

## 잡음 환경하에서의 PSO-NCM을 이용한 거절기능 성능 향상\*

Enhancement of Rejection Performance using the PSO-NCM  
in Noisy Environment

김 병 돈\*\* · 송 민 규\*\* · 최 승 호\*\*\* · 김 진 영\*\*

Byoung Don Kim · Min Gyu Song · Seung Ho Choi · Jin Young Kim

## ABSTRACT

Automatic speech recognition has severe performance degradation under noisy environments. To cope with the noise problem, many methods have been proposed. Most of them focused on noise-robust features or model adaptation. However, researchers have overlooked utterance verification (UV) under noisy environments.

In this paper we discuss UV problems based on the normalized confidence measure. First, we show that UV performance is also degraded in noisy environments with the experiments of an isolated word recognition. Then we observe how the degradation of UV performances is suffered. Based on the UV experiments we propose a modeling method of the statistics of phone confidences using sigmoid functions. For obtaining the parameters of the sigmoidal models, the particle swarm optimization (PSO) is adopted. The proposed method improves 20% rejection performance. Our experimental results show that the PSO-NCM can apply noise speech recognition successfully.

**Keywords:** normalized CM(Confidence Measure), particle swarm optimization, rejection performance

## 1. 서 론

음성인식 기술은 1960년대 중반 이래로 꾸준히 연구되어 실험실 환경에서 상당한 성과를 거두었다. 그러나 음성인식기를 실제 서비스 환경에서 사용하고자 할 때, 주변 잡음 때문에 인식 성능은 크게 저하된다. 따라서 잡음 환경하에서 인식기 성능 향상에 대한 연구가 활발하게 진행되고 있다 [1-6]. 잡음 환경에 강한 음성인식 알고리즘 개발은 주로 잡음에 강인한 특징 파라미터를 결정하는 방법 그리고 인식기의 통계적 모델을 주변 환경에 따라 적응적으로 변경하는 문제를 중심으로 이

---

\* 이 논문은 2007년도 정부재원(교육인적자원부 학술 연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었습니다(KRF-2007-521-D00447).

\*\* 전남대학교 전자정보통신공학과

\*\*\* 동신대학교 컴퓨터학과

루어져 왔다. 환경 변화에 따른 특징 파라미터를 추출하는 CMS(Cepstrum Mean Subtraction)[4] 및 CMN(Cepstral Mean Normalization)[5]과 다중 모델을 이용한 SNR 분류 인식 기법[6] 등이 제안되었다. 위의 방법은 잡음에 강인한 특징 파라미터를 추출하여 인식성능을 보상하는 방안이며, 실제 인식성능을 향상시키는데 유용한 방법이다.

잡음에 강한 파라미터 또는 모델 적용 등의 방법을 사용한다고 하더라도, 잡음 처리 문제는 여전히 해결되지 못한 문제이다. 그런데 한 가지 잡음 환경 하 음성인식에서 간과되어온 것은 발화 검증 기술의 견인성에 관한 것이다. 잡음 환경 하 음성인식기의 오인식률이 증가하기 때문에, 잡음 환경에서의 발화검증은 더 중요한 역할을 하게 된다.

일반적으로 발화 검증시스템에서 거절기능은 오인식 및 인식사전에 등록되지 않은 단어를 발성한 경우에 거절하는 과정이며 신뢰도로 결정된다. phoneme과 tri-phone의 신뢰도는 통계적 가설을 사용하여 구해지나 부가 잡음이 첨가되어 인식된 경우 잡음은 음성 특징 벡터를 변형시키며, 거절기능의 저하를 가져오는 원인이 된다.

이러한 원인을 해결하기 위해 본 논문에서는 백색잡음에서 SNR에 따른 정규화 방안에 대하여 연구하였다. 정규화한 방법론은 다음과 같다. 첫째, 잡음에 따른 음소 신뢰도를 측정하여 음소 신뢰도의 평균과 표준편차 구하고 이를 정규화 하였다. 둘째, 정규화의 최적화를 위하여 적용된 PSO (Particle Swarm Optimization)로 평균과 표준편차를 이용하여 음소 모델을 정규화 하였다. 잡음에 따른 음소 모델을 정규화하여 단어 인식에서 거절기능의 성능 개선에 관한 연구를 하였다.

## 2. 거절기능 성능평가

### 2.1 고립단어 인식기의 구성과 데이터베이스

거절기능 성능평가에 사용된 음성인식 시스템은 고립단어 인식기를 사용하였다. 고립단어 인식기는 전처리부, 인식부, 후처리부로 구성되어 있으며 전처리부는 인식에 필요한 특징 파라미터를 추출하고, 인식부는 단어 사전을 이용하여 인식 단어를 출력한다[7]. 후처리부는 음소 모델과 반음소 모델의 상대적 유사도를 통해서 발화여부를 검증한다.

단어사전의 훈련용 데이터베이스는 국내 지명 및 도로명으로 구성된 600 단어로써 남성화자 52명을 HTK 3.0을 이용하여 훈련하였다. 시험용 데이터베이스는 1,000 단어를 사용하였으며 훈련에 참가한 남성화자 18명이 발성한 단어사전의 600 단어와 단어사전에 없는 400 단어로 구성되었다. 실험에 사용된 잡음 데이터베이스는 clean speech에 white gaussian noise를 식(1)과 같이 SNR에 따라 첨가하여 구성하였다.

$$\hat{S}_n(s) = S_n(s) + WGN_n(s) \quad (1)$$

각 입력 음성에 따라 5, 10, 15, 20 dB로 구성하였다. <그림 1>은 “서강백범로”의 음성에 SNR 15dB 잡음을 첨가한 것이다.

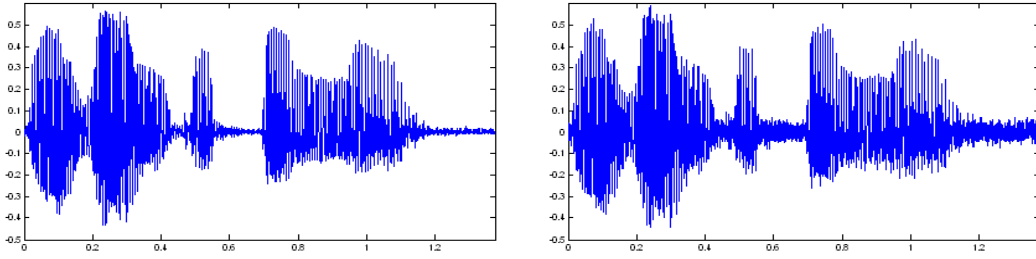


그림 1. “서강백범로”에 15 dB의 잡음을 첨가한 음성

## 2.2 반응소와 정규화 신뢰도

신뢰도의 일반적인 측정방법은 음소 모델과 반응소 모델의 유사도를 이용하여 음소 신뢰도를 계산한 뒤 단어 단위로 확장하여 신뢰도(RLJC-CM)를 측정한다. 반응소 모델은 인식된 음소를 제외한 나머지 음소를 의미하며 실험에 사용된 모든 음소는 51 개이고 반응소는 50 개이다. 식(2)와 식(3)은 RLJC-CM 방법이다[8].

$$\log pr_a = \frac{1}{M} \sum_{i=0}^{M-1} \log pr_{a_i}, \quad cm_p = \frac{\log pr_p - \log pr_a}{|\log pr_p|} \quad (2)$$

$$CM = \frac{1}{f_{cm}} \log \left( \frac{\sum_{p=0}^{n_p-1} \exp(f_{cm} \cdot cm_p)}{n_p} \right) \quad (3)$$

위의 수식에서  $n_p$ 는 구성 음소의 수를 나타내며,  $f_{cm}$ 은 CM에 사용된 음의 상수로서 -5로 세팅되었다.  $cm_p$ 는 음소단위 신뢰도이고,  $\log pr_p$ 는 음소 모델의 로그 확률이며,  $\log pr_a$ 는 반응소 모델의 로그확률이다. CM은 단어 신뢰도로써 음소 신뢰도를 이용하여 구해진다. 본 논문에서는 저자들이 제안한 방법을 통해 얻어진 결과를 활용하여 정규화 방법을 개선하고자한다.

### 2.2.1 반응소 기반 정규화 신뢰도

정규화 신뢰도(Normalized CM)는 기존 신뢰도의 통계적 분포가 불안정한 현상을 해결하기 위해 신뢰도가 정규분포를 가진다는 가정 하에 각 phoneme, tri-phoneme 신뢰도들의 평균과 표준편차를 계산하고 정규화한다[9]. 식(4)와 식(5)는 참고문헌 [9]에 NCM방법이다.

$$ncm_p = \frac{cm_p - mean_p}{sd_p} + \alpha \quad (4)$$

$$NCM = \frac{1}{f_{ncm}} \log \left( \frac{\sum_{p=0}^{n_p-1} \exp(f_{ncm} \cdot ncm_p)}{n_p} \right) \quad (5)$$

위의 수식에서  $n_{cm_p}$ 는 정규화된 음소의 로그확률이며,  $f_{ncm}$ 은 NCM에 사용된 음의 상수이며,  $-0.5$ 로 세팅되었다.  $mean_p$ ,  $sd_p$ 는 해당 tri-phone 신뢰도의 평균과 표준편차를 말한다.  $\alpha$ 는 정규화에 사용되는 가중치이며, 실험에서 사용된 값은 2를 사용하였다.

### 2.3 거절기능 성능 평가 기준 설정

거절기능의 성능평가를 위한 평가 기준으로 MDR(Miss Detection Ratio)과 FAR(False Alarm Ratio)을 사용하며 식(6)과 식(7)로 구해진다. MDR은 인식 과정에서 인식 단어가 정확하게 인식되지 못하는 경우를 뜻하며, FAR은 사전외단어(out-of-vocabulary)가 인식된 경우를 말한다. 기존의 고립단어 인식기에 잡음이 첨가된 음성을 고립단어 인식기에서 거절 성능을 테스트하였다.

$$MDR = \frac{\text{Number of missed detection}}{\text{Total number of recognition word}} \quad (6)$$

$$FAR = \frac{\text{Number of false alarm}}{\text{Total number of recognition word}} \quad (7)$$

#### 2.3.1 SNR에 따른 거절기능 및 인식 성능 평가 실험

잡음에 따른 거절기능 성능을 평가하기 위해 SNR 0에서 40 dB까지 5 dB 간격으로 FAR과 MDR을 측정된 결과 20 dB 이상에서는 잡음이 무의미함을 알 수 있었다(<그림 2>). 잡음에 따른 인식 성능 평가에서는 거절기능 성능 평가와 동일한 조건으로 인식한 결과 5 dB에서의 인식 성능은 잡음으로 인해 매우 낮고, 10 dB와 15 dB 사이에서 급격한 성능 변화를 보이고 있음을 알 수 있다(<표 1>).

표 1. 잡음에 따른 인식 성능 평가

| SNR      | 5   | 10   | 15   | 20   | 25   | 30   | 40   | 잡음 없음 |
|----------|-----|------|------|------|------|------|------|-------|
| 인식 성능(%) | 5.1 | 45.0 | 74.4 | 85.2 | 88.8 | 90.3 | 91.1 | 91.2  |

#### 2.3.2 신뢰도에 따른 거절 성능의 비교 실험

신뢰도에 따른 거절 성능실험은 잡음이 없는 음성(SNR:0dB)과 잡음이 첨가된 음성(SNR:15dB)에 RLJC-CM와 NCM대해 FAR과 MDR을 비교하여 <그림 3>에 나타내었다. 잡음 없는 음성에 대하여 MDR 8.9%에서 RLJC-CM의 FAR은 0.41이며 NCM의 FAR은 0.29이다. 실험 결과에서 잡음이 존재하지 않을 때의 성능은 NCM이 RLJC-CM보다 우수함을 알 수 있다. 그러나 잡음이 첨가된 음성에서 MDR 22.9%에서 RLJC-CM의 FAR은 0.65이며 NCM의 FAR은 0.61이며, 두 신뢰도의 성능은 비슷하다. 이는 잡음의 영향으로 음소 확률이 변형되어 음소 신뢰도에 대한 정규화가 이루어지지 않았음을 의미한다.

잡음에 의한 NCM의 성능 감소를 이를 해결하기 위해 본 논문에서는 잡음 환경에서 음소 신뢰도를 추출하고, PSO를 이용하여 SNR에 따른 음소 신뢰도 정규화 방안에 대하여 연구하였다.

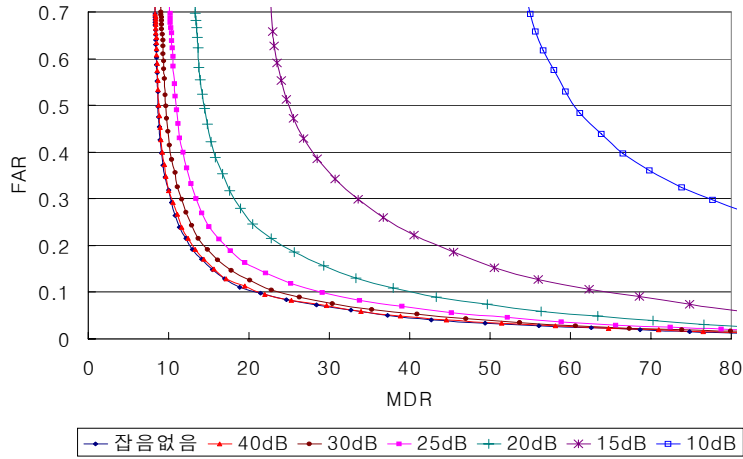


그림 2. 잡음에 따른 거절기능 성능 평가

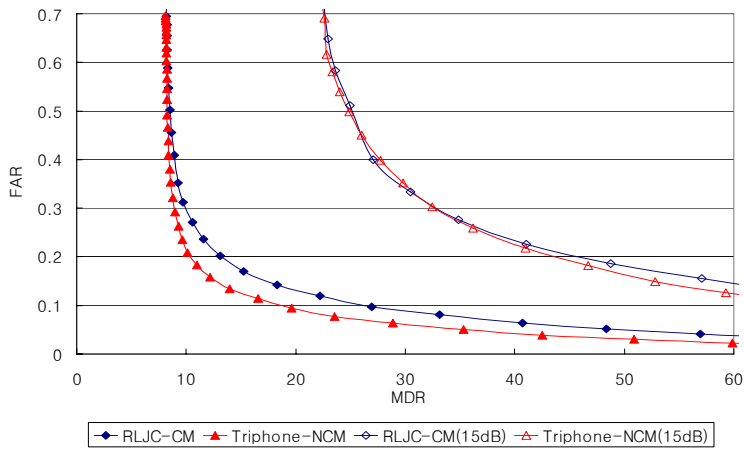


그림 3. 신뢰도에 따른 거절 성능 비교

### 3. 잡음 환경의 거절기능 성능 향상

#### 3.1 잡음환경의 정규화 신뢰도의 측정

앞 절에서 입력된 음성에 잡음이 포함될 경우 음소 신뢰도에 영향을 주어 전체 인식 성능에 영향을 미치게 됨을 실험결과 확인하였다.

따라서 본 논문에서는 SNR에 따른 음소 신뢰도를 구하기 위해 (1)훈련에 사용된 52 명의 훈련 데이터에 잡음(SNR: 10, 15, 20, 25, 30, 40 dB)을 첨가하고, (2)각 SNR의 음소 신뢰도 평균과 표준편차를 구하고 잡음환경에서의 정규화 신뢰도를 측정하였다.

본 논문에서는 각 음소에 따른 SNR를 구하기 위해 <그림 4>와 같은 과정을 거쳐 신뢰도를 측정하였다. 왜냐하면, 실제 음소가 발생하는 구간은 훈련 데이터마다 다르며, 음소구간 내의 SNR은 입력 음성 전체의 SNR과 다른 값을 갖게 된다. 이것은 SNR에 따른 음소 신뢰도를 구성하는데 정확한 표본 추출이 불가능함을 의미한다. 그림에서는 음소 구간내의 정확한 SNR 정보를 얻기 위하여 기존의 clean speech와 잡음이 포함된 입력 신호를 비교하여 음소 구간에 삽입된 잡음의 양을 계산하고 음소 구간별로 그 값을 구한 뒤 SNR에 따른 음소별 신뢰도를 측정하였다.

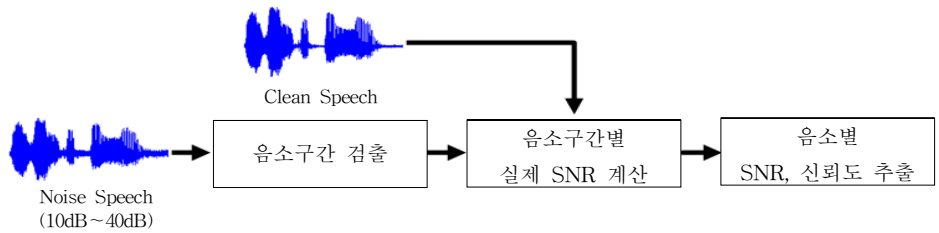


그림 4. SNR 음소 신뢰도 추출 과정

SNR에 측정은 본 논문에서 제안한 방법이다. <그림 4>에서 측정된 음소 신뢰도는 SNR 정보를 가지고 SNR에 따른 음소 신뢰도의 평균과 표준편차를 구하여 <그림 5>와 같은 음소 신뢰도의 분포를 알 수 있다. SNR window size는 평균 계산 과정에서 사용된 4를 사용하였다. <그림 5>는 음소 'aa', 'dd'에 대한 평균과 표준편차이며 잡음이 감소할수록 음소 확률은 증가하지만 특정 구간에서 음소 확률이 증가하거나 감소함을 알 수 있다. 이는 SNR에 따라 음소 신뢰도의 사용된 개수가 각 음소마다 다르고 음소 사용 빈도가 다르기 때문에 나타나는 현상이라고 볼 수 있다. 따라서 사용된 음소개수가 평균적으로 다르다는 것은 정규화 과정의 성능 저하를 가져오게 되는 원인이 된다.

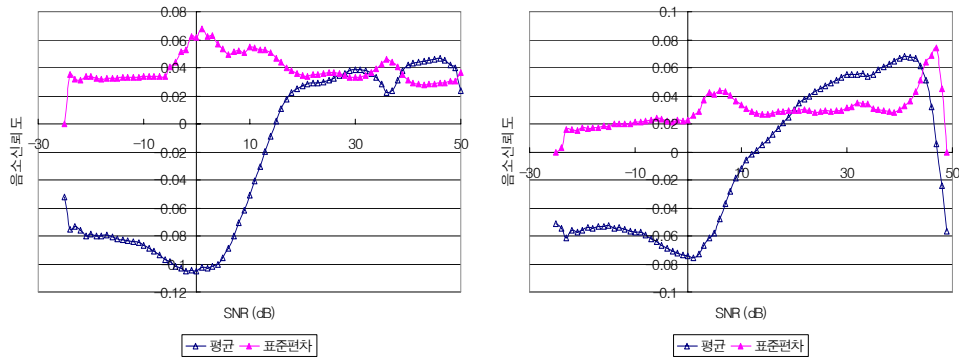


그림 5. SNR에 따른 음소 신뢰도 분포(왼쪽: 'aa', 오른쪽: 'dd')

음소 신뢰도 계산 과정에서 사용된 음소 개수는 <표 2>와 같으며 음소 분포는 <그림 6>과 같이 나타났다. 여기에서 추출된 음소는 적용된 SNR에 따라 0~40 dB 사이의 음소가 많이 존재함을

알 수 있다. 음소 신뢰도의 분포는 최적화 관점에서 보면 일반적이지 못한 특수한 경우이며 확률 분포로 정규화 할 경우 인식 과정에서 음소에 따른 신뢰도 정보를 검색하게 되어 인식 성능이 감소 된다는 사실을 이 실험을 통해서 얻어진 결과이다.

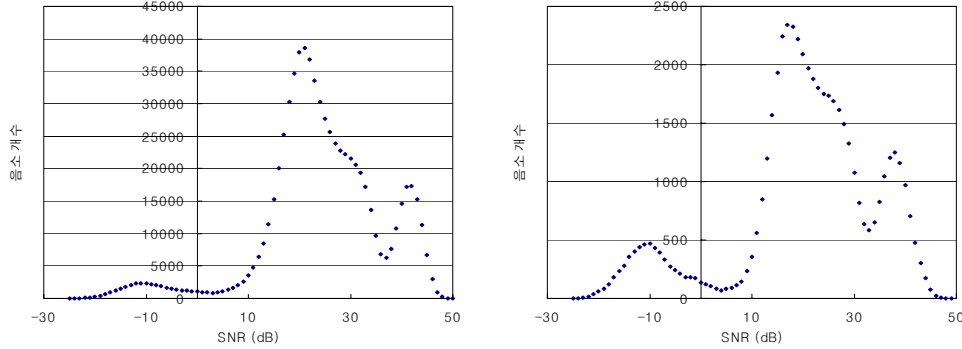


그림 6. 음소 신뢰도 계산에 사용된 음소 개수 비교(왼쪽: 'aa', 오른쪽: 'dd')

표 2. SNR에 따른 음소 개수

| dB  | aa   | dd  | dB | aa   | dd  | dB | aa    | dd   | dB | aa    | dd   | dB | aa    | dd   |
|-----|------|-----|----|------|-----|----|-------|------|----|-------|------|----|-------|------|
| -24 | 19   | 3   | -9 | 2203 | 429 | 6  | 1326  | 93   | 21 | 38546 | 1966 | 36 | 6858  | 1042 |
| -23 | 55   | 8   | -8 | 2086 | 394 | 7  | 1644  | 116  | 22 | 36790 | 1876 | 37 | 6265  | 1204 |
| -22 | 96   | 13  | -7 | 1866 | 331 | 8  | 2101  | 146  | 23 | 33555 | 1800 | 38 | 7653  | 1247 |
| -21 | 181  | 39  | -6 | 1695 | 275 | 9  | 2658  | 232  | 24 | 30318 | 1749 | 39 | 10839 | 1159 |
| -20 | 312  | 64  | -5 | 1509 | 239 | 10 | 3543  | 358  | 25 | 27730 | 1735 | 40 | 14620 | 973  |
| -19 | 470  | 85  | -4 | 1375 | 213 | 11 | 4734  | 557  | 26 | 25579 | 1686 | 41 | 17142 | 706  |
| -18 | 665  | 122 | -3 | 1208 | 181 | 12 | 6412  | 849  | 27 | 23808 | 1615 | 42 | 17265 | 474  |
| -17 | 916  | 183 | -2 | 1283 | 183 | 13 | 8445  | 1194 | 28 | 22823 | 1493 | 43 | 15259 | 306  |
| -16 | 1191 | 235 | -1 | 1157 | 174 | 14 | 11476 | 1566 | 29 | 22245 | 1328 | 44 | 11321 | 174  |
| -15 | 1545 | 283 | 0  | 1048 | 138 | 15 | 15252 | 1931 | 30 | 21564 | 1073 | 45 | 6631  | 74   |
| -14 | 1833 | 353 | 1  | 968  | 119 | 16 | 20082 | 2242 | 31 | 20645 | 815  | 46 | 2939  | 25   |
| -13 | 2065 | 403 | 2  | 962  | 105 | 17 | 25194 | 2338 | 32 | 19318 | 639  | 47 | 982   | 10   |
| -12 | 2253 | 440 | 3  | 843  | 84  | 18 | 30247 | 2327 | 33 | 17154 | 584  | 48 | 225   | 3    |
| -11 | 2349 | 461 | 4  | 927  | 66  | 19 | 34626 | 2223 | 34 | 13693 | 651  | 49 | 43    | 2    |
| -10 | 2302 | 466 | 5  | 1089 | 84  | 20 | 37932 | 2093 | 35 | 9634  | 822  | 50 | 7     | 0    |

### 3.2 PSO(Particle Swarm Optimization)을 적용한 신뢰도

3.1 절에서는 잡음환경에서 음소신뢰도에 대한 적용방법이 적절하지 않음을 본 실험을 통해서 알 수 있었다. 따라서 본 논문에서는 잡음환경에 적합한 음소 신뢰도의 최적화 방안으로 PSO (Particle Swarm Optimization)를 적용하여 음소 신뢰도의 평균을 잡음의 함수로 사용하였다.

참고문헌[10-11]으로부터 PSO는 초기의 불규칙한 해들의 모임으로 시작하여 각 잠재적인 해들이 다시 랜덤한 속도와 이전 잠재적인 해들의 결합으로 구성된다. 잠재적인 해들의 모임을 Particle Swarm이라고 하며, 파라미터의 최적화 함수  $f()$ 가 구해진다. 수학적으로 계산하기 힘든 비선형 문제의 해결방법 중의 하나인 PSO 방법은 함수  $f()$ 가 비선형일 경우 해가 국부적인 최적 값을 피

하면서 전체 최적의 해를 구하는데 사용된다. 다음은 PSO의 해를 구하는 방법이다.

- 1) 랜덤한 잠재적인 해  $\{P_{i0}\}$ 를 결정한다.
- 2) 각  $j$ 에 대하여 아래와 같이 반복한다.
  - 2-1) 각  $P_{ij}$ 에 대하여  $f(P_{ij})$ 를 계산한다.
  - 2-2) 최적  $f()$  값의 변화를 계산하고 수렴한 경우 루프를 빠져나간다.
  - 2-3) 각  $i$ 에 대하여  $\{0, \dots, j-1\}$ 에 대하여 가장 최적인(본 논문에서는 가장 작은)해를 저장한다. 이 값을  $pbest_{ij}$ 라고 한다.
  - 2-4) 모든  $pbest_{ij}$ 를 대상으로 가장 최적인 해를 저장하여 이를  $gbest_j$ 라고 한다.
  - 2-5) 각 입자의 속도는 다음과 같이 계산한다.
 
$$v_{ij} = v_{ij-1} + c_1 r_1 (pbest_{ij} - P_{ij}) + c_2 r_2 (gbest_j - P_{ij})$$
 위 식에서  $c_1$ 과  $c_2$ 는 상수이며,  $r_1$ 과  $r_2$ 는 임의의 수이다.
  - 2-6) 각 particle 값을 갱신한다.
 
$$P_{ij} = P_{ij-1} + v_{ij}$$
- 3)  $gbest_j$ 를 최적의 해로 결정한다.

### 3.2.1 음소 신뢰도 최적화 $M_{phoneme}()$

최적화 함수의 파라미터는 음소에 따라 모두 다른 값을 갖게 되며, 음소 신뢰도의 평균 분포에 따라 재구성되며 식(8)과 같다. 각 음소별 파라미터( $a, b, c, d$ )의 최적화 값은 잡음에 따른 음소 신뢰도 함수로 구해지기 때문에 잡음 환경에서의 정규화 신뢰도의 적용이 가능하다.

$$M_{phoneme}(SNR) = a \left[ \frac{1}{1 + e^{b(snr+c)}} \right] + d \quad (8)$$

여기에서  $a$ 는 스케일 파라미터,  $b$ 는 기울기 파라미터,  $c, d$ 는 x, y축의 이동 파라미터이다.

전체 51개 음소에 대한 음소 신뢰도 분포를 분석한 결과 그 평균은 잡음에 따라 일정하게 증가하거나 감소하며 특정 영역에서 수렴하는 형태를 갖게 됨을 <그림 5>에서 알 수 있었다. 따라서 음소 신뢰도의 최적화 함수는 시그모이드 함수를 사용하였다.

### 3.2.2 최적화 멤버십 함수 $f()$ 의 구성

시그모이드 함수로 표현된 음소 신뢰도의 최적화 함수는 평균에 대응 되는 멤버십 함수를 사용하였으며, 가중치는 사용된 음소 개수에 따른 값이다. 최적화 멤버십 함수  $f()$ 는 식(9)와 같다.

$$f(a, b, c, d) = \frac{1}{m} \sqrt{\sum_{k=1}^m N_{x_k} \left[ a \left[ \frac{1}{1 + e^{b(x_k+c)}} \right] + d - M(x_k) \right]^2} \quad (9)$$



여기에서 사용된 함수의 파라미터는 4개이며  $N_{x_k}$ 는 신뢰도 평균에 사용된 음소의 개수,  $M(x_k)$ 는 실제 음소 신뢰도의 평균값이다.

따라서  $f(a, b, c, d)$  값이 작을수록  $M(x_k)$ 에 가까운 신뢰도 확률 그래프가 <그림 7>과 같이 나타난다.  $f()$ 가 최소화될 때 파라미터  $a, b, c, d$ 를 통하여 최적화된 음소 신뢰도가 구해진다.

### 3.2.3 음소 신뢰도의 비교

기존의 음소 신뢰도와 PSO 적용후 최적화된 음소 신뢰도의 비교 결과를 <그림 7>에 나타내었다. PSO 적용 후 멤버쉽 함수는 일정한 기울기를 갖으며 수렴함을 알 수 있었다. 즉, 음소 확률에서 사용된 음소의 개수가 작은 영역은 식(9)의 가중치가 적용되어 평균 그래프에서 일정하지 않게 큰 변화를 일으키는 영역에 대한 보상이 이루어져 수렴하도록 수정하였다.

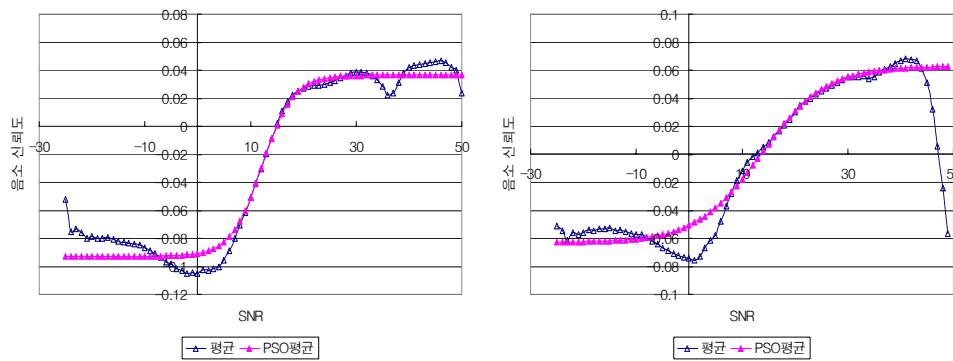


그림 7. 기존의 음소 신뢰도와 PSO 적용 후의 음소 신뢰도 비교(왼쪽: 'aa', 오른쪽: 'dd')

## 4. 실험 및 실험 결과

기존의 RLJC-CM과 NCM, PSO-NCM을 비교하였다. 표준 편차는 평균을 계산하여 일괄적으로 적용하였으며, <표 3>은 15 dB에서의 신뢰도 실험 결과를 나타낸 것이며, <그림 8>은 SNR(15dB, 20dB)에 따른 거절성능에 관한 그림이다.

실험을 통해 두 가지의 사실을 확인 하였다. 첫째, RLJC-CM에 비하여 기존의 정규화 신뢰도의 거절기능 성능이 우수하지만 잡음 음성에서의 정규화 신뢰도는 잡음이 증가할수록 성능이 나빠짐을 알 수 있다. 둘째, 잡음에 적용된 음소 정규화 신뢰도(PSO-NCM)는 잡음이 증가할수록 성능이 우수함을 <그림 8>과 <표 3>에서 확인할 수 있다.

<표 3>에서 MDR 23% 부근에서 NCM의 FAR 0.61이며, PSO-NCM의 FAR은 0.49이다. 이는 NCM에 비해 PSO-NCM의 성능은 20% 향상됨을 의미한다. SNR 20dB를 살펴보면 MDR 22% 부근에서 NCM의 FAR은 0.21이며, PSO-NCM의 FAR은 0.16이다.

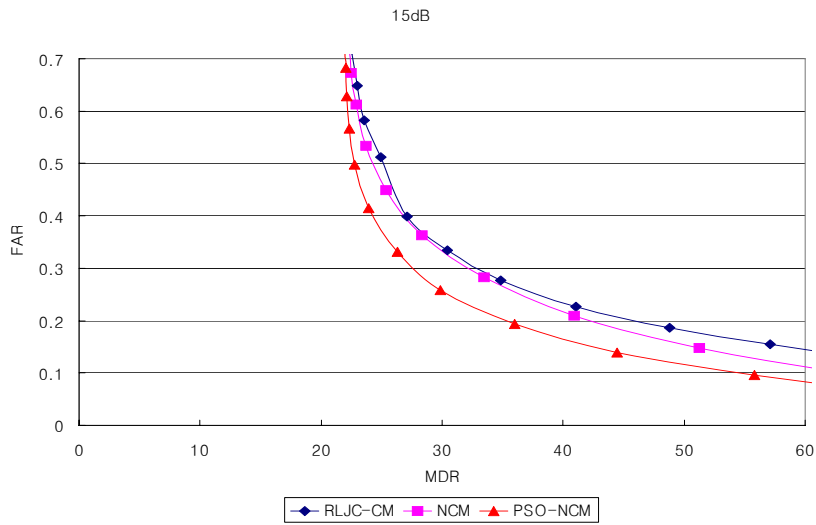
이러한 결과는 (1)잡음이 증가할수록 잡음에 적용된 최적화 정규화 신뢰도를 사용했을 경우 높

은 거절기능 성능 향상을 보임을 알 수 있었다. (2)잡음이 줄어들수록 NCM과 PSO-NCM의 그래프가 일치해감을 알 수 있었다. 이는 잡음이 존재하지 않을 경우에도 PSO-NCM 적용이 거절 기능 향상을 가져올 수 있음을 보여준다.

따라서 PSO-NCM은 경험에서 계산되어진 음소 테이블에 비해 4 개의 파라미터를 이용하여 각 음소에 따라 음소 멤버십 함수 구성이 가능하므로 잡음에 최적화된 신뢰도입을 알 수 있다.

표 3. 15 dB에서의 신뢰도 실험 결과

| RLJC-CM |       | NCM    |       | PSO-NCM |       |
|---------|-------|--------|-------|---------|-------|
| MDR     | FAR   | MDR    | FAR   | MDR     | FAR   |
| 22.583  | 0.709 | 22.037 | 0.810 | 21.972  | 0.722 |
| 22.962  | 0.648 | 22.074 | 0.792 | 22.027  | 0.682 |
| 23.601  | 0.582 | 22.194 | 0.766 | 22.129  | 0.628 |
| 24.944  | 0.512 | 22.324 | 0.726 | 22.333  | 0.567 |
| 27.083  | 0.399 | 22.481 | 0.672 | 22.787  | 0.498 |
| 30.444  | 0.333 | 22.935 | 0.612 | 23.944  | 0.413 |
| 34.870  | 0.276 | 23.731 | 0.533 | 26.287  | 0.331 |
| 41.037  | 0.225 | 25.388 | 0.448 | 29.851  | 0.258 |
| 48.787  | 0.186 | 28.333 | 0.362 | 36      | 0.193 |
| 57.101  | 0.155 | 33.462 | 0.282 | 44.462  | 0.138 |
| 65.342  | 0.127 | 40.935 | 0.208 | 55.805  | 0.095 |
| 73.574  | 0.103 | 51.277 | 0.148 | 67.037  | 0.066 |
| 80.564  | 0.084 | 62.731 | 0.101 | 77.888  | 0.047 |



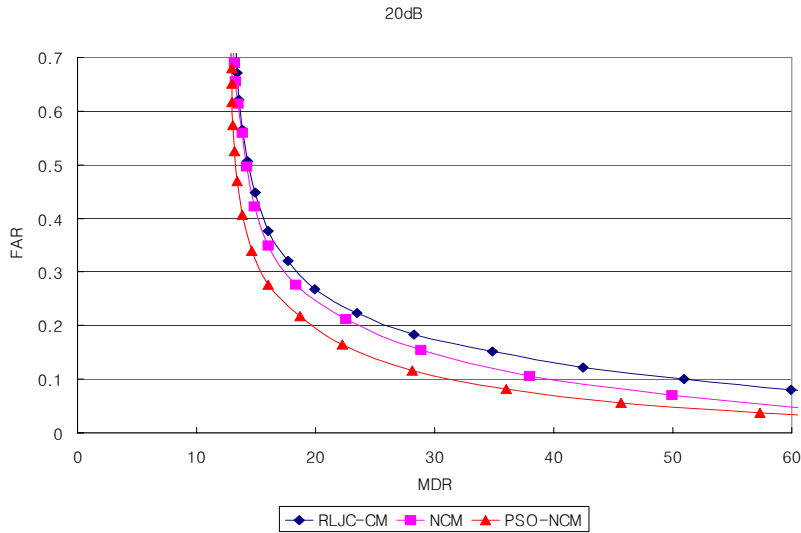


그림 8. 잡음에 따른 거절 성능 평가(15, 20 dB)

## 5. 결 론

본 논문에서는 잡음에 따른 음소 모델을 정규화하여 단어 인식기에서 거절 기능의 성능 개선에 관하여 연구하였다. 실험에서 첫째, 잡음에 대한 정규화는 잡음에 따른 음소 신뢰도의 평균과 표준 편차를 이용하여 SNR에 따른 음소 신뢰도와 음소 개수를 구하고 둘째, 잡음에 대한 음소 신뢰도를 최적화하기 위하여 PSO를 적용하였다. 이러한 두가지 과정을 통하여 NCM과 PSO-NCM에 대해 각각의 거절성능을 MDR과 FAR의 척도로 평가하였다. 이를 단어 인식기에 적용한 결과 PSO-NCM은 NCM의 성능에 비하여 20% 향상이 됨을 확인하였다. 즉, 잡음 환경에 적용된 음소 신뢰도를 적용했을 때 기존의 거절기능 성능에 비해 좋은 결과를 보임을 알 수 있다.

향후, 다양한 잡음 환경에서의 구축된 DB를 바탕으로 음소 신뢰도 모델을 확장하여 거절기능의 성능을 검증하고자 한다.

## 참 고 문 헌

- [1] 민소희, 김진영, 송민규, 나승유, 최승호. 2007. “스코어 기반 관측신뢰도를 이용한 잡음환경하 화자식별.” *음성과학* 14(4), 145-155.
- [2] Yao, K., Paliwal, K. K. & Nakamura, S. 2004. “Noise adaptive speech recognition based on sequential noise parameter estimation.” *Speech Communication* 42(1), 5-23.
- [3] Junqua, J. C. & Haton, J. P. 1996. *Robustness in Automatic Speech Recognition : Fundamentals and Applications*. Kluwer Academic Publishers.
- [4] Rosenberg, A. et. al. 1994. “Cepstral channel normalization techniques for HMM-based

- speaker verification.” *Proc of ICSLP-94*, 1835-1838.
- [5] Moreno, P. J., Raj, B., Gouvea, E. & Stern, R. M. 1995. “Multivariate-Gaussian-based cepstral normalization for robust speech recognition.” *Proc of ICASSP*, 137-140.
- [6] Xu, H., Tan, Z. H., Dalsgaard, P. & Lindberg, B. 2005. “Robust Speech Recognition Based on Noise and SNR Classification - a Multiple-Model Framework.” *Proc of INTERSPEECH-2005*.
- [7] Young, S. J., Russel, N. H. & Thornton, J. H. S. 1989. “Token Passing : a simple conceptual model for connected speech recognition systems.” *Technical report of Cambridge University Engineering Department* TR38, 1-28.
- [8] Rahim, M. G., Lee, C. H., Juang, B. H. & Chou, W. 1996. “Discriminative utterance verification using minimum string verification error(MSVE) training.” *Proc of ICASSP-96*, 3585-3588.
- [9] 이경록, 서현철, 최승호, 최승호, 김진영. 2002. “하이브리드 신뢰도를 이용한 영역 핵심어 검출 성능 향상.” *한국음향학회지* 21(7), 632-640.
- [10] Shi Y. & Eberhart R., 1998. “A modified particle swarm optimizer.” *Proc of the IEEE International Conference on Evolutionary Computation*, 69-73.
- [11] Eberhart, R. & Kennedy, J. 1995. “A New Optimizer Using Particle Swarm Theory.” *Proc of Sixth International Symposium on Micro Machine and Human Science*, 39-43.

접수일자: 2008. 10. 31  
수정일자: 2008. 11. 26  
게재결정: 2008. 12. 10

▲ 김병돈

광주광역시 북구 용봉동 300번지 (우: 500-757)  
전남대학교 공과대학 전자정보통신공학과  
Tel: +82-62-530-0370 Fax: +82-62-530-1750  
E-mail: doni96@empal.com

▲ 송민규

광주광역시 북구 용봉동 300번지 (우: 500-757)  
전남대학교 공과대학 전자정보통신공학과  
Tel: +82-62-530-0370 Fax: +82-62-530-1750  
E-mail: smg686@lycos.co.kr

▲ 최승호(교신저자)

전라남도 나주시 대호동 252번지 (우: 520-714)  
동신대학교 컴퓨터학과  
Tel: +82-61-330-3194 Fax: +82-61-330-3775  
E-mail: shchoi@dsu.ac.kr

▲ 김진영

광주광역시 북구 용봉동 300번지 (우: 500-757)  
전남대학교 공과대학 전자정보통신공학과  
Tel: +82-62-530-1757 Fax: +82-62-530-1750  
E-mail: beyondi@jnu.ac.kr