

연속 음성에서의 신경회로망을 이용한 화자 적응

Speaker Adaptation Using Neural Network in Continuous Speech Recognition

김 선 일*
(Seonil Kim*)

* 연구는 한국과학재단 해외 Post-doc. 연수 지원에 의한 결과입니다.

요 약

RM 음성 Corpus를 이용한 화자 적응 연속 음성 인식을 수행하였다. RM Corpus의 훈련용 데이터를 이용해서 기준화자에 대한 HMM 학습을 실시하고 평가용 데이터를 이용하여 화자 적응 인식에 대한 평가를 실시하였다. 화자 적응을 위해서는 훈련용 데이터의 일부가 사용되었다. DTW를 이용하여 인식 대상화자의 데이터를 기준화자의 데이터와 시간적으로 일치시키고 오차 역전파 신경회로망을 사용하여 인식 대상화자의 스펙트럼이 기준화자의 스펙트럼 특성을 지나도록 변환시켰다. 최적의 화자 적응이 이루어지도록 하기 위해 신경회로망의 여러 요소들을 변화시키면서 실험을 실시하고 그 결과를 제시하였다. 학습을 거쳐 적절한 가중치를 지닌 신경회로망을 이용하여 기준화자에 적응시킨 결과 단어 인식이 최대 2.1배, 단어 정인율이 최대 4.7배 증가하였다.

ABSTRACT

Speaker adaptive continuous speech recognition for the RM speech corpus is described in this paper. Learning of hidden markov models for the reference speaker is performed for the training data of RM corpus. For the evaluation, evaluation data of RM corpus are used. Parts of another training data of RM corpus are used for the speaker adaptation. After dynamic time warping of another speaker's data for the reference data is accomplished, error back propagation neural network is used to transform the spectrum between speakers to be recognized and reference speaker. Experimental results to get the best adaptation by tuning the neural network are described. The recognition ratio after adaptation is substantially increased 2.1 times for the word recognition and 4.7 times for the word accuracy for the best.

I. 서 론

오늘날 음성 인식에 쓰이는 방법은 크게 두 부류로 나뉘어진다. 그 하나가 통계적인 방법을 이용한 것이고 나머지는 비선형망을 이용하는 것이다. 통계적인 방법으로 Hidden Markov Model(HMM)이 대표적이며[1][2], 비선형망으로는 인공 신경회로망(이하 신경회로망)이 이용되고 있다[3][4]. HMM을 이용한 음성 인식은 음성 신호의 시간 특성을 잘 처리해주어 특히 연속 음성의 인식에서 두각을 나타내고 있다. 그러나 통계적 특성을 잘 간직하기 위해서는 학습 단계에서 대량의 데이터가 요구된다. 따라서 가능한 한 대량의 데이터를 수집할 필요가 있으며 이것이 학습에 상당한 부담이 된다. 특히 단어 인식이 아니고 연속 음성의 인식일 때는 대단한 양이 필요하다.

따라서 여러 사람들이 화자 적응을 통해 이를 해결하려는 노력을 하였다. 성도 정규화를 통해 해결하는 방법도 제시되었고[5] Bayesian Learning을 이용한 해결 방법[6] 및 Codebook 변환을 통한 화자 적응[7] 등이 개발되었다. 선형적 방법을 이용해 Feature를 변환하는 방법[8] 및 Joint Feature 공간을 이용한 스펙트럼 영역에서의 변환[7], 선형 변환의 문제점을 보완한 Piecewise Linear 스펙트럼(Spectrum) 변환이 있다[9]. 그러나 어떤 방법이든 기본적으로는 적은 학습 데이터를 이용하여 이미 학습되어진 한 모델과 다른 사람 사이의 스펙트럼을 근사 시키는 것이다. 성도 정규화에서는 사실상 성도 모델 자체를 구하는 것 자체가 쉬운 일이 아니다. 그리고 Codebook 변환을 통한 스펙트럼 Mapping은 Discrete HMM이나 Semi-Continuous HMM을 사용할 때는 유용한 방법이나 [10][11] Continuous HMM(CHMM)에서는 사용할 수 없다. 그리고 선형 변환을 통한 스펙트럼 Mapping은[8] 너무

* 거제대학 전자과
접수일자: 1999년 2월 24일

무리한 가정을 필요로 하므로 좋은 결과를 기대하기가 어렵다. 따라서 비선형 변환이 그 기본 특성인 인공 신경망을 이용하면 스펙트럼 Mapping이 가능해진다. 스펙트럼 Mapping을 하려면 먼저 인공 신경망을 학습시켜야 하며 이 학습 단계에서 비선형 변환에 필요한 각종 계수(가중치)들이 얻어지게 된다.

본 논문에서는 Resource Management(RM) Corpus[12]의 화자종속 연속 음성을 이용하여 기준 화자가 발생한 600개 문장으로 화자 종속 CHMM 모델을 얻어 낸 후 이 화자 종속 모델로 인식 대상화자들의 음성을 인식하는 방법을 선택하였다.

같은 문장을 읽더라도 사람마다 음성 데이터의 길이가 다르므로 Dynamic Time Warping(DTW)을 이용하여 다른 사람이 읽은 문장의 음성 데이터를 기준 화자와 같은 시간 폭을 갖도록 조정된 후 기준 화자의 음성은 신경회로망의 학습 목표로, 인식 대상화자의 음성은 입력 데이터로 설정하여 오차역전파 신경회로망으로 다른 화자의 음성을 기준 화자에 학습시켰다. 신경회로망의 학습에는 RM Corpus의 훈련용 데이터를 사용하였고 여기서 얻은 가중치와 신경회로망을 이용하여 인식 대상화자의 평가용 데이터를 기준 화자의 스펙트럼 특성을 갖는 데이터로 변환시킨 후 이 데이터를 기준 화자의 CHMM 모델로 인식하였다.

II. 음성 인식 시스템

제안된 음성 인식 시스템은 그림 1과 같다. 인식 대상 화자의 일부 음성 데이터를 기준 화자의 음성 데이터와 시간적으로 일치시키기 위해서 DTW 과정을 거친다.

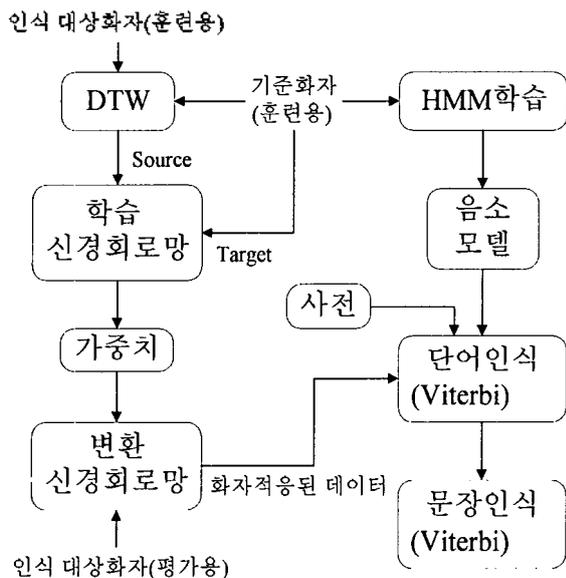


그림 1. 화자 적응 음성 인식 시스템
Fig. 1. Speaker adaptive speech recognition system.

이 과정을 거친 음성 데이터는 신경회로망을 이용한 기준 화자와 인식 대상 화자 사이의 비선형 변환에 사용되는데 여기에 사용되는 데이터는 기준 화자 600개 문장 중 일부이다. 즉 HMM 학습에 사용될 기준 화자 음성 600개 중 일부와 동일한 문장에 대해 인식 대상 화자의 음성 데이터를 가져와서 둘 사이의 비선형 변환 관계를 인공 신경망에 학습시키게 된다. 사용된 신경회로망은 1개의 은닉층을 갖는 오차역전파 신경회로망이다. 기준 화자의 음성 데이터는 신경회로망의 목표값이고 인식 대상화자의 데이터는 입력 신호이다. 음성인식에 사용된 계수는 Feature의 Delta 계수 및 Acceleration 계수까지 포함해 총 39차이지만 화자 적응을 위한 신경회로망의 학습시에는 Delta 계수 및 Acceleration 계수를 제외한 13개의 계수만 이용하였다. 39차 계수를 학습시키면 학습시간이 많이 들뿐더러 Delta 및 Acceleration 계수는 학습이 끝난 후 학습된 데이터로부터 계산하여 쓸 수 있기 때문이다. 학습이 완료되면 기준 화자와 인식 대상 화자 Feature 사이의 비선형 관계를 나타내는 가중치가 얻어지는데 이 가중치와 오차역전파 신경회로망을 사용하면 인식 대상 화자의 평가용 음성의 Feature를 기준화자의 특징 요소들을 갖는 Feature로 변환시킬 수 있다. 이렇게 기준 화자의 특징 요소들을 갖는 데이터를 화자 적응된 데이터라 한다. 따라서 이 데이터는 인식 대상 화자의 음성 주파수 및 기타 특징들을 오차역전파 신경회로망을 사용하여 기준 화자에 가깝게 변환한 것이다. 인식 대상 화자의 Feature를 변환 신경회로망에 입력으로 주면 기준 화자의 특징들과 닮은 Feature가 출력되는데 이를 이용하면 기준 화자의 Feature를 이용해 학습된 HMM 모델을 그대로 쓸 수 있게 되어 대량의 데이터가 필요한 HMM 학습을 개개인에 대해 따로 할 필요가 없어지므로 궁극적으로 소량의 데이터만 사용하여 HMM 학습이 가능해지게 된다. 한편 기준화자의 훈련용 데이터 600개 문장 전체를 사용하여 HMM 모델을 학습시키고 여기서 얻어진 음소 모델을 이용하여 단어를 인식하는데 단어 인식율을 높이기 위해 사전과 비교하였다. 사용된 사전은 CMU 사전을 사용하였고 강세의 표현 발음 기호의 차이등 RM Corpus에서 사용하는 발음 기호와 다른 부분은 일부 수정하여 사용하였다. Viterbi 알고리즘을 사용하여 인식된 단어는 곧바로 문장 인식 Viterbi에 넘겨져 문장 인식 과정을 거치게 된다. 각 문장은 의도적인 쉼 구간 없이 자연스럽게 연속적으로 발음되었고 앞뒤의 단어 종류에 따라 단어가 연결되는 부분의 데이터는 상당히 많은 편차를 보여 줄 수 있다. HMM 학습이 이루어지는 기본적인 모델은 음소이며 사전을 이용하여 단어의 발음 기호를 얻어내고 Viterbi 알고리즘을 사용하여 이 단어의 일치 여부를 계산하여 단어 인식을 수행한다. 문장의 경우는 단어의 연결이므로 인식된 단어를 주어진 기본 문장과 비교하면 문장 인식율을 계산할 수 있다. 문장의 인식에는 Bigram 언어 모델을 사용하였다. Bigram은 궁극적으로 단어의 인식율에 기여하게 되어 단어 인식의 부류에 속하게 되지만 여기서 단어인식과 문장인식을 구분한 것은 인식 대상이 분리되어

있는 단어가 아니라 연속 문장에서의 단어이기 때문이다. 단어 인식의 경우 문장내에 존재하는 단어가 다른 단어로 바뀌어서 인식이 되었다든지 문장내의 특정 단어가 없다는지 기준 문장에는 없는 단어가 들어가 있는 경우 등이 있어서 인식을 계산 때는 식 (1)과 함께 식 (2)도 사용하였다.

III. 결과 및 고찰

연속 음성의 인식과 평가를 위해 제작된 DARPA RM Corpus를 사용하였다. RM Corpus는 RM1과 RM2로 구분되는데 RM1은 4장의 CD-ROM으로 구성되어 있으며 그 중의 2장은 화자 종속(Speaker-Dependent, SD) 학습 데이터, 한 장은 화자 독립(Speaker-Independent, SI) 학습 데이터, 또 한 장은 시험 평가 데이터이다. RM2는 2장의 CD-ROM으로 구성되어 있으며 SD 추가 데이터가 들어 있는데 평가용 데이터도 포함되어 있다. SD 데이터는 12명의 화자로 구성되어 있으며 각 화자당 2개의 사투리 교정 문장과 10개의 고속 적응 문장을 포함해 612개의 문장이 들어 있으나 그 중에 600개의 화자 종속 문장만 HMM의 학습에 쓰였다. HMM 모델은 5상태 Left-to-right 모델이며 시작과 끝 상태를 제외하면 실제로 3상태이다. HMM 학습에 쓰인 초기 모델은 TIMIT Corpus의 학습 모델을 사용하였다. 음성 데이터는 16kHz로 샘플링 되었으며 25msec Hamming 창을 씌웠고 10msec Frame 주기를 적용시켰다. 사용된 Feature는 계수가 12차인 Mel Frequency Cepstral Coefficient(MFCC)이며 여기에 에너지를 추가하여 13차 계수가 되고 이로부터 이들의 Delta 계수 13개 및 Acceleration 계수 13개를 구하여서 총 39차로 이루어져 있다. 연속 HMM을 사용하였고 HTK Toolkit로 Single Gaussian Monophone Mixture를 구현하였다.

표 1. 화자에 따른 인식률
Table 1. Recognition rates by speakers.

화자	인식률(%)		통계
bef0	문장	인식률 : 41.33	H=31, S=44, N=75
	단어	인식률 : 84.13 정인식률 : 83.64	H=509, D=38 S=58, I=3 N=605
dtd0	문장	인식률 : 9.33	H=7, S=68, N=75
	단어	인식률 : 42.64 정인식률 : 28.26	H=258, D=55, S=292, I=87, N=605

신경회로망에서는 13차 계수만 사용하고 하나의 은닉층을 갖는 역전파 신경회로망을 사용하였다.

기준화자로는 bef0 화자의 데이터가 사용되었으며 bef0 화자의 훈련용 데이터 600개 문장으로 HMM을 학습시키고 평가용 데이터 75개 문장으로 인식시켰을 때 인식 대상

단어 수 총 605개 중에 84.13% 인 509개가 인식되었다. 한편 bef0 HMM 모델을 이용하여 인식 대상화자인 dtd0의 음성을 인식했을 때 총 605개의 단어 중 42.64%인 258개가 인식되었다. 정인식률을 따지면 bef0 음성을 인식했을 때 83.64% 이고 다른 화자인 dtd0의 정인식률은 28.26%로 현저히 저하되었다(표 1).

표 1에서 원래 문장과 인식된 문장을 일치시켰을 때 H는 알아맞힌 단어나 문장의 수이고 D는 원래 문장에는 있었으나 인식된 문장에서는 나타나지 않는 단어이며 S는 원래 문장에 있던 단어와 다른 단어가 인식된 경우이며 I는 원래 문장에 없던 단어가 들어간 경우이고 N은 전체 단어 및 문장 수이다. 이 때

$$\text{인식률} = \frac{H}{N} \times 100\% \quad (1)$$

$$\text{정인식률} = \frac{H-I}{N} \times 100\% \quad (2)$$

표 2. 학습율(η) 및 모멘텀(α) 변화에 따른 인식률
Table 2. Recognition rate changes by the learning rate(η) and the momentum(α).

η	α	문장 인식률(%)	단어(%)	
			인식률	정인식률
0.3	0.3	12.00	55.54	49.42
0.3	0.1	13.33	56.20	48.60
0.1	0.3	17.33	63.80	58.68
0.1	0.1	17.33	64.79	60.17
0.1	0.05	17.33	63.80	59.01
0.05	0.1	16.00	65.79	60.83
0.05	0.05	16.00	65.45	60.17

학습용 데이터 중에서 sr001.mfc 부터 sr040.mfc 까지 40개의 문장을 선정하여 bef0의 데이터를 학습 목표값으로 설정하고 dtd0의 데이터에 DTW를 적용시켜 기준 데이터와 시간축을 일치시킨 후 이 데이터를 입력으로 사용하여 신경회로망의 학습을 실시하였다. 입력층의 노드(Node)수와 출력층의 노드 수는 특징 벡터의 수와 같은 13으로 하였고 은닉층을 120으로 설정하여 실험을 실시하였다.

신경망의 학습 능력에 영향을 끼치는 요소는 학습율과 모멘텀, 그리고 은닉층의 노드수 이다. 따라서 최대의 인식율을 낼 수 있는 신경회로망을 구성하기 위해 기초 실험을 몇가지 수행하였다. 각 요소들의 값을 변경시켜 가면서 신경회로망을 학습시킨 후 이 신경회로망으로 다른 화자의 데이터를 변환하여 dtd0 화자의 HMM 모델로 인식하여 그 인식율을 비교하였다.

학습율과 모멘텀은 신경회로망의 학습에 중요한 영향을 끼치는 요소이다. 따라서 적절한 학습율과 모멘텀을 알아내기 위해 먼저 학습 횟수를 3000회로 제한하여 학습율과 모멘텀의 변화에 따른 인식율의 차이를 살펴보고 가장 적절한 학습율과 모멘텀을 선정하였다. 표 2에 제시된 결과에 따르면 학습율 0.05, 모멘텀 0.1일 때 가장 좋

은 결과를 보여주고 있으나 학습을 0.05, 모멘텀 0.05일 때도 근소한 차이로 좋은 결과를 보여주고 있다.

신경망의 학습시에 오차를 설정해 주고 학습 도중에 이 학습 오차가 주어진 값 안에 다다르면 학습을 멈추게 되지만 그렇지 못 할 경우에 적절한 횟수에 강제로 학습을 마쳐야 한다. 따라서 앞의 두 경우에 대해 학습 횟수를 변화시키면서 인식율의 차이를 조사하였다. 표 3은 학습을 0.05, 모멘텀 0.1일 때의 인식율이고 표4는 학습을 0.05, 모멘텀 0.05일 때의 인식율이다.

표 3. 학습 횟수 변화에 따른 인식률 (학습을 0.05, 모멘텀 0.1)
Table 3. Recognition rate changes by the iterations (learning rate 0.05, momentum 0.1).

학습횟수	문장 인식률(%)	단어(%)	
		인식률	정인식률
1,000	16.00	65.29	60.99
2,000	16.00	64.96	60.33
3,000	16.00	65.79	60.83
5,000	18.67	66.94	62.15
8,000	17.33	67.27	61.82

표 4. 학습 횟수 변화에 따른 인식률 (학습을 0.05, 모멘텀 0.05)
Table 4. Recognition rate changes by the iterations (learning rate 0.05, momentum 0.05).

학습횟수	문장 인식률(%)	단어(%)	
		인식률	정인식률
1,000	16.00	65.29	60.50
2,000	16.00	64.63	60.00
3,000	16.00	65.45	60.17
5,000	18.67	67.27	62.81
8,000	17.33	67.93	62.64

표 3과 표 4에 따르면 학습 횟수가 8000회 일 때 단어 인식율이 각각 67.27%와 67.93%로 제일 높지만 학습 횟수가 5000회 일 때는 단어 정인식율이 각각 62.15%와 62.81%로 가장 높다. 따라서 학습 횟수를 얼마로 정하느냐가 그리 쉬운 일은 아닌 것 같다. 그러나 단어 정인식율이 높으면서 학습 횟수가 적은 5000회 쪽이 학습 시간과 함께 고려해 볼 때 더 타당성을 가지는 것으로 생각되어 학습 횟수는 5000회로 제한하였고 학습 횟수가 5000회 일 때는 학습율과 모멘텀이 각각 0.05, 0.05일 때가 가장 높은 인식율을 보여 주었으므로 학습율과 모멘텀을 각각 0.05, 0.05로 선택하였다.

학습을 및 모멘텀 못지 않게 은닉층의 노드 수도 중요하기 때문에 학습을 0.05, 모멘텀 0.05, 학습 횟수 5000으로 설정하여 은닉층의 노드 수를 바꿔가면서 인식율을 조사하였다. 표 5는 은닉층의 노드 수 변화에 따른 인식율 변화를 보여주고 있으며 이 중 은닉층의 노드 수가 120개 일 때 67.27% 로서 최고의 단어 인식율을 나타내고 있다.

표 5. 은닉층의 노드 수 변화에 따른 인식률

Table 5. Recognition rate changes by the number of node of hidden layer.

은닉층의노드 수	문장 인식률(%)	단어(%)	
		인식률	정인식률
30	17.33	66.12	61.66
60	18.67	66.78	61.65
90	18.67	65.95	61.49
120	18.67	67.27	62.81
150	18.67	65.29	59.67
180	18.67	66.78	61.49

은닉층의 노드 120개, 학습을 0.05, 모멘텀 0.05로 설계된 오차 역전파 신경회로망을 사용하여 das1, dms0, dtb0, jws0, pgh0 화자에 대해서도 dtd0에 대한 인식 실험과 같은 방식으로 실험한 결과가 표 6에 나와 있다. das1, dms0, dtd0 화자의 경우는 적응 과정 없이 기준 화자의 모델로 인식하였을 때 인식율이 푹 떨어지는 경우이고 그 나머지 dtb0, jws0, pgh0 화자의 경우는 그런대로 괜찮은 인식율을 보이는 경우이다. 전자를 제 1군이라 하고 후자를 제 2군이라 했을 때 제 1군에 대해서는 최대 2.1배의 단어 인식율(das1,dms0)과 4.7배의 단어 정인식율(das1)을 보이면서 현저한 인식율 향상이 관찰되고 있다. 제 1군은 기준 화자와 음성 특성이 현저히 다른 경우이고 제 2군은 비슷한 경우로 추정되는데 음성 특성이 현저히 다른 경우에 스펙트럼 Mapping에 따른 화자 적응 효과가 잘 관찰되고 있다.

표 6. 화자 적응 음성 인식 비교

Table 6. Comparison for the adapted speech recognition rate.

화자	인식률(%)		
		화자적응 전	화자적응 후
bcf0	문장	41.33	
	단어	C : 84.13 A : 83.64	
das1	문장	2.67	21.33
	단어	C : 33.65 A : 13.46	C : 71.99 A : 63.85
dms0	문장	2.67	26.67
	단어	C : 33.28 A : 24.52	C : 70.86 A : 66.72
dtd0	문장	9.33	18.67
	단어	C : 42.64 A : 28.26	C : 67.27 A : 62.81
dtb0	문장	13.51	28.00
	단어	C : 61.93 A : 57.19	C : 80.91 A : 78.72
jws0	문장	26.67	26.67
	단어	C : 72.77 A : 69.23	C : 74.77 A : 72.31
pgh0	문장	22.67	24.00
	단어	C : 71.51 A : 68.14	C : 74.16 A : 71.37

IV. 결 론

역전파 신경회로망과 Monophone 연속 HMM을 이용한 화자 적응 음성 인식 시스템을 이용하여 RM Corpus에 대해 적용하였다. 최대의 인식률을 얻을 수 있도록 신경회로망의 여러 요소들을 변화시켜 가면서 신경회로망을 설계하였다.

그 결과 최대 2.1배의 단어 인식율과 4.7배의 단어 정인식율을 보여주어 상당히 고무적인 화자 적응 결과를 보여주었다. 이는 한 사람의 목소리로 학습된 HMM 모델을 다른 사람의 음성을 인식하는데도 쓰는 데 무리가 없음을 시사한다. 현존하는 음성 인식 방법 중에 가장 우수하다고 알려진 HMM 모델은 그러나 그 통계적 기반 때문에 충분한 음성 데이터가 주어지지 않으면 신뢰성을 보장받지 못한다. 불특정 화자를 인식하기 위해 엄청난 양의 음성 데이터를 수집하고 이를 모델화 하는 것이나 화자가 바뀔 때마다 그 사람의 음성을 대량으로 수집하여 새로운 HMM 모델을 만드는 것에 비하면 한 특정 화자의 HMM 모델을 만들어 놓고 화자가 바뀌게 되면 그 화자에게 불과 수십개의 문장만을 발음하게 하여 그 문장으로 화자간의 비선형 변환을 시도한 후 기존의 HMM 모델을 이용하여 인식하는 이 방법은 데이터 수집이라는 측면에서 보면 매우 효율성이 높다. 대개의 사람들은 HMM 모델을 형성할 정도의 많은 데이터를 제공하기 어렵고 또한 그 정도의 음성 데이터를 갖고 있기도 사실 어려운 것이다.

역전파 신경회로망을 사용한 음성 데이터의 변환을 통해 한 화자의 HMM 모델로 다른 화자의 음성을 인식하는 화자 적응 연속 음성 인식의 유용성을 이번 연구를 통하여 확인하였다. 앞으로 좀 더 많은 화자에 대해 이 시스템을 적용시켜 이 방법의 유용성을 검증하고 한국어에 대해서도 적용하여 그 유용성을 입증하여야 할 것이다.

참 고 문 헌

1. 김도영, 박용규, 권오욱, 은종관, 박성현, "연속분포 HMM을 이용한 한국어 연속음성 인식 시스템 개발," 한국음향학회지, 13권, 1호, pp. 24-31, 1994. 2.
2. 최인정, 권오욱, 박종렬, 박용규, 김도영, 정호영, 은종관, "대용량 한국어 연속음성 인식 시스템 개발," 한국음향학회지, 14권, 5호, pp. 44-50, 1995. 10.
3. 김선일, 이행세, "분산 신경망을 이용한 고립단어 음성에 나타난 음소 인식," 한국음향학회지, 제14권, 제6호, pp. 54-61, 1995. 12.
4. 최영배, 양진우, 이형준, 김순협, "한국어 음소 인식을 위한 신경회로망에 관한 연구," 한국음향학회지, 13권, 1호, pp. 5-13, 1994. 2.
5. L. Welling, R. Haeb-Umbach, X. Aubert, and N. Haberland, "A study on speaker normalization using vocal tract normalization and speaker adaptive training," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 797-800, 1998.

6. C. Hee, C. Lin, and B. Juang, "A study on speaker adaptation of continuous density HMM parameters," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 145-148, 1990.
7. F. Class, A. Kaltenmeier, P. Regel, K. Trotter, "Fast speaker adaptation for speech recognition systems," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 133-136, 1990.
8. F. Class, A. Kaltenmeier, P. Regel, "Fast Speaker Adaptation combined with soft vector quantization in an HMM speech recognition system," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 461-464, 1992.
9. H. Matsukoto, H. Inoue, "A Piecewise Linear Spectral Mapping for Supervised Speaker Adaptation," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 449-452, 1992.
10. 양태영, 신원호, 김원구, 윤대희, "음성 인식 시스템의 화자 적응 성능 향상을 위한 코드북 설계," 한국음향학회지, 15권, 2호, pp. 5-11, 1996. 4.
11. 황영수, "반연속 HMM의 화자 적응에 관한 연구," 한국음향학회지, 15권, 3호, pp. 97-103, 1996. 6.
12. P. Price, W. Fisher, J. Bernstein, and D. Pallett, "A database for continuous speech recognition in a 1000-word domain," in IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 651-654, 1988.

▲김 선 일(Seonil Kim)



1960년 3월 19일생

1983년 2월 : 아주대학교전자공학과 졸업(공학사)
 1985년 2월 : 아주대학교 대학원 전자공학과 졸업(공학석사)
 1996년 2월 : 아주대학교 대학원 전자공학과 졸업(공학박사)
 1985년 3월~1990년 8월 : 한국기계연구소 자동제어실 선임연구원
 1990년 8월~현재 : 거제대학 전자과 부교수
 1997년 2월~1998년 1월 : 한국과학재단 Post-doc. 프로그램에 의해 미국 Rutgers 대학에서 연구