

카오스차원에 의한 화자식별 파라미터 추출
 Extraction of Speaker Recognition Parameter
 Using Chaos Dimension

유 병 옥* · 김 창 석**
 (Byong Wook Yoo* · Chang Seok Kim**)

ABSTRACT

This paper was constructed to investigate strange attractor in considering speech which is regarded as chaos in that the random signal appears in the deterministic raising system. This paper searches for the delay time from AR model power spectrum for constructing fit attractor for speech signal. As a result of applying Taken's embedding theory to the delay time, an exact correlation dimension solution is obtained. As a result of this consideration of speech, it is found that it has more speaker recognition characteristic parameter, and gains a large speaker discrimination recognition rate.

1. 서 론

음성의 발생 메카니즘은 비선형 동적 시스템이므로 저주파수의 성분 특성이 잘 나타나는 선형 예측계수만으로는 음성의 특징 파라미터로 충분하지 못하다. 따라서 음성 신호처리 알고리즘의 개선과 부가적 특징파라미터가 필요하며, 최근에 음성인식과 화자식별에 적절한 파라미터의 추출에 관한 연구와 비선형인 음성의 발생 메카니즘의 새로운 모델링 접근에 관한 연구가 활발히 보고되고 있다.

특히 화자식별에서는 직교인자, 또는 통계적 처리방법, 퍼지, 신경망이론에 DTW를 이용하여 화자식별을 행한 연구가 보고되고 있으나, 연산량이 많고, 연속발생되는 대단위 화자의 발성에 대해서는 인식율이 저조하다.

한편 과거의 자연현상을 이해하여 이를 이용하고자 하는 카오스(chaos)응용 연구가 많이 연구되고 있다. 결정론적 법칙에 따라 서로 비선형적인 결합으로 인하여 불규칙적이고, 예측 불가능

* 명지대학교 전자공학과 박사과정

** 명지대학교 전자공학과 교수

하게 보이는 현상이 발생할 때, 그 현상을 카오스라고 한다. 불규칙 시계열 신호에 대해서도 카오스 분석을 통하여 신호의 복잡한 변화 속에 숨겨져 있는 동적 정보를 파악함으로써, 신호를 정량화하여 공학에 활용하려는 연구가 활발히 진행 중이다. 음성신호의 발생 메카니즘은 선형 시스템이 아니라 매우 복잡한 비선형 결정론적 동적 시스템이므로 정량화된 동적 정보의 특징을 나타내는 카오스에 대한 차원을 음성인식에 적용하고자 하는 연구보고도 있다. 음성은 개인마다 서로 상이한 발성 습관과 성도의 구조 차이에 따라 주파수 변동과 시간의 변동을 나타내므로 음성인식은 서로 다른 음성을 변별하고, 개인간의 주파수와 시간의 변동을 흡수하는 문제가 중요하다. 그러나 카오스에 대한 차원은 신호의 궤적에 대한 결정론적 규칙 즉, 발성규칙을 찾고 그것에 대한 정량화된 척도를 뜻하므로 개개인의 다른 음성 발생 메카니즘에 대해 개인간의 성도 구조의 차이, 즉 동적시스템에 대한 규칙정도의 차이를 알아내는데 적합하므로 개인간의 변동을 흡수하는 음성인식 보다는 화자인식 파라미터로 유용한 것으로 생각된다. 따라서 이러한 동적 시스템의 특징은 음성 시계열 신호를 상태공간으로 나타난 스트레인지 어트랙터에 잘 나타나므로 스트레인지 어트랙터로부터 화자 특징을 추출하는 것이 가능할 것으로 사료된다.

그러나 이러한 유용한 화자인식 파라미터 추출, 즉 성도의 특징을 나타내는 궤도 특징을 찾기 위해서는 먼저 스트레인지 어트랙터를 올바르게 구성하여야 한다. 스트레인지 어트랙터의 구성에는 여러 가지 방법이 있으나 시계열 데이터로부터 결정론적 규칙을 추정하기 위한 방법인 Takens의 매립정리에 의한 궤도구성이 카오스의 특징량이 위상적으로 매우 양호하게 보존되어 정확한 특징 추정이 가능하다는 점에서 많이 이용되어 지고 있다. 관측된 시계열 데이터로부터 스트레인지 어트랙터를 구성하기 위해서는 일정한 시간 지연마다 차분을 이용하여 어트랙터 상의 궤도를 재 구성하여야 한다. 그러나 음성신호가 관측시계열 데이터인 경우, 주요 주파수 성분이 존재하며 또한, 자음과 모음의 주파수 차이는 매우 크게 존재하므로 현재 사용되고 있는 지연시간의 결정방법으로는 부적절한 어트랙터가 구성되고 부적절한 카오스 차원의 추정이 이뤄진다. 따라서 자음과 모음이 섞여 있는 단어의 경우는 자음과 모음에 대한 최소 2개이상의 주된 주기의 고려가 필요하다. 이 연구에서는 모음만을 음성시료로 하였으나 모음에도 여러 주파수 성분을 갖을 수 있다고 가정하여 음성신호에 대해 AR모델에 의해 파워 스펙트럼을 추정한 후 가장 진폭이 큰 주파수 성분을 그 음성신호의 주요 주기로 하여 지연시간을 결정하였다.

음성신호가 카오스 신호이면 음성은 결정론적 규칙에 의하여 지배된다. 이러한 규칙은 카오스 운동을 특징 지우는 스트레인지 어트랙터(strange attractor)를 찾으므로써 가능하다. 정확한 스트레인지 어트랙터를 구성하기 위해 AR모델에 의해 구해진 파워 스펙트럼으로부터 음성신호의 주요 주기를 찾고, 스트레인지 어트랙터를 구성함으로써 정확한 상관차원의 값이 추정되도록 하였다. 구성된 어트랙터 궤도의 동적 특성을 특징 지우는 상관차원이 개인차가 심한 음성신호에 대해 개인의 역학계의 특징을 매우 잘 나타내어 줄 것이라는 점에 착안하여 상관차원을 추정하고 평가하여 음성의 화자식별의 특징파라미터로서 타당성을 고찰하고자 한다.

2. 지연시간의 설정

카오스의 발생 시스템은 시계열 데이터로서 관측할 경우 그 동작은 복잡하고 초기값에 대해 매우 민감하게 변하는 불안정한 궤도를 나타낸다. 이러한 불안정한 궤도의 정량적 특징을 알아보기 위해서는 데이터를 공간에 재구성한 어트랙터를 관측하여야 하며, 어트랙터를 구성하기 위해서 Takens의 매립정리를 이용한다. 음성시계열을 x_1, x_2, \dots, x_N 이라할 때, Takens의 매립정리에 의하여 식 (1)과 같이 벡터 시계열을 구성할 수 있다.

$$\begin{aligned} X_1 &= (x_1, x_{1+\tau}, \dots, x_{1+(m-1)\tau}) \\ X_2 &= (x_2, x_{2+\tau}, \dots, x_{2+(m-1)\tau}) \\ &\vdots \\ X_t &= (x_t, x_{t+\tau}, \dots, x_{t+(m-1)\tau}) \end{aligned} \tag{1}$$

여기서 $t = N - (m - 1)\tau$ 이다. 위 식에서 τ 는 지연시간으로서 주된 주기값보다 작은 값으로 설정하여야 하며, 어트랙터의 구성시 τ 가 너무 적으면 재구성상태의 공간내의 데이터는 극단적으로 상관성이 너무 크게 된다. 스트레인지 어트랙터 구성방법에 있어서 현재까지 알려진 지연시간 τ 의 최적 결정방법은 τ 를 관측된 시계열 데이터 X_t 의 주요한 주기의 수분의 1로 선택하는 방법과 X_t 의 자기상관함수 R 이 최초로 0이 되는 시각, 즉 주요 주기의 1/2이 되는 시각으로 선택하는 방법, 그리고 X_t 의 상호정보량 I 가 최초로 극소치를 취하는 시각을 선택하는 방법 등이 사용되고 있다.

표 1은 관측된 음성 시계열 데이터 각 5개 모음에 대해 X_t 의 자기상관계수가 최초로 0 이 되는 시각 τ' 와 파워 스펙트럼을 추정하여 진폭이 가장 큰 2개의 주파수 성분에 대한 주기를 τ_1 과 τ_2 로 하여 비교한 것이다. 표를 보면 τ' 의 값이 모음 '아'에 대해 2로서 주요 주기는 샘플수 \times 샘플주기 = 2 msec 로 추정되며 주파수 분석하여 결정한 τ_1 과 τ_2 와 큰 차이가 있다. 따라서 음성신호에 대한 카오스 분석시 정확한 어트랙터를 구성함에 있어서 지연시간은 자기상관계수로 구하는 것 보다 음성 성분에 대한 주파수 분석후 주된 주기의 추정이 올바른 지연시간 추정이라고 사료된다.

3. 상관 차원

상관차원은 어트랙터 궤도상에서 재구성된 데이터와 데이터 사이의 상관관계를 계산하여 구한 변화율이다. 상관관계란 재구성된 데이터와 데이터 사이의 거리, 즉, 임의의 한 데이터값을 중심으로 일정한 반경 안에 존재하는 데이터의 수를 계산하여 구할 수 있다.

표 1. 지연시간 τ 과 τ' 의 추정 비교.

지연시간		모음별				
		아	에	이	오	우
τ'		2	3	2	2	3
τ	τ_1	8	10	10	8	8
	τ_2	7	8	8	7	7

이러한 상관관계의 계산으로 일정한 기울기의 변화율을 얻을 수 있으며 이 일정한 기울기의 변화율은 신호의 카오스적 특징, 즉 규칙을 나타내게 된다. 또한 신호의 발생 시스템의 카오스적 특징을 나타내기도 한다. m차원 공간에 있어서 재구성된 어트랙터상의 한 점을 $X_i = R^m$ 이라 하면 상관적분 $C^m(\epsilon)$ 은 다음과 같이 정의된다.

$$C^m(\epsilon) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{\substack{i, j=1 \\ i \neq j}}^N H(\epsilon - |X_i - X_j|) \tag{2}$$

$$H(\epsilon - |X_i - X_j|) = \begin{cases} 1, & (\epsilon - |X_i - X_j|) \geq 0 \\ 0, & (\epsilon - |X_i - X_j|) < 0 \end{cases} \tag{3}$$

어트랙터 상의 한 점 $X_i (i = 1, 2, \dots, N)$ 을 중심으로 나머지 (N-1)개의 점에 대해 반경 ϵ 의 m차원 구내로 들어가는 점 X_j 를 계산하고, 모든 X_i 에 대해서 반복 계산한다. 위 식 (2)에서 상관적분이 임의의 ϵ 의 한 영역에서 선형적으로 비례관계가 존재한다면 어트랙터상의 선형 비례 범위 $\epsilon_1 < \epsilon < \epsilon_2$ 에서 자기상사성을 갖는다고 할 수 있다.

$$C^m(\epsilon) \propto \epsilon^{l(m)} \tag{4}$$

$$Dc = \lim_{\epsilon \rightarrow \infty} \frac{\ln C^m(\epsilon)}{l(m)} \tag{5}$$

식 (4)에서 공간차원 m을 증가시키므로써 식(5)와 같이 $l(m)$ 이 포화되어 접근하는 값 Dc 를 상관차원이라 하고, 이는 공간차원 m이 증가하여 포화 상태가 되는 $l(m)$ 값이 각 신호의 발생 시스템에 따라 다른 특징이 될 수 있다.

4. 실험 및 결과

음성데이터 수집은 주변잡음이 존재하는 실험실에서 잡음을 최소화 한 상태로 20대 남녀 화자 각각 5명씩으로 구성하여 1명당 10회씩 단어를 발성하여 발성된 음성에 대해 샘플링 주파수를 10 KHz로 하여 A/D 변환하여 음성 시료를 수집하였다.

이 연구에서 사용된 음성 시료는 상관차원을 화자식별에 대한 파라미터로써 평가를 하기 위해 모음만을 고려하였다.

3명의 화자가 모음 5개 아, 에, 이, 오, 우를 10회씩 발성한 150개의 음성시료 중 1회 발성한 음성을 표준패턴으로 나머지 음성을 시험패턴으로 실험하였다.

그림 1은 모음 '아'에 대한 A 화자의 어트랙터이다. 동일한 모음이라도 어트랙터 궤도형태는 각 화자에 대해 공통적인 형태를 보이지 않는다. 그것은 같은 모음을 발성해도 개개인간의 발성 습관의 차이와 주파수 정보의 차이 즉, 각 화자의 비선형적인 발성기관의 차이에 원인이 있다. 따라서 음성 시계열 신호의 어트랙터는 모음 '아'의 특징도 나타내고 있으나 각 화자의 음원의 특징도 나타내고 있다고 생각할 수 있다.

음성인식의 문제는 각 화자의 모음 '아'에 대한 발성시간과 주파수 차이의 특징 함수의 문제이므로 어트랙터의 궤도 모양의 특징을 정량적으로 나타내는 상관차원을 음성인식의 파라미터로 사용하기에는 부적합하다. 그러나 상관차원이 주어진 한 음성 또는 여러 음성에 대해서는 화자의 특징을 궤도의 형태에 대한 특징으로 나타내므로 화자식별의 파라미터로 사용하는 것이 가능할 것으로 생각된다. 따라서 이 연구에서는 각기 다른 형태의 어트랙터가 각 화자의 비선형적인 발성기관의 차이에서 오는 것으로 보고, 그 차이가 화자간의 특징을 포함하고 있다고 가정하여 어트랙터 궤도의 특징을 정량적으로 평가하여 주는 상관차원을 구해 화자식별을 행하였다.

5. 상관차원에 의한 화자식별

상관차원은 식 (5)를 이용하여 구할 수 있으며, 식 (5)에서의 상관적분 $C^m(\epsilon)$ 은 임의의 X_i 와 X_{i+1} 의 유클리안 거리에 대해 수렴될 때까지 반경 ϵ 을 미소량으로 증가시키면서 각각에 따른 식 (2)로부터 구할 수 있다. 상관차원과 상관적분 간의 관계를 고찰하기 위해 차원 m 을 증가시키면서 $l(m)$ 을 구해본 결과 $m=5$ 이상에서 모든 모음이 일정한 접근값을 나타냈다. m 의 증가에 따라 $l(m)$ 이 포화되어 접근하는 값 D_c 를 구할 수 있었으며 각각의 모음에 대해서도 구하였다. 접근값 D_c 가 상관차원이며 각 화자별 모음별 결과를 표 2에 나타내었다.

그림 1. 화자 A에 대한 '아'의 어트랙터

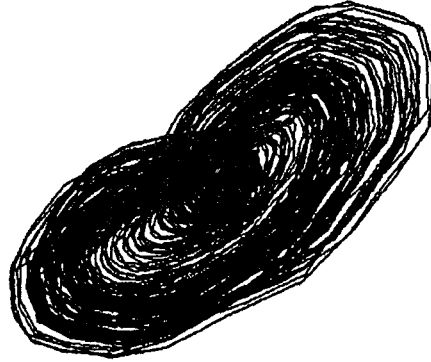


표 2에서 화자의 구별없이 모음에 따라 상관차원의 분포를 나타낸 아래의 분포를 살펴보면, 「아」에 대한 상관차원의 값이 4.2 ~ 4.6으로 「에」의 4.4 ~ 4.7과 겹치는 분포를 갖으며, 다른 모음에 대해서도 모음간의 식별이 곤란한 것을 알 수 있으며 음성인식의 파라미터로서 부적합한 파라미터로 보인다. 화자 간의 차이가 존재하는가를 알아보기 위해 오른쪽의 분포를 살펴보면, A화자의 상관차원의 분포는 4.6 ~ 4.8, B화자의 상관차원의 분포는 4.2 ~ 4.5로서 화자 간 차이를 보이고 있으므로 화자식별이 더 용이함을 알 수 있다.

표 2. 상관차원의 화자별 평균값

상관차원 화자	상관차원의 평균값					
	아	에	이	오	우	분포
A	4.629	4.760	4.604	4.698	4.887	4.6~4.8
B	4.264	4.498	4.336	4.202	4.134	4.2~4.5
C	4.503	4.599	4.603	4.453	4.497	4.4~4.6
분포	4.2~4.6	4.4~4.7	4.3~4.6	4.2~4.7	4.1~4.8	

구해진 상관차원값에 대해 1회 발성한 A화자의 5개 모음을 표준패턴으로하여 각 A,B,C화자의 2,3,4,5회 발성한 모음을 시험패턴으로 하여 다음 식(17)과 같이 정의한 상관차원에 대한 식별오차를 구하였다.

$$\text{식별오차율} = \frac{1}{4} \sum_{i=0}^{i=4} \left| \frac{\text{표준패턴의차원} - \text{시험패턴의차원}}{\text{표준패턴의차원}} \times 100 \right| \quad (6)$$

여기서 i 는 음성 발성 수이다. 식 (6)로 평균 식별오차를 구하여 표 3에 나타내었다.

표 3를 보면 A화자 5개 모음에 대한 식별오차율은 평균 6.5%로서 A화자의 인식율이 매우 높음을 알수 있으며 화자식별 파라미터로 상관차원 값을 이용하는것에 대해 타당하다는 것을 알수 있다. 한편 상관차원에 대한 음성인식 파라미터로의 타당성을 알아보기 위해 1회 발성한 모음을 표준패턴으로 각 화자별 모음에 대한 평균식별 오차율을 구하여 표 4에 나타내었다.

표 3. 표준패턴 A화자에 대한 식별오차율

화 자 \ 모 음	아	에	이	오	우
A	15.97	25.76	20.33	17.50	18.09
B	17.31	18.78	27.35	19.28	19.00
C	12.23	13.93	13.57	18.18	16.29
평균	15.17	19.46	20.42	18.32	17.79

표 4. 각 모음에 대한 평균

화자 \ 표준패턴	A					B					C				
	아	에	이	오	우	아	에	이	오	우	아	에	이	오	우
아	5.10	28.82	15.91	17.30	30.08	15.99	30.36	40.00	18.30	23.33	9.13	13.55	37.67	35.26	27.51
에	27.23	4.50	21.52	18.99	25.77	26.91	11.90	22.17	16.45	30.13	8.88	25.82	24.81	17.68	26.59
이	17.84	33.00	5.98	27.27	26.53	24.68	17.28	17.16	16.36	20.18	26.59	18.37	7.98	40.04	26.53
오	34.43	23.66	18.50	8.53	11.44	27.37	26.88	19.60	14.95	14.74	30.94	26.11	14.24	15.60	19.56
우	22.02	36.63	41.54	16.92	8.39	26.57	37.08	16.49	18.28	13.40	26.42	18.48	29.64	27.64	23.40
평균	6.50					14.68					16.38				

표 4에서 보는 것과 같이 각 모음에 대한 평균 식별 오차율이 매우 크다. 그것은 표2와 같이 각 모음의 상관차원의 값이 서로 겹치는 분포를 가지므로 모음간 식별이 곤란하고 카오스에 대한 차원이 어트랙터케도 형태의 흡수이기 보다는 특징을 척도로 나타내어 케도형태를 분류하므로 표 4의 평균 식별오차가 크게 나타나는 것으로 보이며 카오스의 차원은 음성인식보다는 화자식별에 더 유용한 파라미터임을 알 수 있다. 각 화자마다 1회 발성한 모음을 표준 패턴으로 하여 인식실험을 행한 결과, 표 5를 얻을 수 있었다.

표 5를 보면 평균 91.6 %의 높은 인식율을 얻을 수 있었다. 그러나 화자수가 증가하면 상관차원은 많은 데이터 처리 양이 요구되고 자기상사성을 갖기 위한 선형적 비례범위가 불안정해질

수 있다. 따라서 화자식별에 제한적이라고 할 수 있다. 그러나 단어 또는 연속음인 경우 상관차원의 값의 연속한 변화의 정도를 추정하여 식별파라미터로 사용한다면 우수한 화자식별 인식율을 나타낼 것으로 사료된다.

표 5. 상관차원에 의한 화자 인식율

화자 \ 인식율	인식율 (%)
A	92.5
B	95.4
C	87.1
평균	91.6

6. 결 론

이 논문에서는 음성신호에 비선형적 분석방법인 카오스차원 중 상관차원을 추정하여 화자 인식방법을 제안, 실행한 결과 다음과 같은 결과를 얻었다.

음성신호에 대한 적절한 어트랙터 구성을 위해 AR모델의 파워스펙트럼으로부터 지연시간을 구하였다. 구해진 지연시간을 Takens의 매립정리에 적용하여 정확한 상관차원값을 구할 수 있었다. 궤도의 특징을 나타내는 상관차원값에 대해 음성인식과 화자인식의 특징파라미터로서의 성능을 평가한 결과, 음성인식보다는 화자식별의 특징파라미터로 타당하였으며 높은 화자식별 인식율을 얻을 수 있었다.

참고문헌

- [1] 김지훈·유병욱·김창석. 1996. "Lyapunov 차원을 이용한 화자인식," 제9회 신호처리합동학술대회 논문집, 385 -388.
- [2] 合原一辛. 1992. 脳とカオス. 東京電氣 大學.
- [3] Alan Wolf. 1985. "Determining Lyapunov Exponents from a tim series." *Physica 16D*, 285 - 317.
- Rraserb. A. M. Winney. L. S. 1986. "Independent coordianates for strange attractors from mutual information." *Phys. Rev. A. 33*, 1134-1140.
- [4] Malraison, B., Atten. P., Brege. P and Dubois. M. 1983. "Dimension of strange attractors an experimental." *de Physique Lettres 44(22)*, 897-902.
- [5] MIM, D. S. 1994. "Chaotic Analysis of Speech Signals." 한·일·중 카오스 공동 세미나 논문집, 124-129.

- [6] Takens, F. 1981. "Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence." *Lecture Notes in Mathematics 898*, 366-381.
- [7] Shimada, I. & Nagashima. T. 1979. "A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems." *Prog. Theor. Phys. 61(6)*, 1665-1616.
- [8] Farmer, J. D., Ott, E. and Yorke. J. A. 1983. "The dimension of chatic attractors." *Physics 7*, 153-180.
- [9] Eckman, J. P. & Ruelle D. 1985. "Ergodic theory of chaos and strange attractors." *Rev. Mod. Phys. 57(3)*, 617-656.
- [10] Eckmann, J. P., Kanmhorst. S. O., Ruelle. D and Ciliberto. S. 1986. "Liapunov exponents from a time series." *Phys. Rev. A 6*, 4971-4979.
- [11] Theiler, J. 1986. "Spurious dimension from correlation algorithms applied to limited time-series data." *Phys. Rev. A 34(3)*, 2427-2432.
- [12] Sano, M Sawada, Y. 1985. "Measurement of the Lyapunov spectrum from a chaotic time series." *Phys. Rev. Lett. 55*, 1082-1085.
- [13] Grassberger, P. & Procaccia, I. 1983. "Measuring strangeness of strange attractors," *Physica 9D*, 189-208.
- [14] Frederickson, P., Kaplan, J. L., Yorke, E. D. and Yorke, J. A. 1983. "The Liapunov dimension of strange attractors," *J. Diff. Equations 49*, 185-207.
- [15] May, R. 1976. "Simple mathematical models with very complicated dynamics." *Nature 261*, 459-467.
- [16] Sadaoki Furui. 1992. *Advances in Speech Signal Processing*. Marcel Dekker Inc.
- [17] Thomas S. Parker & Leon O. Chua. 1987. "Chaos: a tutorial for engineers." *Proceeding of the IEEE 75(8)*, 982-1008.
- [18] Li, T. Y., Yorke, J. A. 1975. "Period three implies chaos." *Amer. Math. Monthly 82*, 985-992.

접수일자 : '97. 1. 20.

게재결정 : '97. 2. 18.

▲ 유병욱

경기도 용인시 남동 산 38의 2

명지대학교 전자공학과 (용인캠퍼스) 우편번호 : 449-728

Tel : (0335) 36-5165 FAX : (0335) 30-6374

e-mail: binguki@wh.myongji.ac.kr

▲ 김창석

경기도 용인시 남동 산 38의 2

명지대학교 전자공학과 (용인캠퍼스) 우편번호 : 449-728

Tel: 0335-37-2300 FAX: 0335-30-6374

e-mail: kimcs@wh.myongji.ac.kr