

## 세 국가(國歌)의 복잡도 및 장기기억 속성의 비교

박영선<sup>1)</sup>, 차경준<sup>2)</sup>, 박홍구<sup>3)</sup>

### 요약

우리나라의 애국가(愛國歌), 일본(Kimigayo) 그리고 미국국가(The star-spangled Banner) 등에 대해서 악보가 갖는 고유정보를 카오스적 접근 방법인 근사엔트로피(approximate entropy)와 허스트(Hurst) 지수를 이용하여 각각 음계(scale)의 복잡도(複雜度)와 장기기억속성(長期 記憶 屬性)을 계산하여 비교하였던 바, 애국가가 상대적으로 복잡도에서 가장 높았으며, 세 국가 모두 장기 기억효과가 있는 것으로 나타났다. 지속적인(persistent) 성향은 일본국가가 가장 컸다.

주요용어 : 근사엔트로피, 허스트지수, 복잡도, 장기기억

### 1. 서론

국가(國歌; national anthem)는 국가(國家)를 상징하는 가곡 또는 기악곡. 근대적인 '국가'의 개념이 형성되고 그 건설이 진행됨에 따라 국가에 대한 귀속심(歸屬心)이나 애국심(愛國心)을 고취하는 방책이 필요해져, 표어나 기(旗)·문장(紋章)·꽃·새 등과 함께 노래가 공식적으로 제정되거나 또는 공식적으로 제정된 것이 아니더라도 관례적으로 국민 사이에 널리 불려져 왔다.

음악적으로는 지방의 민요에 바탕을 둔 경우조차 서양적 평균율에 맞춘 선율을 사용하고 있는 것이 대부분이며, 곡조에서는 찬가, 경쾌한 행진곡, 높이 올려 퍼지는 팡파르, 민속소재 등에서 채용한 것으로 나눌 수 있다. 또한, 국가(國歌)는 국위를 과시하거나 민족적 동질성을 상징하는 것으로서, 국제적인 분위기의 자리나, 한 국가 내에서도 많은 국민이 모이는 자리 등에서 노래가 불려지고 연주된다. 따라서 국가(國歌)는 한 국가의 상징으로서 중대한 의미를 가지고 있으며 그에 따른 국민정서나 긍지가 진하게 배어 있다고도 할 수 있다.

이에 본 연구에서는 우리나라의 애국가(愛國歌), 일본국가(Kimigayo) 그리고 미국국가(The star-spangled Banner) 등에 대해서 악보가 갖는 고유정보를 카오스적 접근 방법인 근사엔트로피(approximate entropy)와 허스트지수(Hurst exponent)를 이용하여 각각 음계(scale)의 복잡도(複雜度)와 장기기억속성(長期 記憶 屬性)을 계산하여 비교하여 보았다.

### 2. 세 국가의 자료

우리나라 'Aeguk-ka', 일본국가 'Kimigayo' 그리고 미국국가 'The star-spangled Banner' 등의 악보로부터 음계(scale)를 아래 규칙 1과 2를 적용하여 자료화 하였다.

1) (133-792) 서울특별시 성동구 행당1동 17, 한양대학교 자연과학대학 수학과 강의전담교수

2) (133-792) 서울특별시 성동구 행당1동 17, 한양대학교 자연과학대학 수학과 교수

3) (133-792) 서울특별시 성동구 행당1동 17, 한양대학교 자연과학대학 수학과 부교수

세 국가(國歌)의 복잡도 및 장기기억 속성의 비교

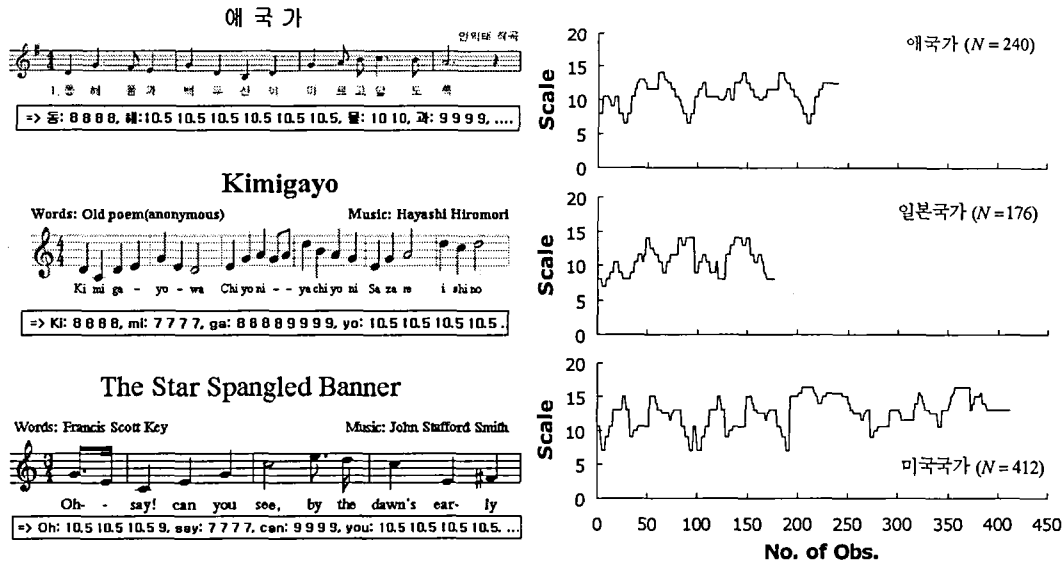


그림 1. 세 국가 음계(scale)의 자료화 예시(Left)와 그에 따른 시계열 변화도(Right)  
 애국가(top), 일본(Kimigayo; middle), 미국(The Star Spangled Banner; bottom)

- 규칙 1. 도-1 레-2 미-3 파-3.5 솔-4.5 라-5.5 시-6.5  
 도-7 레-8 미-9 파-9.5 솔-10.5 라-11.5 시-12.5  
 도-13 레-14 미-15 파-15.5 솔-16.5 라-17.5 시-18.5 도-19
- 규칙 2. 16분음-반복 1회, 8분음-2회, 4분음-4회, 2분음-8회.

먼저, 규칙 1은 음계의 고·저(高·低) 처리를 위해 다장조의 으뜸음 ‘도’를 기준으로 한 옥타브 낮은 ‘도’는 ‘1’로 하였으며, 높은 ‘도’는 ‘13’으로 하였다. 또한 ‘미-파’, 시-도’ 등 사이의 반음은 ‘0.5’로 처리 하였다. 규칙 2는 장·단(長·短) 처리를 위해 16분음을 기준으로 1회 반복, 8분음은 반복 2회, 4분음과 2분음은 각각 4회, 8회로 하였다.

이러한 규칙을 적용하여 세 국가의 음계를 자료화 하여 얻은 결과는 그림 1과 같다. 먼저, 그림 1의 왼쪽은 세 국가 악보의 음계(scale)를 자료화하는 예를 보여 주고 있으며, 오른쪽 그림은 자료화된 음계를 시계열 그림으로 도시한 것이다.

각각의 국가 자료는 미국국가가 가장 긴 자료를(N=412), 그리고 애국가(N=240), 일본(N=176) 등의 순으로 나타났으며, 시계열 그림을 비교해 볼 때, 세 국가 모두가 비교적 ‘sinusoidal’ 패턴이 보이며 미국이 다소 복잡해 보이나 100번째와 200번째 자료가 반복 되어 있다. 또한, 일본의 경우에는 패턴이 자주 반복이 있어 보인다.

### 3. 자료의 복잡도와 장기기억 속성

카오스 이론은 Poincare(1854-1912)에 의하여 제기되었으며 자연의 불규칙성(irregularity), 비예측성(non-prediction) 그리고 복잡성(complexity) 등을 다루며 자연의 복잡성 속에 숨어있는 보편적 질서구조와 생성원리를 이해하는 분야이다 (Freeman 등, 1990).

일반적으로 카오스 정량화 방법에는 불확정성 및 정보이론에 입각한 확률 과정적 엔트로피가 Shannon에 의해 제기되면서 불규칙성과 복잡성 등을 정량화한 근사엔트로피(approximate entropy; ApEn)가 있고(Pincus, 1992; Davidson 등, 1992), 시간지연(time delay) 방법을 이용하여 매립 공간(embedding space)에서 시스템의 자취를 재구성하는 상관차원(correlation dimension)(Packard 등, 1980; Takens, 1980)이 있다. 이밖에 Hurst (1951)에 의해서 제안된 허스트지수와 Peng 등(1994)에 의해 개발된 fluctuation 지수 등이 있다.

### 3.1. 근사엔트로피(approximate entropy)를 이용한 복잡도(複雜度)

본 연구에서는 비선형적인 카오스 지표인 근사엔트로피(ApEn)를 활용하여 자료의 복잡도를 계산하였으며, ApEn은 다음과 같은 연산방법을 이용한다.

$N$ 개의 시계열 자료 점  $u(1), u(2), \dots, u(n)$  이 있다고 하자. 2개의 입력매개변수인  $m$  과  $r$ 이  $ApEn(m, r, N)$ 을 계산하기 위해 고정 한다 (여기서  $m$ 은 매립차원,  $r$ 은 상태공간을  $1/r$ 등분으로 나타내는 변수,  $N$ 은 전체 자료 수).

ApEn을 정의하기 위해 먼저  $\{u(i)\}$ 로부터  $x(i)=[u(i), \dots, u(i+m-1)]$ 로 정의되는  $x(i)$  벡터를 만든다. 이 벡터는  $i$ 번째 점에서 시작되는  $m$ 개의 연속적인  $u$ 값을 나타낸다.  $x(i)$ 와  $x(j)$  사이 거리를 나타내는  $d[x(i), x(j)]$ 는 각 스칼라 성분에 존재하는 최대거리로 정의된다. 이때  $x(1), x(2), \dots, x(N-m+1)$ 을 사용하여 각  $i \leq N-m+1$ 에서

$$C_i^m(r) = \frac{\{d[x(i), x(j)] \leq r \text{이 되는 } x(j) \text{의 개수}\}}{(N-m+1)}$$

가 되며,  $C_i^m(r)$ 는 관용오차(tolerance)  $r$ 내에  $m$ 차원의 기존 벡터의 어떤 형태가 반복되는 빈도(frequency) 또는 규칙성(regularity)의 확률이 된다. 또한  $\Phi^m(r)$ 을 아래와 같이 정의

$$\Phi^m(r) = \left\{ (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^{m+1} \right\}$$

하며, 이때  $\ln$ 은 자연대수이다. 즉,  $\Phi^m(r)$ 은  $\ln C_i^m(r)$ 의 평균으로 정의될 때, 근사 엔트로피 ApEn은 다음과 같이 정의된다.

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r),$$

$$ApEn(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)].$$

즉, 근사엔트로피는  $m$ 개의 관찰점 동안의 어떤 패턴이 다음  $m+1$ 개의 관찰점 동안에도 계속 유사한 확률을 갖고 있느냐 하는 logarithm화한 수이다. 따라서 규칙성의 가능성이 클수록 ApEn은 작고 불규칙성을 갖는 것일수록 ApEn은 크다.

본 연구에서는 세 국가의 자료화된 음계에 대해서 자료 수  $N=40, 80, \dots, 420$  일 때, 각각의 국가에 대해 ApEn ( $m=2, r=0.15 \times SD$ )의 변화도를 살펴보았다 (그림 2).

먼저, 애국가는 전체 자료수가 240개 이므로  $N=40, 80, \dots, 240$ 로 6개의 ApEn을 추출할 수 있었는데,  $N=240$ 에서 ApEn=0.645로 수렴하고 있음을 보여 주었다. 일본의 경우에는 전체 자료수가 176개 이므로  $N=40, 80, \dots, 160$ 로 4개의 ApEn을 추출할 수 있었는데, 전체  $N=176$ 에서 ApEn=0.639로 근사 되었고, 미국국가에서는 전체 자료수가 412개에 대해서  $N=40, 80, \dots, 400$ (10개)의 ApEn가 전체  $N=412$  (ApEn=0.622)에 근사 되었다.

따라서 세 국가의 음계의 복잡도는 큰 차이는 아니지만, 우리나라의 애국가가 가장 높고, 다음으로 일본, 미국국가 등의 순으로 평가해 볼 수 있다.

세 국가(國歌)의 복잡도 및 장기기억 속성의 비교

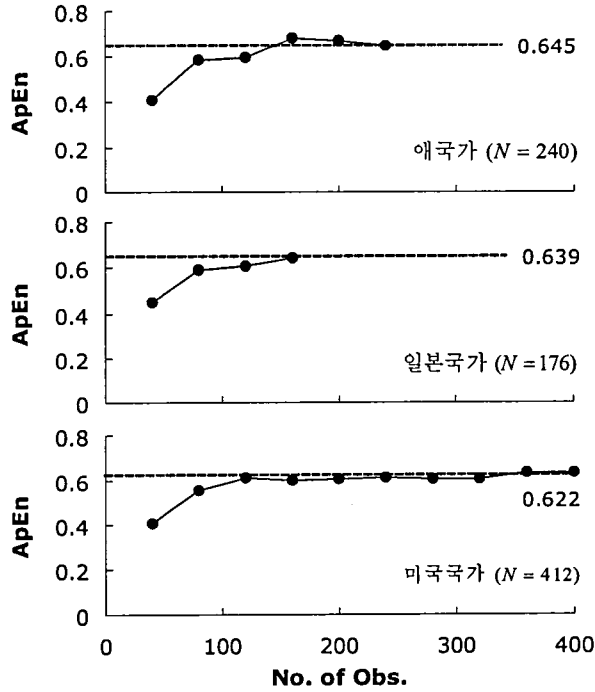


그림 2. 세 국가 음계(scale)의 자료 수(N=40,80, . . . ,420)에 따른 ApEn의 변화  
 애국가(top), 일본(middle), 미국(bottom)

3.2. Hurst 지수를 이용한 장기 기억속성(長期 記憶 屬性)

R/S분석(rescaled range statistical analysis)과 Hurst 지수는 chaos 모형을 직접적으로 검정하기 위한 표준적인 기법 중 하나로서 시간의 흐름에 따라 평균기준으로 양이 어떻게 변하는가를 측정하는 데에서 시작되었다.

어느 한 시계열  $\xi_t$ 에 대한 R/S 통계량은 다음과 같이 정의 된다 (Lo, 1991 참조).

$$X(t, \tau) = \sum_{u=1}^t (\xi_u - \langle \xi \rangle_\tau),$$

$$R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau),$$

$$S(\tau) = \left[ \frac{1}{\tau} \sum_{t=1}^{\tau} (\xi_t - \langle \xi \rangle_\tau)^2 \right]^{\frac{1}{2}}, \quad R/S(\tau) = R(\tau)/S(\tau)$$

일차원 랜덤 워크(random walk)에서 난수를  $\xi_t$ 라 할 때,  $\sum_{t=1}^{\tau} \xi_t$ 는 시간  $\tau$ 후의 보행자의 위치이다. 그리고  $X(t, \tau)$ 에서  $\xi_t$ 의 기댓값이 0이 아닐 때, 척도를 조정하기 위하여 시간지수  $\tau$ 에 걸친 평균  $\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi_t$ 을 차감한다.  $R(\tau)$ 는  $\tau$ 기간 중의  $X(t, \tau)$ 의 최대값과 최소값의 차이로서 시간지수  $\tau$ 에 대하여 보행자가 움직여간 거리를 의미하는 자기 조정 범위(self-adjusted range)이고  $R/S(\tau)$ 는 모형을 일반화하기 위하여 표준편차  $S(\tau)$ 로 나누어주

있는데 이를 자기 재조정 범위(self-rescaled range)라고 한다. 이러한 재조정 범위는 시간이 증가함에 따라 증가하는 경향을 보이며, Hurst는 일반적인 형태로 다음과 같이 제시하고 있다.

$$R/S(\tau) = (a \cdot \tau)^H$$

위의 식에서  $a$ 는 상수이고,  $H$ 를 Hurst 지수(Hurst exponent)라고 하는데, 이 식의 양변에  $\log$ 를 취하고 회귀 분석을 실시하면 Hurst 지수를 추정할 수 있다.

Hurst 지수가  $H=0.5$  이면 장기 기억(long term memory) 효과가 없는 순수한 랜덤 워크며, 현재의 사건은 미래에 영향을 미치지 않음을 의미한다. 그러나 Hurst 지수가 0.5가 아니면 관측값들은 독립적이지 않으며 장기 기억을 가지게 된다. 이는 현재의 사건이 미래의 사건에 영향을 미치며, 단순히 정보의 효과가 급속히 사라지는 연속적 상관관계가 아니라 장기적으로 영향이 지속되는 것을 의미한다. 또한,  $0.5 < H < 1.0$  인 경우, 그 시계열은 랜덤워크 보다 긴 거리에 이르기 때문에 지속적인(persistent) 성향을 의미하고,  $0.0 < H < 0.5$  인 경우는 랜덤워크 보다 짧은 거리에 이르므로 비 지속적인(anti-persistent) 평균 회귀적 시계열(mean reverting series)을 의미한다.

본 연구에서의 Hurst 지수는 각각의 국가음계 자료로부터 먼저  $N=64, 128, \dots, 384$  등 6개의 지수를 추정하고, 각각의 국가 전체자료(애국가  $N=240$ , 일본국가  $N=176$ , 미국국가  $N=420$ )를 최종 구하여 비교하였으며, 이때 lag는 모두  $2^3$ 으로 하였다 (그림 3).

그 결과를 살펴보면, 먼저 애국가의 Hurst지수는 처음  $N=64$ 에서 다소 높았다가 전체 자료 수  $N=240$ 에 와서 안정화된 형태를 띠었으며, 일본의 경우에서도 애국가와 유사한 형태를 보였다. 미국의 경우에는  $N=64$ 에서 애국가와 일본보다는 다소 낮은 값에서 시작하여 점차 중반 이후로 높아지는 형태를 띠었다.

세 국가의 Hurst 지수 ( $H$ )는 모두가  $0.5 < H < 1.0$  의 범위 내에 존재하였으며, 그 중에서도 애국가가 가장 0.50에 근접된 값으로 추정되었으며 ( $H=0.713$ ), 다음으로 미국 ( $H=0.812$ ), 일본 ( $H=0.845$ ) 등의 순이었다.

따라서 세 국가 모두 장기 기억속성이 있는 것으로 나타났는데, 지속성향은 애국가가 가장 낮으며, 다음으로 미국국가 그리고 일본 국가인 'Kimigayo'가 가장 높았다.

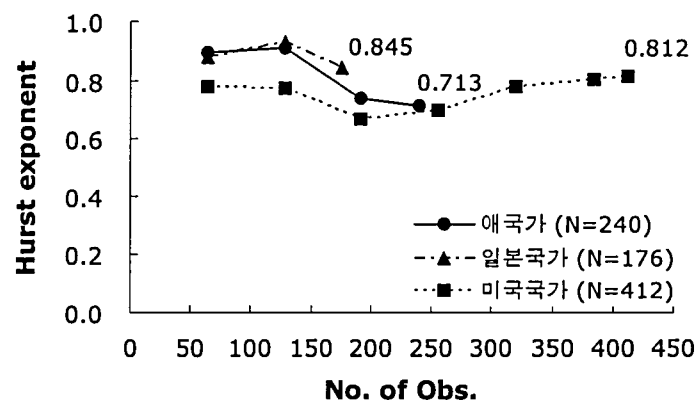


그림 3. 세 국가의 자료 수( $N=64, 128, \dots, 384, 412$ )에 따른 Hurst 지수의 변화

#### 4. 결론

우리나라의 애국가(愛國歌), 일본(Kimigayo) 그리고 미국국가(The star-spangled Banner) 등에 대해서 악보가 갖는 고유정보를 카오스적 접근 방법인 근사엔트로피(approximate entropy)와 Hurst 지수를 이용하여 각각 음계(scale)의 복잡도(複雜度)와 장기기억속성(長期 記憶 屬性)을 계산하여 비교하였던 바, 애국가가 상대적으로 복잡도에서 가장 높았으며, 세 국가 모두 장기 기억효과가 있는 것으로 나타났는데, 지속적인(persistent) 성향은 일본국가가 가장 컸다.

향후에는 좀 더 정확한 국가음계를 계량화하여 분석하고, 좀더 다양한 기법으로 국가에 포함된 고유특성을 추출해야 할 것이다.

### 참고문헌

- Davidson, SR., Rankin, JHG., Martin, JCB., et al (1992), Fetal heart rate variability and behavior state: analysis by power spectrum, *Am J. Obset. Gynecol*, Vol. 167, pp. 717-722.
- Freeman, WJ., Skarda, CA. (1990), *Mind/Brain Science. Neuroscience on philosophy of Mind. Chap.7 in Lepore E, van Gulick R(eds) John Searle and his, Critics*, Oxford, Blackwell, pp. 377-428.
- Hurst, HE. (1951), Long term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, Vol. 166, pp. 770-808.
- Lo, AW. (1991), Long-term memory in stock market prices, *Econometrica*, Vol. 59 No. 5, pp. 1279-1313.
- Packard, NH., Crutchfield, JP., Farmer, JD., Shaw, RS. (1980), Geometry from a Time Series, *Phys. Rev. Let.*, Vol. 45, pp. 712-716.
- Peng, CK., Buldyrev, SV., Havlin, S., et al (1994), On the mosaic organization of DNA sequences, *Phys Rev E*, Vol. 49, pp. 1685-1689.
- Pincus, SM. (1992), Approximating Markov chains, *Proc. Natl. Acad. Sci. USA*, Vol. 89, pp. 4432-36.
- Takens, F. (1980), *Detecting Strange Attractors in Turbulence*, Volume 898 of Lecture Notes in Mathematics (Dynamical Systems and Turbulence, Warwick), pp. 366-381.