# 유전자 알고리즘을 이용한 화자인식 시스템 성능 향상

# Performance Improvement of Speaker Recognition System Using Genetic Algorithm

문 인 섭*, 김 종 교**
(In-Seob Moon*, Chong-Kyo Kim**)

*조선이공대학 정보통신과, **전북대학교 전기전자공학부

(접수일자: 2000년 10월 27일; 채택일자: 2000년 11월 10일)

본 논문에서는 화자인식의 성능향상을 위한 dynamic time warping (DTW) 기반의 문맥 제시형 화자인식에 대해 연구하였다. 화자인식에 있어 중요한 요소인 화자의 특성을 잘 반영할 수 있는 참조패턴을 생성하기 위해 유전자 알고리즘을 적용하였다. 또한, 문맥 종속형과 문맥 독립형 화자인식의 단점을 개선하기 위해 문맥 제시형 화자인식을 수행하였다. Close set에서 화자식별과 open set에서 화자확인 실험을 하였으며 실험결과 기존 방법의 참조패턴을 이용하였을 경우보다 유전자 알고리즘에 의한 참조패턴이 인식률과 인식속도 면에서 우수함을 보였다.

**핵심용어**: 화자인식, DTW, 유전자 알고리즘
**투고분야**: 음성처리 분야

This paper deals with text-prompt speaker recognition based on dynamic time warping (DTW). The Genetic Algorithm was applied to the creation of reference patterns for suitable reflection of the speaker characteristics, one of the most important determinants in the fields of speaker recognition. In order to overcome the weakness of text-dependent and text-independent speaker recognition, the text-prompt type was suggested. Performed speaker identification and verification in close and open set respectively, hence the Genetic algorithm-based reference patterns had been proven to have better performance in both recognition rate and speed than that of conventional reference patterns.

***Key words:*** *Speaker recognition, DTW, Genetic algorithm*
***Subject classification:*** *Speech signal processing*

## I. 서 론

현대 정보화 사회로의 발달로 인한 소비자의 욕구에 맞추어 다양한 서비스가 개발되고 있다. 이러한 서비스에는 은행의 자동 현금지급서비스, 전화 쇼핑서비스, 정보검색서비스 등이 널리 이용되고 있으며, 본인 또는 허가를 얻은 사람만이 접근해서 서비스를 이용하도록 되어 있다. 그러나 만약에 허가를 얻지 않은 사람이 접근해서 사용을 한다면 심각한 사회문제를 발생시킬 수 있다. 따라서 개인의 신분확인은 보안을 필요로 하는 곳에 필수적이다. 과거에는 사람의 신원을 확인하는데 신분증, ID 카드, 도장, 서명 등으로 신원을 확인하였으나 이러한 경우 분실이나 복사 또는 사본을 만들어서 사용하는 문제점이 발생하였다.

따라서 인간의 특성을 이용하여 신원을 확인하고자 하는 연구가 진행되고 있다. 이러한 연구 분야 중에서 인간의 음성을 이용하여 신원을 확인하는 연구가 바로 화자인식 분야이다. 이는 화자의 음성신호에서 발생되는 개인의 특성들을 추출하여 화자를 인식한다. 음성 신호로부터 말하는 사람이 누구인지를 판단하는 화자인식 (speaker recognition) 기술은 task의 성격에 따라 화자식별 (speaker identification)과 화자확인 (speaker verification)으로 나눌 수 있다. 여기서 화자 식별이란 등록된 화자들 중 발화자가 누구인가를 알아내는 것이고, 화자확인은 특정인이라고 자칭하는 인식 대상이 본인인지 여부를 알아내는 과정을 의미한다.

문맥 종속형 화자 인식은 제한된 문장이나 단어를 발성하여 화자를 인식하는 방법이며, 문맥 독립형 화자 인식은 임의의 대화 문장이나 대화를 화자가 발성하여 발성한 화자를 인식하는 방법이다. 이러한 화자인식 방법을 이용한 시스템에서는 비등록자가 등록자의 목소리를 흉내내거나 등록자의 음성을 녹음한 후에 비등록자가 이를

책임저자: 문인섭 (mis@mail.chosun-c.ac.kr)
501-759 광주광역시 동구 서석동 290 조선이공대학 정보통신과
(전화: 062-230-8450; 팩스: 062-230-8451)

이용하여 등록자인 것처럼 모방할 수 있는 단점이 있다.

본 논문에서는 화자 개개인의 특성을 잘 반영할 수 있는 우수한 참조패턴 생성을 위해 유전자 알고리즘을 이용하고,또한 문맥 종속형 및 문맥 독립형 화자인식의 단점인 사칭자의 모방이나 녹음을 통한 접근을 방지하기 위해 등록된 단어나 문장을 랜덤하게 제시하는 방식의 문맥 제시형 화자인식을 수행하였다[1].

## II. 유전자 알고리즘을 이용한 DTW 참조패턴 생성[3,4,5]

### 2.1. DTW 알고리즘의 개요

DTW 알고리즘은 시험 패턴과 기준 패턴을 시간축상에 최적이 되도록 배열한 후, 이 최적 변형 경로를 통한 최적 거리를 얻어내는 방법이다[2].

DTW 알고리즘은 2차원 $d_l(m,n)$ 평면상에서 적절한 제한 조건을 만족하는 $d(1,1)$에서 $d(M,N)$까지의 최적 경로를 구하는 방법이다. 임의의 점 $(m(j),n(j))$까지 축적된 거리를 다음과 같이 정의할 수 있다.

$$C_i(m,n)=\sum_{j=1}^{J} d_i(m(j),n(j))\cdot w(j) \qquad (1)$$

여기에서 $(m(j),n(j))$, $j=1,2,\cdots,J$ 는 주어진 경로이고, $w(j)$는 각 경로에 따른 가중치이다. 그러면 최적 경로는 $d(M,N)$에서의 축적된 거리 $C_i(M,N)$을 최소화하는 경로로

$$D_i(T,R_i)=\min_{path} C_i(M,N) \qquad (2)$$

로 표시된다. 최적 경로를 구하는 과정에서 제한 조건들은 다음과 같은 목적으로 주어진다. 즉, 부분적으로는 경로 기울기의 범위를 제한하며, 전체적으로는 경로의 허용 영역을 제한한다.

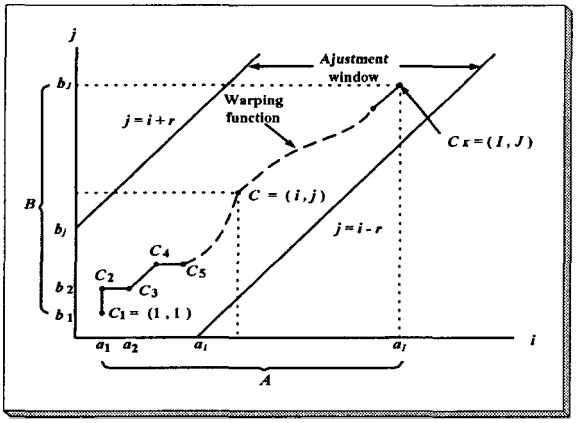
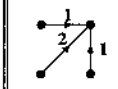
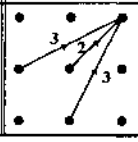



그림 1. 두 패턴 A와 B 사이의 DTW
Fig. 1. DTW between A and B pattern.

본 논문에서는 국부경로제한 조건으로 표 1과 같은 방식의 동적 프로그램 식을 사용하였다. 본 논문에서는 표 1의 Type I은 유전자 알고리즘을 이용한 참조패턴 생성 및 적합도 함수에 사용하고 Type II 화자인식 실험에 사용하였다.

표 1. 국부 경로 조건에 따른 동적 프로그램 식
Table 1. Dynamic programming equation according to local path constraint.

| | 국부경로제한 | 동적 프로그램 식 |
|---|---|---|
| Type I |  | $g(i,j)=\min\begin{cases} g(i-1,j)+d(i,j) \\ g(i-1,j-1)+2d(i,j) \\ g(i,j-1)+d(i,j) \end{cases}$ |
| Type II |  | $g(i,j)=\min\begin{cases} g(i-2,j-1)+3d(i,j) \\ g(i-1,j-1)+2d(i,j) \\ g(i-1,j-2)+3d(i,j) \end{cases}$ |

### 2.2. 유전자 알고리즘의 개요

유전자 알고리즘은 자연의 법칙인 "적자생존의 원리"에 근거를 두고 있다. 즉 자연 환경에 적응하여 진화해 가는 과정을 인공적으로 만들어 알고리즘화한 것이다. 유전자 알고리즘은 최적화하려는 변수를 염색체 형태로 표현한다.

이 염색체들의 집단 (population)중에서 해결해야 할 문제에 적합한 정도에 따라 적합도 (fitness value)가 계산되며, 각 염색체들은 적합도에 비례해서 다음 세대에 재생산될 확률을 가진다. 유전자 알고리즘에는 선택 (selection), 교배 (crossover), 돌연변이 (mutation) 등의 세 가지 기본적인 연산이 있다. 선택 연산은 적합도에 따라 다음 세대를 결정하기 위한 연산이다. 교배 연산은 염색체간에 서로 정보를 교환하기 위한 연산이고, 돌연변이 연산은 염색체 값을 임의의 값으로 바꾸어 전역적인 탐색이 가능하도록 하는 연산이다. 일반적인 유전자 알고리즘의 구조는 그림 2와 같다.

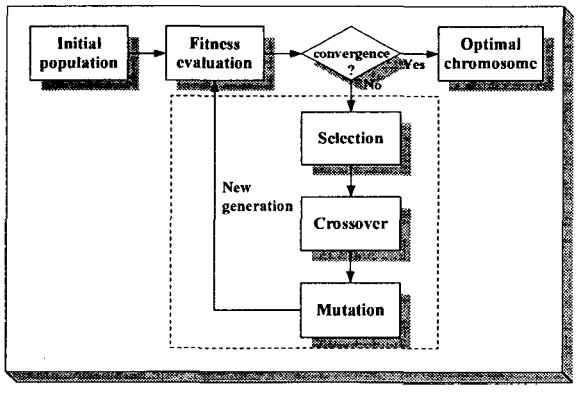



그림 2. 유전자 알고리듬의 구조
Fig. 2. Structure of genetic algorithm.

### 2.3. 참조패턴의 생성

DTW를 이용한 음성인식이나 화자인식에서는 참조패턴이 인식률에 절대적인 영향을 미치므로 가장 적합한 참조패턴의 생성이 중요한 요인으로 작용한다. 그러므로 인식률 향상을 위해 여러개의 참조패턴을 사용하는 방법이 있다. 그러나 이러한 방법은 계산량의 과다 및 사용 메모리의 증가 등이 단점으로 지적되고 있다. 따라서 본 논문에서는 참조패턴의 수를 줄이면서 높은 인식률을 얻기 위해 유전자 알고리즘을 이용하여 보다 우수한 참조패턴을 생성하여 화자인식에 적용하였다.

화자 적응적인 참조패턴을 생성하는 유전자 알고리즘은 다음과 같다. 화자를 특징을 대표할 수 있는 음성으로부터 특징 벡터열을 각각의 화자에 대해 추출하여 한 집합으로 만들고 이를 부모의 집합으로 놓는다. 이 집합의 모든 개체를 일정한 교배율로 서로 교배시켜 자식의 세대를 형성한다. 자식의 세대를 다시 적합도 함수로 평가하여 우월한 개체들을 선택하여 다시 서로 교배를 시킨다. 위의 과정을 반복하여 보다 화자의 특징을 잘 반영할 수 있는 패턴을 생성한다.

구체적인 과정은 다음과 같다.

#### 2.3.1. 적합도 함수

적합도 함수는 유전자 알고리즘에 있어서 중요한 역할을 수행한다. 적절한 적합도 함수의 설정은 개체에 대한 평가를 효과적으로 수행하지만 잘못된 함수는 원하는 결과와는 다른 방법으로 개체들을 평가하여 잘못된 방향으로 진화해 나갈 수 있다. 본 논문에서는 각 염색체의 적합도를 원래의 음성 데이터에서 추출된 특징 벡터열과 개체간의 최적 유클리디안 거리, 즉 DTW 결과의 누적값 $C_k$로 하였다.

$$C_k = \sum_{i=1}^{I}\sum_{j=1}^{J} d_{ik}(m(j), n(j)) \cdot W(j) \qquad k = 1, 2, \cdots, I \tag{3}$$

여기서, $I$는 훈련에 참가한 데이터의 개수, $J$는 최적의 경로, $W$는 경로에 따른 가중치를 나타낸다.

그러므로 적합도 함수의 값이 작을수록 적합도가 크다고 할 수 있다.

#### 2.3.2. 선택 방법

선택방법은 집단의 적절한 진화에 중대한 영향을 미친다. 다양한 방법들이 제시되어 있으나 문제들에 대한 정확한 지침은 없다. 여기서는 round-robin방식에 의한 엘리티즘 (elitism) 우선 순위 선택 (primary rank selection)방법을 이용하였다. 여기서 엘리티즘 우선 순위 선택이란, 한 세대에서 적합도가 큰 순서부터 임의의 개수를 선택하는 방법이다. 본 논문에서는 4개를 선택하였다.

#### 2.3.3. 교배 연산자

교배란 두 개의 염색체에서 서로 정보를 교환하여 새로운 개체를 만들기 위한 방법으로 본 논문에서는 산술교배 (arithmetic crossover) 연산을 이용하였다.

이 방법은 두 개의 염색체에서 각각 하나의 벡터를 선택하여 일점 교배를 하되 음성은 길이가 동적이므로 DTW 경로에 해당하는 두 벡터에 대하여 교배율 0.7을 사용하여 교배하였다.

$$\begin{aligned} \hat{p_n} &= w \cdot p_n + (1-w) \cdot q_n, \qquad 1 \le n \le N \\ \hat{q_n} &= w \cdot q_n + (1-w) \cdot p_n, \qquad 1 \le n \le N \end{aligned} \tag{4}$$

여기에서 $w$는 교배율을 나타내며 N은 DTW 경로에서 마지막 점을 나타낸다.

본 논문에서는 돌연변이 (mutation)를 적용하지 않았다.

#### 2.3.4. 군집의 크기

군집의 크기는 부모세대에서 4개를 이용하였으며 선택된 4개의 데이터를 각각 교배하여 자식세대로 넘어오면서 세대당 12개로 교배하였다. 세대수는 적합도의 결과값이 수렴할 때까지 반복하여야 하나 실험결과 대부분 5세대 이내에서 수렴함을 보여 5세대로 고정하였다.

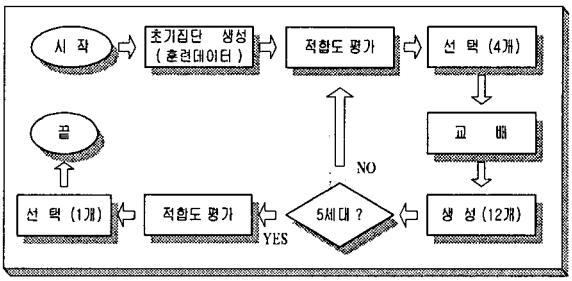



그림 3. 유전자 알고리즘을 이용한 참조패턴 생성 과정
Fig. 3. Process of reference pattern creation using genetic algorithm.

## III. 인식 시스템의 구성

화자인식 시스템에서는 비등록자가 등록자의 목소리를 흉내내거나 등록자의 음성을 녹음한 후에 비등록자가 이를 이용하여 등록자인 것처럼 모방하였을 경우 이를 제지할 마땅한 방법이 존재하지 않는다. 따라서 본 논문에서는 이러한 단점들을 어느 정도 해결할 수 있는 문맥제시형 화자인식을 수행하였다. 화자인식을 위한 구성도는 그림 4와 같다.

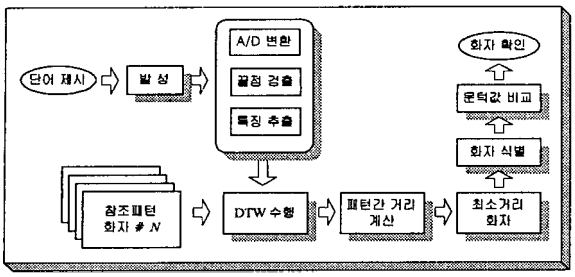



그림 4. 화자인식 시스템 블록도
Fig. 4. Block diagram speaker recognition system.

## IV. 실험 및 결과

본 논문에서는 화자인식 실험을 위해 등록자 6명이 10개의 단어를 20회씩 발성한 데이터와 비등록자 14명이 10개의 단어를 2회씩 발성한 데이터를 수집하였다. close set 화자인식 실험에서는 등록자 데이터 중 4회 데이터를 참조패턴 생성을 위해 훈련데이터로 사용했으며, 나머지 16회 데이터는 시험패턴으로 사용하였다.

Open set실험에서는 close set과 같은 방법으로 참조패턴을 생성하고 등록자 발성데이터 중에서 6회 데이터를 이용하여 문턱값을 결정하였고, 나머지 10회 데이터와 비등록자 14명이 10개 단어를 2회씩 발성한 데이터로 화자인식을 수행하였다. 녹음환경은 주변잡음이 존재하는 일반적인 실험실이며, 10개의 단어는 8kHz 16bit로 녹음한 후 15차 가중 켑스트럼을 추출하였다.

단어는 랜덤하게 시스템에서 제시되며 화자는 제시된 단어를 발성하게 된다. 정확한 화자 인식을 위해서는 화자 개개인의 특성을 잘 반영할 수 있는 어휘선택이 중요하다. 화자인식을 위한 표준어휘세트를 선정하기 위해서 음소별 화자식별을 실험하여 각 음소의 화자특성 반영정도를 비교하고, 실제 단어 내에서의 영향을 고려하기 위해 단어별 화자식별 실험을 통한 단어들 중에서 표 2와 같이 10개를 선정하였다[6].

표 2. 표준 단어 세트
Table 2. Standard vocabulary set.

| | | | |
|---|---|---|---|
| 1 | 공예품 | 6 | 무녕왕릉 |
| 2 | 나그네 | 7 | 사과나무 |
| 3 | 다람쥐 | 8 | 아주머니 |
| 4 | 요사이 | 9 | 정육면체 |
| 5 | 발자취 | 10 | 훈민정음 |

### 4.1. 실험 1(close set)

실험은 참조패턴으로 4개를 사용하였을 경우와 4개의 패턴을 각각 DTW를 수행한 후 최소의 누적값을 갖는 패턴을 참조패턴으로 하였을 경우, 유전자 알고리즘을 통하여 생성된 패턴을 참조패턴으로 하였을 경우에 대해 close set에서의 화자식별 실험을 하였다.

각 단어에 대하여 화자식별 실험을 행한 결과, 유전자 알고리즘을 이용하였을 경우 98.89%, 4개의 패턴을 이용하였을 경우 97.81%, 최소 누적값을 갖는 패턴을 이용했을 경우 95.11%의 인식결과를 얻어 유전자 알고리즘을 이용하였을 경우의 인식률이 각각 1.01%, 3.78% 향상되었다.

표 3. 화자 인식률(%)
Table 3. Speaker recognition rate.

| 방법 / 단어 | Genetic | 4개 패턴 | Min |
|---|---|---|---|
| 1 | 100.00 | 100.00 | 96.67 |
| 2 | 98.89 | 96.88 | 93.33 |
| 3 | 100.00 | 92.71 | 96.67 |
| 4 | 97.78 | 100.00 | 92.22 |
| 5 | 97.78 | 93.75 | 81.11 |
| 6 | 100.00 | 100.00 | 100.00 |
| 7 | 96.67 | 96.88 | 95.56 |
| 8 | 100.00 | 100.00 | 100.00 |
| 9 | 100.00 | 100.00 | 100.00 |
| 10 | 97.78 | 97.92 | 95.56 |
| 평균 | 98.89 | 97.81 | 95.11 |

* Genetic : 유전자 알고리즘을 이용한 참조패턴
4개 패턴 : 4개의 패턴을 이용한 참조패턴
Min : 최소의 누적값을 갖는 패턴

### 4.2. 실험 2(open set)

Close set 화자식별 실험에서 유전자 알고리즘을 이용했을 경우 가장 높은 인식률을 보였기 때문에 이를 참조패턴으로 결정하하고 open set에서 화자인식 실험을 하였다. close set에서는 등록된 화자들만이 시스템에 접근이 허용되어 누구의 발성인지만 찾아내기 때문에 그림 2.2에서 '문턱값 비교' 이전까지만 수행하면 된다. 그러나 open set에서는 등록되지 않은 화자도 시스템에 접근이 허용되고 또한 발성된 음성이 본인의 음성인지의 여부도 판별해야 하기 때문에 문턱값이 필요하다.

본 논문에서는 문턱값 결정을 위해 사용자 거부율 (FR : false rejection)과 사칭자 허용율 (FA : false acceptance)이 같아지는 동일 오류율 (equal error rate)을 고려하여 결정하였다. 실험결과 유전자 알고리즘을 이용한 참조패턴에 대해 FR과 FA가 각각 5.2%, 3.55%이었고 4개의 패턴을 참조패턴을 하였을 경우에 대해서는 6.33%, 3.63%의 결과를 얻었다.

평균 오류율 (verification mean error) 및 전체 오류율 (verification total error)은 식 (5)에 의해 구할 수 있다[7].

$$VM = \frac{FA+FR}{2}(\%)$$
$$VT = \frac{AT+RT}{TA+TC}\times 100\ (\%) \qquad (5)$$

여기서,

AT : falsely Accepted Trial
RT : falsely Rejected Trial
TA : Total number of Autotrial
TC : Total number of Crosstrial
이다.

표 4는 참조패턴의 종류에 따른 오류율을 나타낸다.

표 4. 참조패턴에 따른 오류율(%)
Table 4. Error rate according to reference pattern.

| 참조패턴 | FR | FA | VM | VT |
|---|---|---|---|---|
| Genetic | 5.20 | 3.55 | 4.38 | 4.13 |
| 4개의 패턴 | 6.33 | 3.63 | 4.98 | 4.17 |

표 4에서 보는 바와 같이 유전자 알고리즘에 의해 생성된 참조패턴을 이용하였을 때 인식률 향상되었으며, 또한 1개의 패턴만을 참조패턴으로 설정하기 때문에 4개를 이용하였을 때 보다 인식속도에 향상을 가져올 수 있다.

## V. 결 론

본 논문에서는 화자의 특성을 잘 반영할 수 있는 참조패턴을 생성하기 위해 유전자 알고리즘을 적용하였으며 사칭자의 접근을 방지하기 위해 문맥 제시형 화자인식을 수행하였다.

Close set과 open set에서 실험한 결과 유전자 알고리즘을 이용한 참조패턴이 기존의 방법에 비해 인식률의 향상이 있었으며 이는 교배와 선택의 반복과정에서 각 화자의 특성에 적응하는 참조패턴이 생성되었다고 판단된다. 향후 연구과제로 화자인식에 적합한 특징 파라미터에 대한 연구와 시간에 따라 화자 특성의 변이를 흡수할 수 있는 방법에 대한 연구가 필요하다.

## 참 고 문 헌


1. Chi Wei Chi, "An HMM Approach to Text-Prompted Speaker Verification," *Proc. of the IEEE*, vol. 2, pp. 673-676, 1996.
2. L. R. Rabiner & R. W. *Schafer, Digital Processing of Speech Signal*, Prentice-Hall, Englewood Cliffs, N. J., U.S.A., 1978.
3. 백창흠, 김태우, 이기정, 홍재근, "유전자 알고리듬을 이용한 CDHMM의 최적화," 한국음향학회 학술발표대회 논문집, 제 17권 1호, pp. 71-74, 1998.
4. David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addision Wesley publishing company, INC., 1989.
5. 서광석, "DTW를 이용한 향상된 문맥 제시형 화자인식," 전북대학교 석사학위논문, 1999.
6. 오영환, "음성합성시스템의 평가 및 화자확인시스템을 위한 표준어휘세트의 설계," 한국과학기술원, 1994.
7. Nikos Fakotakis, Anastasios Tsopanoglou, George Kokkinakis, "A Text-Independent Speaker Recognition Sysyem based on Vowel Spoting," Speech Communication, pp. 57-68, Dec., 1993.
8. D. A. Reynolds, R. C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.



▲ 문 인 섭 (In-Seob Moon)

1992년 2월 : 전북대학교 공과대학 전자공학과 (공학사)
1995년 2월 : 전북대학교 대학원 전자공학과 (공학석사)
1997년 2월 : 전북대학교 대학원 전자공학과 (박사과정 수료)
1997년 3월 ~ 1999년 3월 : 한려대학교 정보통신학과 전임강사
1999년 5월 ~ 현재 : 조선이공대학 정보통신과 전임강사
※ 주관심분야 : 음성인식, 음성압축, 음성합성

▲ 김 종 교 (Chong-Kyo Kim)

1966년 2월 : 전북대학교 공과대학 전기공학과 (공학사)
1977년 8월 : 전북대학교 대학원 전기공학과 (공학석사)
1983년 8월 : 전북대학교 대학원 전기공학과 (공학박사)
1983년 ~ 1984년 : 미국 일리노이 공과대학(IIT) 객원교수
1979년 ~ 현재 : 전북대학교 전기전자공학부 교수
※ 주관심분야 : 음성인식, 음성합성, 음성압축, VLSI설계