

벡터 평균값을 갖는 스트레인지 어트랙터 기반 화자인식

A Speaker Recognition Based on Strange Attractor with Vector Average

김 태 식*
Taesik Kim

ABSTRACT

In the area of speech processing, raw signals used to be presented in 2D format and different kinds of algorithms use the format to solve their problems. However, such kinds of presentation methods have limitations to extract characteristics from the signal, even though the algorithms are quiet good. The basic reason is that not much information can be detected from the 2D signal.

Strange attractor in the field of chaos theory provides the 3D presentation method. In the area of the recognition problem, signal construction method is very important because good features can be detected from a good shape of attractors. This paper discusses a new presentation method that can be used to construct strange attractor in a different way. Normal strange attractor uses time-delay idea while the new method uses time-delay and vector average. This method provides us good information to be applied to speaker recognition problem.

Keywords: Chaos, Strange Attractor, Speaker Recognition

1. 서 론

인간의 음성은 시간의 흐름에 따라 연속적으로 발생하는 시계열 데이터이다. 이러한 데이터 속에는 각자의 고유한 음성 정보를 내포하고 있는데 그 정보에 담겨있을 일정한 규칙을 발견한다면 화자 인식에 매우 유용하게 이용될 수 있을 것이다. 음성 데이터는 일반적으로 2 차원 파형 혹은 스펙트럼 등으로 표현 된 후 여러 가지 방법으로 분석하는 것이 컴퓨터분야에서의 전통적인 음성 처리방법이었다. 그러나 다양하고 우수한 알고리즘 개발[1, 11, 13, 14, 16]에도 불구하고 2 차원이라는 음성 데이터 표현 방법의 한계로 인해 음성 특징 추출에는 여러 제약이 따랐었다. 그러나 카오스가 등장하면서 음성과 같은 시계열 데이터를 3 차원 공간에서 운동하는 새로운 기하학적 형태로 제시해 주는 어트랙터(Attractor)라는 방법이 알려졌는데 이를 이용하면 2 차원에서 보다 훨씬 다양한 정보를 얻을 수 있게 된다[2, 10]. 화자 인식의 관점에서는 다른 사람의 어트랙터는 서로 다르게 표현될수록 인식에 유리하게 적용될 수 있

* 계명대학교 공학부 컴퓨터공학전공

기 때문에 어트랙터 표현방법은 매우 중요하다 할 수 있다. 이러한 어트랙터는 이미 여러 시계열 데이터를 분석하는 방법에 비록 단순한 시간지연방법이기는 하지만 긍정적으로 이용되기 시작하였다[3, 4, 12]. 그러나 단순 시간 지연 어트랙터를 이용한 방법은 시간의 변화에 따른 음성 데이터(구체적으로는 음성의 높낮이)가 한 단위 시간 전 혹은 후에 발생하는 데이터에는 어떤 영향을 주고 있는지는 혹은 어떤 카오스 형태를 띠고 있는지 알 수 없는 문제점이 있다. 본 논문에서는 이러한 문제점 해결에 어느 정도 접근하고자 단위 시간에서의 바로 앞뒤 데이터 값을 이용하여 그 변화량을 측정함과 동시에 이를 어트랙터로 표현함으로써 어떤 특징적인 어트랙터가 그려지는지 발견하고자 하였다. 이는 기존의 단순 어트랙터 구성 방법과는 다른 방법으로 음성에서의 카오스 현상을 찾고 인식시스템에 활용하고자 하는 기초 연구가 될 수 있을 것이며 화자 인식뿐만 아니라 소리를 발생시키는 주체 즉, 사람, 동물, 악기 등에 따라 각기 발생시키고 있는 소리의 어트랙터를 통한 카오스 현상 연구에도 도움이 될 수 있을 것이다. 본 논문에서는 음성에 담긴 특징을 주어진 시간 단위에서 앞뒤의 음성 높낮이에 따른 백터값을 찾아 어트랙터에 적용하여 각 개인별로 다른 형태의 어트랙터를 그릴 수 있는 방법을 제시하고자 한다. 이 방법으로 그려진 어트랙터가 다른 사람의 그것과 다름을 보여주기 위해 간단한 신경망을 통해 어느 정도 인식률을 나타내는지 실험할 것이다.

2. 음성의 어트랙터 구성

2.1 어트랙터 이론

음성데이터와 같은 동역학계를 이해하는 중요한 방법은 상태 공간(state space)에서의 운동 양상을 파악하는 것으로서 시간에 따른 계의 변화를 상태 공간의 점의 자취 즉 궤적으로 나타내어 기하학적으로 이해하는 방법이다. 이것을 어트랙터라고 하며 그 종류로는 평형점(equilibrium, fixed point), 폐곡선(limit cycle), 몸통모양(torus, tori)등의 거의 정확히 예측 가능한 형태가 있는가 하면 Chaotic Attractor라는 예측 불가능한 양상을 나타내는 것도 있다 [5, 8, 19]. 카오스 시스템에서 발생하는 어트랙터를 스트레인지 어트랙터(Strange Attractor)라고 하며, 카오스 어트랙터는 카오스에 내재한 질서의 형상으로 카오스가 복잡한 운동의 정적인 측면이라면 스트레인지 어트랙터는 그 복잡성의 동적이며 기하학적인 측면이라고 할 수 있다[6, 7, 17].

카오스 현상을 발견하고 분석하는 중요한 방법의 하나는 어트랙터의 재구성으로 시계열 데이터로부터 계의 운동 양상이 보이는 기하학적인 특징을 3 차원 공간에서 표현, 추출해 내는 과정이다. 이 방법은 물리학자인 N. Packard, J. Crutchfield, D. Farmer 및 R. Shaw가 Floris Takens의 수학적 연구와 함께 개발하였다[15, 18].

2.2 어트랙터 구성방법

어트랙터를 재구성하는 일반적인 방법은 Takens의 지연시간 방법을 이용한다. 지연시간 방법이란 $\psi(t_k)$ 을 1차원 시계열 데이터라 하고, N_{dat} 는 데이터의 전체 개수라 할 때, $k \in K$, $K = \{k \in N_0; k < N_{dat}\}$ 이며, 시계열 데이터 $\psi(t_k)$ 가 관측 가능하다고 한다면, 재구성된 상

태 공간의 벡터는 다음과 같이 주어진다.

$$\psi(t_s) = \begin{pmatrix} \psi(t_s) \\ \psi(t_s + \tau) \\ \vdots \\ \psi(t_s + \tau(D_E - 1)) \end{pmatrix} \quad (1)$$

여기서 $s \in S$, $t_s = sT_a$, $S \in \{N_0; s < N_{dat} - \tau/T_a(D_E - 1)\}$ 이며, D_E 는 임베딩 차원, T_a 는 시계열 데이터의 샘플링 시간, τ 는 지연시간을 나타내고 있다. 음성 데이터를 이용하여 스트레인지 어트랙터를 구성하는 방법은 시계열 음성 데이터에서 적당한 지연시간 T 를 이용하여 3 개의 새로운 벡터를 추출하여 각각의 벡터 값을 각 차원으로 3 차원의 상태 공간에 어트랙터를 구성하면 된다. 시계열 음성 데이터 $S = \{s_0, s_1, s_2, \dots, s_{n-1}\}$ 에서 다음의 벡터 $\Pi(i)$ 를 만들 수 있다.

$$\begin{aligned} \Pi(i) = & (s(0), s(T), s(2T)), \\ & (s(\tau), s(\tau+T), s(\tau+2T)), \\ & (s(2\tau), s(2\tau+T), s(2\tau+2T)), \end{aligned} \quad (2)$$

$$(s(n\tau), s(n\tau+T), s(n\tau+2T)),$$

여기서 τ 는 Sampling 간격, n 은 데이터의 개수를 나타낸다. 위의 벡터를 다음과 같이 일반화할 수 있다.

$$\Pi(i) = (x_i, y_i, z_i), \quad i = 0, 1, 2, \dots, n-1 \quad (3)$$

각각의 x_i, y_i, z_i 를 각각 1차원으로 하여 3차원 공간의 한 점으로 나타내어 시간의 흐름에 따라 이 점들을 선으로 연결하면 스트레인지 어트랙터가 구성된다. 이러한 방법으로 시계열 음성 데이터를 상태 공간상에 나타냄으로써 시간 경과에 따른 데이터의 특징이 보여지게 되며 이 정보로서 어트랙터의 움직이는 법칙성을 발견할 수 있고 이것은 결정론적인 모델이 될 수 있다.

본 논문에서 제시하는 방법은 위와 같은 일반적인 방법을 토대로 벡터의 평균값을 활용, 어트랙터를 재구성하는 방법이다. 이 방법은 어트랙터를 통하여 각 화자의 음성 데이터의 단순한 특징을 추출하는 방법과는 달리, 구성된 어트랙터 내에 음성 데이터의 카오스현상이 포함되어 있다고 볼 수 있을 것이다. 여기서 카오스라 함은 시계열 데이터가 매 시간 단위마다 앞 뒤 값이 서로 어떤 영향 갖는가를 이야기하는 것이다. 따라서 다음 장에서 제시하는 어트랙터에는 각 화자마다 음성의 카오스적 특징이 일반적인 방법보다 더 많이 내재되어 있을 것이다.

3. 벡터의 평균값을 이용한 어트랙터

3.1 어트랙터 구성

시계열 음성 데이터 $S = \{s_0, s_1, s_2, \dots, s_{n-1}\}$ 를 정규화과정을 거쳐 화자마다 입력시 발생하는 음성의 크기 차이를 제거한다. 14 KHz, 8 bit로 입력한 음성 데이터는 시작점 기준 9,000 byte부터 실험 데이터로 이용하였다. 각 화자별 음성 데이터를 시간단위 T 를 3, Sampling 간격 r 를 1로 두고 데이터 길이 n 은 2000이라 할 때 새로운 벡터 X_i, Y_i, Z_i 를 다음과 같이 구성한다. 각 벡터는 일정한 시간단위 T 간격으로 시간의 흐름에 따르는 상관 관계를 지니게 된다.

$$\begin{aligned} X_i &= \{s_r, s_{r+1}, s_{r+2}, \dots, s_{r+n}\} \\ Y_i &= \{s_{T+2r}, s_{T+2r+1}, s_{T+2r+2}, \dots, s_{T+2r+n}\} \\ Z_i &= \{s_{2T+3r}, s_{2T+3r+1}, s_{2T+3r+2}, \dots, s_{2T+3r+n}\} \end{aligned} \quad (4)$$

시간단위 T 의 간격을 두고 각 벡터 X_i, Y_i, Z_i 의 평균값을 구하여 상태 공간상의 좌표를 설정한다. 벡터의 평균값을 구함으로써 다소 흩어져 있는 데이터들의 값을 모아주어 데이터들의 상관 관계를 밀접하게 해주며 이것은 각 화자별 어트랙터의 특징을 잘 드러나게 하여 준다. 구하는 식은 다음과 같다.

$$\begin{aligned} A_i &= \frac{X_i + Y_i + Z_i}{3} \\ B_i &= \frac{X_{i+3} + Y_{i+3} + Z_{i+3}}{3} \end{aligned} \quad (5)$$

식 (5)를 통해 다음과 같이 어트랙터를 구성할 수 있다.

$$X = (A_i, B_i), \quad i = 0, 1, 2, \dots, n-1 \quad (6)$$

다음 그림 1은 위의 식 (5)와 식 (6)에 의하여 임의의 화자의 어트랙터를 구성한 것이다.

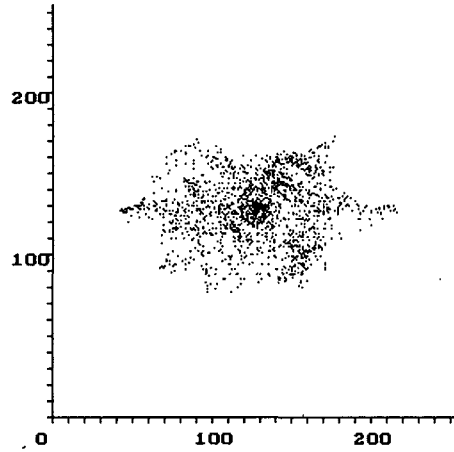


그림 1. 화자의 벡터 평균값 어트랙터

위의 어트랙터를 보면 어느 일정한 영역을 두고 그 영역 이상은 벗어나지 않으면서 특이한 형태를 띄고 있음을 알 수 있다. 이 어트랙터를 시간의 흐름에 의해 나타나는 점들을 선으로 연결한 어트랙터는 다음과 같다.

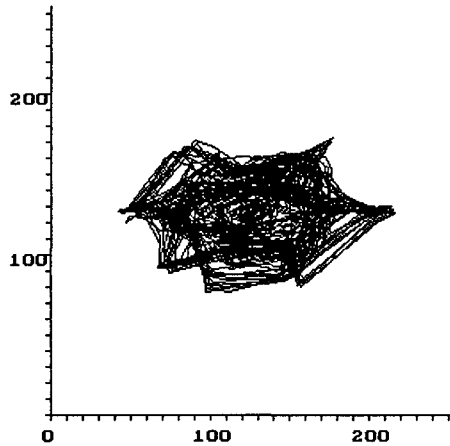
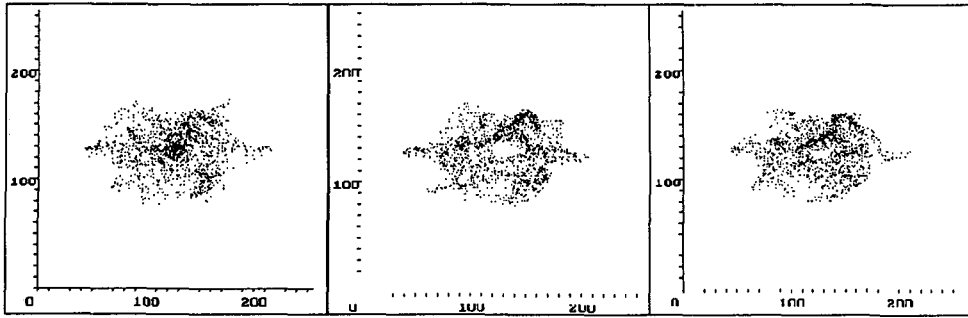


그림 2. 화자의 벡터 평균값 어트랙터(선 연결)

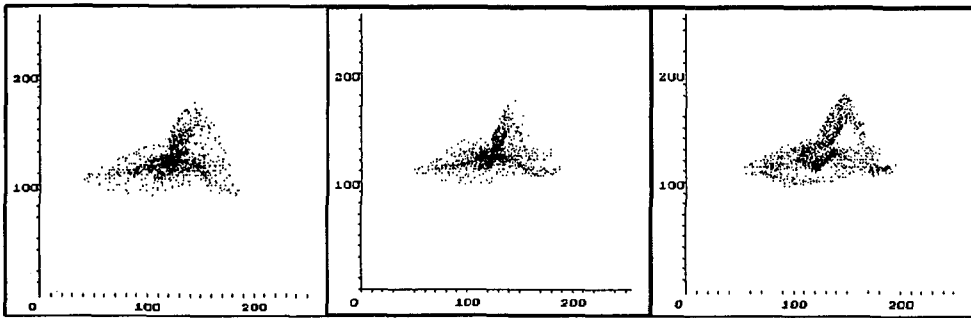
3.2 화자별 어트랙터 구성

다음 그림 3은 4 명의 화자가 발성한 '아' 소리에서 각각 3 개씩을 뽑아 벡터의 평균값을 이용하여 어트랙터로 재구성한 것이다.

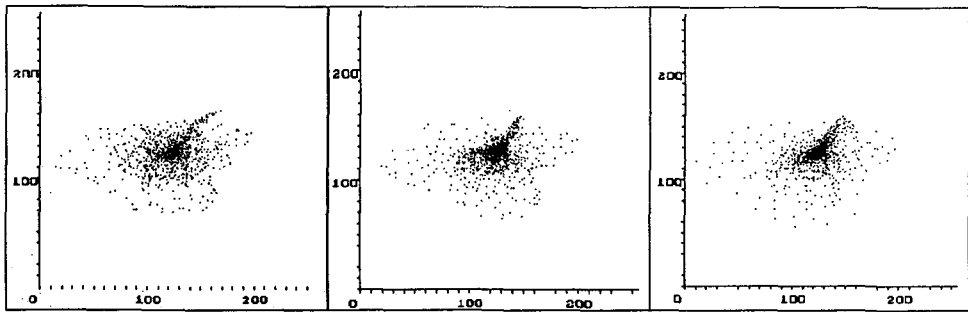
<화자 1>



<화자 2>



<화자 3>



<화자 4>

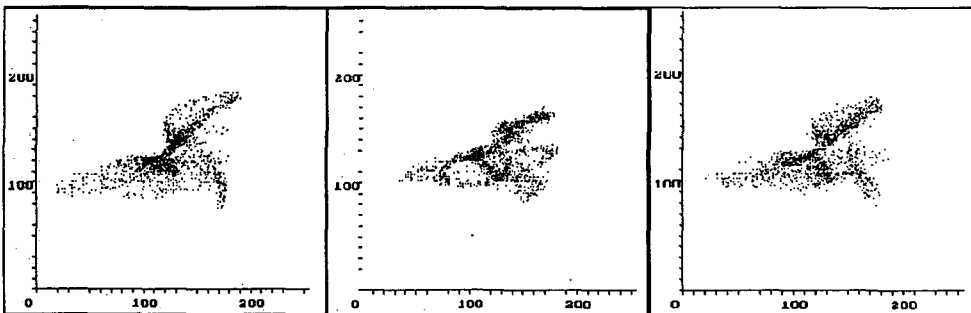
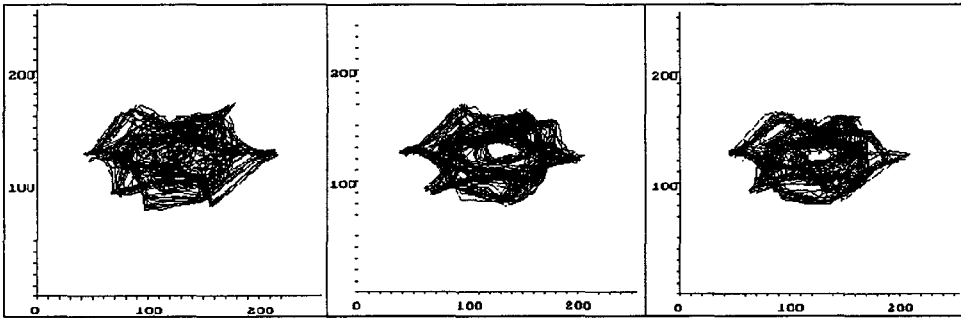


그림 3. 화자별 벡터 평균값 어트랙터

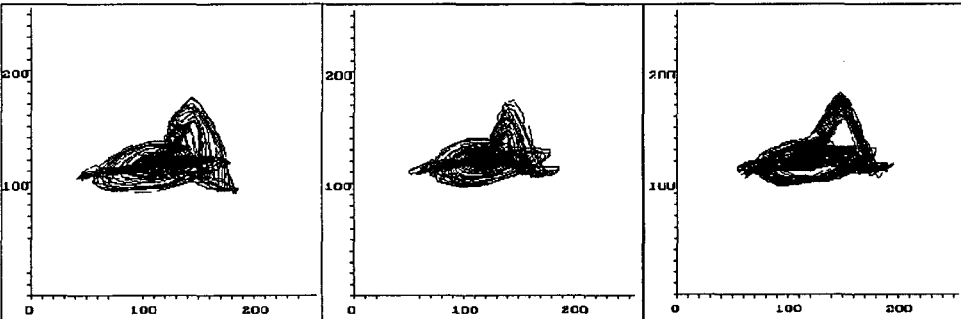
이상의 어트랙터를 분석하면 화자 1과 3, 화자 2와 4는 각각 비슷한 양상을 나타내고 있는데 어트랙터의 점들이 왼쪽에서 오른쪽으로 올라가는 대각선상에 모여 있음은 x, y축 각각의 데이터 값의 차이가 적음으로 인하여 데이터 사이의 상관 관계가 밀접하다는 것을 의미하고, 대각선상에서 떨어져 분포한다는 의미는 x, y축 간에 급격한 변화를 나타내는 부분으로 시계열 데이터 사이의 상관 관계가 적음을 의미한다. 또한 각 화자마다 시각적으로 구별할 수 있는 특이한 모양의 어트랙터를 형성하고 있음을 볼 때 화자 인식에도 적용할 수 있음을 알 수 있다.

다음은 위의 어트랙터를 시간의 흐름에 따라 선으로 연결한 어트랙터를 각 화자별로 나타낸 것이다.

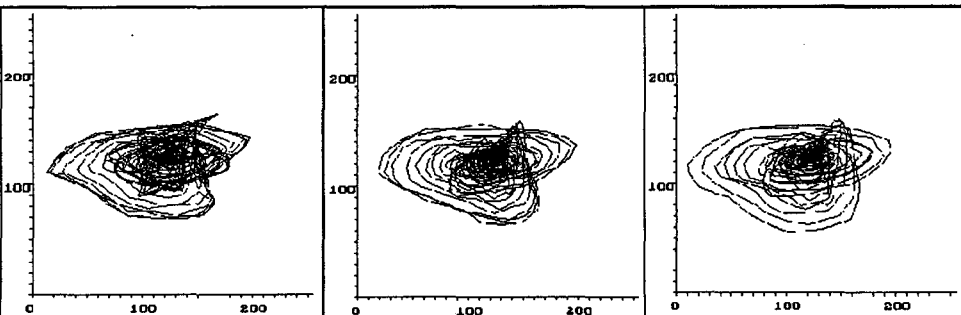
<화자 1>



<화자 2>



<화자 3>



<화자 4>

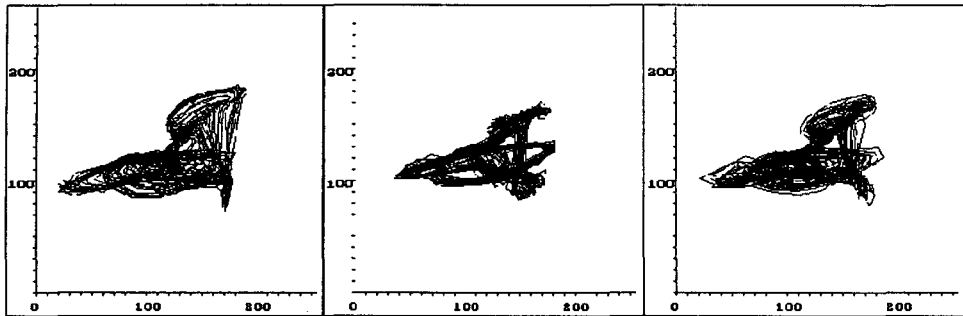


그림 4. 화자별 연결된 벡터 평균값 어트랙터

위의 선으로 연결된 어트랙터를 보면 점 어트랙터에서 볼 수 없었던 일정한 형태의 흐름을 볼 수 있으며 또한 이것은 일정한 사이클을 주기로 하여 반복적으로 그려지고 있음을 알 수 있다. 위의 점과 선의 두 가지 어트랙터는 각 화자별로 상이한 형태로 구성됨을 알 수 있다.

4. 신경망을 이용한 화자 인식 실험

본 논문을 통해 제시한 어트랙터 구성 방법을 이용하여 7 명의 화자에 대한 인식 실험을 하였다. 이 실험은 위 3 절에서 제시한 조건대로 음성데이터를 구성하였으며 신경망을 위한 학습자료는 각 화자에게서 20 번의 '아' 소리를, 인식을 위해서 또 다른 30 번의 소리를 입력하였다. 어트랙터의 형태를 분석하는 방법은 몇 가지가 알려져 있으나 본 실험에서는 Back Propagation Delay 형태의 신경망을 이용하였다. 입력노드값은 상태공간을 250 개의 작은 영역으로 분할한 후 그 영역에 존재하는 점들의 수를 이용하였으며 출력노드는 7 명의 화자를 가리키게 하였고 학습률과 인식률은 각각 95% 이상으로 하였다. 실험결과는 아래 표 1 과 같다.

표 1. 신경망을 통한 화자 인식 실험 결과

화자	1	2	3	4	5	6	7
실험자료수	30	30	30	30	30	30	30
인식수	28	25	29	27	28	26	27
인식률(%)	93.3	83.3	96.7	90.0	93.3	86.7	90.0

5. 결 론

벡터의 평균값을 이용한 어트랙터 구성법은 음성 데이터의 어느 시점과 그 다음 시점과의 발생이 서로 어떠한 연관성을 가지는가를 나타내기 위한 것이다. 이를 통해 각 화자의 어트랙터는 매우 다르게 구분됨을 볼 수 있었고 또한 자신의 목소리는 거의 일정한 형태의 어트랙터로 표현됨을 알 수 있었다. 본 논문은 인식시스템 구현이 목적은 아니지만 간단한 신경망을 통한 실험에서 90.5%의 화자 인식률을 얻을 수 있었다. 비록 많은 경우의 데이터가 아닌 개인별로 50 개의 데이터로 실험한 것이기는 하지만 제안된 방법이 더 많은 연구를 통해 화자 인식에 어느 정도 적용될 수 있을 것으로 판단된다. 음성인식은 다양한 인식 알고리즘의 개발과 더불어 기본 음성데이터의 표현 기법 또한 매우 중요하다고 보아지며 좋은 인식률은 이 두 가지 모두가 중요한 요인으로 작용될 것이다.

현재 다양한 어트랙터 구성 방법이 많이 연구되고 있고 앞으로 새로운 방법에 의한 알고리즘도 등장하게 될 것이다. 인식 시스템의 관점에서 볼 때에는 인식을 위한 알고리즘에 입력되어지는 특징 파라미터 자체가 어떤 의미, 즉 그 음성에 내재된 특별한 정보를 보유하고 있느냐가 없느냐가 중요한 문제가 될 것이다. 그리고 전통적인 어트랙터의 기하학적 형태에 의한 단순한 패턴 매칭 방법보다는 표현된 어트랙터가 그 음성의 특징을 충실히 표현되었느냐 하는 점이 인식률을 높이는데 중요한 역할을 할 것이다.

본 연구의 범위는 어트랙터를 새롭게 구성하여 간단한 신경망을 통해 인식률을 실험한 것으로 인식률 자체보다는 새로운 방법의 음성 데이터의 어트랙터 구성을 제안하였음에 그 의미가 크다 하겠다. 어트랙터에서 보듯이 각 화자별로 구성된 어트랙터는 그 모양이 다를수록 각각의 고유한 음성 정보가 잘 표현되어 있다고 볼 수 있겠다. 따라서 어트랙터를 효과적으로 분석할 수 있는 방법 등을 더욱 연구하면 음성 인식 연구 분야에 이용될 수 있을 것이다.

참 고 문 헌

- [1] 김재홍, 조관선, 이철희. 1999. "연속음성으로부터 추출한 CVC 음성 세그먼트 기반의 음성합성." *한국음향학회지*, 18(7), 10-16.
- [2] 김태식. 1993. "카오스 이론의 어트랙터와 신경회로망을 이용한 화자인식." *계명대학교 산업기술 연구소 논문집*, 16(2), 65-66.
- [3] 김혜경. 1995. *음성신호의 어트랙터에서 사이클 추출을 이용한 화자인식*. 석사학위논문. 계명대학교.
- [4] 김혜경, 성보현, 김태식. 1994. "Strange Attractor를 이용한 화자인식의 음성특징추출에 관한 연구." *한국 정보처리 응용학회 '94 추계 학술발표논문집*, 340-343.
- [5] 도다 모리가즈. 1993. *카오스-혼돈속의 법칙*. 대광서림.
- [6] 박경범. 1998. *선형예측분석법에 의한 음성의 압축과 재생*. 도서출판 하늘소.
- [7] 이강승. 1999. *디지털신호처리*. 21세기사.
- [8] 아이하라 가즈유키. 1995. *쉽게읽는 카오스*. 한빛출판사.
- [9] 이병채. 1995. *카오스 이론을 이용한 생체 비선형 동역학 시스템의 특성 해석*. 박사학위논문. 연세대학교.

- [10] 임진식. 1995. *신경회로망과 카오스 이론의 어트랙터를 이용한 필기자 인식*. 석사학위 논문. 계명대학교.
- [11] 전선도, 강철호. 1999. "잡음에 강한 음성인식을 위한 성문 가중 캡스트럼에 관한 연구." *한국음향학회지*, 18(5), 78-82.
- [12] 정성용, 김태식. 1999. "카오스 어트랙터를 이용한 음성데이터의 특징분석기법에 관한 연구." *한국 정보처리학회 춘계 학술발표 논문집*. 625-628.
- [13] Sun, Fan & Huisheng. Chi. 1992. "Speaker Recognition with Recurrent Network." *International Joint Conference on Neural Networks*, 5(3), 264-267.
- [14] Sun, G. Z. & Y. C. Lee. 1992. "Chaotic Signal Classification with Neural Network Synchronizer." *International Joint Conference on Neural Networks*. 5(2) 68-73.
- [15] Bourlard, H., N. Morgan. & C. Woosters. 1995. "Connectionist Approaches to the use of Markov Models for Speech Recognition." *Neural Information Processing System*, 3, 213-219.
- [16] Tebelskis, J., A. Waibel., B. Petek. & O. S. Bauer. 1995. "Continuous speech Recognition by Linked Predictive Neural Networks." *Neural Information Processing System* 3, 199-205.
- [17] Lee, K., C. Moon. & M. Zaghoul. 1992. "Collective Chaos In Neural Networks." *International Joint Conference on Neural Networks*, 2, 595-602.
- [18] Intrator, Nathan. 1995. "Exploratory Feature Extraction in Speech Signals." *Neural Information Processing System* 3, 241-247.
- [19] Peitgen, Jürgens, Saupe. 1992. *Chaos and Fractals*. Springer-Verlag,

접수일자: 2001. 7. 24.

게재결정: 2001. 9. 5.

▲ 김태식

대구광역시 달서구 신당동 1000 (우: 704-701)

계명대학교 공학부 컴퓨터 공학전공

Tel: +82-53-580-5251

E-mail: tkim@keimyung.ac.kr