화자 적응 방법들의 비교

The Comparison of Speaker Adaptation Methods

황 영 수*
(Young Soo Hwang*)

※본 논문은 한국과학재단(과제번호 95-0100-22-01-3)의 연구 지원에 의해 연구된 것입니다.

요 약

본 논문은 화자 적응 방법 제안과 그 방법들의 성능을 검토한 것이다.

본 논문에서 제안 검토한 방법들은 최대사후확률추정(MAPE)방법, 음성 선형 특성을 이용한 방법, 다층 퍼셉트론(MLP)을 이용한 방법과 ARTMAP을 이용한 방법들이다.

각 방법들의 성능 평가를 위하여 한국어 숫자음으로 실험한 결과, 최대사후확률추정 방법과 반연속 HMM의 출력확률 적응, 음성 선형 특성 등 3방법을 결합한 방법이 가장 우수한 결과를 보였으며, 이와 비슷한 실험 결과를 ARTMAP을 이용한 화자 적응 방법에서 보였다.

ABSTRACT

In this paper, we proposed various speaker adaptation methods and studied the performance of these methods.

Methods which were studied in this paper are MAPE(Maximum A Posteriori Probability Estimation), Linear Spectral Estimating, Multi-Layer Perceptron and ARTMAP.

In order to evaluate the performance of these methods, we used Korean isolated digits as the experimental data, the hybrid speaker adaptation method, which unified MAPE, linear spectral estimating and output probability of SCHMM, showed the better recognition result than those which performed other methods. And the method using ARTMAP showed the similar result to above hybrid method.

I. 서 론

음성은 인간과 기계 사이의 정보를 교환하기 위한 자연스럽고 편리한 수단이다. 대부분의 음성 인식 시스템은 화자 독립이거나 화자 종속 시스템으로 분류되며, 이 중 화자 독립 시스템은 사용자의 학습 단계를 요구하지 않으며, 많은 응용 분야에서 유용한 시스템이다. 그러나 사용 화자의 음향 특성의 변동 때문에 화자 종속 시스템보다 그 성능이 떨어지고 있는 실정이다. 또한 화자 종속 시스템도 성능이 우수하기 위해서는 사용 화자의 많은 학습 시간을 요구하게 된다. 그러므로 가장 이상적인 음성 인식 시스템은 사용함에 따라 사용자의 변화에 적응할 수 있는 시스템이다.

이와 같은 화자 특성 변동을 적응화하기 위하여, 음성 인식 시스템에 화자 적응 기능을 갖게 하는 방법에 대한 연구는 성대파 스펙트럼과 성도의 길이를 정규화하는 방법[1], 일부의 음소로부터 개인차에 적응하는 모든 음소의 스펙트럼을 추정하는 방법[2], 화자에 적응하는 표준 패턴의 집합을 선택하는 방법[3], 벡터 양자화에 의한 코드북의 매핑(mapping) 방법 [4], 매핑 코드북을 확률 모델에 도입하는 방법[5] 등이 있다.

음성 인식 시스템의 화자 적응 방법은 위와같이 여러 가지 방법이 있으나 그 기본 개념은 동일하다. 즉, 인식기의 표준 패턴 인자들에 미지 화자에 해당되는 인자들을 정규화 시킨다는 기본 개념은 같다.

본 연구에서는 표준 패턴 인자들에 미지 화자의 인자

*관동대학교 전자공학과
접수일자 : 1998년 10월 1일

를 정규화시키기 위하여,여러 화자 적응 방법 즉, 최대사후 확률추정법 (Maximum A Posteriori Probability Estimation), 선형 스펙트럼 추정 방법, 다층 퍼셉트론 (Multi-Layer Perceptron), Fuzzy ARTMAP 등을 이용하였다. 선형 스펙트럼 추정 방법은 음향 특성을 추출한 후, 화자 특성을 제거시킨 방법이고, 최대 사후 확률추정법은 최대사후확률을 이용하여 최적 코드워드를 추출한 후, 화자 적응을 수행하였으며, 다층퍼셉트론과 Fuzzy ARTMAP을 이용한 방법은 두 화자 사이의 데이터에 비선형 관계를 이용하기 위한 것이다. 이 방법과 첫 번째 방법을 결합시켜 화자 적응을 검토하였다. 그리고 신경 회로망은 두 화자 사이의 비선형 관계를 이용하여 화자 적응을 하기 위한 것이다.

본 논문에서는 위의 방법들에 대한 화자 적응 방법을 수행한 후, 그 결과들을 상호 비교하였다. 본 논문의 구성은 2장에서 위의 화자 적응 방법들을 소개하고, 3장에서는 위의 방법들을 이용한 화자 적응 방법의 실험 결과를 검토한 후, 4장에서 본 논문의 결론을 맺는다.

II. 본 논문에서 검토한 화자 적응 방법들

본 논문에서는 인식기로 반연속 HMM(Hidden Markov Model)을 사용하였기 때문에, 본 논문의 화자 적응 방법들을 반연속 HMM에 적용시켜 전개를 한다.

II-1. 최대사후확률추정 방법[6]

N 개의 샘플 벡터 X_1 - X_N 이 주어질 경우, 사후 확률은

$$P(v \mid X_1, X_2, \cdots, X_N) = \frac{P(X_N \mid X_1, \cdots\cdots, X_{N-1}, v)P(X_1 X_2 \cdots\cdots, X_{N-1})}{\int P(X_N \mid X_1, \cdots\cdots, X_{N-1}, v)P(X_1 X_2 \cdots\cdots, X_{N-1})dv} \quad (2\text{-}1)$$

이 되고, 여기에서 는 평균 벡터이다.

(2-1) 식에서 1 개의 입력 벡터 X_1 이 정규 분포 상태에 있다고 가정할 경우, 입력 벡터 X_1 에 따른 예측 평균 v_1 는,

$$v_1' = \frac{a v_0 + X_1}{a+1} \quad (2\text{-}2)$$

로 유도되고, 연속된 N 개의 샘플 벡터에 의한 평균 예측값은,

$$v_N' = \frac{a v_0 + \sum_{i=1}^{N} X_i}{a+N} \quad (2\text{-}3)$$

으로 유도된다. 여기에서 X_i 는 인식기에 입력되는 샘플 벡터, a는 상수, v_0 는 표준 모델의 평균 벡터이다.

본 연구에서는 (2-3) 식을 반연속 HMM에 적용시키기 위하여, v_0 를 표준 화자 음성을 이용하여 구성한 반연속 HMM 의 격리 단어 모델 전체의 코드북내 코드워드로 설정하여 다음과 같은 식으로 변경시켰다.

$$v_k' = \frac{a v_{0k} + \sum_{i=0}^{N_k} X_i}{a+N_k} \quad (2\text{-}4)$$

(2-4) 식에서 X_i 는 미지 화자의 입력 벡터, v_k' 는 코드북내 k번째 코드워드의 예측값, N_k 는 미지 화자의 연속된 입력 벡터중 k 번째 코드워드와 가장 유사도가 큰 입력 벡터의 갯수이다.그러므로 (2-4) 식을 이용하여 반연속 HMM의 확률 분포 평균값을 미지 화자 음성에 적응을 시키고, 동시에 출력 확률은 Viterbi 알고리듬을 이용하여, 다음 식에 의해 적응시킨다.

$$B' = (w * B) + (1\text{-}w) * b \quad (2\text{-}5)$$

여기에서 b 는 (2-4) 식을 이용하여, 미지 화자 입력 단어에 따른 평균을 적응시킨후, Viterbi 알고리듬을 이용하여, 각 상태에서의 최적 코드워드가 발견된 상대적 횟수, w는 가중치이다.

이와같은 방법으로 화자 적응을 수행한 음성 인식 시스템을 그림 1 에 나타 내었다.

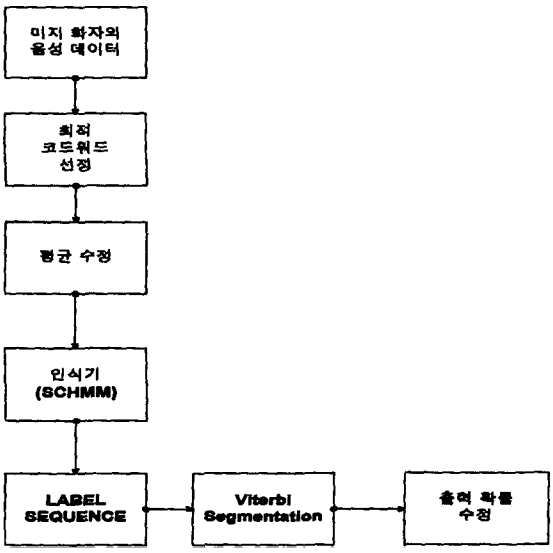

그림 1. 최대사후확률방법을 이용한 화자 적응 시스템
Fig. 1. Speaker adaptation system using MAPE.

II-2. 음성 선형 특성을 이용한 화자 적응[7]

임의의 화자 A 의 음성 특성을, 표준 패턴 화자 B 의 음성 스펙트럼의 선형 변화에 의해 다음과 같이 나타낼 수 있다.

$$X_i^{(A)} = H^{(A)} L^{(A)} X_i^{(B)} \tag{2-6}$$

여기에서 $H^{(A)}$ 는 A 화자의 음향학적 특성, $L^{(A)}$ 는 A 화자의 i 번째 음소 특성 변화식이다.

이와같은 스펙트럼 변화의 양변을 log 화 한 후 선형 특성으로 변화시키면,

$$X_{i,t}^{(A)} = h^{(A)} + l^{(A)} + X_{i,t}^{(B)} \tag{2-7}$$

가 되고, 여기에서 $h^{(A)}$ 는 각 화자의 spectrum bias 라 할 수 있다. 그러므로,

$$h^{(A)} = \frac{1}{T}\sum_{t=1}^{T(A)} (X_t^{(A)} - U_{i=(t)}) \tag{2-8}$$

에서 구할 수 있다. 여기에서 T(A) 는 화자 A가 발음한 음성의 프레임 수, $U_{i=0}$ 는 표준 패턴 코드워드중 화자 A 가 시간 t에서 발성한 $x_t^{(A)}$ 에 가장 유사도가 적합한 것이 된다.

이와같이 구한 화자 A 의 발성 특성 $h^{(A)}$ 를 A 화자의 음성에서 제거함으로써, 미지 화자와 표준 패턴 화자 상호간의 발성 특성을 제거할 수 있다. 이와같은 음성 인식도를 그림 2에 나타내었다.

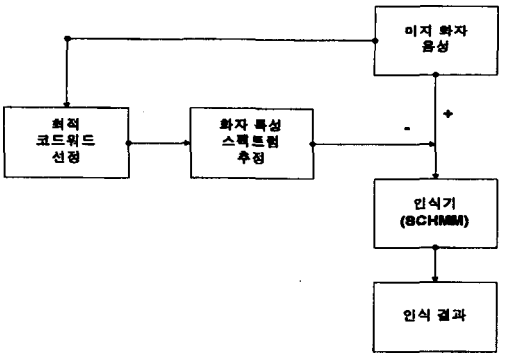

그림 2. 화자 스펙트럼 변동에 따른 음성 인식
Fig. 2. Speech recognition according to the variation of speaker spectrum.

또한 계속되는 미지 화자 A 의 음성 데이터를 이용하여, 먼저 발성한 음성 데이터로부터 추정한 화자 A 의 특성을 연속적으로 수정을 하기 위하여 다음과 같은 연속 스펙트럼 추정식을 이용하였다.

$$h_n^{(A)} = \frac{\sum_{k=1}^{n-1} T_k^{(A)}}{\sum_{k=1}^{n} T_k^{(A)}} h_{n-1}^{(A)} + \frac{T_n^{(A)}}{\sum_{k=1}^{n} T_k^{(A)}} H_n^{(A)} \tag{2-9}$$

여기에서 $H_n^{(A)}$ 는 화자 A가 n 번째 단어만을 발성하였을 경우 추정한 A 화자의 음향학적 특성이고, $h_n^{(A)}$ 는 화자 A 가 n번째까지 발성한 모든 단어들로부터 추정한 A화자의 음향학적 특성으로 화자 A가 단어를 발성할 때마다 그 음향학적 특성의 수정이 연속적으로 처리할 수 있다.

그림 3. 에 첫 번째 화자 적응 방법과 두 번째 화자 적응 방법을 결합시킨 전체 화자 적응 음성 인식 블록도를 나타내었다.

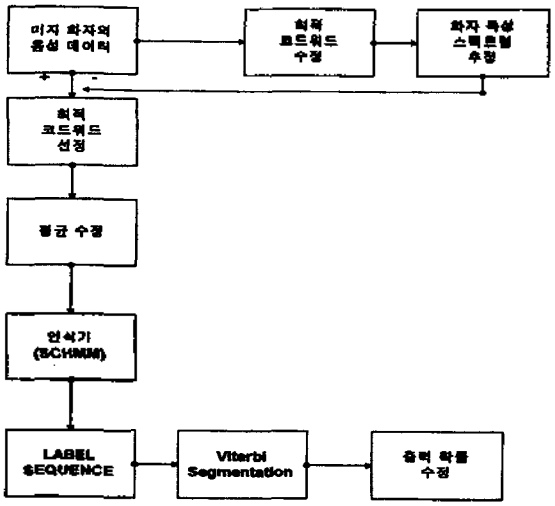

그림 3. 두 화자 적응 방법을 결합시킨 음성 인식 시스템
Fig. 3. Speech recognition system, which unfied two speaker adaptation methods.

II-3. 다층퍼셉트론을 이용한 화자 적응 방법

신경 회로망을 이용한 화자 적응 방법을 인식기로 사용한 반연속 HMM 에 결합시키기 위하여, 반연속 HMM 에서 사용하는 코드북을 이 신경 회로망에 적용시켰다. 즉, 신경 회로망을 이용하는 목적은, 다음 식에 나타낸 것과 같이 두 데이터간의 차를 신경 회로망의 가중치(weight)를 조절하여, 비선형적으로 최소화시키는 것이다.

$$\sum D(F(X^A) - X^B) \tag{2-10}$$

해당 프레임 쌍

여기에서 F() 는 신경 회로망을 이용한 변환식이고, X^A는 미지 화자의 음성 프레임, X^B는 표준 패턴 화자의 음성 프레임이다.

그러므로 거리가 최소화할되는 두 데이터를 선택하는 데에 있어서, 전체 학습 데이터와 미지 화자 음성 데이터를 대상으로 하지 않고, 반연속 HMM 코드북내의 코드워드와 미지 화자 음성 데이터를 상호 비교하였다.

또한 학습 데이터가 많은 경우, 1 개의 신경 회로망으로 구성하면, 만족된 결과를 쉽게얻을 수 없고, 만족된 결과를 얻는데 소요되는 시간도 많이 필요하기 때문에, 본 연

구에서는 반연속 HMM 코드북내의 각 코드워드당 1 개의 신경 회로망을 구성하였다. 이때 비교되는 대상은 각 코드워드를 구성하는데에 속했던 학습 데이터와 미지 화자 음성 데이터이다. 이와같은 신경 회로망 구조를 그림 4. 에 나타내었다.

그림 5. 에는 그림 4.에 나타낸 신경 회로망 구조를 학습시키는 블록도를 나타 내었다.

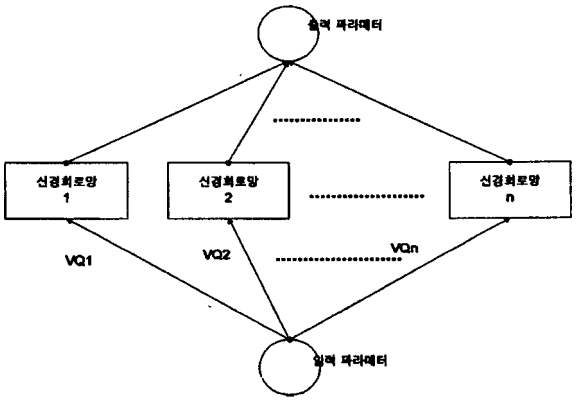

그림 4. 코드워드를 이용한 신경회로망 구조
Fig. 4. N. N architecture using codeword.

그림 5. 의 학습 블록도에서는 입력 미지 화자와의 거리가 최소인 코드워드를 구성한 학습 데이터들중에서 입력 미지 화자와의 거리가 최소인 학습 데이터를 선정하여, 출력 파라메터로 선택하여 상호 비교하였다.

그리고 그림 5. 에 나타낸 신경 회로망 구조를 이용한 화자 적응 음성 인식 시스템 블록도를 그림 6. 에 나타내었다.

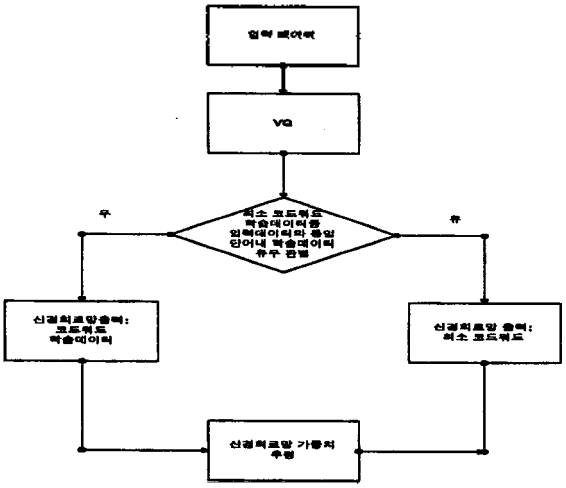

그림 5. 그림 4. 의 신경회로망의 학습 블록도
Fig. 5. Training block diagram using Fig. 4.

II-4. Fuzzy ARTMAP를 이용한 화자 적응 방법

Fuzzy ARTMAP[9]을 이용한 화자 적응 음성 인식 시스템을 그림 7 에 나타내 었다.

그림 7에 나타낸 방법은 ARTMAP을 이용하여, 미지 화자 음성과 표준 패턴시 구성된 화자들의 음성 특성을 비교하여, 화자 인식을 수행한 후, 가장 최적인 화자 인식 모델을 미지 입력 음성을 인식하기 위한 입력 모델로 사용한다. 즉, ARTMAP을 이용하여 미지 화자에 가장 적합한 표준 모델을 먼저 선택한 후, 인식기의 표준 모델로 설정하는 것이다.

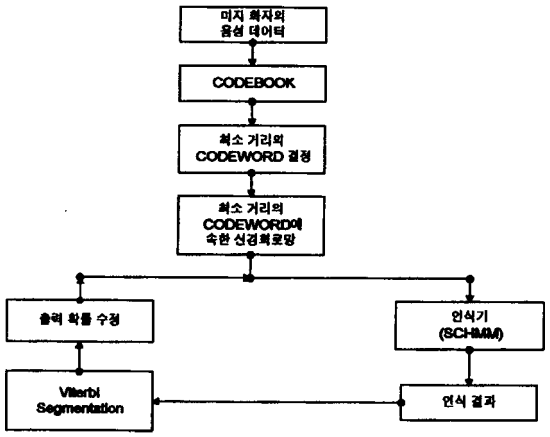

그림 6. 신경 회로망을 이용한 화자 적응 인식 블록도
Fig. 6. Speaker adaptation block diagram using N. N.

ARTMAP을 불특정 화자 음성 인식 모델의 전처리 과정으로서, 입력 데이터로 그 화자를 판별하기때문에, 미지 입력 화자에 가장 적합한 특성 표준 패턴 모델을 선택한 후 음성 인식을 수행한다. 그러므로 본 방법에서는 다른 방법에 비해 표준 데이터를 저장할 기억 용량이 많이 소요되지만, 인식시의 사간 소요는 다른 방법에 비해 차이는 없게 된다.

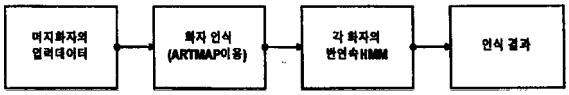

그림 7. ARTMAP을 이용한 화자 적응 인식 시스템
Fig. 7. Speaker adaptation system using ARTMAP.

III. 실험 및 결과 고찰

III-1. 실험 데이터와 인식 모델

본 실험에 이용된 음성 데이터는 6 명의 남성 화자가 한국어 격리 숫자음 ('영' - '구') 을 10번씩 반복 발음하였다. 그리고 이 반복 발음한 음성들을 10KHz (16비트) 샘플링하여, 분석창 길이 25.6ms, 프레임 간격 12.8ms 의 해밍창 (Hamming Window) 으로 추출한 후, 13차 LPC 켑스트럼 (Cepstrum) 계수를 특징 인자로 사용하였다.

Fuzzy ARTMAP과 다층 퍼셉트론을 제외한 다른 화자 적응 방법과 화자 적응을 하지않은 실험시 6명중 2명이 반복 발음한 숫자음을 표준 데이터로 사용하였고, 나머지 4명의 반복 발음한 숫자음을 실험 데이터로 이용하였다.

그리고 Fuzzy ARTMAP 실험시에는 각 4명이 발음한 음성 중 1회 발성한 숫자음의 각 모음 부분을 화자 인식에 사용하였고, 4명의 데이터 중 5회를 학습에 5회를 인식시 사용하여 랜덤(random)하게 학습과 실험 데이터를 조합하여 수행하였다.또한 다층퍼셉트론 실험시에는 6명중 2명의 화자를 표준 데이터에, 화자 적응 과정은 표준 데이터 구성시 제외한 나머지 4명 중 1명을, 인식 실험시에는 적응 과정 1명을 포함 한 4명의 데이터로 수행하였다.

또한 실험에 사용된 인식 모델은 반연속 HMM 으로서, 본 연구에 사용되는 데이터는 격리 숫자음이기 때문에, 상태(state) 개수를 적게하여도 되므로, 본 연구에서 사용한 반연속 HMM 의 상태 개수를 3 개로 구성하였고, 각 상태간 천이 이동은 1 상태 사이의 이동을 허용하였다. 그리고 이산 HMM 과는 달리 반연속 HMM 은 모든 단어의 모든 상태에 쓰이는 코드워드가 한 코드북에 가우시안 분포로 묶여 있으므로, 본 연구에서는 코드워드의 수를 16 개로 한정시켰다.

Ⅳ-2. 실험 결과 고찰

본 논문에서 실험한 각 방법의 인식 결과를 표 1과 그 적응 방법의 평균 인식률을 그림8에 나타내었다. 표 1과 그림 8에서 확률은 반연속 HMM의 출력 확률을 (2-5)식을 이용한 것이며, MLP는 다층퍼셉트론을 이용한 것이다.

표 1. 각 화자 적응 방법에 따른 인식률(단위:%)
Table 1. Recognition rate of each speaker adaptation method.

화 자	확 률	확률+ MAPE	확률+ 선형	확률+MA PE+선형	MLP	확률+ MLP	ARTMAP	적응안함
A	68	72	71	74	67	69	82	62
B	72	78	73	79	68	69	81	61
C	85	92	89	91	75	82	81	71
D	83	86	84	92	73	86	86	72
평 균	77	82	79.3	84	70.8	76.5	83	66.5

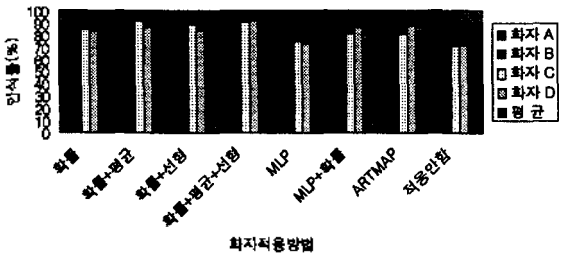

그림 8. 화자적응 방법에따른 인식률
Fig. 8. Recognition rate of each speaker adaptation method.

표 1과 그림 8의 결과를 살펴보면, 여러 화자 적응 방법 중 반연속 HMM의 출력 확률, MAPE와 음성 선형 특성을 결합 시킨 화자 적응 방법의 결과가 84%로 가장 좋은 인식률을 보였으며, 신경 회로망 방법중 ARTMAP을 이용한 방법이 83%로 앞의 화자 적응 방법과 비슷한 인식 결과를 보였다. 또한 가장 낮은 인식 률을 얻은 방법은 신경 회로망중 MLP 방법을 이용한 것으로서 70.8%의 인식률을 나타내고 있다. 이와같이 MLP방법의 인식률이 낮은 이유는 화자 적응을 하기 위하여, 어떤 특정의 화자에 따른 학습을 수행한 결과 다른 화자의 인식률이 급격히 저하된 것으로 사료된다. 또한 신경 회로망중 ARTMAP을 이용한 방법이 다른 방법보다 우수한 결과를 나타낸 것은 각 화자별 인식 모델에 따른 결과로 생각되며, 이때 발생한 오인식 결과는 화자 인식의 오류에 의한 것으로 사료된다. ARTMAP을 이용한 화자 인식 결과는 표 2에 나타낸 것과 같다. 표 2에 나타낸 화자 인식시 사용한 파라메터는 인식시 사용된 13차 ㅋ스트럼을 이용한 것이다. 표 2에 나타난 것과 같이 ARTMAP을 이용한 화자 인식시 화자 인식이 뛰어난 '이' 모음을 사용할 경우에는 전체 인식률이 향상되겠지만, 그러나 음성 인식시에 사용되는 인식 단어를 이용한 화자 인식을 수행할 경우에는 그 인식률이 저하되게 된다. 그 이유는 음성 인식시 화자 적응을 위한 전처리 과정으로 화자 인식을 수행하게 되는데, [표 2]에 나타난 것과 같이 전처리 과정에서 오인식이 발생할 경우에는, 다른 화자의 인식 모델을 인식기로 사용하게됨에 따라 인식률 저하가 발생하는 것으로 사료된다. 그러므로 이 방법을 이용한 화자 적응을 수행할 경우 에는 전처리 과정의 화자 인식 과정의 면밀한 검토가 필요하다.

표 2. ARTMAP을 이용한 화자 인식 결과 (단위:%)
Table 2. Speaker recognition rate using ARTMAP.

	오	이	아	유	여	우
A	82	98	88	80	90	81
B	82	98	96	84	72	82
C	75	96	70	72	98	75
D	78	96	82	80	90	81

그리고 MAPE방법과 음성의 선형 특성을 이용한 화자 적응 방법은 적응시 교사없는(unsupervised) 화자 적응이 가능하지만, 교사없는 화자 적응을 할 경우에는, 그 인식률은 다른 방법에 비해 변동이 심한 것으로 나타났다. 표 3에 교사없는 화자 적응 결과를 나타내었다. 이 실험에 사용된 데이터는 5명의 화자 중 1명의 화자 음성을 학습 데이터로 다른4명의 데이터를 실험 데이터로사용한 것이다.

표 3. 입력 순서에 따른 인식률(MAPE, 선형스펙트럼추정, 단위:%)
Table 3. Recognition rate accodring to the input order.

화 자	입력 순서	확률+MAPE	확률+선형	확률+MAPE+선형
A	(1)	50	70	70
	(2)	63	53	63
	(3)	73	63	70
B	(1)	70	67	87
	(2)	77	83	87
	(3)	80	60	73
C	(1)	90	80	90
	(2)	93	90	93
	(3)	93	90	93
D	(1)	85	82	85
	(2)	82	78	84
	(3)	86	83	86

IV. 결 론

본 논문은 여러 화자 적응 방법(MAPE, 음성 선형 특성, MLP, ARTMAP)들의 성능 평가를 검토한 것이다.

인식기를 반연속 HMM을 이용하여 실험한 결과, MAPE+음성 선형특성+출력확률을 결합시킨 화자 적응 방법이 가장 뛰어난 결과를 보였으며, 이 방법과 비슷한 결과를 보인 것은 신경 회로망 방법중 ARTMAP을 이용한 것이다. 그러나 ARTMAP을 이용한 방법은 화자 구분을 하기위하여, 인식시 전처리 과정으로 화자 인식기로 사용한 것이기 때문에 이 전처리 과정에서의 오인식에 따른 인식률 차이가 많은 것으로 사료된다. 그러므로 본 논문에서 사용한 방법과 같이 ARTMAP을 전처리 과정에 사용할 경우에는, 화자 인식 과정의 면밀한 검토가 필요할 것으로 생각된다. 그리고 인식기로 반연속 HMM을 사용하였기 때문에 출력 확률에 따른 인식률의 성능 향상은 큰 것으로 생각되어, 응용 분야에서 인식기를 반연속 HMM을 사용할 경우에는 다른 방법보다 확률에 따른 화자 적응 방법을 고려하는 것이 인식률 향상에 많은 도움을 줄 것으로 사료된다.

향후 데이터를 보완하여 상기의 결과에 대한 다른 검토를 수행할 것이며, 연구 대상으로는 학습 시간과 패턴 인식이 뛰어난 ARTMAP을 화자 인식을 위한 처리과정으로 사용치 않고, 직접 화자 적응을 수행할 수 있는 방법을 연구할 것이다. 또한 효과적인 출력 확률 방법과 효과적인 교사없는 화자 적응 방법도 연구 대상으로 할 것이다.

참 고 문 헌

1. H.Matsumoto et.al, " Vowel Normalization by Frequency Warped Spectral Matching," Speech Comm., Vol.5, No.2, pp.239-251, 1986.
2. S.Furui, " A Training Procedure for Isolated Word Recognition Systems," IEEE Trans. Acoust., Speech Signal Processing, Vol.ASSP-28, No.2, pp.128-136, 1980.
3. 木下," セット化音韻テンプレットニ基つくん不特定話者單語音聲認識システム," 新學論 J67-A, 6, 1984.
4. K.Shikano et.al, "Speaker Adaptation through Vector Quantization," Proc. ICASSP 86, 49.5, 1986.
5. H.Tseng et.al, " Fuzzy Vector Quantization Applied to Hidden Markov Modeling," Proc. ICASSP 87, 15.5, 1987.
6. M.Tonomura, T.Kosaka and S.Matsunaga, "Speaker Adaptation Using Maximum a Posteriori Probability Estimation Estimation and data Size Dependent Parameter Smoothing," 전자정보통신공학회논문집, Vol.J81-D-II, No.3, pp.465-471, 1998.
7. Young Soo, Hwang, "A Study on the Speaker Adaptation of SCHMM," The Journal of the Acoustical Society of Korea, Vol.15, No.3, pp.97-103, 1996.
8. G.A.Carpenter, Grossberg, N.Markuzon, J.H.Reynolds and D.B.Rosen,"Fuzzy ARTMAP:A neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Neural Networks, NN-3, pp.698-713, 1992.
9. 양태영, 윤대희, 차일환 외 6인, "관찰 확률 최대화를 이용한 화자 적응," 제15 회 음성 통신 및 신호 처리 워크샵 논문집, Vol.15, No.1, pp.324-327, 1998.

▲황 영 수(Young Soo, Hwang)
한국음향학회지 제 16권 8호 참조