에서 필연적으로 발생하는 음성인식기의 성능 저하 문제를 해결하거나 감소시키기 위한 강인한 음성인식 기술 중에서, 특징영역에서 이루어지는 특징보상 기법의 하나로 제안된 클래스 기반 히스토그램 등화 기술에 대해서 다루고자 한다. 본 논문의 전체 구성은 다음과 같다. 먼저, 1장의 서론에 이어 2장에서 기존의 히스토그램 등화 기법에 의한 특징보상에 대해 살펴본다. 3장에서는 클래스 히스토그램 등화 기법을 제안하고 4장에서 실험 및 성능 평가에 대해서 기술한다. 마지막으로 5장에서 결론을 맺도록 한다.

2. 기존의 히스토그램 등화 기법을 이용한 특징보상

2.1. 히스토그램 등화의 원리

히스토그램 등화 기법은 원래 디지털 화상처리 분야에서 디지털 화상의 콘트라스트를 개선하거나 그레이 레벨 신호의 다이나믹 레인지를 최적화하기 위해 제안되었다[12]. 히스토그램 등화의 원리는 시험 환경에 대한 랜덤변수의 확률밀도함

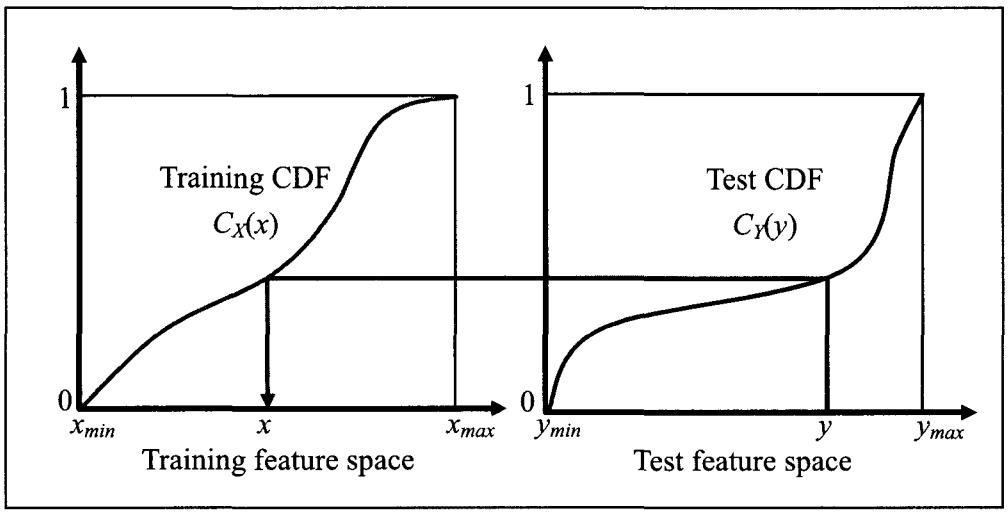
수를 구하고 이를 훈련 환경에 대한 랜덤변수의 확률밀도함수와 동일한 분포가 되도록 변환하면, 시험 환경에서 존재하는 음향 부정합을 제거할 수 있다는데 있다[15][16]. 히스토그램 등화 기법을 이용한 특징보상에 대해 수식적으로 설명하면 다음과 같다.

주어진 시험 환경에 대한 랜덤변수 y와 이 변수의 확률밀도함수를 $p(y)$라 하면, 시험 확률밀도함수 $p(y)$를 훈련 확률밀도함수 $p(x)$로 사상(mapping)시키는 변환 함수 $x=F(y)$는 다음 식과 같이 주어진다[16].

$$x = F(y) = C_X^{-1}[C_Y(y)] \tag{1}$$

여기서 $C_X^{-1}(x)$는 훈련 누적분포함수인 $C_X(x)$의 역함수를 의미한다. 또한, $C_Y(y)$는 시험 랜덤변수 y의 누적분포함수이다.

<그림 1>에서는 히스토그램 등화의 비선형적인 특징보상 원리를 나타내고 있다. 이 그림과 같이, 히스토그램 등화는 훈련과 시험 환경의 두 특징 축 상에서 동일한 클래스 순서정보를 이용하여, noisy 시험 음성의 특징 계수를 개선된 음성의 특징 계수로 비선형적으로 변환시킨다. 그러나, 식 (1)에서 정의된 누적분포함수의 실제적인 구현에서는 유한한 수의 히스토그램 빈들로 구성된 누적 히스토그램을 대체하여 사용하고, 이에 따라서 이 보상 기술을 히스토그램 등화 기법이라고 한다[16].



<그림 1> 히스토그램 등화에 의한 음성인식 특징보상의 원리.

2.2. Order-Statistics 기반의 누적분포함수 추정

히스토그램 등화 기법을 이용한 특징보상에서는 식 (1)에서도 나타난 바와 같이, 누적분포함수를 사용하여 보상함수를 구한다. 따라서, 이 기법에서 가장 중요한 과제들 중의 하나는 정확한 훈련 및 시험 누적분포함수의 추정이다. 히스토그램에 기반한 전통적인 누적분포함수 추정법에서는 먼저 입력 랜덤변수들에 대하여 유한한 수의 히스토그램 빈들로 구성된 빈도 히스토그램을 구한 다음, 이 히스토그램에서 빈 값들의 누적을 통해서 누적 히스토그램을 추정한다. 일반적으로 음성인식기의 훈련에서는 인식기의 음향 모델들을 훈련시키기에 충분한 분량의 훈련 음성 데이타를 사용한다고 가정할 수 있다. 따라서, 대개의 경우에 훈련 누적분포함수는 전통적인 누적 히스토그램 추정 방법을 적용하여도 신뢰성있는 훈련 누적분포함수의 추정치를 구할 수 있다. 이와 같은 이유로 인하여 히스토그램 등화 기법에 의한 특징보상에서의 훈련 누적분포함수 추정은 전통적인 누적 히스토그램 추정방법을 많이 사용한다. 그러나 길이가 충분히 길지 않은 시험 발화에 대한 시험 누적분포함수의 추정에서는 샘플 데이터의 부족에 의한 추정 모델의 신뢰도 문제가 발생될 수 있다. 따라서, 이 경우에 시험 누적분포함수의 정확한 추정이 매우 중요하게 되며, 이는 결국 특징보상의 효과와 직접적으로 연결된다. 일반적으로, 샘플 데이터의 양이 적을 경우에는 전통적인 누적 히스토그램 추정 방식보다 order-statistics에 기반한 추정 방식이 더 정확하게 누적분포함수를 추정할 수 있어서 선호되고 있다[15].

Order-statistics에 기반한 누적분포함수의 추정 과정은 다음과 같다. 먼저, noisy 음성신호로부터 추출된 특정한 차수의 특징계수들의 열인 S를 다음 식과 같이 정의한다.

$$S = \{y_1, y_2, \cdots, y_n, \cdots, y_N\} \tag{2}$$

여기서 y_n은 n번째 프레임의 특징계수를 의미한다.

식 (2)의 order-statistics는 특징계수열 S의 샘플들을 오름차순(ascending order)으로 나열했을 때의 서열(rank) 또는 순서(order)로서 다음과 같다.

$$y_{T(1)} \le y_{T(2)} \le \cdots \le y_{T(r)} \le \cdots \le y_{T(N)} \tag{3}$$

여기서 $T(r)$은 특징계수 $y_{T(r)}$이 특징계수열 S에서 r번째 order-statistic 일 경우에, 오름차순으로 재배열하기 전의 원래의 프레임 정보를 의미한다.

이와 같은 순서정보를 이용한 order-statistics 기반의 시험 누적분포함수 추정치는 다음과 같이 정의된다[15].

$$\hat{C}(y_n) = \frac{R(y_n) - 0.5}{N} \tag{4}$$

여기서 $R(y_n)$은 n번째 프레임에서의 특징계수인 y_n의 오름차순 서열을 의미하며 범위는 1에서 N 까지의 정수값 이다.

식 (4)와 같이 구해지는 order-statistics에 기반한 시험 누적분포함수 추정법과 누적 히스토그램에 기반한 훈련 누적분포함수 추정법을 이용하는 히스토그램 등화 기법을 적용하였을 때, noisy 음성의 특징계수인 y_n으로부터 보상된 특징계수는 다음과 같이 구해진다[15].

$$\hat{x}_n = C_X^{-1}[\hat{C}_Y(y_n)] = C_X^{-1}[\frac{R(y_n) - 0.5}{N}] \tag{5}$$

일반적으로, 히스토그램을 이용하는 경우에는 유한한 수의 히스토그램 빈들에 의하여 사상오차가 필연적으로 발생한다. 따라서, 보다 정확한 특징보상을 위하여, 식 (5)에서 훈련 누적분포함수의 역함수에 의한 변환 과정에서 히스토그램 빈의 대표값에 대한 사상이 아니라 보간법 (interpolation)을 이용하여 빈 내에서의 상대적인 위치에 대한 사상이 이루어진다. 본 실험에서는 다음과 같이 일차 보간법을 사용하여 보상되는 특징계수의 정확도를 개선하였다[18].

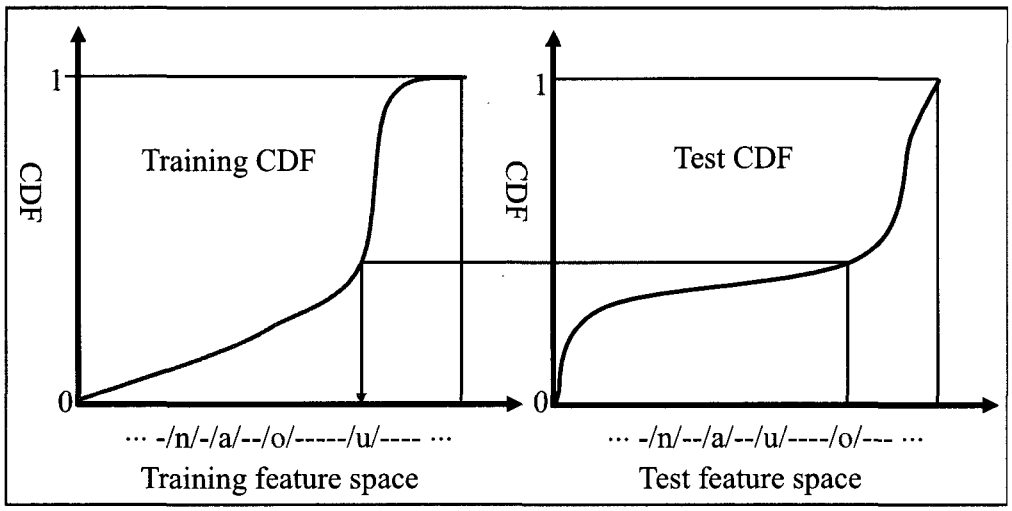
$$\hat{x}_n = \begin{cases} \triangle \times \left[\left(\dfrac{R(y_n) - 0.5}{N \times \hat{C}_X(1)} \right) - 0.5 \right] + x_{\min}, & \text{if } b = 1 \\ \triangle \times \left[\dfrac{0.5\left(\dfrac{R(y_n) - 0.5}{N} - \hat{C}_X(b-1) \right)}{1 - \hat{C}_X(b-1)} - 0.5 \right] + x_{\max}, & \text{elseif b} = \mathrm{N_{BIN-X}} \\ \triangle \times \left[\dfrac{\left(\dfrac{R(y_n) - 0.5}{N} - \hat{C}_X(b-1) \right)}{\hat{C}_X(b) - \hat{C}_X(b-1)} + b - 0.5 \right] + x_{\min}, & \text{else} \end{cases} \tag{6}$$

여기서 $\triangle$는 훈련 누적 히스토그램에서 빈의 폭을 의미하고, b는 훈련 누적 히스토그램에서 $\hat{C}_Y(y_n)$이 속한 빈의 인덱스를 나타낸다. 또한, $\hat{C}_X(b)$는 b번째 빈에서 훈련 누적분포함수의 추정치 이며, N_{BIN_X}는 훈련 누적 히스토그램에서 전체 빈들의 수를 뜻한다. 마지막으로, x_{min}과 x_{max}는 각각 훈련 데이터로부터 구한 특징 계수들의 최소값과 최대값을 의미한다.

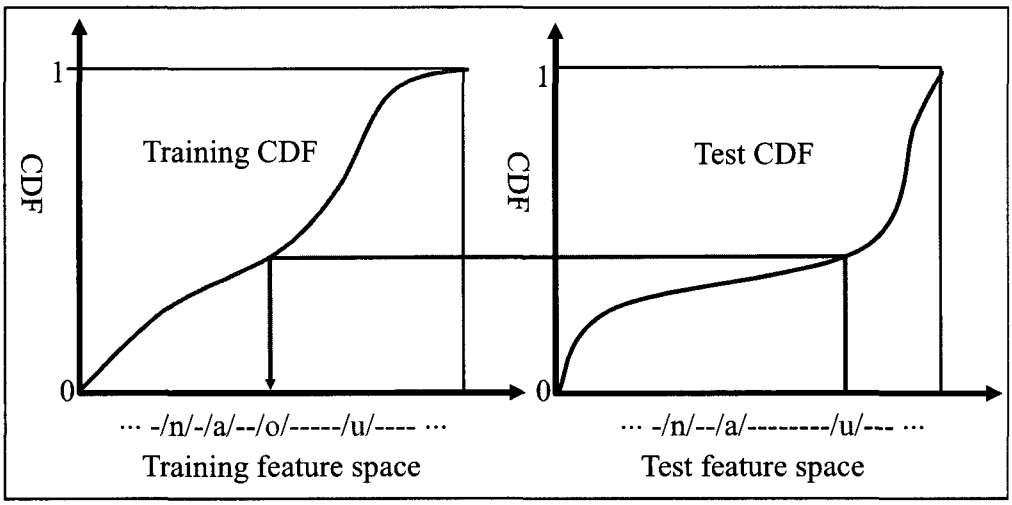
2.3. 기존의 히스토그램 등화 기법의 단점

음성인식의 특징영역에서 음향 부정합을 효과적으로 제거하기 위해서 히스토그램 등화 기법이 최근 많이 연구되고 있다. 그러나, 서론에서도 언급한 바와 같이 기존의 히스토그램 등화 기법에서는 효과적인 특징보상을 위해서 몇 가지 근본적인 가정이 충족되어야 한다[18]-[21]. 첫번째 가정으로서, 음향 부정합은 음성 특징 영역에서 단조 변환(monotonic transformation)으로만 작용해야 한다. 즉, 음향 부정합이 발생하여 음성특징값이 비선형적으로 변할 경우에도 이 특징들의 음향 클래스에 대한 순서정보는 훈련 환경에서와 동일하게 유지가 될 경우에 한하여 히스토그램 등화 기법의 보상효과가 제대로 나타난다. <그림 2>는 이와 같은 단조 변환이 아닐 경우에 대해 히스토그램 등화에 의한 특징보상의 부작용을 나타낸다. 이 그림에서 가로축은 음성특징축 상에서 통계적으로 구해진, 여러 음향 클래스들의 영역을 나타낸다. 이 그림을 보면, 시험 환경에서 나타난 음향 부정합으로 인하여 훈련과 시험 환경에서 음향 클래스 /o/와 /u/의 순서정보가 서로 바뀌어졌으며, 이러한 경우에 기존의 히스토그램 등화에 의한 특징보상이 다른 음향 클래스의 특징값으로 잘못 사상시킬 수 있음을 알 수 있다. 두번째 가정으로, 음성인식의 음향 모델링에서 정의된 음향 클래스들의 확률분포가 훈련과 시험 환경 간에 동일하여야 한다는 점을 들 수 있다. 만일, 훈련 환경에 존재하였던 특정한 음향 클래스들이 시험 환경에서 관측되지 않았을 경우에는 시험 환경에서의 음향 클래스들의 순서정보가 훈련 환경에서와 달라지게 되며, 이 경우에도 히스토그램 등화를 통해 보상된 특징값은 다른 음향 클래스의 값으로 사상될 수 있는 문제점이 발생한다. <그림 3>은 이와 같은 경우를 나타낸 그림이다. 즉, 시험 음성발화에서 음향 클래스 /o/의 부재로 인하여 이 특징보상이 다른 음향 클래스의 값으로 사상시킬 수 있음을 알 수 있다. 따라서, 위의 두 가지 가정들 중에서 어느 하나라도 만족되지 않을 경우에는 히스토그램 등화가 음성특징을 원래의 동일한 음향 클래스 영역으로 일관성 있게 변환시키지 못하게 되고 이는 음성특징에 대한 클래스 분리도의 감소를 가져와 결과적으로 음성인식에서의 특징보상 효과를 감소시킨다. 그러나, 랜덤 특성이 있는 배경 잡음의 부가는 비단조 변환 (non-monotonic transformation)을 야기할 수도 있다. 또한, 지속시간이 길지 않은 시험 음성 발화에서는 특정 음향 클래스들이 관측되지 않을 수도 있기 때문에, 이 경우에 시험 음향 클래스들의 확률분포는 훈련 음향 클래스들의 확률분포와 상이해질 가능성이 크다. 따라서, 랜덤 잡음이 부가되고 비교적 짧게 발성되는 시험 환경에서는, 기존의 히스토그램 등화 기법이 제대로 동작하지 않음으로써 음향 부정합 제거 효과를 감소시키는 문제가 발생된다. 따라서, 히스토그램 등화 기법에 의한 음성인식의 특징보상이 다른 방법에 비해서 더 우수한 계산상의 효율성과 특징 보상효과를 제공하지만, 위에서 언급한 근본적인 제한 사항들로 인하여 일반적인 상황에서

최대한의 보상효과를 얻기 어렵다는 단점이 존재한다. 본 논문에서는 이러한 기존의 히스토그램 등화 기법에 내재된 단점을 해결하거나 감소시키기 위하여 클래스 히스토그램 등화 기법을 제안하였다.



<그림 2> 비단조 변환에서 발생하는 기존 히스토그램 등화의 단점.



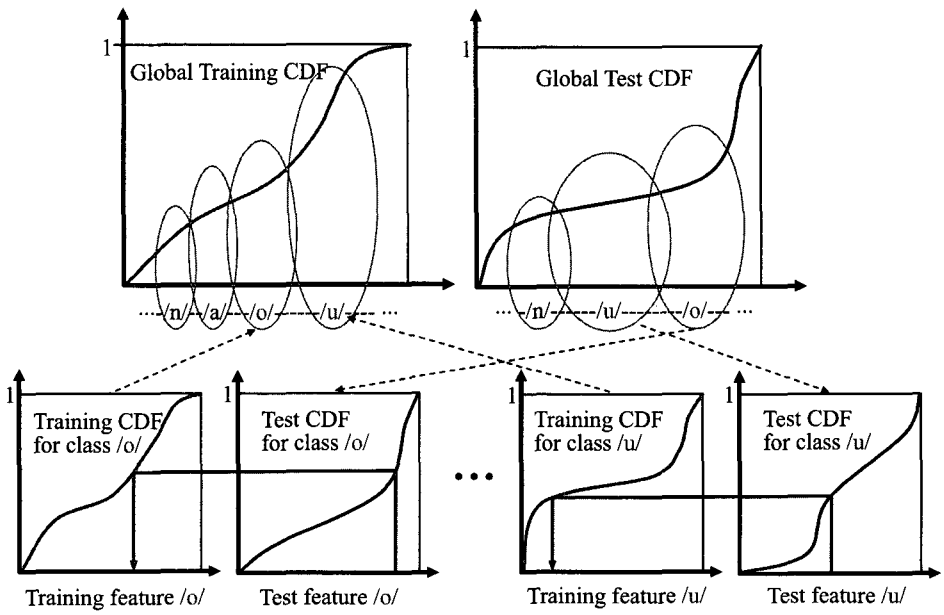
<그림 3> 훈련과 시험 클래스 분포 간의 상이함에서 발생하는
기존 히스토그램 등화의 단점.

3. 클래스 히스토그램 등화 기법

3.1. 클래스 히스토그램 등화의 개념

기존의 히스토그램 등화 기법이 가지는 두 가지 제한 사항들은 이 기법이 각각 하나의 훈련 및 시험 누적분포함수를 전역적으로 사용하고 있기 때문에 발생된다. 따라서, 음성인식의 음성특징 축을 따라서 존재하는 음향 클래스 별로 훈련과 시험 누적분포함수를 별도로 사용할 경우에는 그 두 가지 제한 사항들이 감소되고 결과적으로 히스토그램 등화에 의한 특징보상의 효과를 최대화할 수 있다. 이 점을 근거로, 제안된 클래스 히스토그램 등화 기법은 복수의 클래스 기반 훈련 및 시험 누적분포함수들을 이용하여 잡음이 부가된 시험 음성특징을 클래스별로 나누어 보상하는 구조를 취하고 있다.

히스토그램 등화 기법에 음향 클래스 개념을 도입할 경우에는 클래스 정보의 추출이 필수적으로 요구된다. 현재 특징보상을 위한 히스토그램 등화 기법에서는 특징벡터 단위보다는 개별적인 특징계수 단위로 등화 변환이 수행되고 있다[16]. 이는 시험 발화의 샘플 데이터 양이 부족한 상황에서도 보다 높은 신뢰도의 시험 누적분포함수를 추정하기 위한 것 이거나 벡터 단위로 히스토그램 등화를 할 경우에 수반되는 알고리즘의 복잡도 문제에서 벗어나고자 함이다. 이와 더불어, 음성인식의 특징 파라메터가 캡스트럼일 경우에 다른 특징계수들 간의 공분산의 값



<그림 4> 클래스 히스토그램 등화 기법을 이용한 특징보상의 원리.

이 작다는 점도 특징계수 단위의 등화를 가능하게 하는 이유이다. 그러나, 현재 음성인식에서 프레임 단위의 음성 신호를 특징 파라메터화 하거나 음향모델로 모델링할 경우에 대부분 벡터 단위로 처리한다. 따라서, 본 논문에서는 클래스 히스토그램 등화에서 요구되는 음향 클래스 정보를 특징계수 단위가 아니라 특징벡터 단위로 추출하여 사용하였다[18]-[21]. 클래스 정보의 추출과 이용 관점에서 접근하면, 제안된 클래스 기반 히스토그램 등화 기법은 벡터 양자화 (vector quantization)에 의한 하드-클래스 기반[19],[20]과 GMM (Gaussian mixture model)으로부터 유도된 확률값에 기반한 소프트-클래스 기반[21]의 두 가지 방법들로 나누어질 수 있다. <그림 4>는 음향 클래스 정보를 이용하여 클래스 별로 히스토그램 등화를 수행하는 제안된 방법의 원리에 대한 설명도이다.

3.2. 하드-클래스 기반의 히스토그램 등화 기법

벡터 양자화에 의한 하드-클래스 기반의 히스토그램 등화는 음향 클래스 정보의 추출을 위해 벡터 양자화를 수행하고 그 결과로서 얻어지는 최적의 코드워드에 해당하는 하나의 음향 클래스 만을 분석 프레임에서의 클래스 정보로 사용하는 방식이다. 이와 같은 하드-클래스 기반의 히스토그램 등화 기법에 대한 보다 자세한 설명은 다음과 같다.

n번째 프레임에서 K차원의 계수들로 구성된 시험 특징벡터 V_n을 다음 식과 같이 정의한다.

$$V_n = [y_n^{(1)} y_n^{(2)} \cdots y_n^{(k)} \cdots y_n^{(K)}]^T \tag{7}$$

여기서 $y_n^{(k)}$는 k차 시험 특징계수를 나타내고 T는 벡터의 전치(transpose)를 의미한다.

본 논문에서는 가산성 또는 채널 잡음이 부가된 음성신호에서 강인한 음성인식을 위한 특징보상에 대해 다루고 있다. 따라서, 입력으로 사용되는 음성신호에는 불가피하게 잡음이 부가되어 있는 상황이라고 가정할 수 있다. 이 경우에 noisy 음성특징으로부터 정확한 음향 클래스 정보를 추출하여야 하는 어려움에 직면하게 된다. 따라서, 보다 정확한 음향 클래스 정보를 추출하기 위해서는 잡음의 영향을 제거하거나 감소시켜야 한다. 이를 위해서, 음향 클래스 정보의 추출 과정에서 noisy 시험 음성의 특징벡터를 사용하는 대신에 잡음의 영향을 일차적으로 감소시킨 시험 음성의 특징벡터를 사용하였다. 일차적인 잡음감소 방법으로는 기존의 히스토그램 등화 기법을 적용하였다. 이에 따른 음향 클래스 추출 과정은 다음과 같다.

식 (7)에서 정의된 프레임별 시험 음성의 특징벡터에 대해 기존의 히스토그램

등화 기법을 적용한, 일차적으로 잡음이 감소된 특징벡터는 다음 식과 같이 주어진다.

$$\begin{aligned}\hat{V}_n &= [\hat{x}_n^{(1)}\hat{x}_n^{(2)}\cdots\hat{x}_n^{(K)}]^T \\ &= [C_X^{-1}(\hat{C}_Y(y_n^{(1)}))\,C_X^{-1}(\hat{C}_Y(y_n^{(2)}))\cdots C_X^{-1}(\hat{C}_Y(y_n^{(K)}))]^T\end{aligned} \tag{8}$$

이 특징벡터를 벡터 양자화의 입력으로 사용하여 얻어지는 하드-클래스 정보는 다음과 같이 추출된다.

$$\hat{i} = \underset{i}{\operatorname{argmin}}\, d(\hat{V}_n, z_i),\ 1 \le i \le I_H \tag{9}$$

여기서, $d(\cdot,\cdot)$는 마할라노비스 (Mahalanobis) 거리척도이고, z_i는 기존의 히스토그램 등화 기법에 의해 정규화된 훈련 데이터에 대해 k-means 알고리즘으로 구한 i 번째 하드-클래스의 중심 벡터를 의미한다. 또한, I_H는 하드-클래스의 전체 수이다.

식 (9)로부터 구해지는 하드-클래스 정보를 이용한 하드-클래스 기반의 히스토그램 등화 기법에 의해 변환된 훈련 특징계수 추정치는 다음과 같이 구해진다.

$$\begin{aligned}\hat{x}_{H,n} &= C_{H,X(\hat{i})}^{-1}[\hat{C}_{H,Y(\hat{i})}(y_n)] \\ &= C_{H,X(\hat{i})}^{-1}\left(\frac{R_{\hat{i}}(y_n)-0.5}{N_{\hat{i}}}\right)\end{aligned} \tag{10}$$

여기서, $\hat{C}_{H,Y(\hat{i})}(y_n)$ 과 $R_{\hat{i}}(y_n)$은 각각 하드-클래스 기반의 시험 누적분포함수 추정치와 $\hat{i}$ 번째 하드-클래스에서 yn의 서열을 의미한다. 또한, $C_{H,X(\hat{i})}^{-1}(\cdot)$와 $N_{\hat{i}}$는 각각 훈련 데이터로부터 구해진 하드-클래스 기반의 훈련 누적분포함수인 $C_{H,X(\hat{i})}(\cdot)$의 역함수와 $\hat{i}$번째 하드-클래스에 속한 프레임들의 수를 나타낸다.

3.3. 소프트-클래스 기반의 히스토그램 등화 기법

소프트-클래스 기반의 히스토그램 등화에서는 주어진 특징벡터에 하나의 음향 클래스만 할당하는 하드-클래스 방식과는 달리 모든 음향 클래스로부터의 상관관계를 고려한다. 따라서, 소프트-클래스 방식은 하드-클래스 방식의 일반화된 형태라고 할 수 있다. 여기서, 각 음향 클래스로부터의 기여도는 GMM 기반의 사후확률(posterior probability)로서 정해진다.

특징벡터 $\hat{V}_n$이 주어졌을 때, 소프트-클래스 ω_i에 대한 사후확률은 다음과 같이 정의된다.

$$P(\omega_i|\hat{V}_n) = \frac{\alpha_i N(\hat{V}_n;\mu_i,\Sigma_i)}{\sum_{m=1}^{I_S}\alpha_m N(\hat{V}_n;\mu_m,\Sigma_m)} \tag{11}$$

여기서, I_S는 소프트-클래스들의 수이고, a_i는 i번째 소프트-클래스의 혼합성분계수를 나타낸다. 또한, $N(\hat{V}_n;\mu_i,\Sigma_i)$는 i번째 소프트-클래스에서 기존의 히스토그램 등화 기법에 의해 일차적으로 잡음이 감소된 특징벡터 $\hat{V}_n$에 대해 평균 벡터 μ_i와 공분산 행렬이 Σ_i인 정규확률분포의 확률값을 나타낸다. 이 평균 벡터와 공분산 행렬은 기존의 히스토그램 등화 기법에 의해 정규화된 훈련 데이터로부터 구해진다.

식 (10)에서 정의된 클래스 히스토그램 등화의 원리와 식 (11)에서 정의된 소프트-클래스의 개념을 통합한 소프트-클래스 기반의 히스토그램 등화에 의한 훈련 특징계수 추정치는 다음과 같이 구해진다.

$$\hat{x}_{S,n} = \sum_{i=1}^{I_S} P(\omega_i|\hat{V}_n)\, C^{-1}_{S,X(i)}[\hat{C}_{S,Y(i)}(y_n)] \tag{12}$$

여기서, $\hat{C}_{S,Y(i)}(y_n)$ 는 i번째 소프트-클래스에서의 시험 누적분포함수 추정치를 나타내며 식 (4)의 order-statistics에 의한 시험 누적분포함수 개념을 소프트-클래스에 적용함으로써 구해지는데 다음 식과 같이 유도된다.

$$\hat{C}_{S,Y(i)}(y_n) = \frac{\left(\sum_{r=1}^{R(y_n)-1} P(\omega_i|\hat{V}_{T(r)})\right) + 0.5P(\omega_i|\hat{V}_n)}{\sum_{r=1}^{N} P(\omega_i|\hat{V}_n)} \tag{13}$$

i 번째 소프트-클래스에서 훈련 누적분포함수 추정치는 훈련 과정에서 구해진 누적 히스토그램의 해당 히스토그램 빈을 참조함으로써 얻을 수 있는데 다음과 같이 정의된다.

$$C_{S,X(i)}[x_n] = \sum_{b=1}^{B_{X(i)}(x_n)-1} P_{X(i)}(b) + 0.5P_{X(i)}(B_{X(i)}(x_n)) \tag{14}$$

여기서 $B_{X(i)}(x_n)$은 i 번째 소프트-클래스의 훈련 히스토그램에서 x_n이 속한 히스토그램 빈의 인덱스를 의미하고 $P_{X(i)}(b)$는 b번째 빈에서 i 번째 소프트-클래스의 훈련 누적분포함수의 확률 추정치로서, 클래스 가중치 개념을 적용하였을 때에 전체 훈련 특징계수들 중에서 b번째 히스토그램 빈에 포함되는 훈련 계수들의 비율을 의미하는데 다음과 같이 유도된다.

$$P_{X(i)}(b) = \frac{\sum_{u=1}^{U_X}\sum_{n=1}^{N_{X(u)}} CF[P(\omega_i|\hat{V}_n), (x_n \in HST_{X(i)}(b))]}{\sum_{u=1}^{U_X}\sum_{n=1}^{N_{X(u)}} P(\omega_i|\hat{V}_n)} \tag{15}$$

여기서 U_X는 훈련 발화의 수를 나타내고, $N_{X(u)}$는 u번째 훈련 발화에서 프레임의 수이다. $HST_{X(i)}(b)$는 i번째 소프트-클래스의 훈련 히스토그램에서 b번째 빈을 의미한다. 마지막으로 $CF(\cdot,\cdot)$는 조건 함수로서 다음과 같이 정의된다.

$$CF(fn, cond) = \begin{cases} fn, & \text{if } cond \text{ is true} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

4. 실험 및 성능 평가

4.1. 음성 데이터베이스

실험에서는 TI-DIGITS 데이터베이스로부터 변환된 Aurora 2 데이터베이스를 사용하였다[22]. 이 데이터베이스는 훈련 데이터와 평가 데이터로 구성되어 있으며 모두 8kHz의 표본화율과 16bits의 양자화 단계로 디지털화 되었다. 훈련에서는 clean 음성 데이터 (clean-condition training)를 사용하였는데 이 훈련 데이터는 모두 8,440개의 clean 음성 발화로 구성되어 있다. 평가에서는 세 가지 데이터 셋인 test sets A, B 및 C를 사용하였다. Test sets A와 B에는 각각 다른 4가지의 가산성 잡음이 부가되었으며 test set C는 두 가지 종류의 가산성 잡음과 훈련 환경과 다른 채널 잡음 (MIRS)이 부가된 음성 데이터이다. Test sets 데이터는 각각의 잡음에 대해서 7 가지 (clean, 20dB, 15dB, 10dB, 5dB, 0dB, -5dB)의 신호대잡음비로 구분되는 세부 데이터로 구성되어 있으며 각 신호대잡음비에서는 한 자리부터 일곱 자리까지의 숫자음으로 이루어진 1,001개의 발화로 구성되어 있다. Aurora 2 데이터베이스를 이용한 음성인식 성능평가에서 noisy 음성 데이터는 이중에서 20dB에서 0dB까지의 데이터들로 구성된다[22].

4.2. 특징추출 및 음성인식 조건

특징 파라메터의 추출은 Aurora 2의 표준 MFCC (Mel-frequency cepstral coefficient) 추출 과정을 적용하였다[22]. 먼저 음성신호에 대해서 프리엠프시스(0.97) 처리를 하고, 25ms 길이의 해밍(Hamming) 창을 10ms마다 취하여 음성 프레임을 추출한다. 추출된 음성 프레임에 대해 FFT를 통해 얻어진 진폭 스펙트럼으로부터 23차의 필터뱅크 에너지 계수를 구한다. 이로부터 12차 멜-켑스트럼 계수와 프레임 기반 로그 에너지로 구성된 13차 정적 계수를 얻는다. 이 정적계수에 대해서 22차 sine 창함수 형태의 켑스트럼 리프터링[23]을 취한 후 이들의 일차 및 이차 델타 계수를 추가로 추출하여 최종적으로 39차원의 멜-켑스트럼 기반 특징벡터를 추출하여 실험에서의 특징벡터로 사용한다.

음성인식 실험에서는 Aurora 그룹에서 규정한 음성인식기의 구조와 훈련 및 평가 과정을 채용하였다. 기본 음성인식기는 전체단어(whole-word) 모델을 취하고 있으며 발음사전은 13개의 단어들(11개의 숫자음, 묵음 및 단구간 휴지음)로 구성되어 있다. 각 숫자음 HMM (hidden Markov model)은 16개의 상태(state)로 구성되어 있고 묵음과 단구간 휴지음 HMM은 각각 3개와 1개의 상태로 이루어져 있다. 모든 모델에서 각각의 상태는 3개의 혼합(mixture)으로 구성되어 있으며 이 혼합모델들은 대각선 공분산 행렬로 모델링 되었다.

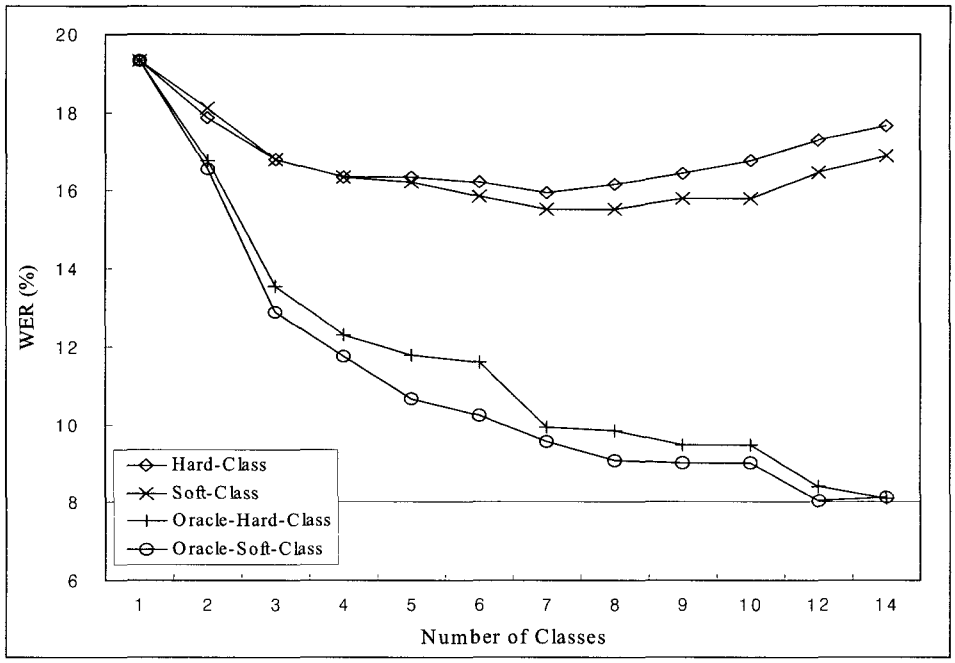
음성인식 평가에서는 Aurora 2에서 규정한 MFCC 음성특징 파라미터에 대한 음성인식 성능과 이 MFCC에 기존의 히스토그램 등화 기법을 적용하였을 때의 성능 및 제안된 클래스 히스토그램 등화 기법을 적용하였을 때의 성능 측정과 비교가 수행되었다.

히스토그램 등화를 위한 누적분포함수의 추정에서는 히스토그램 빈의 수를 실험을 통해서 64로 정하였다. 히스토그램 등화에 의한 특징보상 효과를 확인하기 위한 음성인식 실험에서는, 훈련 및 평가용 MFCC 특징 데이터에 발화 단위로 히스토그램 등화 기법을 적용하여 얻어진 히스토그램 등화된 훈련 및 평가 데이터를 사용하였다. 이 히스토그램 등화 과정에서 필요한 훈련 누적분포함수는 전체 훈련 데이터로부터 사전에 구해진 추정치가 사용되었다. 반면에, 시험 누적분포함수로는 입력 발화마다 order-statistics 방식에 의해 새롭게 구해진 값을 사용하였다. 이러한 히스토그램 등화는 MFCC의 39차 계수 전체에 대해서 계수별로 적용하였으며 별도의 CMN이나 CMVN은 적용하지 않았다.

4.3. 성능 평가

<그림 5>는 Aurora 2 데이터베이스의 test sets A, B 및 C 의 noisy 음성 데이터에 대해, 클래스 수에 따른 제안된 클래스 히스토그램 등화기법의 음성인식 특징

보상 결과를 단어 오인식률 (WER: word error rate)로 나타낸다.



<그림 5> 음향 클래스 수에 따른 클래스 히스토그램 등화 기법의 음성인식 특징보상 성능.

이 그림에서 Oracle-Hard-Class와 Oracle-Soft-Class는 이상적인 클래스 정보가 제공되었을 때의 방법으로서, 음향 클래스 정보를 추출하기 위한 벡터 양자화 또는 GMM 기반의 확률 추정에서 noisy 음성 대신에 clean 음성을 사용했을 경우에 대한 하드-클래스 및 소프트-클래스 기반 히스토그램 등화 기법을 각각 의미한다. 이 그림에서 Oracle 클래스 정보를 적용한 두 결과를 보면, 클래스의 정보가 거의 정확할 경우에 제안된 클래스 히스토그램 등화 기법이 매우 뛰어난 음성특징 보상 효과를 가져옴을 알 수 있다. 클래스의 수가 증가할수록 보상효과도 따라서 증가함을 알 수 있는데, 이는 제안된 클래스 히스토그램 등화 기법의 원리와도 일치한다. 즉, 클래스 히스토그램 등화에서 더 많은 수의 클래스를 도입할수록 기존의 히스토그램 등화 기법이 가지는 두 가지 제한 조건을 더욱 효과적으로 감소시킬 수 있다. 반면에, 클래스 정보의 추출에서 noisy 음성을 사용한 Hard-Class와 Soft-Class의 경우를 보면, 처음에는 클래스의 수가 증가할수록 음성특징 보상 효과가 증가하지만 특정한 클래스의 수 (본 실험에서는 7) 이상에서는 오히려 보상 효과가 감소하는 패턴을 보임을 알 수 있다. 이는 noisy 음성을 클래스 정보의 추출에 사용할 경우에 클래스의 수가 증가함에 따라서 급격하게 저하되는 클래스 정보 추출의 정확도에서 기인한다. 즉, 클래스의 수가 7일 경우까지는 클래스 수의

증가에 따라 비례하는 클래스 히스토그램 등화의 장점에 의한 특징보상 효과의 증가폭이 클래스 정보 추출 정확도의 감소에 의한 특징보상 효과의 감소폭보다 크게 작용하지만, 그 보다 많은 클래스들에 대한 경우에서는 반대로 작용하여 전체적인 특징보상의 효과가 오히려 감소된다고 분석할 수 있다. 하드-클래스와 소프트-클래스에 대한 성능 결과를 보면, 예상한 바와 같이 소프트-클래스의 경우에서 특징보상 성능이 약간 우수함을 알 수 있다. 결론적으로, 이 그림이 나타내는 결과로부터 주목할 점은 정확한 클래스 정보의 추출이 제안된 클래스 히스토그램 등화 기법의 성능에 매우 중요한 요인으로 작용함을 알 수 있다. 따라서, 본 논문에서 제안된 기존의 히스토그램 등화기법에 의한 일차적인 잡음 감소 방법에 따른 클래스 정보 추출보다 우수한 새로운 클래스 정보 추출 방법이 제안되면 보다 개선된 특징보상 효과를 얻을 수 있음을 알 수 있다.

<표 1>은 Aurora 2 데이터베이스에 하드-클래스 (H-CHEQ) 및 소프트-클래스 (S-CHEQ) 기반의 히스토그램 등화 기법을 적용하여 음성인식의 특징을 보상하였을 때의 인식 결과를 단어 오인식률로 나타낸 결과이다. 이 실험에서, 하드-클래스 및 소프트-클래스 기반의 히스토그램 등화 기법에서 사용한 음향 클래스의 수는 <그림 5>의 결과에 따라서 7로 정하였다. 이 표에서, test sets A (sets A와 B의 clean 음성 데이터는 서로 동일)와 C의 clean 음성 데이터에 대해서, 하드-클래스 기반의 히스토그램 등화 기법은 MFCC에 비해 각각 63.20%와 72.28% 및 전체 평균으로 66.13%의 오인식률 증가를 나타내었고 소프트-클래스 기반의 히스토그램 등화 기법이 44.34%와 37.62% 및 전체적으로 42.17%의 오인식률 증가를 보였다. 반면에, test sets A, B 및 C의 noisy 음성 데이터에 대해서는, 하드-클래스 방식의 히스토그램 등화 기법은 MFCC에 비해 각각 58.95%, 65.25% 및 49.28%, 그리고 전체 평균에서 60.14%의 오인식률 감소 효과를 나타냄을 알 수 있다. 소프트-클래스 기반의 기법의 경우에는 MFCC에 비해 test sets A, B 및 C의 noisy 음성 데이터에서 각각 60.01%, 66.10% 및 50.63%, 그리고 전체 평균으로 61.17%의 오인식률 감소 효과를 나타낸다. 기존의 히스토그램 등화 기법 (HEQ)의 성능과 비교하면, 먼저, test sets A 와 C의 clean 음성 데이터에 대해서, 하드-클래스 방식의 경우는 각각 71.28%, 58.18%와 전체적으로 66.67%의 오인식률 증가를 보였으며 소프트-클래스 방식은 각각 51.48%와 26.36%, 전체적으로 42.62%의 오인식률 증가를 나타내었다. 반면에, test sets A, B 및 C의 noisy 음성 데이터에 대해서, 하드-클래스 방식의 경우는 각각 17.48%, 15.35%, 및 21.25%, 그리고, 전체적으로 17.49%의 오인식률 개선 효과를 보이고 있으며, 소프트-클래스 방식은 각각 19.60%, 17.43%, 및 23.35%, 그리고, 전체적으로 19.62%의 오인식률 개선 효과를 보인다. 이 결과를 분석하면 기본 음성특징인 MFCC에 비해 제안된 특징보상 방법이 clean 음성 데이터에 대해서 비록 절대적인 인식률 측면에서는 미미한 수준의 저하를 보이지만 상대적인 인식률 변동 측면에서 무시할 수 없는 성능 저하를 나타냄을 알 수 있다.

동시에, 음성인식의 실제 응용 측면에서 훨씬 큰 비중을 차지하는 잡음 환경을 반영하는 noisy 음성 데이터에 대해서는 MFCC에 비해서 현저한 인식성능의 개선을 가져옴을 알 수 있다. 또한, 이 noisy 음성 데이터에서 기존의 히스토그램 등화 기법에 비해서도 제안된 히스토그램 등화 기법은 의미있는 성능 개선을 보인다. 특히, 후자의 경우에는, 가산성 잡음만 부가된 test sets A와 B의 경우에 비해 가산성 잡음과 채널 잡음이 동시에 부가된 test set C에서 제안된 방식의 성능 개선 효과가 더 우수함을 알 수 있다. 이는 여러 종류의 상이한 특성을 가진 잡음이 혼재된 경우의 특징보상에서 제안된 클래스-기반 방식이 기존의 히스토그램 등화 기법에 비해 더 효과적으로 작용함을 의미한다.

<표 1> Aurora 2 데이터베이스에 대해 클래스 히스토그램 등화 기법을 적용한 음성인식 결과.

Test Noise			WER			
			MFCC	HEQ	H-CHEQ	S-CHEQ
A	Subway	Clean	1.17	1.20	1.93	1.57
		Noisy	30.14	18.67	15.92	15.14
	Babble	Clean	1.03	0.97	1.57	1.54
		Noisy	49.76	18.54	16.59	16.18
	Car	Clean	1.19	0.89	1.79	1.58
		Noisy	40.13	18.42	13.61	13.33
	Exhibit.	Clean	0.86	0.96	1.64	1.42
		Noisy	35.47	21.72	17.78	17.53
	Average	Clean	**1.06**	**1.01**	**1.73**	**1.53**
		Noisy	**38.88**	**19.34**	**15.96**	**15.55**
B	Rest.	Clean	1.17	1.20	1.93	1.57
		Noisy	48.49	17.99	17.26	16.87
	Street	Clean	1.03	0.97	1.57	1.54
		Noisy	38.48	18.16	15.21	14.85
	Airport	Clean	1.19	0.89	1.79	1.58
		Noisy	46.67	17.50	14.27	13.94
	Station	Clean	0.86	0.96	1.64	1.42
		Noisy	44.08	19.31	15.06	14.57
	Average	Clean	**1.06**	**1.01**	**1.73**	**1.53**
		Noisy	**44.43**	**18.24**	**15.44**	**15.06**
C	Subway + MIRS	Clean	0.98	1.11	1.84	1.41
		Noisy	32.77	22.28	17.93	17.46
	Street + MIRS	Clean	1.03	1.09	1.63	1.36
		Noisy	33.87	20.64	15.88	15.44
	Average	Clean	**1.01**	**1.10**	**1.74**	**1.39**
		Noisy	**33.32**	**21.46**	**16.90**	**16.45**
Overall Average		Clean	**1.04**	**1.04**	**1.73**	**1.48**
		Noisy	**39.99**	**19.32**	**15.94**	**15.53**

5. 결 론

잡음이 내재된 환경에서 음성인식의 성능 저하를 방지하기 위한 강인한 음성인식 기술의 하나로서 음성인식의 특징영역에서 효과적으로 음향 부정합을 보상하는 히스토그램 등화 기법이 제안되었으며 최근 활발히 연구되고 있다. 그러나, 기존의 히스토그램 등화 기법이 본래 의도한 특징보상 효과를 가지기 위해서는 단조 변환과 훈련 및 시험 클래스 확률분포의 동일성이라는 두 가정을 동시에 만족하여야 한다. 그러나, 대부분의 실질적인 음성인식 환경에서 이 두가지 조건은 만족되기 어렵다. 이와 같은 제한점에서 기인하는 기존의 히스토그램 등화 기법의 특징보상 저하를 해결하기 위하여, 본 논문에서는 클래스 정보를 이용하는 히스토그램 등화 기법을 제안하였다. 제안된 방식은 훈련 및 시험 누적분포함수의 모델링과 이들을 이용한 음성특징의 보상에서 복수의 음향 클래스를 도입하고 있으며, 음향 클래스 정보의 추출방법에 따라서 하드-클래스 정보와 소프트-클래스 정보를 이용하는 히스토그램 등화 기법으로 나누어진다. Aurora 2 데이터베이스를 이용한 음성인식 성능평가에서 제안된 방식은 noisy 음성 데이터에 대해서 MFCC 음성특징에 비해 61.17%, 그리고 기존의 히스토그램 등화 기법에 대해서는 19.62%의 인식률 개선을 각각 보였다. 또한, 하드-클래스 정보를 이용하는 방식에 비해 소프트-클래스 정보를 이용하는 방식이 특징보상과 인식률 개선에서 더 우수한 효과를 나타내었다. 그러나, clean 음성 데이터에 대해서는 각각 42.17%와 42.62%의 인식률 저하를 보였다.

실험결과로부터, 제안된 클래스 히스토그램 등화 기법의 특징보상 효과는 음향 클래스 정보 추출의 정확도와 밀접한 비례 관계가 있음을 확인하였다. 따라서, 향후에 잡음환경에서 좀 더 정확하게 음향 클래스 정보를 추출할 수 있는 기술에 대한 연구가 필요하다. 또한, 클래스 개념의 도입에 따라, 개별 클래스에 할당되는 샘플 데이터 양의 감소에서 기인하는 클래스 시험 누적분포함수 추정의 신뢰성 저하 문제의 개선도 요구된다. 추가적으로, clean 음성에 대한 성능저하를 보완할 수 있는 방법에 대한 연구도 필요하다. 마지막으로, 제안한 방법은 multi-condition training 환경에 대한 내부적인 성능평가에서는 기존의 히스토그램 등화 기법에 비해 개선된 결과를 나타내지 못했다. 이는 음향 클래스 추출을 위한 훈련 과정에서 multi-condition training 환경의 noisy 음성 데이터를 사용할 경우에 clean 음성에 대한 정확한 음향 클래스의 추출이 어려워지는데서 기인한다. 따라서, 제안된 방법이 multi-condition training 환경에서 기존의 히스토그램 등화 기법 보다 개선된 성능을 얻기 위해서는 음향 클래스 정보 추출과 관련된 훈련과 실제 추출 과정에서 잡음이 감소된 음성 데이터의 사용이 필요하다고 본다. Multi-condition training 환경에서 이러한 접근방법을 제안된 히스토그램 등화 기법에 적용하였을 때의 성능평가도 향후 필요한 연구항목들 중의 하나이다.

참 고 문 헌

[1] J. C. Junqua and J. P. Haton, *Robustness in Automatic Speech Recognition*, Boston, MA: Kluwer Academic Press, 1996.

[2] A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*, Boston, MA: Kluwer Academic Press, 1992.

[3] B.-H. Juang, "Speech recognition in adverse environments", *Computer Speech, Language*, Vol. 5, pp. 275-294, 1991.

[4] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Upper Saddle River, NJ: Prentice-Hall, 2001.

[5] A. Sankar and C.-H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition", *IEEE Trans. Speech, Audio Processing*, Vol. 4, No. 3, pp. 190-202, May 1996.

[6] N. S. Kim, "Speech recognition under noisy environments", *Telecommunications Review*, Vol. 13, No. 5, pp. 650-661, Oct. 2003.

[7] B. S. Atal, "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, 1974.

[8] A. E. Rosenberg, C.-H. Lee, and F. K. Soong, "Cepstral channel normalization techniques for HMM-based speaker verification", *Proc. ICSLP*, pp. 1835-1838, 1992.

[9] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition", *Speech Communication*, Vol. 25, pp. 133-147, 1998.

[10] P. J. Moreno, B. Raj, and R.M. Stern, "A vector Taylor series approach for environment independent speech recognition", *Proc. ICASSP*, pp. 733-736, 1996.

[11] N. S. Kim, "Statistical linear approximation for environment compensation", *IEEE Signal Processing Letters*, Vol. 5, No. 1, pp. 8-10, Jan. 1998.

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Upper Sabble River, NJ: Prentice-Hall, 2002.

[13] S. Dharanipragada and M. Padmanabhan, "A nonlinear unsupervised adaptation technique for speech recognition", *Proc. ICSLP*, pp. 556-559, 2000.

[14] F. Hilger and H. Ney, "Quantile based histogram equalization for noise robust speech recognition", *Proc. EUROSPEECH*, pp. 1135-1138, 2001.

[15] 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 Processing Letters*, Vol. 11, pp. 517-520, May 2004.

[16] 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 and Audio Processing*, Vol. 13, pp. 355-366, May 2005.

[17] S.-T. Kim, H.-R. Kim, "Robust histogram equalization using compensated probability distribution", *말소리*, 55호, pp. 131-142, 2005.

[18] Y. Suh, "*Acoustic feature compensation by class-based histogram equalization for robust*

speech recognition", Ph. D. dissertation, Information and Communications University, Korea, 2006.

[19] Y. Suh, S.-B. Kwon, H. Kim, "Feature compensation with class-based histogram equalization for robust speech recognition", *Proc. WESPAC IX*, Jun. 2006.

[20] Y. Suh and H. Kim, "Class-based histogram equalization for robust speech recognition", *ETRI JOURNAL*, Vol. 28, No. 4, pp. 502-504, Aug. 2006.

[21] Y. Suh, M. Ji, H. Kim, "Probabilistic class histogram equalization for robust speech recognition", *IEEE Signal Processing Letters*, Vol. 14, No. 4, Apr. 2007 (to be published).

[22] D. Pearce, H.-G. Hirsh, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions", *Proc. ICSLP*, Oct. 2000.

[23] S. Young et al., *The HTK Book for HTK version 3.2.1*, Cambridge University Engineering Department, 2002.

접수일자 : 2006년 11월 8일
게재결정 : 2006년 12월 22일

▶ 서영주 (Youngjoo Suh) : 교신저자
주소: 305-732 대전광역시 유성구 문지로 119번지
소속: 한국정보통신대학교(ICU) 공학부
전화: 042) 866-6221
E-mail: yjsuh@icu.ac.kr

▶ 김회린 (Hoirin Kim)
주소: 305-732 대전광역시 유성구 문지로 119번지
소속: 한국정보통신대학교(ICU) 공학부
전화: 042) 866-6139
E-mail: hrkim@icu.ac.kr

▶ 이윤근 (Yunkeun Lee)
주소: 305-700 대전광역시 유성구 가정동 161번지
소속: 한국전자통신연구원(ETRI) 음성언어정보연구센터 음성처리연구팀
전화: 042) 860-1869
E-mail: yklee@etri.re.kr