

技術解説

화자 독립 음성 인식 및 화자 적응에 관한 연구

오 광 철* · 김 흥 국* · 김 희 린** · 이 황 수*

(*한국과학기술원, 정보및 통신공학과, **한국전자통신연구소)

I. 서 론

1. 연구의 목적 및 중요성

음성 인식 시스템에서는 개개인의 음성으로부터 특징을 추출할 때 가능하다면 개인차를 줄일 수 있는 방향으로 파라미터를 추출하여 인식에 이용하여야 한다. 그러나 모든 개인차를 극복할 수 있는 파라미터 정규화 방법이 제시되지 못하기 때문에 음성 인식 시스템은 인식 대상에 따라 특정 화자의 음성만을 인식하는 화자 종속 시스템과 화자에 관계없이 인식할 수 있는 화자 독립 시스템으로 나뉘게 되며, 현재의 기술로는 화자 종속 시스템의 인식률과 비슷한 화자 독립 시스템은 개발되어 있지 않다.

화자 종속 시스템의 경우 화자가 바뀔 때마다 특정 화자에 맞게끔 새롭게 화자 종속 시스템을 재구성해야 하기 때문에 많은 데이터와 시간이 필요하게 되므로 시스템을 재구성하지 않고서 기존의 시스템 정보를 충분히 이용하고 새로운 화자의 음성 정보를 약간만 사용하여서 화자 종속 시스템의 성능과 비슷하게 하는 방식이 화자 적응이다. 실제로 화자 적응이란 완벽한 화자 독립 시스템을 만들 수 없으므로 화자 종속 시스템을 화자에 독립적으로 사용할 수 있도록 하는 방법이지만, 동일한 화자라도 화자의 건강 상태, 기분, 또는 발음할 때의 주변 환경에 따라 음성의 특징이 다양하게 변하기 때문에 특정 화자에 맞게 구성된 음성 인식 시스템을 이러한 조건에 따라 시스템 내부의 정보를 바꾸어 줄 때에도 이용될 수 있다. 이러한 생각은 인간이 새로운 환경, 혹은 새로운 화자

의 음성에 적응하면서 음성을 인식한다는 사실과 유사하다.

화자 적응의 기본적인 목적은 한명의 화자에 대해서만 음성 데이터를 분석하여 음성 인식을 수행하는 것이 아니라 그 대상이 다수 또는 모든 사람에 대해서도 음성 인식이 가능하게 그 인식 시스템을 화자가 바뀔 때마다 그 새로운 화자에 맞게 변형시켜 주는 것이다. 즉, 화자 종속 음성 인식기를 입력 화자가 바뀔 때마다 다시 구성하게 되면 학습하는데 많은 시간이 필요하게 되고 또 어떤 경우에는 짧은 시간으로 음성 인식을 수행해야 하는 경우도 있으므로, 어떻게 하면 적은 음성 데이터를 가지고 짧은 시간에 현재의 음성인식 시스템을 새로운 화자에 맞도록 변환하는 문제가 된다.

2. 연구의 방향

기본적으로 화자 적응은 화자 독립 인식 시스템으로부터 새로운 화자에 대하여도 음성 인식이 가능하도록 인식 대상을 넓혀주는 기술이나, 최근들어 화자 독립 인식 시스템에도 화자 적응 기술을 사용하여 인식률을 높이려는 시도가 연구되고 있다. 따라서 본 논문은 적절한 화자 독립 인식률을 가지는 기본 인식 시스템을 2장에서 기술하고, 특정 화자에 대한 인식 성능을 높이기 위한 화자 적응 알고리즘들을 3장에서 기술한다. 4장에서는 이들 화자적응 알고리즘중에서 정적 지도적응에 적합한 확률적 스펙트럼 사상법에 의한 적응 알고리즘을 설명하고, 마지막으로 5장에서 이들을 요약 정리한다.

II. 음성 인식 알고리즘

1. 음성인식 알고리즘의 개요 및 분류

음성인식 기술은 1970년대 초부터 연구가 진행되어 많은 진전이 이루어졌지만 현재까지 인간과 비슷한 능력을 지닌 인식 기술은 아직 개발되어 있지 않다. 그러나 최근 10여년간의 급속한 기술 개발로 미국, 일본 등지에서는 격리단어를 인식할 수 있는 상업용 제품이 나와 있으며 연속단어를 인식할 수 있는 시스템들도 연구실에서 개발되고 있다. 이러한 음성인식 시스템을 개발하기 위하여 그동안 많은 기술들이 음성 인식에 적용되어 효과적으로 사용되고 있다. 예를들면, 자연스러운 연속 음성을 인식하기 위해서는 수의 인식 단위를 이용하여 문장 단위의 음성을 모델링하는 방법, 사용하는 언어의 문법이나 구문등을 이용하는 방법, 자연스러운 음성에서 나타나는 다양한 음가의 변화를 수용할 수 있는 인식 단위의 선정 방법등이 연구되었다. 다양한 화자의 음성을 인식하기 위해서는 다수의 표준 패턴을 선택하는 방법과 새로운 화자에 대하여 인식 시스템을 적응시키는 화자 적응 방법, 화자 독립적인 음성 특징의 추출 방법등을 연구하였다. 이상에서 언급한 것 이외에도 여러 가지 방법이 음성 인식에 이용되었고 음성을 인식하기 위한 기본적인 접근 방식도 많은 발전을 이룩하였으며 새로운 방법들이 제안되었다¹¹⁾.

지금까지 개발된 음성 인식 알고리즘은 크게 네가지로 나누어 볼 수 있다. 첫째로는 동적 프로그래밍 기법을 이용하여 기준이 되는 음성과 입력 음성을 비교하여 최소거리 법칙에 의하여 입력 음성을 인식하는 dynamic time warping (DTW) 방법이다¹²⁾. 이 방법은 한 사람 혹은 여러 사람으로 하여금 인식 대상 어휘를 여러번 발음하도록 한 후 clustering 알고리즘¹³⁾에 의하여 기준 패턴을 선택하는 훈련 과정과 이 기준 패턴과 입력 음성을 비교하는 음성 인식 과정으로 나뉘어 진다. 음성은 동일한 사람이 동일한 단어를 발음하여도 발음 속도가 매번 다르므로 패턴 사이의 유사도를 측정하려면 이 시간축 상의 변화를 고려하여야 한다. 이 시간축 상의 변화를 보상해주는 효과적인 방법이 DTW이다. 이 DTW 알고리즘을 이용하여 격리 단어뿐만 아니라 연결 단어까지도 인식

하기 위하여 one-stage dynamic programming (OSDP)¹⁴⁾ 방법이나 level-building dynamic time warping(LBDTW)¹⁵⁾ 방법등이 제안되었다. DTW를 이용하는 음성 인식 방법은 높은 인식률을 얻을 수 있다는 장점이 있으나 많은 수의 기준 패턴을 구성하기 위해서는 많은 시간과 노력이 필요하며 인식 과정에 걸리는 시간이 길기 때문에 100여 단어 정도의 어휘를 화자 종속으로 인식하는 경우에 적합하나 대용량 단어의 인식에는 적합하지 못하다는 단점이 있다.

두번째로는 hidden Markov model(HMM)을 이용하여 음성을 확률적으로 모델링하는 방법이 있다¹⁶⁾¹⁷⁾. 이 방법은 음성을 상태 천이 확률 및 각 상태에서의 출력 심볼과 관찰 확률을 갖는 Markov process로 가정한 후에 훈련용 데이터를 통하여 상태 천이 확률 및 출력 심볼 관측 확률을 추정하는 훈련 과정과 추정된 모델에서 입력 음성이 발생할 확률을 계산하는 인식 과정으로 나누어 진다. 이러한 HMM을 이용한 음성 인식 시스템의 성능을 향상시키기 위하여 음성에 관한 여러가지 정보(음소의 지속 시간, 음운의 탈락, 변화, 첨가등)와 HMM 과정에 사용되는 벡터 양자화 과정의 왜곡을 없애거나 줄이기 위한 방법(continuous or semi-continuous HMM)등 여러가지 방법이 제안되고 연구되어 왔다¹⁸⁾. 특히, HMM이 갖는 장점중의 하나인 단어 이하의 음성 인식 단위, 예를들면 음소, diphone등으로 부터 단어나 문장등의 모델을 쉽게 구성할 수 있다는 점을 이용하면서 인식률을 향상시킬 수 있도록 인식 단위를 선정하는 방법에 관한 연구도 수행되었다¹⁹⁾. HMM을 이용한 음성인식 시스템은 인식 소요 시간이 짧고 적은 수의 음성 인식 단위로 부터 쉽게 어휘들을 모델링할 수 있다는 장점을 가지고 있어서 DTW보다 앞선 방법으로 평가받고 있다.

세번째로 신경망(neural network)을 이용하여 음성을 인식하는 방법이 최근에 제안되어 주목을 받고 있다. 신경망은 인간의 신경 조직을 모방한 것으로 입력을 받아들여 입력의 합에 비선형성을 가한 후 이를 출력으로 내보내는 node들을 상호 연결함으로써 구성된다. 이 node들의 연결방식에 따라서 여러가지 형태의 신경망이 구성되며 그 동작은 node간의 연결의 세기를 결정하여 주는 가중치(weight)에 따라 큰

영향을 받는다¹¹⁶⁾. Minsky와 Papert에 의하여 perceptron이 제안된 후에 나중 구조를 갖는 multi layer perceptron(MLP)의 기층적층 훈련시킬 수 있는 back propagation 훈련 알고리즘¹¹⁷⁾이 발표된 후에 이를 음성 인식에 적용하는 연구가 활발히 진행되고 있다. 신경망이 갖는 장점은 병렬처리가 가능하므로 계산 속도를 높일 수 있다는 점과 신경망의 성능을 입력 데이터가 주어짐에 따라서 점진적으로 향상시킬 수 있도록 가중치를 학습시킬 수 있다는 것이다¹²¹⁾. 이러한 이유로 신경망을 이용하여 음성을 인식하는 연구가 많이 진행되어서, 지도(supervised) 학습에 해당하는 MLP, learning vector quantization 2(LVQ2), time-delayed neural network(TDNN)등을 이용하거나 독자(unsupervised) 학습에 해당하는 feature map등을 이용하여 음소등을 인식하는 연구가 수행되었다^{118) 119)}. 그러나 신경망은 정적(static) 패턴의 인식에는 우수한 성능을 보이지만 시간에 따라 변화하는 동적(dynamic) 패턴의 인식에 취약한 점이 있다. 이를 해결하기 위하여 최근에는 여러가지 recurrent neural network¹²⁰⁾을 이용하거나 DTW나 HMM등과 결합하여 음성 인식을 수행하고자 하는 연구가 진행되고 있다¹²¹⁾. 또한 음성 인식에 적합한 신경망의 구조나 학습 시간의 단축에 관한 연구도 수행되고 있다.

마지막으로 음성에 관한 지식을 이용하여 음성을 인식하는 knowledge-based 음성 인식 방법이 있다. 이 방법은 음성에 관한 지식을 구축하고 이를 여러가지 규칙에 의하여 구성되는 inference engine을 이용하여 입력 음성을 인식하도록 하는 방법이다. 이러한 방법중 가장 대표적인 것은 spectrogram reading 과정에서 얻은 지식을 바탕으로 전문가 시스템(expert system)을 구성하여 음성을 인식하는 Zue의 연구이다¹²²⁾. 이 방법은 전문적인 spectrogram reader가 spectrogram을 보고 음성을 인식할 수 있다는 점에 착안하여 연구되었으나, 인간이 시각 정보를 처리하는 과정을 전문가 시스템으로 효과적으로 구현할 수 없기 때문에 다른 방법에 비하여 음소의 분할이나 labeling등의 영역에서 사용되고 있다¹²³⁾.

이상에서 살펴본 네가지의 음성 인식 방법은 각기 장단점을 가지고 있어서 이를 독립적으로 구현하기 보다는 각각의 장점을 살릴 수 있도록 혼합하여 구현

하는 것이 인식 성능을 향상시킬 수 있는 방법이다. 다음 섹션에서는 앞서 살펴본 인식 알고리즘중에서 HMM을 이용하는 인식 알고리즘을 기본으로 하여 여기에 화자 적응 알고리즘은 적용할 수 있도록 화자 독립 인식 시스템의 성능 향상에 관하여 몇가지 방법을 제안한다.

2. HMM에 기초한 음성 인식 알고리즘의 개선 방법

HMM을 이용한 음성 인식은 앞서 기술한 바와 같이 출력 확률 분포의 정의에 따라 크게 두가지로 분류할 수 있다. 하나는 이산(discrete) HMM이고 다른 하나는 연속(continuous) HMM이다. 연속 HMM은 이산 HMM에 비하여 훈련시 필요로 하는 데이터량이 많을 뿐만 아니라, Gaussian 분포와 같은 인위적인 확률 분포를 가정해야 한다. 하지만 충분한 데이터량에 대해서는 이산 HMM보다 우수한 성능을 보여준다. 기본적으로 이산 HMM시스템은 두가지의 문제점을 가지고 있다. 첫번째는 HMM을 훈련할 때 maximum likelihood estimation(MLE)를 사용한다는 것이고, 두번째로는 이산 HMM의 전처리기로 벡터 양자화를 사용함으로써 생기는 문제이다. 본 논문에서는 이 두가지 문제점에 대한 기존의 해결 방법들에 수정을 가하여 보다 성능이 우수한 알고리즘을 제안한다.

먼저 HMM을 훈련할 때 일반적으로 사용하는 MLE 방법¹²⁴⁾을 고려한다. 이 MLE 방법에 의한 모델 파라미터 추정 방법에는 전향 변수(forward parameter)와 후향 변수(backward parameter)를 이용하는 Baum-Welch 알고리즘과 모델의 most-likely state sequence 정보를 이용하는 segmental k-means 알고리즘이 있다. 이 방법들은 훈련용 데이터와 모델사이의 likelihood를 최대로 하도록 파라미터를 추정하는 방법으로서 가정된 모델과 실제의 음성 발생 모델이 동일한 때에 정확한 모델 파라미터를 얻을 수 있다. 그러나 실제로는 이 두 모델들이 일치하지 않기 때문에 이를 보완하기 위한 새로운 파라미터 추정 방법이 필요하게 된다. 이를 위해 제안된 방법들로 대표적인 것이 훈련용 데이터와 모델사이의 상호 정보(mutual information)을 최대화하는 maximum mutual information estimation(MMIE)¹²⁵⁾ 방법과 훈련용 데이터에 대하여 인식 과정을 수행한

후 오인착이 발생하기 경우에 파라미터를 수정하여 주는 corrective training(CT)^[21] 방법등이 있다. 이 두가지 방법들의 비교연구 결과 CT 방법이 더 우수한 것으로 발표되었다. 기존의 CT 방법은 모델 파라미터 교정시에 Baum-Welch 알고리즘을 사용하므로 전향-후향 알고리즘이 가지고 있는 단점들을 그대로 내포하고 있다. 이에 대한 대안으로 본 과제에서는 전향-후향 알고리즘 대신에 segmental k-means 알고리즘을 CT 방법에 적용하는 것을 제안한다. 이렇게 함으로써 개선된 CT 방법은 기존의 방법에 비하여 첫째, 전향-후향 알고리즘에 필수적인 변수의 scaling 과정을 피할 수 있고, 둘째, 모델링과 decoding 과정이 동일한 criteria에서 수행되므로써 연속 음성 인식에 적합하며, 셋째, 음성 신호내의 공통의 통계적 특성을 지니는 모델 상태 정보를 이용함으로써 인식 성능을 향상시킬 수 있다. 이 방법을 우리는 segmental corrective training(SCT) 방법이라 한다^[22].

이산 HMM이 가지고 있는 두번째 문제로서 벡터 양자화기에 의한 정보의 손실이 있다. 기존의 벡터 양자화기는 각 입력 스펙트럼을 거리상 가장 가까운 하나의 코드워드(codeword)에 대응시킴으로써 다른 코드워드와의 관계에 대한 정보를 잃어버리게 된다. 이에 대한 해결책으로 제안된 방법이 fuzzy vector quantizer(FVQ)이다^[23]. FVQ는 각 입력 특징 벡터를 각각의 코드워드와의 거리에 따라 각 코드워드를 발생시킬 확률로 대응시킴으로써 기존의 VQ가 가지고 있는 정보 손실을 보상할 수 있다. 이 FVQ를 사용하여 기존의 Baum-Welch 알고리즘을 수정하는 방법이 제안되어 기존의 VQ/HMM 시스템보다 우수한 성능을 보여주었다. FVQ를 이용하는 또다른 연구로서, FVQ를 훈련 과정이 아니라 추정된 모델 파라미터를 smoothing하는 과정에 적용하는 방법이 있다^[24]. 이러한 방법들은 VQ 대신에 FVQ를 이용함으로써 HMM 인식 시스템의 성능을 향상시키고, 또 훈련에 필요한 데이터량을 줄일 수 있지만 훈련에 소요되는 시간이 길어지는 단점이 있다. 이를 보완하기 위하여 본 연구과제에서는 FVQ를 Baum-Welch 알고리즘 대신에 segmental k-means 알고리즘에 적용하는 방법을 제안한다. 이렇게 함으로써 앞서 기술한 바와 같은 segmental k-means 알고리즘의 장점을

그대로 살릴 수 있고 또한 인식 시스템의 성능도 향상시킬 수 있다. 또한 FVQ에서 각 입력 특징 벡터에 대한 각 코드워드의 fuzziness를 모두 구하지 않고, 거리상으로 가까운 단지 몇개의 코드워드에 대한 fuzziness를 구하여 이를 segmental k-means 알고리즘에 적용할 수도 있다.

Ⅲ. 화자 적응 알고리즘 분석

1. 화자 적응 알고리즘의 분류

화자 적응 알고리즘은 적응 과정의 존재 여부에 따라서 정적 적응(static adaptation)과 동적 적응(dynamic adaptation)으로 나눌 수 있으며, 적응 데이터의 시스템 인지 여부에 따라 지도 적응(supervised adaptation)과 독자 적응(unsupervised adaptation)으로 구분 된다.

정적 지도 적응 알고리즘은 새로운 화자가 기존의 화자 종속 인식 시스템을 이용하기 위하여 적응 과정이란 특정한 모드에서 미리 지정된 어휘를 발음하여 새로운 화자에 적합하도록 적응시키는 알고리즘으로 그동안 가장 많이 연구되어 왔다. Y.Grenier는 새로운 화자가 발음한 음성과 기존의 화자가 발음한 음성 사이의 상관도(correlation)가 최대가 되도록 적응시키는 정규 상관 해석(canonical correlation analysis)에 의한 알고리즘을 개발하였으며^[25], G. Rigoll은 화자 마코프 모델(speaker Markov model) 개념을 도입하여 화자간의 음성 변화를 마코프 모델로 모델링하고 기존의 음성 정보를 이 마코프 모델을 통해서 새로운 화자에 적응하도록 하였다^[26]. 이러한 알고리즘은 음성 특징 벡터상에서 직접 적응이 가능하고 적응 상태도 높은 장점이 있지만 적응 시간이 길어지는 단점이 있다. 한편 BBN 연구실에서는 BVBLOS라는 화자 종속 음성 인식 시스템에 사용하기 위하여 수년간에 걸쳐 정적 지도 적응 알고리즘을 개발하였는데 VQ 코드북(codebook), HMM 파라미터 영역에서 확률적 스펙트럼 사상(probabilistic spectral mapping) 방식으로 적응을 수행하는 방법^[27]과 텍스트 종속 확률적 스펙트럼 사상(text dependent probabilistic spectral mapping) 방식이 있다^{[28][29]}. 또한 최근에는 화자 독립 인식 시스템에서도 화자 단위로 화자 적응을 수행하여 인식률을 향상

시킨 SPHINX 시스템의 deleted interpolation adaptation 알고리즘이 개발되었다³¹⁾.

정적 독자 적응 알고리즘은 적응 과정이런 특정한 모우드가 존재하지만 화자 적응을 위한 어휘가 미리 지정되어 있지 않고 새로운 화자가 인식 대상 어휘중에서 임의의 어휘를 발음하여도 화자 적응이 되는 알고리즘이다. 이 알고리즘은 새로운 화자가 적응 과정 모우드에서 특정한 어휘를 발음할 필요성이 없으므로 편리하지만 적응도가 떨어지는 단점이 있다. NTT의 Sadaoki Furui는 VQ 코드워드 사이의 관계를 계층 구조로 만들어서 새로운 화자가 임의의 어휘를 발음하더라도 적응할 수 있는 알고리즘을 개발하였다³⁰⁾.

동적 독자 적응 알고리즘은 새로운 화자를 위한 특정 적응 모우드가 없이 시스템 내부에서 자동적으로 새로운 화자에 적응하는 알고리즘이다. 그러므로 이 알고리즘을 사용하면 외형적으로 화자 독립 인식 시스템과 같은 역할을 수행하지만, 내부적으로는 화자가 바뀌거나 혹은 동일한 화자일 경우에도 인식 작업을 수행할 때마다 인식 시스템을 사용하는 화자에 게로 적응하게 된다. Richard M. Stern은 FEATURE 시스템에 MAP(maximum a posteriori probability) 추정 이론을 이용하여 동적 독자 적응 알고리즘을 개발하였으나 적응 영역의 확장과 인식률의 개선이 필요하다³³⁾.

이상의 분류는 화자 적응 시스템의 특징에 따른 분류이지만, 이를 화자 적응의 방법론에 따라 스펙트럼 특징에 기초를 둔 방법과 학습된 파라미터에 기초한 방법의 두 가지 부류로 크게 나눌 수 있다. 다음절에서는 화자 적응을 방법론적으로 분류하고 설명한다.

2. 스펙트럼 특징에 기초한 방식

스펙트럼 특징에 기초한 방식은 음성 신호의 주파수 특성을 화자에 따라 차이가 심하지 않도록 조작하는 것으로, 주로 템플레이트 정합(template matching) 방식에서 사용된다. 임의의 화자 종속 인식 시스템에 새로운 화자에 대한 음성 신호가 들어오면, 새로운 음성 신호의 주파수 특성이 원 화자의 주파수 특성과 일치하지 않으므로 인식 시스템이 제대로 인식할 수가 없다. 따라서 새로운 음성 신호의 스펙트럼을 원 화자와 비슷하도록 하거나, 인식 시스템의

템플레이트를 화자간의 스펙트럼 변화 형태에 따라 변화시켜서 화자 적응을 한다. 그 스펙트럼을 변화시키는 방식에 따라 다음과 같은 기법들이 있다.

- 평균(averaging)법
- 정규화(normalization)법
- 변환 함수(transform function)법
- 화자 마코프 모델(speaker Markov model) 방법

가. 평균법

이미 구성된 기준 템플레이트를 새로운 화자에 맞도록 재구성하는 것으로 새로운 화자로부터 얻어진 템플레이트들을 평균하여 새로운 기준 템플레이트로써 대치시키는 것이다³²⁾. 이 방법은 DTW를 이용한 패턴 정합 방식에서 초기의 화자 적응 방식으로 개발되었으며, 비교적 많은 적응 데이터가 필요하고 기본적으로 음성 인식 시스템을 학습하는 방법과 비슷하다.

나. 정규화법

최적의 특징에 상관하게 인식할 수 있도록 화자의 특성을 제거하는 방식이다. 화자에 따른 변화는 화자 고유의 피치(pitch)와 성도(vocal tract)의 길이 또는 모양에서 기인한다. 피치는 성(sex)에 따른 변화가 크지만 같은 성에서는 변화가 작다. 그러나 성도는 화자에 따른 변화가 많으며, 이러한 성도의 변화는 음성 신호의 포먼트(formant)에 반영이 된다. 따라서 화자에 따른 포먼트의 변화를 정규화함으로써 화자 적응을 수행한다³⁶⁾³⁷⁾. 이와 같은 방식은 동적 지도 적응에도 적용할 수 있으나 인식 시간이 길어지는 단점이 있다.

다른 방법으로 부트스트래핑(bootstrapping) 방식이 있는데, 이 방법은 새로운 화자의 음성 신호로부터 기준 템플레이트를 바꾸어 주는 방법으로 다음과 같다³⁸⁾³⁹⁾.

- 기준 패턴을 얻는데 사용된 학습 데이터를 새로운 화자가 발음한다.
- 동적 패턴 정합을 이용하여 기준 화자의 음성 패턴과 새로운 화자의 음성 패턴 사이의 일치(alignment)를 이룬다.
- 새로운 화자의 음성 패턴에서 기준 템플레이트와

일치하는 부분을 새로운 화자의 템플레이트로서 얻는다.

부트스트래핑 방법은 정적 지도 적응 방식이며, 대체로 정규화 화자 적응 방식은 DTW에 기초한 음성 인식 시스템의 화자 적응에 사용되었다.

다. 변환 함수법

새로운 화자의 스펙트럼 차이를 선형 또는 비선형 변환 함수를 사용하여 화자 적응을 수행하는 방법이다. 이러한 방식의 화자 적응은 변환 함수를 찾는 것이 중요하며, 그 방식은 음성 인식 시스템의 특성에 따라 다양하다. 그 하나로 정규 상관 해석(canonical correlation analysis) 방식이 있다^{[26][27]}. 이 방법은 그림 1과 같이 기준 스펙트럼 패턴(X_i)과, 같은 단어(w)를 새로운 화자가 발음한 스펙트럼 패턴(Y_i)를 사영 연산자(projection operator) P_r 과 P_n 을 이용하여 새로운 스펙트럼 공간(spectral space)로 변환시킨다.

$$x_i = P_r(X_i), \quad y_i = P_n(Y_i) \tag{1}$$

여기서 X_i 와 Y_i 는 각각 $p \times n$ 행렬 형태이고 p 와 n 은 각각 프레임(frame) 수와 스펙트럼 특징 벡터의 dimension이다. 따라서 least mean square적인 면에서 거리(dimension) $d(x_i, y_i)$ 가 최소가 되도록 하는 P_r, P_n 을 구하여 기준 패턴을 다음과 같이 변환시킨다.

$$X'_i = P_n^{-1}(X_i - X_0) + Y_0 \tag{2}$$

여기서 X_0 와 Y_0 는 각 화자에 대한 평균 벡터이다.

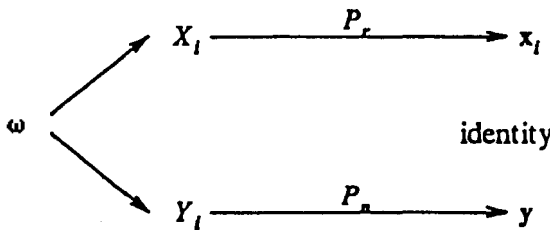


그림 1. 정규 상관 해석.

라. 화자 마코프 모델

새로운 화자의 스펙트럼을 변화시키는 방식은 인식 시간을 길어지게 만드는 단점이 있으며, 템플레이

트를 변화시키는 방식은 HMM에 적당하지 않다. HMM은 학습 데이터로부터 그 파라미터들을 예측하여야 하므로 많은 데이터가 필요하다. 화자 적응은 비교적 적은 데이터로부터 비교적 정확한 파라미터를 구하는 것으로, 모자라는 데이터를 원 화자의 음성 신호로부터 얻어내고자 하는 방식이 화자 마코프 모델이다^[26].

이 방법은 정적 지도 적응에서 가능하며 새로운 화자의 음성 신호는 적어도 되는 장점이 있다. 그러나 원 화자의 음성 신호를 기억해 두어야 하며, 화자 적응을 하는데 걸리는 시간이 학습하는데 소요되는 시간보다도 더 길어지는 단점이 있다.

3. 학습된 파라미터 방법

학습된 파라미터 방법은 음성 신호의 스펙트럼을 직접 조작하는 것이 아니라, 그 중간에 적절한 파라미터를 찾아서 사상등의 과정을 통하여 새로운 화자의 음성도 인식할 수 있도록 하는 방법이다. 따라서 이 방법은 음성특징을 이용한 음성 인식 시스템(MAP 추정)이나 HMM을 이용한 음성 인식에서 사용한다(확률적 스펙트럼 사상).

가. MAP 추정

MAP 추정 방법은 템플레이트의 화자 독립 mean과 variance의 priori knowledge를 이용한다^[31]. 특징에 근거한 인식 시스템에서 사용된 특징이 화자 독립적이라도 특징값이 화자내보다 화자간 변화가 크을 이용해서 특징들의 통계적 파라미터를 새로운 화자에 적응시키는 방법이다. 즉 특징들과 그의 평균값들의 확률 밀도 함수가 주어졌을 때 mean vector M에 대한 MAP 추정은 다음의 a posteriori probability를 최대로 하는 값으로 추정된다.

$$P(M|x) = \frac{P(x|M) P(M)}{p(x)} \tag{3}$$

여기서, $P(x|M)$ 은 특징 벡터들의 mean vector가 주어졌을 때 특징들의 조건화 확률 밀도 함수이고, $p(x)$ 와 $P(M)$ 은 특징 벡터와 mean vector의 a priori 확률 밀도 함수이다. 이를 이용한 적응 방법에서 독자 적응 방식에는 인식 문자의 posteriori probability 값

이 매우 클 때 오차일 가능성이 적음을 이용하는 confidence-based tuning 방식과 화자내에서 무자간 특징의 상관을 이용한 correlation-based tuning 방법이 있다.

나. 확률적 스펙트럼 사상

HMM을 이용한 음성 인식 시스템에서 모델의 파라미터를 새로운 화자에 적용시키는 방법으로 상태 s 에서 기존 화자의 음성 패턴 L_i 에 대한 새로운 화자의 음성 패턴 L'_i 이 나타날 확률 $P(L'_i|s)$ 는 다음과 같이 표현된다.

$$P(L'_i|s) = \sum_{L_i}^N P(L_i|s) P(L'_i|L_i, s) \quad (4)$$

여기서 N 은 상태의 수이다. 따라서 $P(L'_i|L_i, s), 1 \leq i, j \leq N$ 을 구하여 기존 HMM 파라미터를 새로운 화자에 적용시킨다.

IV. 최적 화자 적응 알고리즘

1. 확률적 스펙트럼 사상을 이용한 HMM 파라미터 적응

HMM의 관측 심볼 확률 행렬(observation symbol probability matrix) B 는 그 모델의 모든 상태에서의 임의의 심볼이 관측될 확률을 나타낸다. 이중 하나의 상태 s_k 에서의 관측 심볼 확률을 행 벡터 $\mathbf{P}(s_k)$ 로 표시한다.

$$\mathbf{P}(s_k) = [P(L_1|s_k), P(L_2|s_k), \dots, P(L_M|s_k)] \quad (5)$$

$1 \leq k \leq N$

여기서, N 은 HMM의 상태 수이고, M 은 관측 심볼의 수이다.

원 화자의 관측 심볼을 $L_i, 1 \leq i \leq M$ 이라 하고, 새로운 화자의 관측 심볼을 $L'_i, 1 \leq i \leq M$ 이라 하자. 그러면 원 화자의 관측 심볼이 L_i 일 때 새로운 화자의 관측 심볼 L'_i 이 발생할 확률을 $P(L'_i|L_i)$ 로 표시할 수 있다. 이상의 표현식을 이용하여 새로운 화자에

대해 HMM의 하나의 상태 s_k 에서의 관측 심볼 확률 $P(L'_i|s_k)$ 를 나타내면 아래와 같다.

$$P(L'_i|s_k) = \sum_{L_i}^M P(L_i|s_k) P(L'_i|L_i, s_k) \quad 1 \leq L'_i \leq M \quad (6)$$

여기서, 관측 심볼들과 상태 s_k 가 독립적이라고 가정하면 위식은

$$P(L'_i|s_k) = \sum_{L_i}^M P(L_i|s_k) P(L'_i|L_i) \quad 1 \leq L'_i \leq M \quad (7)$$

로 된다. 확률 $P(L'_i|L_i)$ 는 $M \times M$ 행렬을 형성하는 데, 이 행렬이 원 화자의 스펙트럼 공간(spectral space)으로부터 다른 화자의 스펙트럼 공간으로 변환시키는 확률적 스펙트럼 사상 행렬 \mathbf{T} 이다. 따라서, 새로운 화자의 상태 s_k 에서의 심볼 확률 $\mathbf{P}'(s_k)$ 는 행 벡터 $\mathbf{P}(s_k)$ 와 행렬 \mathbf{T} 의 곱(product)으로 얻어진다.

$$\mathbf{P}'(s_k) = \mathbf{P}(s_k) \mathbf{T}, \quad \mathbf{T}_{ij} = P(L'_j|L_i) \quad (8)$$

여기서,

$$\mathbf{T} = \begin{bmatrix} P(L'_1|L_1) & P(L'_2|L_1) & \dots & P(L'_M|L_1) \\ P(L'_1|L_2) & P(L'_2|L_2) & \dots & P(L'_M|L_2) \\ \vdots & \vdots & \dots & \vdots \\ P(L'_1|L_M) & P(L'_2|L_M) & \dots & P(L'_M|L_M) \end{bmatrix} \quad (9)$$

HMM에서 화자 적응을 한다는 것은 모든 상태에서의 관측 심볼 확률 $\mathbf{P}'(s_k), 1 \leq k \leq N$ 를 구하는 것이다. 위 식(8)을 보면 $\mathbf{P}(s_k)$ 는 원 화자의 모델 파라미터이므로 이미 알고있는 값이다. 따라서, $\mathbf{P}'(s_k)$ 를 구한다는 것은 확률적 스펙트럼 사상 행렬 \mathbf{T} 를 구하는 문제로 압축된다. 전향-후향 알고리즘, DTW 알고리즘, 그리고 Viterbi 알고리즘을 이용하여 행렬 \mathbf{T} 를 구할 수 있으며, 이들 각 방식에 대해서는 각각 다음 절에서 설명한다.

2. 전향-후향 알고리즘 접근법

전향 후향 알고리즘은 중간 파라미터인 시간-위치 벡터 $x(s_k)$ 와 후향 확률 $\beta_i(s_k)$ 를 이용하여 그 상태에서의 score를 구하거나, 모델 파라미터를 재추정하는 데 사용된다. 새로운 화자에 대한 HMM을 학습할 때 s_k 에서의 전향 및 후향 확률을 사용하여 $\alpha_i(s_k)$ 와 $\beta_i(s_k)$ 를 구한다. 그림 2(a)와 같다. 이것은 그 상태 s_k 에서의 전향 확률 벡터 $P(L')$ 로 표시된다. 이 모델을 그림 2(b)와 같이 확장시켜서 단일 상태를 M개의 병렬 path로 표현할 수 있다. 그림 2(b)의 한 path를 이용하여 설명된다.

$$\begin{aligned}
 P(L_i|s_k) P(L'_i|L_i, s_k) &= \frac{P(L_i|s_k)}{P(s_k)} \cdot \frac{P(L'_i|L_i, s_k)}{P(L_i, s_k)} \quad (10) \\
 &= \frac{P(L'_i|L_i, s_k)}{P(L_i|s_k)} \\
 &= P(L'_i|L_i, s_k)
 \end{aligned}$$

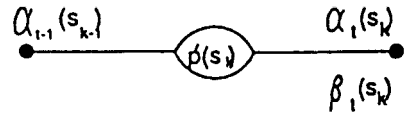
즉, 상태 s_k 에 있을 때, 적응 화자의 심볼 L' 과 원 화자의 심볼 L 가 동시에 나타날 확률이 된다. 그리고, 이들은 모든 path에 대해 더하면 그림 2(a)과 그림 2(b)는 동일한 표현이 됨을 알 수 있다. 따라서, 적응 화자에 대한 관측 심볼 확률들을 각각의 구성 요소로 분해할 수 있으며(expanded HMM), 전향-후향 알고리즘을 이용하여 반복적으로 확률적 스펙트럼 사상 행렬을 구할 수 있다.

다음과 같이 정의된 t번째 반복 횟수에서의 통계적 횟수(probabilistic count) $C_t(L', L)$ 를 구하면

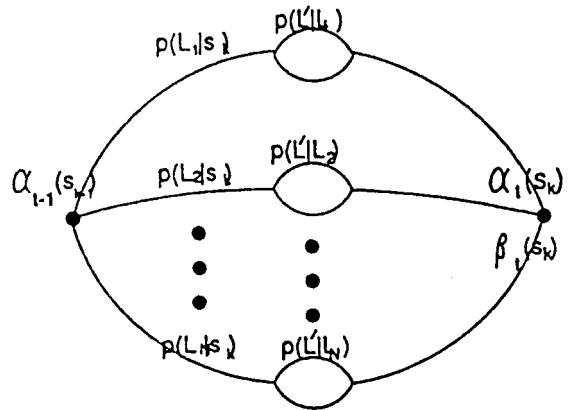
$$C_t(L', L) = P(L_i|s_k) \cdot P_{t-1}(L'_i|L_i) \cdot \alpha_i(s_k) \cdot \beta_i(s_k) \quad (11)$$

이 되고, 확률적 스펙트럼 사상 행렬은 통계적 횟수로부터 다음 식과 같이 t번째 반복횟수에서 구해진다. 이 과정을 반복하여 최종 사상 행렬을 얻는다.

$$P_t(L', L) = \frac{C_t(L', L)}{\sum_{L_i} C_t(L', L_i)} \quad (12)$$



(a) 단일 상태에서의 HMM



(b) 확장된 HMM

그림 2. 전향-후향 표현으로의 HMM.

3. DTM 접근법

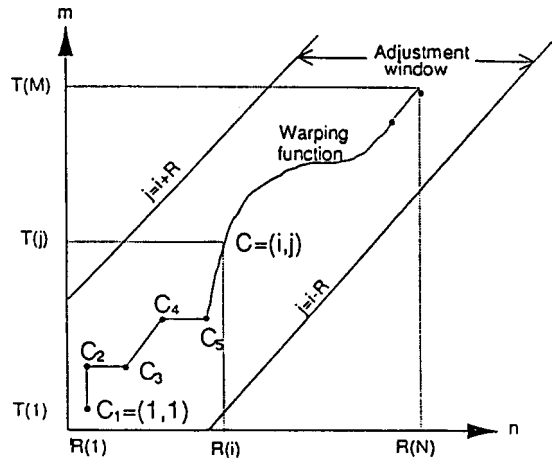


그림 3. Contour of DTW 대응도.

DTM은 두 패턴 사이의 유사도를 비선형 시간 warping을 통하여 측정하는 방식이다. 그림 3에 이를 도시하였다. R(N)은 기준 패턴이고 n은 기준 패턴의 시간 색인(index)이다. T(m)은 시험 패턴이고 m은 시험 패턴의 시간 색인(index)이다. DTW는 두 패턴 사이의 유사도를 비선형적으로 측정하며, 부산물로

두 패턴이 어떻게 유사한가를 warping 함수를 통해서 알 수 있다. DTW 접근 방법은 이 부산물을 이용하여 원 화자의 패턴과 적응 화자의 패턴 사이의 대응도를 찾아서, 이를 통계적 횡수로써 누적시키는 것이다. 이 대응도는 그림 3에서와 같이 $C(i, j)$ 로 나타나며, 이들을 누적하여 통계적 횡수 $C(L', L_i)$ 가 구해진다. 이 통계적 횡수를 정규화하여 확률적 스펙트럼 사상행렬을 구한다.

4. Viterbi 접근법

Viterbi 알고리즘은 전향-후향 알고리즘과 같은 목적으로 사용된다. 즉 decoding 문제를 풀 수 있는 방법으로, 전향-후향 알고리즘은 모든 path를 고려하여 Viterbi 알고리즘은 최적 path만을 고려하는 차이가 있다. 따라서, 최적 path가 다른 모든 path에 대해 우세하다면 두 알고리즘에 의한 decoding은 거의 비슷한 결과를 가져온다. HMM에서는 이러한 가정이 적절하기 때문에 HMM을 이용한 음성 인식에서의 decoding은 대부분 Viterbi 알고리즘을 사용한다. Viterbi 알고리즘은 다음과 같이 구현할 수 있다.

초기화(initialization)	$\delta_1(i) = \pi_i \cdot b_i(o_1), \quad 1 \leq i \leq N$
	$\Psi_1(i) = 0,$
반 복(recursion) for $2 \leq t \leq T, \quad 1 \leq i \leq N$	$\delta_t(i) = \max_{1 \leq j \leq N} [\delta_{t-1}(j) \cdot a_{ji}] \cdot b_i(o_t),$
	$\Psi_t(i) = \operatorname{argmax}_{1 \leq j \leq N} [\delta_{t-1}(j) \cdot a_{ji}]$
종 료(termination)	$p^* = \max_{1 \leq i \leq N} [\delta_T(i)],$
	$i^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)].$
Path(상태 순서)backtracking	$i^* = \Psi_{t+1}(i+r),$

여기서, i_t 는 시간 t 에 해당하는 상태의 색인이고, p^* 는 Viterbi 알고리즘으로 구한 확률값이다. 그리고 $\delta_t(j)$ 는 전향 파라미터와 같은 의미를 가지는 것으로 시간 t 에서 그 시간까지의 관측 심볼들을 관측하며 상태 j 까지 올 확률이며, $\Psi_t(j)$ 는 바로이전 상태, 즉

시간 $t-1$ 에서 있었던 상태를 의미한다.

마지막 단계인 상태 순서 backtraking을 통하여 주어진 객체에 해당하는 상태를 추적할 수 있다. 따라서, 이러한 정보를 통하여 시간 $t, 1 \leq t \leq T$ 에서 적응 화자에 대한 data $w, 1 \leq w \leq W$ 의 관측 심볼 $L'(w, t)$ 와 모델 상태 s_k 와의 관련성을 표시할 수 있다. 이 대응 관계를 Viterbi 일치(alignment)라 하며, 다음 식(13)으로 표현 된다.

$$s_k = V[L'(w, t)] \tag{13}$$

여기서, W 는 적응 데이터의 수이고, T 는 관측 심볼의 프레임 길이이며, V 는 Viterbi 일치 함수이다.

Viterbi 일치를 통해 원 화자의 관측 심볼 L 와 적응 화자의 관측 심볼 L' 사이의 통계적 횡수 $C(L', L)$ 를 식(14)에서와 같이 구한다.

$$C(L', L) = \sum_{L'(w, t) = L'} P(L' | V[L'(w, t)]) \tag{14}$$

여기서, 우변은 모델 파라미터 $P(L' | s_k)$ 이므로 알고 있는 값이다. 이 통계적 횡수를 전향-후향 접근법에서와 같은 정규화를 거쳐 확률적 스펙트럼 사상행렬을 구한다.

V. 결 론

음성 인식 시스템을 인식 방법에 따라 동적 프로그래밍 기법, HMM 기법, 신경망을 이용한 기법, 그리고 knowledge-based 음성 인식 기법으로 분류하였고 이들의 장·단점을 비교하였다. 그 결과 HMM 기법이 화자 독립, 대응량 및 연속 음성 인식에서 가장 우수한 성능을 가졌음을 알게되었고, 이산 HMM이 가지는 단점을 극복할 수 있는 방식들을 제시하였다.

화자 적응 방식에는 정적 지도 적응 방식과 정적 독자 적응 방식 그리고 동적 독자 적응 방식의 세가지로 분류할 수 있으며, 이들중 동적 독자 적응 알고리즘이 사람이 이용하기 가장 편리하다. 즉, 동적 독자 적응 알고리즘은 외형적으로 화자 독립 인식 시스템과 같은 역할을 하고, 내부적으로는 화자가 바뀌거나 동일한 화자일 경우 발음 상태 및 환경이 바뀔 때마다, 인식 시스템의 파라미터들을 연속적으로 조정

하이 퍼치 서랍의 인지와정과 비슷한 동작을 하게 한다. 그러나 상대적으로 연구가 더디고 성능 좋은 화자 독립 시스템을 필요로 하므로 단시일 내에 이 방식의 알고리즘을 개발하기는 어려운 것으로 보인다. 반면 성적 지도 적응 알고리즘은 성능 좋은 여러가지 방식이 많이 제안되었으며, 본 연구팀이 채택한 HMM 음성 인식 시스템의 적용에 적합한 알고리즘도 연구중이므로 이를 집중 연구하였다. 이에 적합한 정적 지도 적응 알고리즘으로 확률적 스펙트럼 사상법에 의한 적응 방식을 연구하였으며, 전향-후향, DTW, Viterbi 알고리즘에 의한 접근법을 생각할 수 있었다.

참 고 문 헌

1. J. Mariani, "Recent Advances in Speech Processing," Proc. of ICASSP, s9.1, 1989.
2. C. S. Myers, L. R. Rabiner, and A. E. Rosenberg, "Performance Trade-offs in Dynamic Time Warping Algorithms for Isolated Word Recognitions," IEEE Trans. on ASSP, Vol.28, pp.623-635, Dec. 1980.
3. J. G. Wilpon and L. R. Rabiner, "A Modified K-Means Clustering Algorithm for Use in Isolated Word Recognition," IEEE Trans. on ASSP, Vol.33, pp.587-594, June 1985.
4. H. Ney, "The Use of an One-Stage Dynamic Programming Algorithm for Connected Word Recognition," IEEE Trans. on ASSP, Vol.33, pp.67-72, Feb. 1975.
5. C. S. Myers and L. R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," IEEE Trans. on ASSP, Vol.29, pp.284-297, Apr. 1981.
6. L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," IEEE ASSP Magazine, Jan. 1986.
7. L. R. Bahl, F. Jelinek, and R. L. Mercer, "A Maximum Likelihood Approach to Continuous Speech Recognition," IEEE Trans. on PAMI, Vol. 5, pp. 179-190, 1983.
8. B. H. Juang and L. R. Rabiner, "Mixture Autoregressive HMM for Speech Signals," IEEE Trans. on ASSP, Vol. 33, pp.1404-1413, 1985.
9. K. F. Lee, *Automatic Speech Recognition*, Kluwer Academic Publishers, 1989.
10. D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing*, the MIT Press, 1986.
11. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representation by Error Propagation," in D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing*, The MIT Press, 1986.
12. R. P. Lippman, "An Introduction to Computing with Neural Nets," IEEE ASSP Magazine, Apr. 1987.
13. E. McDermott and S. Katagiri, "Shift-Invariant, Multi-Category Phoneme Recognition Using Kohonen's LVQ2," Proc. of ICASSP, s3.1, 1989.
14. T. Kohonen, "The Neural Phonetic Type-Writer," IEEE Computer, Vol.21, No.3, pp.11-22, Mar. 1988.
15. H. Bourlard and C. J. Wellekens, "Speech Dynamics and Recurrent Neural Networks," Proc. of ICASSP, s1.9, 1989.
16. L. T. Niles and H. F. Silverman, "Combining Hidden Markov Model and Neural Network Classifiers," Proc. of ICASSP, s8.2, 1990.
17. V. W. Zue, "The Use of Speech Knowledge in Automatic Speech Recognition," Proc. of IEEE, Vol.73, No.11, Nov. 1985.
18. V. W. Zue, et al., "Acoustic Segmentation and Phonetic Classification in the SUMMIT system," Proc. of ICASSP, s8.1, 1989.
19. L. E. Baum, et al., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Ann. Math. Stat.*, Vol.41, No.1, pp.164-171, 1970.
20. L. R. Bahl, et al., "A Maximization Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition," Proc. of ICASSP, pp.47-52, 1986.
21. L. R. Bahl, et al., "A New Algorithm for the Estimation of Hidden Markov Model Parameters," Proc. of ICASSP, pp.493-496, 1988.
22. H. R. Kim and H. S. Lee, "Segmental Corrective Training for Hidden Markov Model Parameter Estimation in Speech Recognition," *Electronics Letters*, Vol.27, No.18, pp.1633-1635, Aug. 1991.
23. H. P. Tseng, M. J. Sabir, and E. A. Lee, "Fuzzy Vector Quantization Applied to Hidden Markov Modeling," Proc. of ICASSP, pp.15.5.1-4, 1987.
24. J. M. Koo and C. K. Un, "Fuzzy Smoothing of HMM Parameters in Speech Recognition," *Electronics Letters*, Vol.26, No.11, pp.743-744, 1990.

25. Y. Grenier, et al. "Spectral Transformation Through Canonical Correlation Analysis for Speaker Adaptation in ASR," Proc. of ICASSP, pp.2659-2662, 1986.
26. G. Rigol, "Speaker Adaptation for Large Vocabulary Speech Recognition System Using Speaker Markov Models," Proc. of ICASSP, pp.5-8, 1989.
27. R. Schwartz, et al., "Rapid Speaker Adaptation Using a Probabilistic Spectral Mapping," Proc. of ICASSP, pp.633-636, 1987.
28. M. W. Feng, "Improved Speaker Adaptation Using Text Dependent Spectral Mappings," Proc. of ICASSP, pp.131-134, 1988.
29. M. W. Feng, et al., "Iterative Normalization for Speaker-Adaptive Training in Continuous Speech Recognition," Proc. of ICASSP, pp.612-615, 1989.
30. S. Furui, "Unsupervised Speaker Adaptation Method Based on Hierarchical Spectral Clustering," Proc. of ICASSP, pp.286-289, 1989.
31. R. M. Stern, "Dynamic Speaker Adaptation for Feature-Based Isolated Word Recognition," IEEE Trans. on ASSP, pp.751-763, 1987.
32. B. T. Lowerre, "Dynamic Speaker Adaptation in the Harpy Speech Recognition System," Proc. of ICASSP, pp.788-790, 1977.
33. R. Zelinek and F. Class, "A Learning Procedure for Speaker-Dependent Word Recognition Systems Based on Sequential Processing of Input Tokens," Proc. of ICASSP, pp.1053-1056, 1983.
34. K. Shikano and K. F. Lee, "Speaker Adaptation Through Vector Quantization," Proc. of ICASSP, pp.2643-2646, 1986.
35. K. Sugawara and M. Fumi, "Speaker Adaptation for a Hidden Markov Model," Proc. of ICASSP, pp.2667-2670, 1987.
36. A. Buzo, A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Speech Coding Based upon Vector Quantization," IEEE Trans. on ASSP, Vol.28, No. 5, pp.562-574, Oct. 1980.
37. B. H. Juang, D. Y. Wong, and A. H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," IEEE Trans. on ASSP, Vol. 30, No.5, pp.294-303, Apr., 1982.
38. G. D. Forney, Jr., "The Viterbi Algorithm," Proc. of IEEE, Vol.61, No.3, pp.268-278, Mar, 1973.
39. S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Function of a Markov Process to Automatic Speech Recognition," B.S.T.J., Vol. 62, No.4, pp.1035-1074, Apr., 1983.
40. Y. Grenier, "Speaker Adaptation Through Canonical Correlate Analysis," Proc. of ICASSP, pp. 888-891, 1980.

▲오 광 철(정회원)

1965년 4월 13일생

1988년 2월 : 한양대학교 공과대학 전자공학과 졸업

1990년 2월 : 한국과학기술원 전기및 전자공학과 졸업(M.S)

1990년 3월 ~ 현재 : 한국과학기술원 전기및 전자공학과 박사과정 재학중.

▲김 흥 국 : 제10권 5호 참조

▲김 회 린 : 제6권 3호 참조

▲이 활 수 : 제6권 3호 참조