

# Bicoherence와 Tricoherence를 이용한 음성 신호의 비선형성 검출

김영인, 임성빈

송실대학교 정보통신전자공학부

156-743 서울시 동작구 상도동 1번지

E-mail: sbi@interpia.net

## Detection of Nonlinearities in Speech Signal Using Bicoherence and Tricoherence

Youngin Kim and Sungbin Im

School of Electronic Engineering, Soongsil University

※이 연구는 97년도 한국과학재단 연구비 지원(과제번호: 971-0917-107-2)에 의한 결과입니다.

### 요약

본 논문의 목적은 한국어 모음의 음성 신호에 대하여 2차 및 3차 비선형성의 존재에 대한 정량적인 분석을 수행함에 있다. 음성 신호의 비선형성을 분석하기 위하여 표본화한 음성 신호에 대하여 bicoherence 및 tricoherence를 측정하였다. 실험 결과에 의하면 한국어 모음의 음성 신호의 발생과정에 상당히 강한 2차 및 3차 비선형성이 존재함을 알 수 있었다. 특히 음성 신호의 3차 비선형성에 대한 연구는 처음 수행되는 것으로 음성 신호 분석에 있어서 매우 중요한 결과로 사료된다.

### 1. 서론

음성 신호의 분석은 음성의 인식, 합성, 코딩에 필요한 정보를 제공하는 기본적인 과정이다. 기존의 음성 신호에 대한 분석 모델들은 주로 선형 모델을 기초로 하고 있다. 선형 모델은 성문을 모델링하는 필터의 계수와 피치의 주기를 추정함으로써 이루어진다. 최근, 음성 파형을 결정하는 성도와 성대의 비선형 특성이 있음을 기초로 하여 음성 신호의 비선형성에 대한 연구가 활발히 진행되고 있다. 비선형성의 영향의 예로는 선형 모델을 사용하여 유성음을 분석하는 과정에서 발생하는 피치 바이어스를 들 수 있다 [1].

비선형 음성 신호에 대한 대표적인 연구로는 다음과 같다. 참고문헌 [2]에서는 Higher-Order Statistics (HOS)를 이용하여 음성 신호의 비선형 특성을 이해하고 그 검출 방법을 보고하였다. 참고문헌 [3, 4]에서는 각각 QPC (Quadratic Phase Coupling) 검출기와 Bispectrum을 측정하여 2차 위상 결합의 검출에 관한 연구가 발표되었다. 참고 문헌 [5, 6, 7]에서는 chaotic 신호 처리, 신경망, Volterra 필터, Bilinear 필터를 사

용하여 음성신호의 모델링을 수행하고 있다. 이러한 연구들은 주로 음성 신호의 2차 비선형성에 대하여 초점을 맞추고 있으며, 또한 2차 비선형성의 검출하는 방법론에 있어서도 간접적으로 비선형성의 존재 여부를 분석하고 있다.

본 논문의 목적은 한국어 모음의 음성 신호에 대하여 비선형성의 존재를 2차 비선형성과 3차 비선형성으로 구분하여 검출하는데 있다. 이러한 목적을 위하여 본 논문에서는 HOS에서 사용되는 bicoherence test와 tricoherence test를 사용하여 각각 2차와 3차 비선형성을 정량적으로 측정한다. 음성 신호에 대한 2차 및 3차 비선형성에 대한 연구는 음성 신호 처리에서 새로운 정보를 제공할 것이며 비선형 필터를 사용하는 음성 분석 연구에 대하여 중요한 토대를 제공한다.

다음 절에서는 HOS의 이론적 측면에서 bicoherence와 tricoherence에 관련된 정의들을 소개하고 3절에서는 2절에서의 정의를 음성 신호의 비선형 검출에 적용하기 위해 bicoherence와 tricoherence의 추정 방법에 대하여 설명한다. 4절에서는 한국어 모음 /아, 에, 이, 오, 우/의 표본에 bicoherence와 tricoherence를 적용하여 측정된 실험 결과를 보고하고 끝으로 5절에서 결론을 맺는다.

### 2. Bicoherence와 Tricoherence의 정의

본 절에서는 불규칙 신호의 Higher-Order Statistics [8]로부터 정의되는 bicoherence와 tricoherence에 대하여 살펴본다. 먼저, 랜덤 과정  $x(k)$ 에 대한 1차 및 2차 cumulant 수열을 각각  $c_1^*$  와  $c_2^*(\tau)$ 로 표시하면 다음과 같다.

$$c_1^* = E\{x(k)\} = m_x \quad (1)$$

$$c_2^x(\tau) = E\{(x(k) - m_x)(x(k + \tau) - m_x)\} \quad (2)$$

2차 cumulant 수열  $c_2^x(\tau)$ 의 푸리에 변환을 2차 cumulant 스펙트럼  $C_2^x(\omega)$ 는 다음과 같이 정의된다.

$$C_2^x(\omega) = \sum_{\tau=-\infty}^{+\infty} c_2^x(\tau) \exp\{-j\omega\tau\} \quad (3)$$

$C_2^x(\omega)$ 는 power spectrum을 나타낸다.

3차 cumulant 수열  $c_3^x(\tau_1, \tau_2)$ 와 푸리에 변환 bispectrum  $C_3^x(\omega_1, \omega_2)$  (혹은 3차 cumulant 스펙트럼)은 다음과 같다.

$$c_3^x(\tau_1, \tau_2) = E\{(x(k) - m_x) \cdot (x(k + \tau_1) - m_x)(x(k + \tau_2) - m_x)\} \quad (4)$$

$$C_3^x(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{+\infty} \sum_{\tau_2=-\infty}^{+\infty} c_3^x(\tau_1, \tau_2) \exp\{-j(\omega_1\tau_1 + \omega_2\tau_2)\} \quad (5)$$

Bispectrum의 크기는 랜덤 과정  $x(k)$ 의 2차 비선형성을 반영하므로 그 특성을 이해하는데 매우 유용하다. 이를 신호의 파워 스펙트럼들로 정규화한 것을 bicoherence라고 하며 다음과 같다.

$$b^x(\omega_1, \omega_2) = \frac{C_3^x(\omega_1, \omega_2)}{\sqrt{C_2^x(\omega_1) \cdot C_2^x(\omega_2) \cdot C_2^x(\omega_1 + \omega_2)}} \quad (6)$$

4차 cumulant 수열  $c_4^x(\tau_1, \tau_2, \tau_3)$  및 trispectrum  $C_4^x(\omega_1, \omega_2, \omega_3)$  (혹은 4차 cumulant 스펙트럼)을 다음과 같이 나타낼 수 있다.

$$c_4^x(\tau_1, \tau_2, \tau_3) = E\{(x(k) - m_x)(x(k + \tau_1) - m_x)(x(k + \tau_2) - m_x)(x(k + \tau_3) - m_x)\} \quad (7)$$

$$C_4^x(\omega_1, \omega_2, \omega_3) = \sum_{\tau_1=-\infty}^{+\infty} \sum_{\tau_2=-\infty}^{+\infty} \sum_{\tau_3=-\infty}^{+\infty} c_4^x(\tau_1, \tau_2, \tau_3) \exp\{-j(\omega_1\tau_1 + \omega_2\tau_2 + \omega_3\tau_3)\} \quad (8)$$

Trispectrum의 크기는 랜덤 과정  $x(k)$ 의 3차 비선형성의 유무를 검출할 수 있다. 또한 이를 신호의 파워 스펙트럼들로 정규화한 것을 tricoherence라고 하며 다음과 같이 정의한다.

$$t^x(\omega_1, \omega_2, \omega_3) = \frac{C_4^x(\omega_1, \omega_2, \omega_3)}{\sqrt{C_2^x(\omega_1) \cdot C_2^x(\omega_2) \cdot C_2^x(\omega_3) \cdot C_2^x(\omega_1 + \omega_2 + \omega_3)}} \quad (9)$$

위에서 언급된 것처럼 bispectrum과 trispectrum도 불규칙 신호에 존재하는 2차 및 3차 비선형성을 반영하지만 그 크기가 신호 표면의 크기에 비례한다. 따라서 같은 표면 신호에 대하여 비교함에 있어서는 bispectrum이나 trispectrum의 크기를 비교하여 상대적인 비선형성의 정도를 나타낼 수 있지만 서로 다른 표면 신호에 대하여서는 그 크기를 가지고 비선형성의 정도를 비교할 수가 없다. 따라서, 서로 다른 신호에 대한 비교에 있어서는, 정규화를 통하여 신호 크기에 대한 의존도를 제거한 bicoherence나 tricoherence의 사용이 비선형성의 정도를 측정하는데 있어서 유리하다.

### 3. 음성 신호에 대한 비선형성 검출 방법

본 논문에서는 음성 신호의 2차 및 3차의 비선형 성분을 검출하기 위하여 식 (6)과 (9)에서 정의된 bicoherence와 tricoherence를 이용한다. 위에서의 정의는 시간 영역에서 구한 상관함수들에 대하여 다차원의 푸리에 변환을 적용시키고 있다. 따라서, 상당한 계산량이 요구된다. 이러한 문제를 피하기 위하여 본 논문에서는 먼저 이산 시간 영역의 표본 신호들에 대하여 segment로 나누고 각 segment에 대하여 DFT를 적용하여 주파수 영역으로 변환한다. 변환된 값을 이용하여 다음의 식에 의하여 bicoherence와 tricoherence를 추정하게 된다. bicoherence의 추정은 다음과 같다 [9].

$$\hat{b}^2(f_1, f_2) = \frac{|\hat{B}(f_1, f_2)|^2}{E(|X_i(f_1)|^2)E(|X_i(f_2)|^2)E(|X_i(f_1 + f_2)|^2)} \quad (10)$$

여기서  $\hat{B}(f_1, f_2)$ 은 bispectrum의 추정치로서 다음과 같다.

$$\hat{B}(f_1, f_2) = E\{X_i(f_1 + f_2)X_i^*(f_1)X_i^*(f_2)\} \quad (11)$$

식 (10)과 (11)에서,  $X_i(\cdot)$ 는 주어진 신호의  $i$ 번째 segment의 DFT를 나타내며,  $f_1, f_2$ 는 각각의 주파수 성분을 표시한다. 또, \*은 공액 복소수임을 표시하고  $f_1$ 과  $f_2$ 의 쌍을 bifrequency라고 부른다.  $E\{\cdot\}$ 는 segment에 대한 평균을 의미한다.

Tricoherence는 다음의 식을 사용하여 추정한다.

$$\hat{t}^3(f_1, f_2, f_3) = \frac{|\hat{T}(f_1, f_2, f_3)|^2}{[E(|X_i(f_1)|^2)E(|X_i(f_2)|^2)E(|X_i(f_3)|^2)E(|X_i(f_1 + f_2 + f_3)|^2)]} \quad (12)$$

여기서 trispectrum  $\hat{T}(f_1, f_2, f_3)$ 은 다음과 같다.

$$\hat{T}(f_1, f_2, f_3) = E\{X_i(f_1 + f_2 + f_3)X_i^*(f_1)X_i^*(f_2)X_i^*(f_3)\} \quad (13)$$

Schwartz부등식을 사용하면 다음의 조건이 성립한다 [10].

$$0 \leq \hat{b}^2(f_1, f_2) \leq 1 \quad (14)$$

$$0 \leq \hat{t}^3(f_1, f_2, f_3) \leq 1 \quad (15)$$

따라서, bicoherence의 경우, 주파수  $f_1, f_2$ 와  $f_1 + f_2$  사이에 강한 2차 위상 결합이 존재한다면 추정된 bicoherence 값은 1에 가까운 값이 되고, 이들이 서로 독립적이라면 0에 가까운 값을 갖는다. 예를 들면, 가우시안과 같은 랜덤 신호는 거의 0에 가까운 값을 갖는다. 또한, tricoherence도 신호의 3차 위상 결합의 정도에 따라 0에서 1사이의 값을 갖는다.

### 4. 실험 결과

본 연구에서는 한국어 모음 /아, 에, 이, 오, 우/에 대하여 bicoherence 및 tricoherence 실험을 수행하였다. 임의로 20세의 여성 화자 2명과 남성 화자 3명으로부터 데이터를 채취하였으며 각 데이터는 표본화 주파수

11025Hz, 모노 8bit로 표본화하였다. 512-point DFT 변환시 각 segment간에 50%의 overlap을 적용하였다.

그림 1은 위에서 언급된 조건하에서 여성 화자 1의 발음 /우/에 대하여 bicoherence를 측정된 결과이다. 512-point FFT를 적용한 400개의 segments를 사용하였다. Bifrequency 237Hz와 237Hz 그리고 두 주파수 성분의 합인 474Hz에서 최대 bicoherence값 0.8646이 검출되었다. 본 실험 결과에서 한국어 모음 /우/의 bicoherence의 추정치가 상당히 큼을 알 수 있었다. 따라서 2차 비선형성의 존재를 확인할 수 있다.

그림 2는 남성과 여성 화자 각각 2인으로부터 102400개씩 채취한 모음 /아 에 이 오 우/의 표본에 대하여 bicoherence를 측정된 결과이다. 남성과 여성 모두 상당히 큰 bicoherence가 pitch 주파수와 pitch 주파수의 harmonics에 관련되어 검출되었으며 bicoherence의 평균값은 약 0.88정도임을 알 수 있었다. 따라서 한국어 모음에는 2차 비선형 성분이 존재함을 알 수 있다.

그림 3은 segment수에 따른 bicoherence 추정치의 변화를 비교한 것이다. 남성 화자 3의 모음 /아 에 이 오 우/에 대하여 각각 102400개와 614400개의 표본을 사용하였다. 서로 다른 길이의 표본에 대하여 각각 평균값은 0.88과 0.89로 거의 유사한 값이 검출되었다.

그림 4는 남성 화자 3의 발음 /아/에 대하여 2400개의 512-point segments를 사용하여 tricoherence를 측정된 결과의 일부를 보여주고 있다. 주파수  $f_3$ 의 각각에 대하여 최대의 tricoherence 값을 도시한 것이다. 약 20Hz부터 1000Hz까지 추정된 tricoherence는 평균 약 0.5정도이다. 그림에서 peak와 음성 신호의 power 스펙트럼의 peak에 해당하는 주파수 성분들이 일치하고 있다. 그리고 3000Hz이상의 고주파 영역에서는 음성 신호의 성분들이 거의 존재하지 않기 때문에 낮은 tricoherence를 보여 주고 있다.

그림 5는 그림 4에서 보여준 실험에서  $f_3$ 가 151Hz일 때의 tricoherence의 단면 (slice)을 도시하고 있다. 주파수가 각각 151Hz, 129Hz, 302Hz 및 이들의 결합 주파수인 582Hz에서 최대 tricoherence 값인 0.8334가 검출되었다.

그림 6은 남성 화자 3의 한국어 모음에 대한 tricoherence의 최대값을 도시하였다. 평균적으로 약 0.68이다. 따라서 3차 비선형성이 존재함을 알 수 있다.

## 5. 결론

본 논문에서는 bicoherence와 tricoherence를 이용한 한국어 모음에 대한 2차 및 3차 비선형성 검출을 수행하였다. 실험을 통하여 한국어 모음의 음성 신호의 비선형성을 검출하였고 그 결과로 우리는 한국어 모음에 2차 및 3차의 비선형성이 존재함을 알 수 있었다.

음성 신호는 그 동안 선형이라는 가정에서 연구가 이

루어져왔다. 그러나 본 연구의 결과에 의하면 기존의 선형 모델 방법보다 비선형 모델을 사용하여 음성 신호를 좀 더 충실하게 모델링을 할 수 있을 것으로 사료된다.

## 참고 문헌

- [1] J. R. Deller, J. G. Proakis, and J. H. L. Hasen, *Discrete-Time Processing of Speech Signals*, Prentice-Hall, New Jersey, 1993.
- [2] J. W. A. Fackrell, S. McLaughlin, "Detecting phase coupling in speech signals", *IEE Colloquium Digest 1995/091 on Speech and Image Processing* pp. 4/1-4/8, London, England May 2, 1995.
- [3] J. W. A. Fackrell, S. McLaughlin, and P. R. White, "Practical issues in the application of the bicoherence to the detection of quadratic phase coupling," in *IEEE/ATHOS Signal Processing Workshop on Higher-Order Statistics*, Costa Brava, Spain, June 1995.
- [4] J. Thyssen, H. Nielsen, and S. D. Hansen, "Non-linearities in Speech," *Proc. of IEEE Workshop on Nonlinear Signal and Image Processing*, pp. 662-665, 1995.
- [5] B. Townshend, "Nonlinear Prediction of Speech Signals," in *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank, pp. 443-452, Addison-Wesley, 1992.
- [6] J. Thyssen, H. Nielsen, and S. D. Hansen, "Nonlinear Short-Term Prediction in Speech Coding," *Proc. of ICASSP 94*, vol. 1, pp. 185-188, 1994.
- [7] 강영태, 이기용, "비선형 예측기에 의한 음성 신호 해석," *한국음향학회 학술발표대회논문집*, pp. 159-162, 1995.
- [8] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [9] E. J. Powers and S. Im, "Introduction to Higher-Order Statistical Signal Processing and Its Applications," *Higher-Order Statistical Signal Processing*, edited by Boashash, Powers, and Zoubir, pp. 3-25, Wiley, 1995.
- [10] S. Elgar and R. T. Guza, "Statistics of bicoherence," *IEEE Transactions in Acoustics, Speech and Signal Processing*, vol. 36, pp. 1667-1668, October 1988.

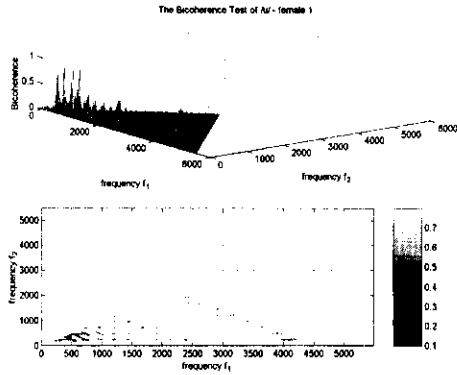


그림 1. 여성 화자 1 의 /우/에 대한 bicoherence 실험.

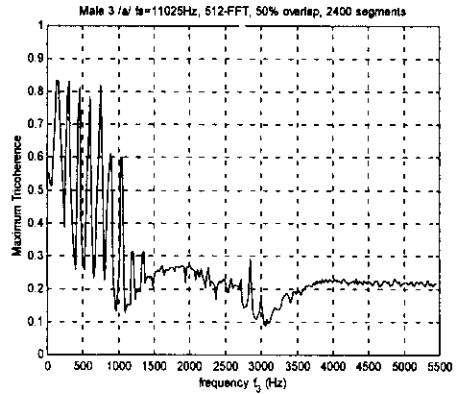


그림 4. 남성 화자 3 의 /아/에 대한 tricoherence 실험.

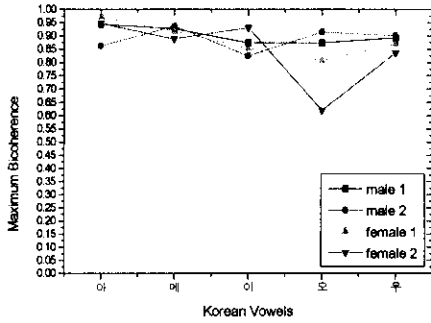


그림 2. 국어 모음의 남성, 여성 각 2인에 대한 bicoherence 실험 결과.

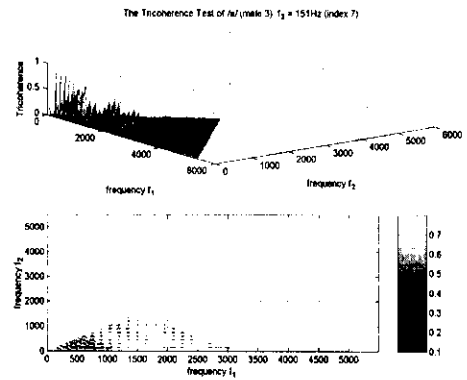


그림 5. 그림 4에서  $f_3=151\text{Hz}$  일 때 tricoherence 실험.

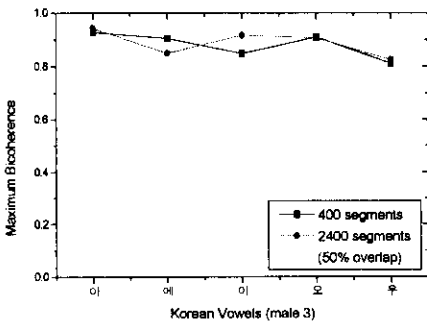


그림 3. 국어 모음의 남성 화자 3에 대한 bicoherence 실험 결과.

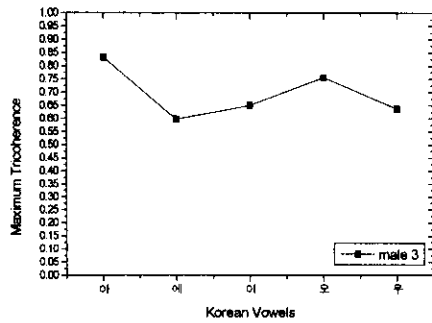


그림 6. 국어 모음의 남성 화자 3에 대한 tricoherence 실험 결과.