음성을 이용한 화자인식 시스템 기술의 현황과 전망

영남대학교 정현열

1. 화자인식의 원리

1.1 인간과 컴퓨터에 의한 화자인식

화자인식은 음성인식과 가장 밀접한 관계가 있는 것으로 말하는 사람(화자)의 음성파에 포함되어 있는 개인성 정보를 이용하여 누구의 음성인가를 자동적으로 판정하는 것을 말한다[4,9,32]. 화자인식연구는 화자독립음성인식기술의 원리와 매우 흡사하다. 넓은 의미에서 화자인식연구는 사람이 음성을 청취하여 그 스펙트럼을 이용하여 화자를 식별하는 원리를 연구하는 분야를 포함한다[13,40].

화자인식의 역사를 살펴보면 1660년대 Charles I세의 죽음을 둘러싼 정황을 청취하기 위해 열린 재판에서 피고를 식별하기 위해 녹음된 증인의 음성을 이용한 경우로 거슬러 올라간다[25]. 그 후 약 2세기 동안은 화자인식이 과학적 증거로 이용되지는 않았으나 전화가 발명되면서부터 거리 및 시간에 관계하지 않는 화자인식 연구가 재개되어 1940년대에 들어서면서부터 음성의 스펙트로그램이 화자인식에 사용되기 시작했고 1966년에는 법원이 음성의 스펙트로그램에 기반을 둔 화자인식 결과를 증거로 채택하는 것을 인정하기에 이르렀다.

이와 같은 음성 또는 스펙트로그램으로 나타낸 음성의 시각자료에 의한 방법과 더불어 기계에 의한 화자인식에 관한 연구도 계속되어 최근의 컴퓨터 및 패턴인식 기술의 발달과 더불어 놀랄만한 진전을 보이고 있다. 실제로 화자인식 시스템은 음성을 이용하여 고객의 요구를 만족시키는 여러 분야에서 중요한 역할을 할 것으로 기대되는 데 예를 들면 전화선 또는 인터넷을 이용한 은행간 송금, 잔고조회, 쇼핑, 음성메일, 개인정보조회, 예약, 컴퓨터의 원격접근, 극비장소에의 접근통제 등에 편리하게 이용될 수 있기 때문이다. 특히 음성을 이용한 개인확인은 카드, 키 등과 같은 인공적인 수단보다는 매우 편리할 뿐만 아니라 음성은 분실위험이나 도난위험이 전혀 없어 매우 안전하다. 또, 화자인식은 손이나 다른 도구를 전혀 필요로 하지 않으므로 급속히 발전하는 정보화시대의 각종 시스템 구현에 중요한 기술로서 각광받고 있다. 이와 같은 요구에 따라 화자인식기술이 인터넷을 이용한 각종 개인의 인증방법에 이용되기 시작되었으며 법정 증빙도구로서도 이용되기 시작되고 있다[15].

그러나, 이와 같은 음성을 이용한 화자인식의 단점은 음성의 물리적 특성은 항상 불변이 아니며 마이크 특성, 전송선로 또는 배경잡음 등에 의해 쉽게 변할 수 있다는 것이다. 만약 시스템이 어떤 고객 음성의 다양한 변화를 수용한다고 하면 이는 또한 목소리가 비슷한 다른 사람의 음성을 수용한다는 의미가 되므로 화자식별은 실패하게 된다. 따라서 화자인식에 있어서 무엇보다 중요한 것은 쉽게 모방할 수 없으며 전송특성에 영향을 받지 않는 안정된 물리적 특징을 특징파라미터로 사용하는 것이라고 할 수 있다.

1.2 개인특성

음성의 개인정보는 음성의 질, 높이, 강도, 속도, 템포, 억양, 액센트, 어휘의 사용 등에 따라 다르게 나타난다. 이들 특성들은 각종 물리적 특징들이 복잡한 상호작용을 거쳐 나타나는데, 성도의 길이, 성대특성 등과 같이 선천적으로 타고나는 조음기관의 개인적 차이로부터 나타나며 말하는 습성 등으로부터도 나타난다. 또, 각 개인의 가장 중요한 청각정보인 음성의 질과 높이는 스펙트럼 포락선과 기본주파수(피치)의 정적, 순시적 특성에 의존한다.

음성의 순시적 특성을 나타내는 스펙트럼 포락선의 시간함수, 기본주파수, 에너지 등은 음성인식에서와 마찬가지로 화자인식에 많이 이용된다. 그러나, 고성능의 화자인식 시스템 구성을 위해서는 안정된 개인특징을 얻기 위한 조치가 뒤따라야 한다. 스펙트럼 특징의 시간함수로부터 얻어지는 통계적 특성도 화자인식에 많이 이용되는데 통계적 특성을 사용함으로서 표준패턴의 차원을 줄일 수 있으며 결과적으로 표준패턴 작성을 위한 메모리와 계산시간을 줄일 수 있다[7]. 피치와 에너지 등과 같은 기본적 특징의 순시적 패턴만을 이용한 화자인식 시스템은 모방음성에는 강건하지 못하므로 성도특성(스펙트럼 포락선 특징파라미터)과 같은 특징을 결합할 필요가 있다[29].

2. 화자인식 방법

2.1 화자인식방법의 분류

화자인식 시스템은 응용방법에 따라 화자검증(speaker verification) 및 화자식별(speaker identification) 시스템으로 나누어진다. 화자검증은 검증을 요구하는 화자의 발성과 그 화자의 등록된 기준패턴을 비교하여 미리 정해놓은 임계값(발생확률값)을 넘어서면 승인(accepting)결과를 출력하고 그렇지 않으면 거절(rejecting)결과를 출력한다. 화자식별은 고립단어 인식과정과 유사한 것으로 등록된 표준패턴 가운데 어떤 화자의 패턴과 입력 음성의 패턴이 가장 유사한가를 비교하여 화자를 결정하는 것이다. 화자검증은 음성을 열쇄와 같이 이용할 수 있는 여러 서비스에 응용 가능하다. 화자식별은 범죄수사(현장에서 녹음된 범죄자의 음성을 혐의자의 음성과 비교하여 범죄여부를 판단), 출퇴근 관리(회사의 직원들의 음성을 녹음하여 등록한 후 이를 출퇴근시 비교) 등에 이용된다. 여기서 화자식별은 화자검증과 화자식별과정이 결합되어 이루어진다.

화자인식은 발성방법에 따라 문맥종속형과 문맥독립형으로 나눌 수 있다. 전자는 미리 정해진 문장을 발성케 하는 반면 후자는 특별히 정하지 않는다. 일반적으로 음성은 음향학적, 음성학적으로 변화가 많기 때문에 문맥독립형은 문맥종속형에 비해 많은 훈련 데이터를 필요로 한다. 문맥종속형은 일부 시스템에서 비음 등과 같은 특별한 음소를 이용하는 경우도 있으나 대부분 단어(핵심어, 이름, ID 등) 또는 임의로 선정한 문장을 이용한다. 문장을 이용하는 경우, 독립단어를 이용하는 경우에 비해 문장 가운데의 단어발성 또는 문장발성이 화자에 따라 달라지는 예가 많기 때문에 인식률 향상을 가져올 수 있다.

화자인식에 있어서 어려운 점은 화자가 협조적이냐 아니냐에 따라 인식률이 크게 달라진다는 것이다. 예를 들어 범죄자인 경우는 의도적으로 발성방법, 발성속도를 변화시킴으로써 인식을 방해할 수 있기 때문이다. 문맥종속 시스템 및 문맥독립 시스템 공히 최대의 약점은 녹음기를 이용하여 등록된 화자의 음성에 의해 발성된 핵심어 또는 지정한 문장을 녹음하여 마이크로 입력할 경우 등록된 화자로 승인되는 문제이다. 이를 해결하기 위한 방법으로 핵심어들로 구성된 소규모의 단어세트(예를 들어 숫자음들로 구성된)를 이용하여 이들을 랜덤하게 제시하여 발성하게 하는 방법이 있다[12,31]. 그러나 이 방법도 요구된 순서대로 단어를 발성시킬 수 있는 최신 전자녹음장비를 이용하면 의미가 없어진다. 이 문제를 해결하기 위한 새로운 방법으로 문장제시형(text-prompted) 화자인식 시스템이 제안되어 있다[21].

2.2 화자인식 시스템의 구조

일반적인 화자인식시스템의 구조를 그림 1에 보인다. 입력된 음성으로부터 추출된 특징파라미터들은 등록된 화자의 표준패턴 또는 모델과 비교되는데 이때 판단척도로서는 두 패턴간의 거리값(혹은 유사도값)이 많이 이용된다. 화자 검증에서는 위에서도 언급한 바와 같이 입력음성이 등록된 화자의 발성패턴과 비교되어 두 패턴간의 거리가 특정 임계값 이하인 경우 승인되나 그 이상인 경우는 거절된다. 화자식별의 경우는 입력음성이 등록된 여러 화자의 발성패턴과의 거리를 측정하여 가장 작은

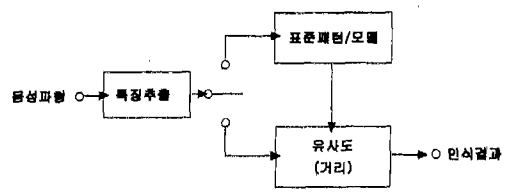

그림 1 일반적인 화자인식시스템의 구조

거리를 가진 등록화자가 선택된다.

화자검증시스템을 평가하는데는 정신생리학으로부터 도출된 수용자작용특성곡선(Receiver operating curve, ROC)이 이용된다. 입력발성이 고객(customer)에 속할 경우(s)와 그렇지 않을 경우, 즉 고객을 사칭하는 사칭자(impostor)인 경우(n)로 나누었을 때 시스템에 의한 결정조건은 입력발성이 고객으로 승인될 경우(S)와 그렇지 않을 경우(N)로 나누어진다. 이들 조건의 결합에 따라 다음과 같은 네 개의 조건부 확률을 얻을 수 있다.

P(S|s) = 바르게 승인될 확률

P(S|n) = 잘못 승인(False Acceptance, FA)될 확률

P(N|s) = 잘못 거부(False Rejection, FR)될 확률 즉, 진짜 고객을 잘못하여 거부할 확률

P(N|n) = 바르게 거부할 확률

따라서

$$P(S|s) + P(N|s) = 1$$
$$P(S|n) + P(N|n) = 1 \quad (1)$$

이 성립하며 화자검증 시스템은 확률 P(S|s)와 P(S|n)으로 평가될 수 있다. 이 두 값의 변화를 2차원 평면에 나타내면 입력음성을 고객의 음성으로 승인하는 결정기준(Decision Criterion, threshold)을 나타내는 ROC는 그림 2와 같이 변화한다. 그림 2는 세 종류의 시스템 A, B, D를 나타내고 있는데 곡선 B의 성능이 A, D의 성능보다 일관되게 우수함을 알 수 있다. 한편, 결정기준과 2종류의 에러(FR, FA)와의 관계를 그림 3에 나타낸다. 그림 2와 그림 3에서의 a 위치는 명확한 결정기준을 적용한 경우이고 b 위치는 그렇지 않은 경우이다. 적절

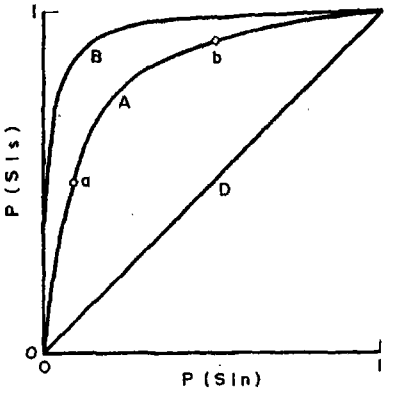

그림 2 수용자작용특성곡선(ROC)

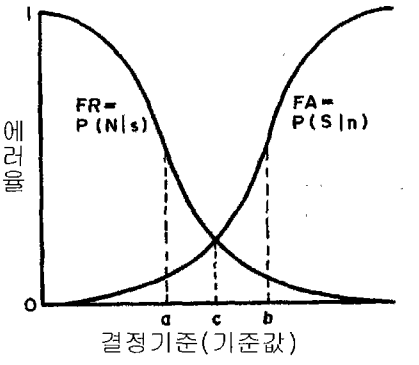

그림 3 에러율과 결정값과의 관계

한 수준의 고객거부(사칭자 승인) 기준값을 정하기 위해서는 기본자료로서 고객 및 사칭자의 점수의 분포를 알 필요가 있다. 실제로 결정기준은 결정에러의 효과에 따라 결정되는 데 이는 결정결과에 대한 비용, ROC 곡선 등을 고려한 정합의 사전확률, P(s),에 기반하여 판단되어진다. 실제로는 그림 3에서와 같이 각 화자에 대한 두 종류의 에러 FR, FA가 교차하는 점 c를 기준으로 한다.

2.3 에러율과 화자 수와의 관계

Z_N을 N 등록화자의 집합이라 가정하고 $X = (x_1, x_2, \cdots, x_n)$을 표본음성을 나타내는 n차원 특징벡터라고 하면 $Pi(X)$는 화자 $i(i \in Z_N)$에 대한 확률밀도함수이다. 확률밀도함수 X가 집합 Z_N 내에 존재할 확률은 식 (2)와 같다.

$$P_Z(x) = \underset{i \in Z_N}{E}[P_i(X)] = \sum_i P_i(X) P_{r[i]}, \; i \in Z_N \quad (2)$$

여기서 $Pr[i]$는 화자 i의 사전확률이다[4].

화자검증의 경우 고객 i의 음성으로 승인되어야 할 X의 영역은 식 (3)과 같이 나타난다.

$$R_{Vi} = [X | P_i(X) \rangle C_i P_Z(X)] \quad (3)$$

여기서 Ci는 FA와 FR 에러 사이의 균형을 맞추기 위해 적절히 선택된 값이다. Z_N을 랜덤하게 선택한 화자들로 구성하고 사전확률이 화자와 독립이면 $Pr[i] = 1/N$이 되고 $P_Z[X]$는 N이 커질수록 Z_N에 독립인 한계밀도함수에 접근한다. 따라서 $Pr[FA]$와 $Pr[FR]$은 상대적으로 집합의 크기 N이 크면 그 크기에 영향을 받지 않는다. 실제로

$Pz[X]$는 정확히 추정할 수 없기 때문에 일정한 상수값으로 한다. 그리고

$$R_{Vi}=[X|P_i(X)\rangle k_i] \quad (4)$$

는 승인영역으로만 이용된다.

화자식별에서 화자 i의 음성으로 판정되어야 하는 영역 X는 다음과 같이 나타나며

$$R_{Ii}=[X|P_{i(X)}\rangle P_j(X), \forall j\neq i] \quad (5)$$

화자 i에 대한 에러확률은 다음과 같이 나타난다.

$$P_{Ei}=1-\prod_{\substack{k=1\\k\neq i}} P_r(P_i(X)\rangle P_k(X)) \quad (6)$$

Z_N을 랜덤하게 선택한 화자들로 구성한 경우

$$E[P_{Ei}]=1-\frac{E}{Z_N}\left[\prod_{\substack{k=1\\k\neq 1}} P_r(P_i(X)\rangle P_k(X))\right]$$

$$=1-\frac{E}{i}\prod_{\substack{k=1\\k\neq 1}} E[P_r(P_i(X)\rangle P_k(X))]$$

$$=-\frac{E}{i}[P_{Ai}{}^{N-1}] \quad (7)$$

이 얻어진다. 여기서 P_{Ai}는 화자 i가 다른 화자와의 혼돈되지 않을 기대확률이다. 따라서 화자를 바르게 식별할 기대확률은 집합의 크기에 지수함수적으로 감소한다.

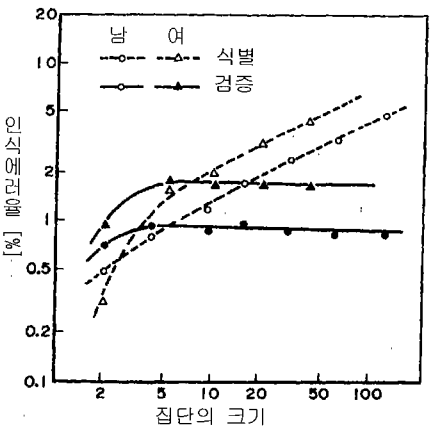

그림 4 화자식별, 검증에서의 집단의 크기와 인식 에러율과의 관계

이것은 무한개의 점들의 분포는 유한 파라미터 공간내에서 분리되지 않는다는 사실에 기인한 자연스러운 결과이다. 좀 더 구체적으로 말하면 화자의 집단이 증대되면 둘 혹은 그 이상의 화자의 분포가 근접할 확률은 증가하므로 화자식별시스템의 유효성은 집단의 크기에 따라 평가되어야 한다. 그림 4는 화자식별, 화자검증에 있어서 집단의 크기와 인식 에러율과의 관계를 나타낸다[6]. 이 결과는 발성된 단어들로부터 추출된 스펙트럼 파라미터들로 구성된 통계학적 특징을 이용한 인식시스템으로부터 얻어진 것이다.

2.4 화자특성의 변화와 특징파라미터의 평가

화자인식에 있어서 가장 어려운 점은 화자특성의 변화이다. 화자인식 성능에 영향을 미치는 가장 중요한 요소는 시간이 변함에 따라 변하는 화자특성을 나타내는 특징파라미터이다. 이 변화는 화자 자신의 특성, 녹음환경, 전송환경, 잡음 등이 시간에 따라 변하기 때문에 나타나는 현상으로 화자는 매 발성마다 정확히 꼭 같은 발성을 반복할 수 없기 때문이다. 또한 화자간 물리적 파라미터의 변화는 발성된 문장 또는 단어의 음소간 변화보다 매우 작다. 따라서 화자인식시스템은 이와 같은 변화를 잘 배려하여 설계할 필요가 있다. 여러 해 동안에 걸쳐 이루어진 연구결과, 화자에 대한 장기간 특징파라미터의 변화를 줄여 인식성능을 높일 수 있는 방법들이 개발되었는 데 이를 소개하면 다음과 같다.

1. 스펙트럼 등화법 적용, 즉, 1차 혹은 2차의 시간평균필터를 통과시키는 방법. 이 효과는 켑스트럼 평균차감(Cepstral Mean Subtraction, CMS)이나 켑스트럼 평균정규화(Cepstral Mean Normalization, CMN)와 유사함[2,8].
2. 장기간의 음성발성 중 통계적 평가에 기반한 안정적 특징 파라미터 선택
3. 다양한 단어로부터 추출한 특징파라미터와의 조합
4. 장기간에 걸쳐 녹음한 훈련데이터로부터 거리척도, 표준패턴(모델) 구성
5. 적절한 시간 간격을 두고 각 고객의 표준패턴의 재 작성
6. 각 화자에 대한 표준패턴(모델) 및 기준값 적응

스펙트럼 등화법(spectrum equalization, blind equalization)은 발성된 단어로부터 추출된 통계적 특징을 이용한 화자인식에 유효한데 Soong과 Rosenburg는 스펙트럼등화의 유/무에 따른 결과를 단기간, 장기간 훈련과 비교하였는데 이 결과를

그림 5에 나타낸다. 여기서 단기간훈련 세트는 10일간에 걸쳐 녹음하는 과정을 2-3일간의 간격을 두고 3-4차례 반복한 것을 말하고 장기간 훈련세트는 10개월간에 걸쳐 녹음하는 과정을 3개월의 간격을 두고 4차례 반복한 것을 말한다. 여기서 녹음기간 간격은 짧게는 2-3일에서 길게는 5년에 걸쳐 이루어졌다. 그림 5로부터 스펙트럼 등화법은 장기간 단기간 훈련 모두에 효과가 있는 것으로 나타나며 특히 단기간 훈련에 더욱 효과가 있음을 알 수 있다. 이 결과를 음성생성 메커니즘의 관점으로부터 살펴보면 성도특성이 성대 스펙트럼의 전체적 패턴보다 훨씬 안정되어 있음을 의미한다[37].

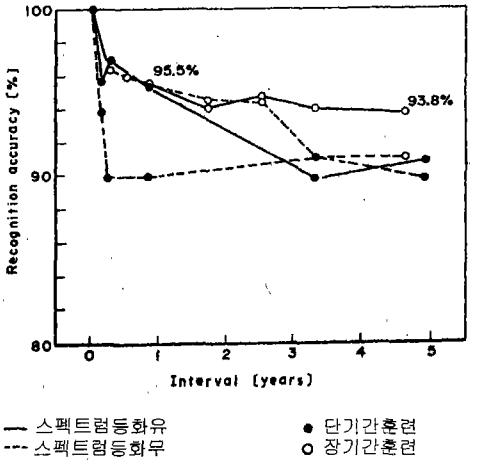

그림 5 스펙트럼등화의 유무에 따른 화자검증결과

또 다른 방법으로는 CMS 법이 있는 데 이 방법은 추출한 컵스트럼 계수를 발성된 음성의 전구간에 걸쳐 평균한 다음 이 값을 각 프레임의 컵스트럼 계수로부터 빼는 과정을 추가한 것이다. 이 방법은 로그 스펙트럼 영역에서의 부가적인 변화를 잘 보상하지만 약간의 문맥정보와 화자정보도 함께 제거되는 것을 피할 수 없게된다.

파라미터 영역에서의 정규화방법 외에 거리/유사도 영역에서의 유사도비, 사후확률도 많이 사용된다. 잡음 환경하에서의 HMM을 적응화시키기 위해 PMC(parallel model combination) 방법도 유효하다고 알려져 있다[12]. 가장 유효한 특징 파라미터를 선택하기 위해 다음과 같은 파라미터 평가법이 사용된다.

1. 여러 파라미터를 조합하여 인식실험 실시
2. 각 파라미터의 F비(급내-급간 분산비, F Ratio) 측정[6].
3. F비를 다차원공간으로 확장한 divergence 측정[1].
4. 인식에러율에 기반한 Knockout 기법[33].

이밖에 정보의 양(파라미터의 수)을 효과적으로 줄이기 위해 F비를 최대로 하는 판별분석에 의한 공간투영법도 가끔 사용된다.

2.5 유사도 정규화

동일 화자내의 특징파라미터 변화를 감소시키기 위해 Higgins 등은 식 (8)과 같은 유사도비(likelihood ratio)를 이용한 거리 정규화법을 제안하였다[12].

$$\log L(X) = \log p(X|S = S_c) - \log p(X|S \neq S_c) \quad (8)$$

유사도비는 검증하려는 대상이 올바르다고 가정한 경우의 관측 측정값에 대한 조건부 확률과 화자가 사칭자로 주어진 경우의 관측된 측정값의 조건부 확률의 비를 말한다. 일반적으로 $log\ L$의 값이 양일 경우 올바른 검증을, 음일 경우는 사칭자를 나타낸다. 식 (8)의 우측의 2번째항을 정규화항이라 부른다. 기준화자 세트가 모든 화자들을 대표한다고 가정할 때 진짜 화자 S를 제외한 모든 화자에 대한 점 X에서의 밀도는 가장 가까운 기준화자에 대한 밀도에 의해서 지배된다. 그러므로 이것은 유사도비 정규화가 Bayes 정리의 최적화 개념을 근사화한다고 할 수 있다. 그러나, 이 정규화법은 가장 가까운 기준화자의 것이 사용된다 해도 모든 기준화자들과의 조건부 확률 계산이 이루어져야 하기 때문에 계산비용이 많이 소요되어 비현실적이다. 따라서, 식 (8)의 정규화항을 계산하기 위해 '화자그룹(cohort speakers)'을 선택하여 이용하는 방법이 Rosenburg에 의해 제안되었다[32].

화자그룹의 크기를 결정하기 위해 그 크기를 1 - 5로 변화시켜가면서 실시한 실험결과 화자검증 성능은 화자그룹의 크기와 비례함을 알 수 있었고 정규화방법은 마이크의 종류에 따른 차이를 잘 보상해 주고 있음을 알 수 있었다.

또다른 정규화법으로 Matui와 Furui는 다음과 같은 사후확률에 기저를 둔 정규화법을 제안하였다[22].

$$\log L(X) = \log p(X|S = S_c) - \log \sum_{S \varepsilon Ref} p(X|S) \quad (9)$$

이 방법과 유사도비를 이용한 정규화법과의 차이점은 검증을 요구하는 화자가 사칭자 화자 세트에 포함되는가 아닌가에 있다. 즉, 유사도비를 이용한 정규화법에서는 화자그룹에 검증을 요구하는 화자를 포함하지 않으나 사후확률 기반의 정규화법의 정규화항은 이를 포함한 화자 세트를 이용하여 계산한다. 검증화자의 모델만 사용한 경우와의 비교실험결과에 의하면 두 방법 모두 거의 대등하게 화자 분별력을 향상시키고 화자종속 또는 문장종속 기준치의 필요성의 감소시키는 결과를 얻을 수 있음이 확인되었다.

화자그룹을 이용한 정규화법은 검증화자와 유사한 목소리에 대한 선택성을 증가시키는 효과가 있으나 이성(opposite gender)의 사칭자에 의한 불법접근에 허약한 심각한 문제를 가지고 있다. 화자그룹은 동성(same gender)의 화자들로 모델링되기 때문에 이성의 사칭자에 대해서는 잘 모델링되지 않아 유사도비가 신뢰할 수 없는 값이 되어버리기 때문이다. 이를 해결하기위해 성등화(gender balanced) 화자집단을 이용하는 방법이 제안되었는 데 성능평가결과 검증화자에 가까운 화자집단을 이용한 경우보다 나은 결과를 나타내었다[31].

그외 정규화법에 관해 정규화항을 일반집단을 대표하는 세계모델(world model)의 유사도로 근사하는 방법이 Carey 등에 의해 제안되었으며[3], Matui 등은 여기에 tied-mixture HMM을 부가하는 방법을 제안하였다[21]. 이들 정규화 모델은 입력음성과 검증화자 사이의 절대편차는 고려하고 있지 않기 때문에 전혀 다른 화자는 구별하지 못하는 문제가 있다.

3. 화자인식시스템의 예

3.1 문맥종속 화자인식시스템

실용가능한 화자인식시스템으로서는 문맥독립방식보다 문맥종속방식이 유리하므로 이에 관해 많은 연구가 이루어졌다[8,41,24,31]. 일 예로 Bell연구소에서는 전화회선을 이용한 화자검증 시스템을 개발하여 남성과 여성화자 100명을 대상으로 실험을 실시하여 어느 정도의 성능을 나타내었다[8]. 그림 6에 전형적인 문맥종속 화자검증 시스템의 기본적인 구성을 나타낸다. 이 시스템에 입력되는 음성은 1.5초 정도의 정해진 짧은 문장으로 구성되며 음성파는 먼저 LPC 켑스트럼계수의 시계열로 변환된다. 다음에 각 차수의 LPC 켑스트럼 계수의 시계열로부터 동적 특징으로서 90ms프레임 주기마다 1차 및 2차의 다항식전개계수를 계산한다. 이 1차계수는 △켑스트럼, 2차계수는 △△켑스트럼의 정의와 동일하다. 이들 전개계수 중 화자인식에 비교적 유효성이 높은 계수의 세트(18개의 요소)를 다수화자의 음성을 이용하여 미리 조사한 다음 이들 시계열을 LPC켑스트럼의 시계열과 조합하여 인식에 이용한다. 추출된 LPC 켑스트럼 계수는 각 차수별로 음성구간 전체에 대해 평균값을 구한 후 CMS처리한다. CMS처리는 전송시스템에 의해 발생되는 주파수왜곡보상과 화자내의 장시간 스펙트럼 변화를 줄이기 위해서이다.

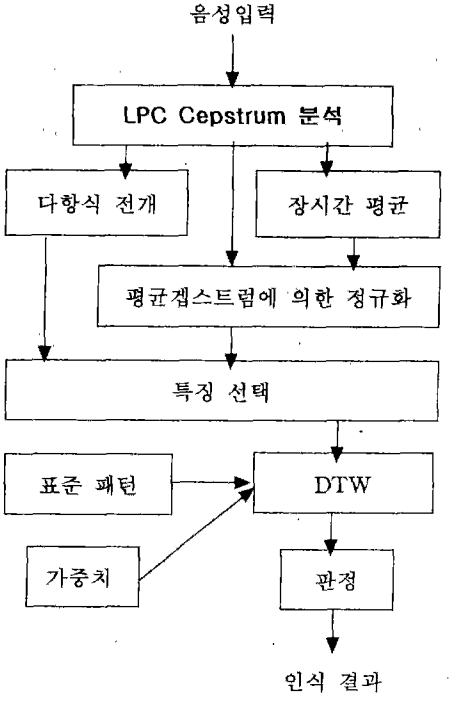

그림 6 캡스트럼 계수의 시계열과 다항계수전개를 이용한 문맥종속화자인식시스템

그 다음에 단 문장의 음성구간 전체에 대해 DTW를 이용하여 입력음성과 표준패턴과의 유사도를 계산하여 이것을 기준치(각 화자의 변동허용치)와 비교한 후 화자검증 판정을 실시한다. 기준치와 표준패턴은 화자간 거리 분포를 사용하여 매 2주마다

갱신된다.

실험결과, 비록 표준발성과 입력발성이 ADPCM 또는 LPC vocoder와 같은 다른 전송시스템을 통하여 전송된 경우에도 매우 높은 검증율을 나타내었다. 6개월에 걸친 60명의 남성과 60명의 여성화자에 의해 발성된 전화 음성을 사용한 online 실험결과에서도 이 시스템의 유효성을 확인할 수 있었다.

HMM은 스펙트럼 특징의 통계적 변화를 효과적으로 모델링할 수 있다. 그러므로, 학습을 위한 충분한 화자들의 발성을 확보할 수 있는 경우 HMM에 기반한 방법은 DTW 방법보다 훨씬 우수한 인식성능을 얻을 수 있음이 확인되었다[10].

3.2 문맥 독립 화자인식시스템

문맥독립화자인식에서는 인식실험에서 사용되는 단어나 문장을 미리 예측할 수 없다. 단어나 문장수준에서 음성의 발성을 모델링하거나 매칭시킬 수 없기 때문에, 그림 7에 나타낸 것과 같은 세 가지 방법들이 활발히 연구되고 있다. 이하 각 방법들에 관해 간략하기로 한다.

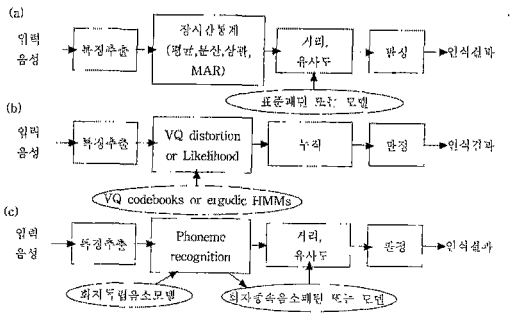

그림 7 문맥독립화자인식방법의 기본구조

(a) 장시간(Long-term)통계에 기반한 방법

문맥독립특징으로서 장시간 발성 시계열에 대한 스펙트럼특징의 평균, 공분산과 같은 다양한 스펙트럼 특징들이 많이 사용되었다[5,16,17](그림 7(a)). 그러나, 장시간 스펙트럼 평균은 화자의 발성, 문맥종속방법에서 모델로 사용되는 단구간 스펙트럼 특징의 열에 포함되어 있는 정보 등과 같은 스펙트럼 특징들을 극단적으로 함축시켜버리기 때문에 개인성이 상실되어 인식성능에 한계가 있다.

통계적 동적 특징을 이용한 연구로서 Montacie 등은 화자특성을 나타내는 켑스트럼 벡터의 시계열에 multivariate autoregression(MAR) 모델을 적용하여 좋은 화자인식결과를 얻었다[23]. Griffin 등은 MAR 모델에 기반한 방법에 거리척도를 이용하는 방법을 제안하고 10문장으로 학습한 후 1문장을 사용하여 시험하였을 때, 화자검증, 화자식별율 모두 HMM에 기반한 방법과 거의 동일한 결과를 얻을 수 있음을 보고하였다. 또한, 최적 MAR model 차수는 2차 또는 3차라는 것과 화자검증에서 좋은 결과를 얻기 위해서는 사후확률을 이용한 거리정규화가 필수라는 것을 보고하였다[11].

(b) VQ에 기반한 방법

화자의 단구간 훈련특징벡터는 그 화자의 기본특성을 표현하는데 직접 사용될 수 있으나 직접적 표현방법은 훈련벡터의 수가 많아지면 메모리와 계산량이 극단적으로 증가하여 비현실적이다. 따라서 벡터양자화(VQ)기법을 이용하여 훈련데이터를 압축하는 방법을 시도하게 되었다.

이 방법(그림 7(b))에서는 적은 수의 특징벡터로 구성된 VQ 코드북이 화자의 고유 특징을 나타내는 효과적인 방법으로 사용된다[14,18,19,30,36,38]. 이 방법에서는 각 등록화자에 대해 임의의 문장을 발성할 때 단시간 스펙트럼의 집합을 분류(cluster)하여 각 클러스터의 중심(centroid)을 코드북의 요소로 저장한다. 미지의 입력음성에 대해서는 각 등록화자의 코드북으로 벡터양자화하여 입력음성 전체에 대해 평균양자화 오차를 구한다. 화자 식별의 경우는 평균양자화 오차가 가장 작은 코드북의 화자를 선택하고, 화자검증의 경우는 평균양자화오차를 기준값과 비교하여 승인 또는 거절여부를 판정한다.

(c) Ergodic-HMM에 기반한 방법

이 방법의 기본 구조는 VQ에 기반한 방법(그림 7(b))과 동일하지만 VQ 코드북 대신에 ergodic HMM을 사용한다. 음성의 장시간 구간에 대한 음성신호 파라미터의 순시적 변화는 각 상태간 통계적 Markovian 천이로 표현된다. Poritz는 음성편을 다양한 음성학적 카테고리 각각에 대응하도록 하기 위해 각 상태간 모든 천이가 가능한 5상태 ergodic HMM을 제안하고 출력 확률 함수를 표현하기 위해 선형 예측 HMM을 채용하였다[42]. 또

한, 자동적으로 얻어진 각 카타고리를 이용하여 강한 발성, 묵음, 비음/류음, 파열음/폐쇄음, 파찰음 등을 표현하였다. Tishby는 Poritz의 방법을 mixture autogressive(AR) HMM으로 확장하였다[39].

(d) 음성인식에 기반한 방법

VQ 와 HMM에 기반한 방법은 음소-음소군 인식과 화자인식이 동시에 이루어진다. 그러나, 음성인식에 기반한 방법(그림 7(c))에서는 음소 또는 음소군은 별도로 인식된 다음 입력음성의 각 음소(군)편은 그 음소(군)에 해당되는 화자모델이나 표준패턴과 비교된다.

Savic 등은 광범위 음소분류를 위해 5상태 ergodic 선형 예측 HMM을 사용하였다[34]. 이 방법은 특정 음소 카타고리에 속한 프레임이 인식된 후에 특징 선택이 이루어진다. 또, 학습단계에서는 표준패턴이 생성되고 각 음소 카테고리별 검증 기준치가 계산된다. 검증단계에서는 음소분류 후 표준패턴과 비교하여 검증스코어를 얻는다. 마지막 검증 스코어는 각 카테고리별 가중선형 조합으로 나타내고 가중치는 화자판별에 있어서의 특정 음소카테고리의 유효성이 반영될 수 있도록 선택되며 검증 성능을 최대화되도록 조정된다. 실험결과 검증 정확도는 이와같은 카테고리 의존 가중선형 조합 방법에 의해 상당히 개선될 수 있음을 보여주었다.

Rosenberg 등은 화자검증 시스템을 4연속숫자를 사용하여 은행조회용으로 시험하였다[31,35]. 이 시스템에서 입력음성은 화자독립HMM을 사용하여 각각의 숫자들로 분리된 후 각 숫자는 대응되는 화자 특정 HMM 숫자 모델과 비교되고, Viterbi 유사도가 계산된다. 이를 입력발성의 4숫자에 대해 수행한다. 검증 스코어는 입력발성에 포함된 모든 숫자에 대해 평균정규화 로그 유사도로 계산한다.

Newman 등은 화자 검증을 위해 대어휘 음성인식시스템을 사용하였다[26]. 화자적응을 위해 화자독립음소모델 세트를 이용하였다. 이 방법에서 화자 검증과정은 두 단계로 구성된다. 즉, 먼저 각 시험발성음성에 대해 음소분할을 위한 화자독립음성인식이 수행한 다음 분할된 음성편은 특정 목표화자에 적응화된 모델을 이용하여 인식한 후 인식률이 계산된다. 그 후 인식률은 그 음성의 화자독립모델을 이용하여 정규화된다. 이 시스템을 전화교환 음성데이터로 구성된 1995 NIST 화자 검증 데이터베이스를 이용하여 평가한 결과, Gausian Mixture model을 이용한 방법에 미치지 못함을 확인하였다.

3.3 텍스트 제시형 화자인식시스템

앞에서 언급한 바와 같이 녹음된 음성 또는 다른 사람이 듣는 가운데 비밀번호 등과 같은 자신들의 ID 숫자를 발성하는 것을 기피하는 것에 대한 대책으로 텍스트 제시형 화자 인식방법이 제안되었다. 이 방법에서는 텍스트의 핵심문장이 매번 바뀐다[20,21]. 또 시스템은 등록된 화자가 발성한 문장이 제시된 텍스트 문장이라는 결정이 내려졌을 때만 입력 발성을 받아들인다. 사용되는 어휘가 제한되어 있지 않기 때문에 사칭을 시도하려고 하는 사람은 발성되어야 할 문장들을 미리 알지 못한다. 따라서, 이 방법은 정확하게 화자를 인식할 뿐만 아니라, 등록된 화자에 의해 발성된 문장이라도 제시된 문장과 다르면 거절한다.

이 방법에서는 기본 음향 단위로서 화자특정음소모델을 사용한다. 여기서 신중히 고려해야 할 점은 제한된 크기의 학습발성을 이용하여 어떻게 적절하게 화자특정음소모델을 생성하느냐 하는 것이다. 음소모델은 tied-mixture HMM이나 Gaussian-mixture continuous HMM으로 표현되는 데, 화자독립음소모델을 각 화자의 음성에 적응화하여 구성한다. 학습발성을 위한 텍스트 문장은 알려져 있기 때문에 이 발성은 음소모델을 연결하여 모델링할 수 있으며 음소모델은 반복알고리즘에 의해 자동으로 적응화시킬 수 있다.

인식단계에서 시스템은 제시된 텍스트에 따라서 문장 HMM을 생성하기 위해 등록된 화자의 음소모델을 연결한 다음 문장 모델에 대한 입력음성의 유사도를 계산한 후 이것을 화자인식 결정을 위해 사용한다. 만약 화자와 원문의 유사도가 모두 높으면 발성화자가 요구된 화자로 승인된다. 실험결과, tied-mixture 기반 음소 모델에 대한 적응화방법과 유사도 정규화 방법을 사용한 경우 높은 화자 및 원문 검증율을 나타내었다.

3.4 실시간 문맥독립화자인식시스템의 예

화자인식시스템의 한 예로서 본 연구실의 화자인식시스템을 소개한다[43]. 시스템은 실시간문맥독립화자인증 및 식별시스템으로서 크게 두 단계로 구분된다. 첫째는 화자의 등록단계이며, 두 번째 단계는 검증 및 식별단계이다.

화자등록단계는 다시 전처리, 코드북 생성, GMM 모델생성단으로 나눠진다. 화자의 검증 및 식별단계에서는 ML(Maximum Likelihood) Test를 거쳐 인식결과를 출력하게 된다. 이러한 과정을 그림 8에 나타내었다. 그림 9는 실제로 구성한 시스템의 화자의 입력단계를 나타낸 것이며, 화자등록시와 화자검증 및 화자판별시에 사용된다. 그림 10은 검증된 결과를 화면에 나타낸 그림이다.

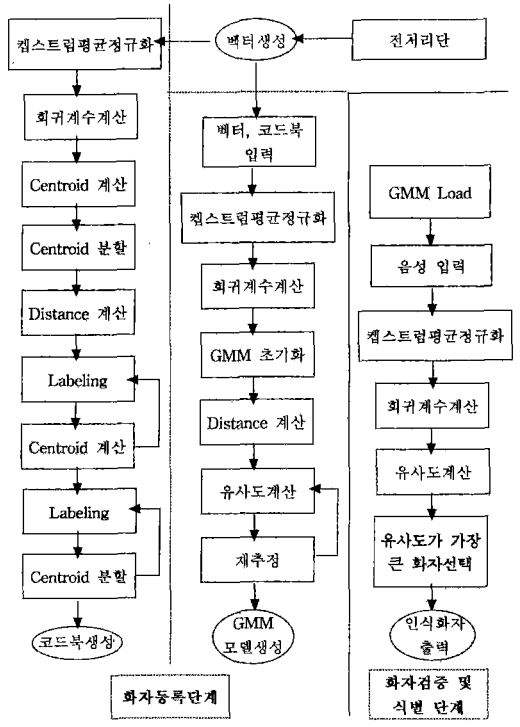

그림 8 실시간 문맥독립화자인식시스템의 처리과정

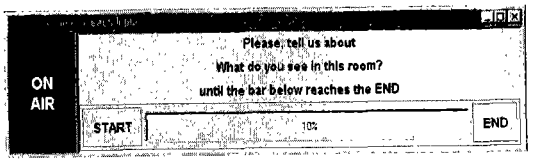

그림 9 화자발성 입력단계

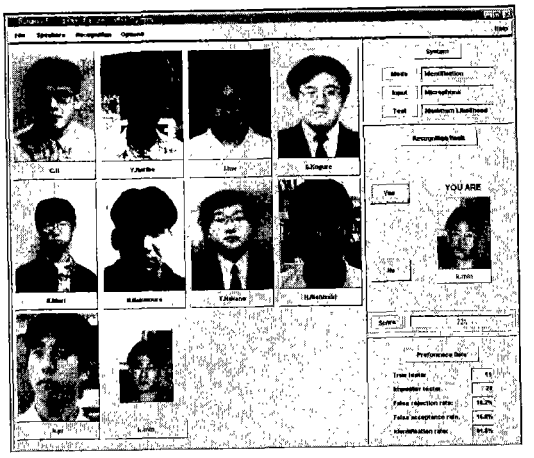

그림 10 실시간 문맥독립화자검증/식별시스템 검증결과화면

5. 국내외연구동향 및 전망

화자인식에 관한 연구결과는 음성인식에 비해 그리 많지 않다. 홍콩의 City 대학은 최근 국제 학술지에 발표한 논문에서 벡터 양자화를 이용한 화자 인식 시스템을 이용하여 200명의 프랑스인이 5개 단어로 이루어진 문장을 5회 발성한 음성 데이터베이스로 인식실험한 결과 화자식별에서는 0.4%의 에러율을 보였고, 화자검증에서는 0.7%의 에러율을 보였다고 보고하고 있다[44]. 독일의 Ulm 대학은 630명으로 구성된 음성데이터베이스를 이용하여 프레임 당 36.16%의 인식률을, 문장당 97.3%의 인식률을 달성하고 있다. 이때 이용된 음성 데이터베이스는 MIT에서 음성 인식을 위해서 만들어진 것이지만 많은 화자가 포함되어 있으므로 화자인식실험에 이용되었다. 미국의 럭커스(Rutgers) 대학에서는 YOHO 데이터베이스(남성 106명, 여성 32명으로 구성)를 사용하여 남성 화자에 대해서 0.09%, 여성 화자에 대해서는 0.31%이라는 화자식별 에러율 얻었다고 발표하였다. 미국의 AT&T의 Services and Speech Technology Department에서도 YOHO 데이터베이스를 사용하여 평균 EER(Equal Error Rate)이 0.4%의 성능을 얻었다고 보고하였다. 미국의 커즈웨일사(Kurzweil Applied Intelligence Inc)에서도 YOHO 데이터베이스를 사용하여 EER이 0.22%, 화자식별 에러가 0.29%의 성능을 얻고 있다. 미국의 공군 공과 대학(Air Force Institute of Tech-

nology)에서도 YOHO 데이터베이스를 사용하여 EER이 0.28%, 화자 식별 에러가 남성의 경우 0.19%, 여성의 경우 0.31%의 성능을 얻었다.

이상과 같이 많은 연구기관에서 거의 비슷한 성능의 화자 인식 시스템을 개발하고 있음을 알 수 있다. 또한 개발한 화자 인식 시스템의 성능 평가를 위해서 YOHO 음성 데이터베이스를 사용하고 있으며 해마다 약간씩 향상된 결과를 얻고 있어 향후 수년이내 1,000명 이하의 소규모 화자인증을 위한 실용 가능한 시스템이 출현될 것으로 전망된다.

한편, 국내의 화자인식기술은 인식대상 화자의 수나, 인식의 정확도 면에서 아직까지 외국의 수준에는 미치지 못하는 실정이다. 그러나, 인터넷 관련 서비스의 급격한 증가와 더불어 음성을 이용한 고객인증의 필요성이 급증함에 따라 이미 여러 기업체에서 화자시스템을 상용화하고 있으며, 여러 연구기관 및 대학에서도 이에 관한 연구결과를 발표하고 있다. 한국과학기술원(KAIST)에서는 기존의 이산형HMM의 문제점인 벡터 양자화에 의한 왜곡을 줄이고, 보다 정확한 출력 확률 분포추정을 위한 방법으로서 여러 개의 근접코드워드를 고려한 복수관측 이산형 HMM을 사용하고, 각 상태별 코드북의 왜곡거리를 출력확률로 갖는 HMVQM(Hidden Markov VQ-codebook Model)을 이용하여, 문장종속화자검증 및 식별시스템을 구성하여 남녀 486명을 대상으로 한 문장독립 화자식별 실험에서 98.6%을 얻고 있다[45]. 영남대에서는 GMM 모델과 DMR 가중치를 이용한 실시간 문맥독립화자인식 시스템을 개발하여 168명을 대상으로 한 화자식별 실험에서 검증에러률 0.28%, 식별율 95.4%를 얻고 있다[43].

웹프로텍(주)은 현재 온라인 상에서 화자인증기술을 시연하고 있으며, 이 기술을 향후 인터넷 뱅킹이나 사이버 트레이딩 같은 전자금융결제에 적용할 계획을 갖고 있다. 또한, 전자개폐장치에 화자인증기술을 적용한 시제품을 완성하여 2001년에 각 가정의 자물쇠를 음성으로 개폐하는 제품을 선보일 계획이다. L&H(주)는 최근 화자인증 보안시스템을 함경남도 금호지구 소재 한반도에너지개발기구(KEDO) 대북 원전 공사현장의 보안시스템으로 공급했다. 브레이크마인드(주)는 최근 인터넷기반의 화자인증보안솔루션인 [시큐어보이스]를 개발, 상용화하였다. 이 솔류션은 음성을 통해 웹사이트 및 네트워크의 사용자를 식별할 수 있는 것으로 화자가 식별단어들을 설정해 음성을 입력하면 네트워크 접속시 서버측에서 기존에 입력한 단어들 중 하나를 임의로 전송, 발성하게 하는 단어지시형 화자검증방법을 사용하고 있다. 이외에도 현재 T-NETIX, ITT, VeriVoice, Veritel, Keyware, Nuance, VSS(Voice Security Systems) 등이 화자 인증 시스템을 상용화하고 있다.

그러나 인터넷 뱅킹, 사이버트레이딩 등과 같은 대규모 시스템에 음성을 이용한 화자인증을 위해서는 대규모의 데이터를 이용하여 99.9%이상의 인식률을 달성해야만 실용화 가능할 것으로 보여 향후 화자인식에 관한 연구가 보다 더 적극적으로 이루어 져야 할 것으로 생각된다.

6. 맺음말

최근 인터넷 관련 서비스의 급격한 증가와 더불어 음성을 이용한 고객인증의 필요성이 급증함에 화자인식을 이용한 고객인증에 관한 관심이 증대되고 있다. 음성을 이용한 개인확인은 카드, 키 등과 같은 인공적인 수단보다는 매우 편리할 뿐만 아니라 음성은 분실위험이나 도난위험이 전혀 없어 매우 안전하다. 또, 화자인식은 손이나 다른 도구를 전혀 필요로 하지 않으므로 급속히 발전하는 정보화시대의 각종 시스템 구현에 중요한 기술로서 각광받고 있다. 이와 같은 요구에 따라 화자인식기술이 인터넷을 이용한 각종 개인의 인증방법에 이용되기 시작되었으며 법정 증빙도구로서도 이용되기 시작되고 있다. 특히, 전화선 또는 인터넷을 이용한 은행간 송금, 잔고조회, 쇼핑, 음성메일, 개인정보조회, 예약, 컴퓨터의 원격접근, 극비장소에의 접근통제 등에 편리하게 이용될 수 있다.

본고에서는 이와 같은 요구에 따라 화자인식에 관한 기본이론, 화자인식 방법, 시스템의 종류, 화자인식에 관한 연구의 국내외 동향 등에 관해 살펴보았다. 현재까지 개발된 화자인식 시스템은 한정된 소규모 인증을 위한 시스템 구성에는 별 문제없으나 인터넷 뱅킹, 사이버 드레이딩과 같은 대규모 시스템의 고객인증을 위해서는 대규모의 데이터를 이용하여 99.9%이상의 화자검증/식별률을 달성해야만 실용화 가능할 것으로 보여 향후 화자인식에 관한 연구가 보다 더 적극적으로 이루어져야 할 것으로 생각된다.

참고문헌

[1] Atal, B. S. (1972) 'Automatic speaker recognition based on pitch contours, 'J. Acoust. Soc. Amer., 52, 6(Part 2), pp. 1687-1697.

[2] Atral, B. S. (1974)'Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, ' J.Acoust. Soc. Amer., 55, 6, pp. 1304-1312.

[3] Carey, M. and Parris, E. (1992) 'Speaker verification using connected words,' Proc. Institute of Acoustic, 14, 6, PP. 95-100.

[4] Doddington, G. (1985) 'Speaker recognition-Identifying people by their voices,' Proc. IEEE, 73, 11, pp. 1651-1654.

[5] Frui, S., Itakura, F., and Saito, S. (1972)'Talker recognition by longtime averaged speech spectrum,' Trans. IEC더, 55-A, 10, pp. 549-556.

[6] Frui, S. (1978) Research on Individuality Information in speech pp. 549-556.

[7] Frui, S. (1981) 'Comparision of speaker recognition methods using statistical features and dynamic features,' IEEE Trans. Acoust., Speech, Signal processing, ASSP-29, 3, pp. 342-350.

[8] Frui, S. (1981) 'Cepstral analysis technique for automatic speaker verification,' IEEE Trans. Acoust., Speech, Signal Processing, ASSP-29,2, pp. 254-272.

[9] Frui, S. (1996)'An overview of speaker recognition technology,' in Automatic Speech and Speaker Recognition (eds. C.-H.Lee, F. K. Soong and K.K. Paliwal), Kluwer, Boston, pp. 31-56.

[10] Frui, S. (1997) 'Recent advance in speaker recognition,' Pattern Letters, 18, pp. 859-872.

[11] Griffin, C., Matsue, T. and Frui, S.(1994) 'Distancemeasures for text-independent speaker recognition based on MAR model,' Proc. IEEE Int. C ONF. Acoust. Speech, Signal Processing, Adelaide, 23. 6. pp. 309-312

[12] Higgins, A., Bahler, L. and Porter, J. (1991)' Speaker verification using randomized phrase prompting,' Digital Signal Processing, 1, pp. 89-106.

[13] Kersta, L. G. (1962) 'Voiceprint identification,' Nature, 196, pp. 1253-1257.

[14] Li, K. P. and Wrench, Jr., E. H. (1983) 'An approach to text-independent speaker recognition with short utterances,' Proc. IEEE Int, Conf. Acoust., Speech, Signal Processing, Boston, MA, 12.9, pp. 558-558.

[15] Kunzel, H. (1994) 'Current approaches to forensic speaker recognition,' ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 135-141.

[16] Markel, J., Oshika. B. and Gray, A. (1997) 'Long-term feature averaging for speaker recognition,' IEEE Trans. Acoust. Speech Signal Processing, ASSP-25, 4, pp. 330-337.

[17] Markel, J., and Davi, S.(1979) 'Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base ,'IEEE Trans. Acoust. Speech Signal Processing, ASSP-27, 1, pp. 74-82.

[18] Matsui, T. and Frui, S. (1990) 'Text-independent speaker recognition using vocal tract and pitch information,' Proc. Int. Conf. Spoken Language Processing, Kobe, 5.3, pp. 137-140.

[19] Matsui, T. and Frui, S. (1991) ' A text-independent speaker recognition method robust against utterance variation ,'Proc. IEEE Int. Conf . Acoust. Speech Signal Processing, S6.3, pp. 377-380.

[20] Matsui, T. and Frui, S. (1993) 'Concatenated phoneme models for

text-variable speaker recognition,' Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, Minneapolis, pp. II-391-394.

[21] Matsui, T. and Frui, S. (1994a) 'Speaker adaptation of tied-mixture-based phoneme models for text-fprompt speaker recognition,' Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, Adelaide, 13.1.

[22] Matsui, T. and Frui, S. (1994b) 'Similiarity normalization method for speaker verification based on a posteriori probabillity,' ESCA Workshop on Automatic Speaker Recoggnition, Identification and Verification, pp. 59-62.

[23] Montacie, C., Deleglise, P., Bimbot, F. and Caraty, M. -J.(1992)'Cinematic techniqes for speech processing: Temporal decomposition and multivariate linear prediction,' Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, San Francisco, pp. I-153-156.

[24] Naik, J., Netsch, M. and Dodding, G. (1989)'Speaker verification over long distance telephone lines, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,' S10b.3, pp. 524-527.

[25] National Research Council (1979) On the Theory and Practice of Voice Identification, Washington, D. C.

[26] Newman, M., Gillick, L., Ito, Y., McAllaster, D. and Peskin,B (1996) 'Speaker verification through large vocaburary continuous speech recognition,' Proc. Int. Conf. Spoken Language Processing, Philadelphia, pp. 2419-1294.

[27] Reynolds. D. (1994) 'Speaker identification and verification using Gaussian mixture speaker models,' ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 27-30.

[28] Rose, R. and Reynolds, R. (1990) 'Text independent speaker identification using automatic acoustic segmentation,' Proc . IEEE Int. Conf. Acoust. Speech, Signal Processing, S51. 10, pp. 293-296.

[29] Rosenberg, A. E. and Sambur, M. R. (1975)' New techniques for automatic speaker verification,' IEEE Trans. Acoust., Speech, Signal Processing, ASSP-23,2, pp. 169-176.

[30] Rosenberg, A. and Soong, F. (1987) 'Evaluation of a vector quantization talker recognition system in text independent and text dependent modes,' Computer Speech and Language, 22, pp. 143-157.

[31] Rosenberg, A., Lee, C. and Goken, S. (1991) 'Connected word talker verification using whole word hidden Marcov models,' Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, Toronto, S6.4, pp. 381-384.

[32] Rosenberg, A. and Song, F. (1991) 'Recent research in automatic speaker recognition,' in Advances in Speech Signal Processing (eds S. Frui and M. Sondhi), Marcel Dekker, New York, pp. 701-737.

[33] Rosenberg. A. (1992) 'The use of cohort normalized scores for speaker verification,' Proc. Int. Conf. Spoken Language Processing, Banff, Th. sAM.4,2, pp. 599-602.

[34] Savic, M. and Gupta, S. (1990) 'variable parameter speaker verification system based on hidden Markov modeling,' Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, S5.7, pp. 281-284.

[35] Setlur, A. and Jacobs, T. (1995) 'Result of a speaker verification service trial using HMM models,' EUROSPEECH' 95, Madrid , pp. 639-642.

[36] Shikano, K.(1985) 'Text-independent

speaker recognition experiment using codebooks in vector quantization,' J. Acoust. Soc. Am. (abstract), Suppl. 1, 77, S11.

[37] Soong, F. K. and Rosenberg, A. E. (1986) 'On the use of instantenous andtransitional spectral information in speaker recognition,' Proc IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 877-880.

[38] Soong, F., Rosenberg, A. and Juang, B. (1987) ' A vector quantization approach to speaker recognition,' AT&T Technical Journal, 66, pp. 14-26.

[39] Tishby, N .(1991) 'On the application of mixture AR hidden Marcov models to text independent speaker recognition,' IEEE Trans. A COOUST. Speech,Signal Processing, ASSP-30,3, pp. 563-570.

[40] Tosi, O., H., Lashbrook, W., C., Nicol, J.M and Nash, E. (1972) 'Experiment on voice identification,' J. Acoust. Soc. Amer., 51,6(part2), pp. 2030-2043.

[41] Zheng, Y. and Yaun, B.(1988) 'Text-dependent speaker identification using circular hidden Marcov models,' Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, S13.3, pp. 580-582.

[42] Poritz, A.(1982)'Linera predictive hidden Markov models and the speech signal', Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing, S11.5, pp. 1291-1294

[43] 김민정, 석수영, 정현열,(2001) 'Gaussian Mixture Model을 이용한 실시간 문맥독립화자인식에 관한 고찰' 한국음향학회 하계학술대회논문집, 2001.7

[44] 한국전자통신연구원, 연구동향보고서, 2000.

[45] 김세현, 장길진, 오영환, "개인성정보의 가중화에 의한 화자확인의 성능향상", 한국정보과학회 춘계학술발표대회 논문집, 제2권 pp. 539~541 1999, 10.

정 현 열

1975 영남대학교 전자공학과 졸업
1989 일본 동북대학교 정보공학과 수료(공학박사)
1989. 3~현재 영남대학교 전자정보공학부 교수
1992. 7~1993. 7 Carnegie-Mellon 대학 Robotics 연구소 객원연구원
1994. 12~1995. 2 일본 토요하시기술과학대학 외국인 연구자
2000. 6~2000. 8 미국 Qualcomm Inc. 수석 엔지니어

학회활동:한국음향학회(영남지회장), 한국통신학회, IEEE, 일본음향학회, 일본 전자정보통신학회 각 회원
관심분야:음성인식, 화자인식, 음성합성 및 DSP 응용분야
E-mail:hychung@ynucc.yeungnam.ac.kr