자동차 제어용 음성 인식시스템 구현

이광석*, 김현덕*

An Implementation of Speech Recognition System for Car's Control

Kwang-Seok Lee, Hyun-Duck Kim

요 약

본 연구는 자동차내의 각종 제어장치들을 음성으로 실시간 제어하기 위한 음성제어 시스템을 제안하고 실험적으로 검증하였다. 실시간 제어음성 인식시스템은 8bit-10㎑로 A/D변환된 음성 데이터를 실시간으로 시작점과 끝점을 검출한 후, One Pass DP법으로 인식하였으며 그 결과를 모니터에 문장으로 출력하며 제어용 인터페이스에 제어데이터를 보내도록 구성하였다. HMM모델은 자동차내의 장치들을 제어하기 위한 제어음성 및 숫자음들로 구성되는 연속음성을 학습 및 모델링 하였다. 단어·제어문들의 인식률은 평균 97.3%, 숫자음의 경우는 평균 96.3% 정도의 인식률을 얻을 수 있었다.

ABSTRACT

In this paper, we propose speech control system for a various control device in the car with real time control speech. A real time speech control system is detected start-end points from speech data processing by A/D conversion, and recognize by one pass dynamic programming method. The results displays a monitor, and transports control data to control interfaces. The HMM model is modeled by a continuous control speech consists of control speech and digit speech for controlling of a various control device in the car. The recognition rates is an average 97.3% in case of word & control speech, and is an average 96.3% in case of digit speech.

*진주산업대학교 전자공학과

Ⅰ. 서 론

최근 지능시스템에 대한 요구는 각종 컴퓨터시스템, 정보통신시스템 및 각종 산업분야로의 확산이 활발해짐에 따라 인간-기계간의 인터페이스(Man-Machine Interface)에 대한 기대가 더욱더 높아지게 되었다. 음성(인간)-기계의 인터페이스 시스템은 속도가 빠르며 특별한 훈련 없이 이루어질 수 있다는 점에서 최근 실용화 기술연구의 확립은 각국의 중요한 연구과제가 되고 있다.

기존의 음성 인식 방법으로는 시계열 패턴을 비선형 압축으로 표준패턴과 유사도를 측정하는 DP(Dynamic programming)매칭이 주로 연구되어 왔으나, 최근 연속음성의 문장인식이나 불특정화자 음성인식에서 인식시간이 길어지는 단점 때문에 확률적으로 모델링하는 HMM(Hidden Markov Model)이 연구되고 있다.[1]-[3],[7]-[8]

본 연구에서는 자동차내의 각종 제어장치들을 음성으로 실시간 제어하기 위한 음성제어 시스템을 제안 구현하고 제한된 연속음성의 실시간 인식을 위한 가능성을 검토하였다. 먼저, HMM모델에 대해 제어용 음성인식 시뮬레이션을 행하고 그 결과를 검토하여 되도록 적은 계산량으로 실시간으로 연속음성을 인식할 HMM학습모델과 인식 알고리듬을 선택한다. 실시간 제어용 연속음성 인식을 위하여 발성한 음성을 시작점과 끝점을 실시간으로 검출한 후, 제안한 제어음성 인식 시스템으로 인식하고 결과를 확인하였다. 이로써 구현시스템에 대하여 인간-기계간의 인터페이스에 연속음성 인식 시스템을 적용할 경우의 적용 가능성을 검토하였다.

1. 연속출력분포 HMM

Left-to-right형 HMM은 그림 1과 같은 유한 오토마타(Finite State Automata)로 정의되며 일반적으로 HMM을 이용한 음성인식은 미리 인식에 필요한 만큼의 표준패턴을 학습해 두고 입력패턴에 대하여 그 출력확률이 최대가 되는 표준패턴을 인식결과로 한다.

연속 출력분포 HMM의 경우, 상태 i에서 j로의 천이확률 a_{ij} 및 천이경로에서 심벌 k의 출력확률 b_{ijk}를 학습 데이터에서 구하기 위해 Baum-Welch 알고리듬을 이용하였다.[4]

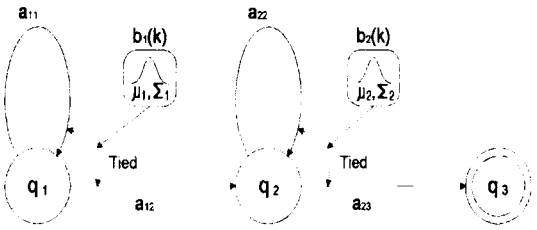

그림 1 연속 출력확률 HMM

Fig. 1 HMM model

상태수를 N, 심벌계열 길이를 T, 전향확률을 $\alpha(i,t)$,$(i=1,2,\cdots,N;\ t=1,2,\cdots,T)$, 후향확률을 $\beta(j,t)$,$(j=1,2,\cdots,N;\ t=T,T-1,\cdots,0)$모델 M의 심벌 계열 $o = o_1, o_2, \cdots, o_T$ 를 출력하는 확률을 $P(o \mid M)$ 및 상태 i에서 상태 j로의 천이가 시간 t에서 발생할 확률을 식(1)로 정의하면 천이확률의 추정식은 식(2), (3)과 같다.

$$\gamma_t(i,j) = \frac{\alpha(i,t-1)a_{ij}b_{ij}(o_t,\mu_{ij},\sum_{ij})\beta(j,t)}{P(o \mid M)} \quad \cdots\cdots (1)$$

$$a_{ij} = \sum_t \gamma_t(i,j) / \sum_t \sum_j \gamma_t(i,j) \quad \cdots\cdots (2)$$

$$b_{ijk} = \sum_{t:\,o_t=k} \gamma_t(i,j) / \sum_t \gamma_t(i,j) \quad \cdots\cdots (3)$$

출력벡터 o_t가 n차원의 정규분포에 따른다고 할 때 출력확률 밀도함수는 다음과 같이 주어진다.

$$b_{ij}(o_t,\ \mu_{ij},\ \Sigma_{ij}) = \frac{\exp\{-(o_t-\mu_{ij})^t \Sigma_{ij}^{-1}(o_t-\mu_{ij})/2\}}{(2\pi)^{n/2}\,|\Sigma_{ij}|^{1/2}} \quad \cdots\cdots (4)$$

여기서, μ_{ij}는 출력벡터의 평균치, Σ_{ij}는 공분산 행렬을 나타내며, 또한 μ_{ij} 및 Σ_{ij}의 추정식은 각각 식(5), 식(6)으로 주어진다.

$$\mu_{ij} = \sum_t \gamma_t(i,j) o_t / \sum_t \gamma_t(i,j) \quad \cdots\cdots (5)$$

$$\Sigma_{ij} = \frac{\sum_t \gamma_t(i,j)(o_t-\mu_{ij})(o_t-\mu_{ij})^t}{\sum_t \gamma_t(i,j)} \quad \cdots\cdots (6)$$

II. HMM에 의한 연속음성 인식

연속음성의 경우 단어경계가 명확하지 않으며 단어내부 및 단어경계 부근의 음이 전후 단어의 영향으로 변화하는 조음결합 현상이 발생하며 발성시간이 짧아지고 발음도 불명확해지는 등의 문제점으로 인해 인식 시스템은 매우 복잡한 구성을 필요로 하게 된다. 연속음성 인식기술은 기본적으로 패턴인식의 문제이며 인식하고자 하는 단어·음절 수 만큼의 표준 시계열 패턴을 준비하고 표준패턴과 입력패턴을 비선형으로 매칭 하는 DP매칭방식이 연구되어 왔다. 현재는 확률적인 모델로 표현하는 HMM이 음성인식 기술의 주류를 이루고 있으나 HMM방법도 음향적 우도계산에 Viterbi 알고리듬을 이용하고 있다는 점을 제외하면 기본적으로 DP매칭법과 같은 방식이다.

그림 2는 연속음성 인식 시스템의 구성도로써 음성 DB부, HMM모델 학습부, O(n) DP법 및 One Pass DP법에 의한 인식부로 대별할 수 있다. 음성 DB부는 제어음성을 샘플링, 특징파라미터 추출 및 라벨링을 통하여 음성 DB를 작성하고 학습 및 인식 처리부에 음성 데이터를 제공한다. HMM모델의 학습은 Baum-Welch 알고리듬에 의해 음절단위로 학습되며 단어사전과 연결기에 의해 단어 또는 음절단위로 학습 된 모델이 저장된다. 인식 알고리듬으로는 One Pass DP알고리듬과 O(n)DP알고리듬을 이용하였으며 전자는 일반적으로 1단 DP(One Stage DP)법으로 부르며 유한상태 오토마타에 의한 구문의 제약을 통하여 효율적인 탐색을 실시하는 연속 음성인식 알고리듬이며, 후자는 One Pass DP법의 오토마타를 1상태로 감소시켜 구문제약이 없이 최적 단어계열을 구하는 연속인식 알고리듬이다. 음성인식 시스템에서 이러한 언어 모델을 사용하는 목적은 언어적인 제약을 통해서 최적인 단어 열을 탐색하는 위함이다.[5]-[6]

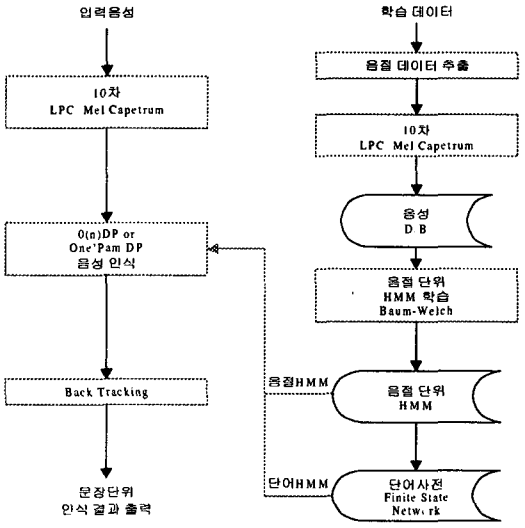

그림 2 연속음성 인식 시스템의 구성
Fig. 2 Configuration of recognition system for continuous speech

III. 실시간 제어용 연속음성 인식시스템

1. 시스템 구성

제어용 연속음성 인식 시스템의 알고리듬을 검토하고, 자동화 부분의 인간-기계 인터페이스를 음성으로 실시간 처리하기 위한 인식 시스템 구현을 목적으로 HMM모델들에 대해 음성인식 시뮬레이션 결과, 실시간 처리를 위해 비교적 계산량이 적고 인식률도 크게 떨어지지 않는 연속 출력분포 HMM모델을 선정하고, 인식 알고리듬으로는 유한상태 오토마타에 의한 구문제어 One Pass DP알고리듬을 선정하여 실시간 연속음성 인식 시스템을 구성하였다.

그림 3(a), 3(b)에서 연속음성 학습과정용과 연속음성 실시간 인식용 시스템의 구성도를 각각 나타내었다.

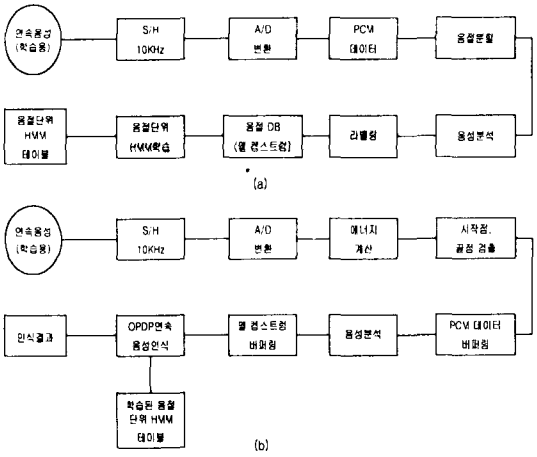

그림 3(a) 연속음성 학습시스템
(b) 연속음성 인식시스템

Fig. 3 (a) Traning system for continuous speech
(b) Recognition system for continuous speech

2. 실시간 시작점·끝점 검출

실시간 제어용 연속음성 인식 시스템에서 8bit, 10㎑로 샘플링 되어 입력 연속음성에서 음성시작 전의 무음구간과 연속음성이 끝나고 지속되는 무음구간을 제거하고 실시간으로 실제의 제어용 음성부분만의 버퍼링을 위하여 실시간 시작점·끝점 검출이 필요하다. 따라서 실시간으로 시작점과 끝점을 검출하기 위해서는 샘플링 후, 입력음성을 그림 4와 같은 환상형 버퍼에 순환적으로 DMA(Direct Memory Access)에 의해 저장하는 시스템을 구성하며 이 환상형 버퍼를 4개의 영역으로 나누어 음성의 버퍼링 동안 4개의 영역 중, 첫 영역부터 버퍼링이 완료되었는가를 폴링하고 버퍼링 완료이면 두 번째 영역의 버퍼링 동안, 첫 버퍼의 음성 데이터를 64프레임으로 나누고 프레임별로 식(7)과 같은 방법으로 에너지를 계산하고 에너지 값이 임계치 이하이면 무음으로 간주하며 무음이 아니면, 제어용 음성 저장용 버퍼에 보관한다. 그리고 계속해서 다음 프레임에 대해서도 같은 방법으로 진행하며 64프레임 분(1개 버퍼영역)이 끝나면 다음 버퍼영역에 대해서도 위와 같은 방법으로 진행해 나간다.

여기서 최초로 유음(에너지가 임계치 이상 : 시작점)이 발견되면 저장용 버퍼에 저장함과 함께 끝점을 검출하기 시작한다. 끝점 검출은 에너지 계산으로 하며 시작점이 검출된 후, 다시 에너지 값이 임계치 이하인 프레임의 개수를 체크해 가면서 임계치 이하인 무음 프레임24 프레임 이상 연속적으로 지속되면 끝점으로 간주하여 한문장의 연속음성 입력이 완료된 것으로 간주하고 입력을 종료하며 음성분석 단계로 넘어간다. 만일 24프레임 미만인 경우는 연속음성 중간의 묵음구간으로 간주하고 과정을 반복한다.

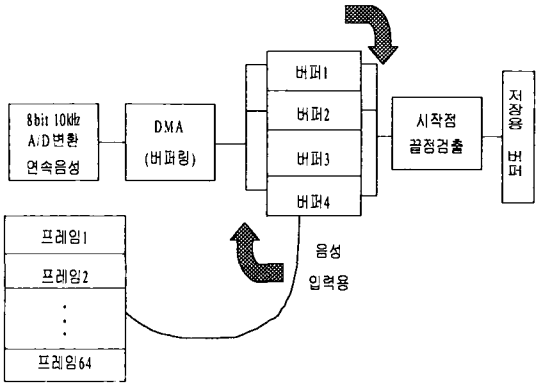

그림 4 실시간 시작점·끝점 검출
Fig. 4 Detection of start·end points

음성데이터는 DMA장치에 의해 입력용 버퍼에 입력(버퍼1→·····→버퍼4→버퍼1)된다.

음성이 입력되고 있는 동안에 버퍼1이 채워졌는가를 검사(DMA 어드레스 레지스터 검사)하며 각 입력용 버퍼영역은 다시 그림 4와 같이 64프레임(256 샘플)으로 나누며 계산된 에너지 $P(i)$가 임계치 512보다 작으면 무음 프레임으로 저장하지 않는다.

$$P(i) = \sum_{i=1}^{256} S(i)^2 \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (7)$$

IV. 인식 실험 및 고찰

1. 실험방법

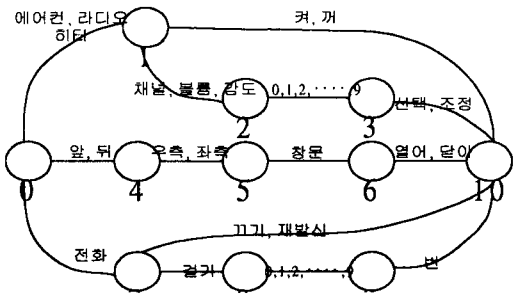

그림 5 제어음성용 유한상태 오토마타
Fig. 5 Finite automata for speech control

자동차내의 장치들을 음성으로 제어하기 위해 참조 패턴은 그림 5와 같이 유한상태 오토마타로 구성하고 One Pass DP법 인식시스템으로 구현하였다. 참조패턴을 구성을 위해 그림 5로 표현되는 유한상태 오토마타의 제어문장들을 표 1과 같이 구성하고 여기에 나오는 문장들을 각각 10회씩 발성한 후, 학습시스템으로 학습하여 HMM 테이블을 표 2와 같이 그룹으로 나누어 구성하였다.

표 1. 유한상태 오토마타에서의 제어음성문
Table 1. Control speech in finite automata

- 에어콘 켜, 에어콘 꺼
- 라디오 켜, 라디오 꺼
- 히터 켜, 히터 꺼
- 에어콘 강도 (0~9)에 조정
- 라디오 채널 (0~9)에 조정
- 라디오 볼륨 (0~9)에 조정
- 히터 강도 (0~9)에 조정
- 앞 우측 창문 열어, 앞 우측 창문 닫아
- 앞 좌측 창문 열어, 앞 좌측 창문 닫아
- 뒤 우측 창문 열어, 뒤 우측 창문 닫아
- 뒤 좌측 창문 열어, 뒤 좌측 창문 닫아
- 전화 걸기 (0,1,2, ·····, 9 : 전화번호)번
- 전화 끄기, 전화 재발신

표 2에서 그룹별 참조패턴의 음절수가 가장 많은 경우는, 묵음과 10가지의 숫자음으로 구성된 참조패턴 그룹이다.

표 2. 그룹화된 참조패턴
Table 2. Reference pattern's groups

그룹 / 오토마타	그룹 1	그룹 2	그룹 3

상태 0	에, 라, 히 전, 앞, 뒤	어, 디, 터,우 좌, 화,"묵음"	콘, 오, 측 "묵음"
상태 1	채, 블, 강, 켜, 꺼,"묵음"	널, 륨, 도 "묵음"	-
상태 2	0, 1, 2,…,9 "묵음"	-	-
상태 3	선, 조, "묵음"	택, 정, "묵음"	-
상태 4	우, 좌, "묵음"	측,"묵음"	-
상태 5	창, "묵음"	문,"묵음"	-
상태 6	열, 단, "묵음"	어, 아, "묵음"	-
상태 7	결, 끄, 재 "묵음"	기, 발, "묵음"	신, "묵음"
상태 8	0, 1, 2,…,9 "묵음"	-	-
상태 9	번, "묵음"	-	-

이상과 같은 실시간 제어용 연속음성 인식시스템 성능평가를 위해 표 2의 음절들에 대해 그룹별 음절인식을 실험하고 검토하였다.

2. 결과 및 고찰

음성분석 시, 프레임 간격을 5ms단위로 10회 발성한 숫자음을 학습한 후, 30회 인식실험한 결과를 표 3에 나타내었다. 인식결과, 평균 96.3%의 인식률을 보였으나 "사"의 경우는 "삼"으로 인식되는 경우가 많았으며 77%정도로 인식하였다. "삼"과 "사"의 오 인식이 다른 음에 비해 높은 것은 음성 스펙트럼 특징이 유사한 이유인 것으로 판단되며 이는 음절의 끝부분 가중치 처리로써 해결될 수 있으리라 생각된다.

표 3. 숫자음 실시간 인식결과(프레임 간격 : 5ms)
Table 3. A real time recognition rates of a digits

출력 / 입력	0	1	2	3	4	5	6	7	8	9	오인식	인식률(%)
0(영)	30											100
1(일)		30										100
2(이)			30									100
3(삼)				30								100
4(사)				7	23						7	77
5(오)						28				2	2	93
6(육)							30					100
7(칠)								28			2	93
8(팔)									30			100
9(구)										30		100
평균												96.3

표 4는 자동차 내의 장치 제어음성을 실시간 인식하기 위하여 표 2와 같이 그룹별로 구분하고 음절단위로 학습하여 그룹별로 30회씩 발성하여 인식 실험한 결과를 나타내었다. 그 결과, 그룹별로 가장 많은 음절패턴을 가진 경우는 묵음과 숫자음 10가지를 갖는 경우로써, 인식률이 다른 그룹보다 다소 낮았지만 전반적인 인식률은 양호하였다.

표 4. 제어음성패턴 그룹별 인식률/각 30회
Table 4. Recognition rates of an each control speech patterns

(단위: %)

상태 / 그룹	0	1	2	3	4	5	6	7	8	9
그룹1	95	95	100	95	100	100	95	95	85	100
그룹2	95	95		100	100	100	100	100		
그룹3	100							100		

이상과 같이 구문제어에 의한 제어용 연속음성 인식실험 결과로 미루어 제한된 소규모의 문장으로 음성제어 응용분야에 적용 가능함을 알 수 있었으며 오 인식 문제는 인식결과의 음성합성에 의한 확인 등의 후처리 또는 음절의 끝부분(유사부분)을 가중치 처리함으로써 해결될 수 있으리

라 생각된다. 연속음성 인식 시스템 적용이 가능하리라 생각된다.

V. 결 론

본 연구는 자동차 내부장치의 실시간 음성제어를 위하여 실시간 음성인식시스템을 제안하고 실험을 통하여 검증하였다. 음성인식은 확률모델을 이용하는 HMM방법을 택했으며 인식률, 인식 시 계산속도 및 메모리 사용량 등을 고려한 HMM모델과 인식 알고리듬 선택을 위해 여러가지 HMM 모델별로 인식 시뮬레이션 한 결과, 소규모의 제한된 문장에서 인식률이 높으며 가장 효율적인 것으로 판단한 One Pass DP알고리듬으로 실시간 연속음성 인식 시스템을 구현하고, 실제 응용 분야의 음성인터페이스에 적용할 수 있는가를 검증키 위해 인식 실험을 하였다. 실험결과, 본 실험에서 모델로 한 장치제어 음성에서 사용된 음절 패턴들은 음성제어용 숫자음과 단어·명령문에 포함될 수 있는 음절들로서 단어·명령문 음절들의 인식률은 평균 97.3%로써 양호하였으며, 숫자음의 경우는 /사/가 /삼/으로 인식되는 경우가 있었으나 전체적인 인식률은 96.3% 정도를 얻을 수 있었다. 본 실험의 인식률은 실제 응용 분야에서 적용되는 경우와 같은 상황인 실시간 인식시스템에 의해 실시간으로 인식 실험한 결과로써, 같은 내용의 음성의 경우도 학습 시 발성할 때의 주위 환경, 발성상태 등의 조건이 인식 시와 는 다른 조건일 수도 있는 경우를 감안하더라도 중·소규모로 제한된 음성이 사용되는 음성제어 분야 등에 적용하면 유용하리라 판단된다.

참고문헌

[1] Josè A. Martins and Fàbio Violaro, "Hybrid Recognizers Combining Hidden Markov Models and Multilayer Perceptron", IEEE, pp. 146-150 (1998)

[2] Katagiri S., Lee C. H., "A New Hybrid Algorithm for Speech Recognition Based on HMM Segmentation and Learning Vector Quantization ", IEEE Trans. on Speech and Audio Processing, vol. 1, No 4, pp. 421-430, Oct. 1993.

[3] Rabiner L. R., Wilpon J. G., Soong F. K., "High Performance Connected Digit Recognition using Hidden Markov Models", IEEE Trans. on Acoustics, Speech, Signal Processing, vol. 4 ASSP-37, pp. 1214-1225, Aug. 1989.

[4] Joe Tebelskis, "Speech Recognition using Neural Networks" CMU-CS-95-142, 1995.

[5] Rabiner L. R., Juang B. H., "Fundamentals of Speech Recognition", Englewood Cliffs, Prentice-Hall, 1993.

[6] S.H.Kim, S.Y.Koh, K.I.Hur: "A Study on the Recognition of the Isolated Digits Using Recurrent Neural Predictive HMM", TENCON '99 Vol. I, pp. 593-596, 1999.

[7] 이광석, 심장엽, 이영재, 고시영, 허강인, "HMM에 의한 연속음성인식 시스템의 구현", 제13회 음성통신및신호처리워크샵 논문집 제13권 1호, pp. 325-330, 1996. 8.

[8] 김수훈, 이광석, 고시영, 이영재, 허강인, "신경망 예측 HMM을 이용한 음성인식", 제9회 신호처리합동학술대회논문집, pp. 239-242, 1996. 10.

이 광 석(Kwang-Seok Lee)
1983년 2월 동아대학교 전자공학과 졸업(공학사)
1985년 2월 동아대학교 대학원전자공학과 졸업(공학석사)
1991년 2월 동아대학교 대학원 전자공학과 졸업(공학박사)

1995년~현재 진주산업대학교 전자공학과 조교수
관심분야 : 음성신호처리 및 인식, 지능시스템, 통신시스템

김 현 덕(Hyun-Duck, Kim)
1976년 2월 동아대학교 전자공학과 졸업(공학사)
1985년 2월 동아대학교 대학원 전자공학과 졸업(공학석사)
1996년 8월 경남대학교 대학원 전기공학과 졸업(공학박사)
1989년~현재 진주산업대학교 전자공학과 부교수
관심분야 : 신경회로망, 디지털신호처리