

# LMS를 이용한 TSIUVC의 음성신호처리에 관한 연구

이시우<sup>1\*</sup>

## A Study on Speech Signal Processing of TSIUVC using Least Mean Square

See-Woo Lee<sup>1\*</sup>

**요 약** 유성음원과 무성음원을 사용하는 음성부호화 방식에 있어서, 프레임 안에 모음과 무성자음이 있는 경우에 음성 파형에 일그러짐이 나타난다. 본 논문에서는 LMS를 적용한 새로운 TSIUVC 근사합성법을 제시하였다. 실험결과, LMS를 적용하여 양호한 TSIUVC 근사합성 파형을 얻을 수 있었으며, 오차신호가 일그러짐이 적은 근사합성 파형에 중요한 역할을 한다는 것을 알 수 있었다. 이 방법은 음성합성, 음성분석, 새로운 Voiced/Silence/TSIUVC의 음성 부호화 방식에 활용할 수 있을 것으로 기대된다.

**Abstract** In a speech coding system using excitation source of voiced and unvoiced, it would be a distortion of speech waveform in case of exist a voiced and an unvoiced consonants in a frame. In this paper, I propose a new method of TSIUVC(Transition Segment Including UnVoiced Consonant) approximate-synthesis by using Least Mean Square. As a result, a method by using Least Mean Square was obtained a high quality approximation-synthesis waveform. The important thing is that the frequency signals in a maximum error signal can be made with low distortion approximation-synthesis waveform. This method has the capability of being applied to a new speech coding of Voiced/Silence/TSIUVC, speech analysis and synthesis.

**Key words** : Speech Signal Processing, 음성신호처리, Approximate-Synthesis, 근사합성

### 1. 서론

음성신호를 유성음(Voiced)/무성음(Unvoiced) 혹은 유성음(Voiced)/무성음(Unvoiced)/무음(Silence)과 같은 선택정보에 의해서 음성신호를 재생하는 방식[1][2]에서는 음성신호를 수십 ms의 고정된 프레임으로 분할하여 처리하는데, 프레임내의 음성신호는 유성음, 무성음, 무음으로 각기 독립적으로 존재하는 것이 아니라 무음(S)+무성음(UV) 또는 무음(S)+유성음(V), 유성음(V)+무성음(UV)의 형태로 존재한다. 이러한 형태의 음성신호를 유성음원 혹은 무성음원 어느 한쪽의 음원을 일방적으로 사용하여 음성신호를 재생하는데 문제점이 있다. 또한 음성신호는 연속적이며 전후의 모음과 자음의 결합에 의하여 과도기적으로 변화하게 되는 천이구간내의 음성신호를 유성음원 또는 무성음원으로 재생하는 것도 한계가 있다고 볼 수 있다. 이

러한 문제점을 해결하는 방법으로는 유성음과 무성자음이 같은 프레임에 존재하지 않도록 프레임의 길이를 동적으로 할당하는 것도 고려해 볼 수 있으나, 이것은 디지털 신호처리의 특성상 상당히 어려운 처리 과정이라 생각된다. 따라서 본 연구에서는 연속음성에서 TSIUVC(Transition Segment Including Unvoiced Consonant)에 최소 자승법(LMS: Least Mean Square)을 적용하여 재생하는 새로운 근사 합성법에 관하여 기술하고자 한다. 이 방법은 V/S/TSIUVC 음성부호화 방식에 응용하기 위한 방법이다.

### 2. LMS에 의한 음성신호처리 모델

기존의 Voiced/Unvoiced 형태의 음성신호처리 [3]~[5] 방식에서는 음성신호를 Clustering하여 생성된 프레임 안에 유성음 혹은 무성음이 같이 존재할 수 있으며, 이러한 경우에 TSIUVC를 유성음원 혹은 무

<sup>1</sup>상명대학교 정보통신공학과

\*교신저자: 이시우(swlee@smu.ac.kr)

성음원의 어느 한쪽의 음원을 사용하여 재생하게 된다. 그러나 TSIUVC의 특성은 유성음도 무성음도 아닌 모음과 자음의 결합에 의해 발생한 과도기적인 음성신호이기 때문에 유성음원 또는 무성음원 어느 한쪽의 음원을 사용하여 TSIUVC를 재생하는 데는 한계가 있다.

이러한 문제점을 해결하고자 본 논문에서는 최소 자승법을 이용한 음성신호처리 모델을 그림 1에 제시한다. 10kHz로 표본화된 연속음성신호에서 추출한 TSIUVC를 스펙트럼 상에서 신호처리하기 위해 256 point Hamming Window와 FFT를 사용하였다. 일반적으로 TSIUVC 신호는 유성음과 무성자음의 신호와 달리 신호의 진폭이 급격하게 변화하는 특성을 갖고 있기 때문에 최소 자승법의 처리 과정에서 발생하는 최대 오차신호  $e(x_{ij})$ 의 위치에 있는 주파수 신호(k)를 사용하여 파형의 일그러짐을 보상하게 된다. 사용하는 k의 수에 따라서 근사합성 파형의 보상정도가 달라진다.

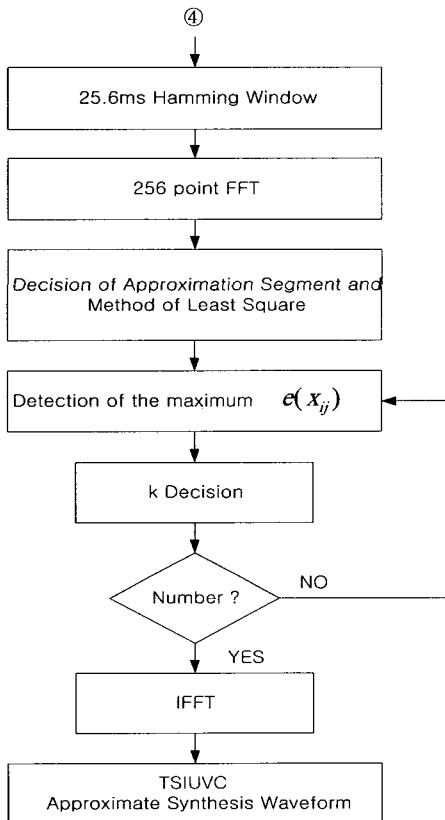


그림 1. 최소 자승법(LMS)에 의한 TSIUVC 근사합성 방법

최소 자승법의 대표적인 적용 알고리즘인 Widrow-Hoff는 분석하고자 하는 파라미터가 시간과 함께 변동한다는 가정하에서 추정한다[6]. 이때, 파라미터 변화의 추이에 따라서 최소 자승법의 추정속도가 급격히 떨어지는 경우가 있는데, 이 문제를 Clark, G. A는 입력한 데이터를 블록으로 처리하여 신호처리 속도의 향상을 꾀하였다[7]. 이러한 것을 고려하여 본 연구에서는 그림 2와 같이 연속음성에서 무성자음과 천이구간, 유성음이 존재하는 프레임의 음성신호에서 TSIUVC를 추출하고(그림2 중앙) FFT하여 최소 자승법에 의하여 근사곡선을 구하였다(그림2 하단).

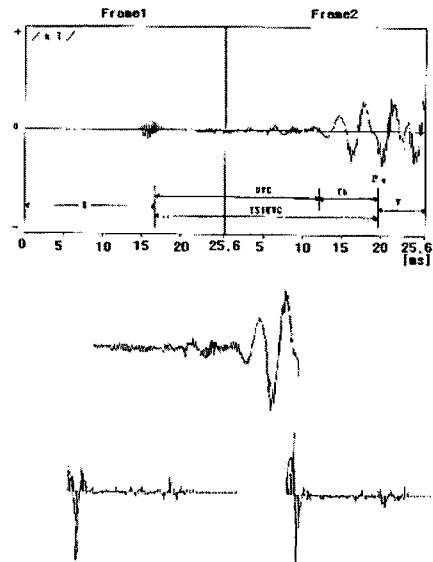


그림 2. 최소 자승법(LMS)에 의한 TSIUVC 추출과 신호 처리

이때 FFT한 TSIUVC신호는 3.4kHz 주파수 대역에 87개 주파수신호가 존재하게 되며, 이 87개의 주파수 신호  $x_{ij}$ 를 다음 식과 같이 블록화 하였다.

$$\begin{aligned}
 x_{ij} = & (x_{00}, x_{01}, \dots, x_{0M}) \\
 & + (x_{10}, x_{11}, \dots, x_{1M}) \\
 & + (x_{m0}, x_{m1}, \dots, x_{mM})
 \end{aligned} \tag{1}$$

블록화한 신호에 대한 근사신호  $y(x)$ 를  $x$ 의  $n$ 차 다항식으로 나타내면 다음과 같이 나타낼 수 있다.

$$\begin{aligned}
 y(x) = & a_{00} + a_{01}x + a_{02}x^2 + \dots + a_{0M}x^n \\
 & + a_{10} + a_{11}x + a_{12}x^2 + \dots + a_{1M}x^n \\
 & + a_{m0} + a_{m1}x + a_{m2}x^2 + \dots + a_{mM}x^n
 \end{aligned} \tag{2}$$

이때, 각 신호에 있어서 실제 측정된 신호인 측정신호( $f_{ij}$ )와 최소 자승법에 의하여 추정된 신호인 근사신호( $y(x_{ij})$ )의 차가 오차( $e(x_{ij})$ )신호가 된다.

$$e(x_{ij}) = y(x_{ij}) - f_{ij} \quad (3)$$

이때  $a_{ij}$ 는 오차신호 평가값(A)이 최소가 되도록 하여 얻는다.

$$A = \sum_{i=0}^m \sum_{j=0}^M e^2(x_{ij}) \quad (4)$$

오차신호가  $e=0$ 이라면 측정신호( $f_{ij}$ )에 대하여 오차 없이 근사신호( $y(x_{ij})$ )를 얻은 결과로서, 오차신호( $e(x_{ij})$ )에는 근사신호를 보정하기 위한 정보가 포함되어 있다고 볼 수 없다. 만약,  $e \neq 0$ 이라면 오차신호( $e(x_{ij})$ )에는 근사신호를 보정할 수 있는 정보가 포함되어 있다고 볼 수 있다. 따라서 TSIUVC의 신호를 최소 자승법으로 처리하는 과정에서 발생하는 오차신호를 이용하여 근사합성 파형의 일그러짐을 보상할 수 있다.

### 3. 최소 자승법(LMS)의 오차신호에 의한 보정

TSIUVC를 FFT하여 얻은 3.4kHz 주파수 대역에 존재하는 주파수신호에 최소 자승법을 적용하여 오차신호의 평가 값을 추정하였다. 이 주파수 신호에  $n \geq 3$ 의 낮은 근사 다항식으로 추정하면 연립방정식의 계수행렬이 감소하여 적은 계산으로 처리할 수 있기 때문에 본 연구에서는 근사다항식을  $n=3$ 으로 하고 최소 자승법에 의하여 측정치에 대한 근사 값을 구하였다.

근사 곡선을 얻기 위해서는 사용하는 주파수 신호를 최소 2개에서 최대 29개로 할 경우에, 분할 가능한 블록 수는 최소 3블록에서 최대 43블록이 된다. 블록화한 TSIUVC신호에 최소 자승법을 적용하여 계산한 오차의 평가 값을 그림 3에 나타내었다. 그림 3에는 무성 파열자음/kI/, 무성 마찰자음/sO/, 무성 파찰자음/ts/이 포함된 TSIUVC의 오차 평가 값으로서 실선은 TSIUVC 주파수 신호를 2~29개의 범위 내에서 점차 늘렸을 경우의 오차 평가 값이며, 이때 오차신호( $e(x_{ij})$ )가 최대인 위치에 있는 주파수신호(k)는 사용하지 않았다. 또한 TSIUVC 주파수 신호에 k를

1~29개 범위 내에서 점차 늘렸을 경우의 오차 평가 값을 파선으로 나타냈다.

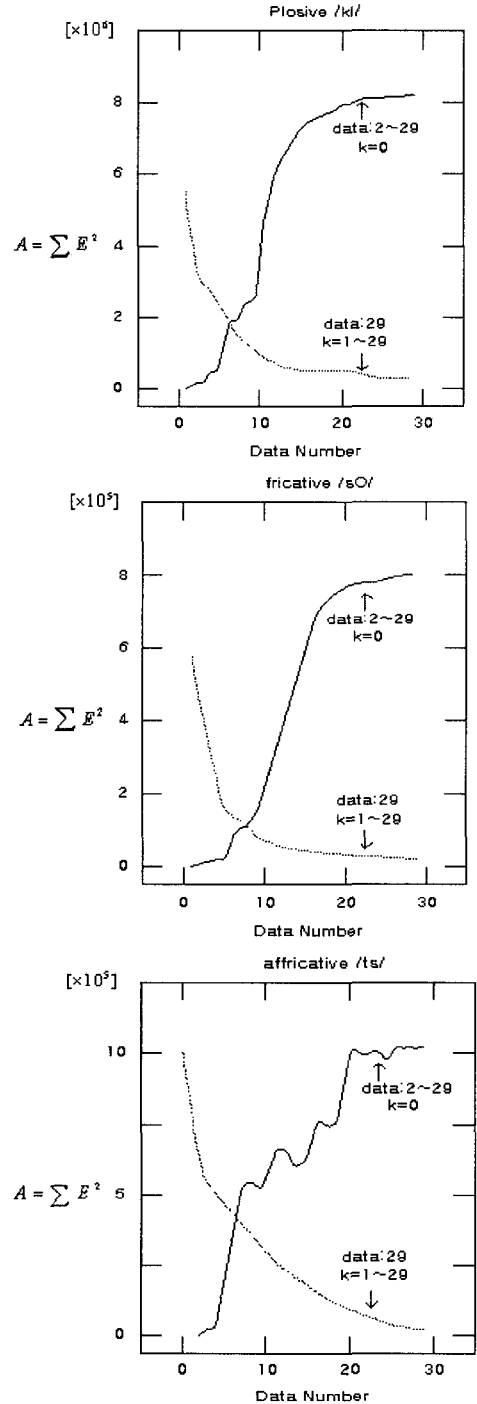


그림 3. 최소 자승법(LMS)에 의한 에러평가

실험결과, 그림 3에서 볼 수 있듯이 k를 사용하지 않는 경우에는 오차 평가 값이 증가하나 k를 점차 늘릴수록 오차 평가 값이 현저히 감소하는 것을 알 수 있다. 오차의 평가 값이 큰 경우 /kI/에서 약  $8.5 \times 10^6$ , /sO/에서 약  $7.9 \times 10^5$ , /ts/에서 약  $12.8 \times 10^5$  로 나타났으나, 이 경우에 k=1를 사용한 경우에 오차 평가 값이 /kI/에서 약  $5.5 \times 10^5$ , /sO/에서 약  $3.8 \times 10^5$ , /ts/에서 약  $8.1 \times 10^5$  로 나타났다. 결과적으로 k를 사용하지 않는 경우에 비해 사용한 경우의 오차 평가 값이 /kI/에서 약93.5%, /sO/에서 약51.9%, /ts/에서 약36.7% 정도 감소한 것으로 나타났다. 그리고 이번 실험에서 최대 29개의 신호를 사용한 경우, 0~1.13kHz의 낮은 주파수 신호에서 근사곡선의 발산하는 현상이 가끔 보여지는데, 이것은 낮은 주파수 대역에서는 주파수 신호의 진폭이 급격하게 변화하기 때문에 발생하는 것으로 추정된다.

#### 4. 근사합성

연속음성에서 TSIUVC를 근사합성 함에 있어서 주파수신호(k)를 사용하지 않은 경우와 k를 0~10개 사용한 경우의 근사합성 파형을 그림 4에 나타내었다. 그림 4에서 (a)는 연속음성에서 추출한 TSIUVC 파형이고 (b)는 k=0인 경우의 근사합성파형, (c)~(l)은 각 k=1~10을 사용하였을 경우의 근사합성 파형을 나타내었다. k=0인 경우에는 원래의 TSIUVC 파형에 근접한 파형을 얻을 수 없었으며, 파형의 형태로 보아 비교적 높거나 낮은 주파수를 적절하게 재생하지 못한다는 것을 알 수 있다. 이에 반해 주파수신호(k)를 사용한 경우에는 k의 수를 늘릴수록 파형의 일그러짐이 개선되는 것을 알 수 있었으며, 천이구간의 파형은 적은 수의 k로도 적절하게 근사합성할 수 있었다. 그러나 무성자음부의 파형은 비교적 많은 수의 k를 사용하여야만 원래에 가까운 근사합성 파형을 얻을 수 있다. 특히, 천이구간의 파형을 원래의 파형에 가깝게 재생하기 위해서는 5개 이상의 k를 사용하여야 한다는 사실과  $k \geq 10$ 에서는 파형의 일그러짐이 개선되는 정도가 둔화된다는 것을 알 수 있었다. 이 방법을 낮은 bit rate의 음성부호화 방식에 적용하고자 한다면 단순히 k를 늘리는 것보다 적절한 수의 k를 사용하는 것이 유리하다고 생각된다.

일반적으로 음성부호화 방식에서의 음질 평가에 사용하는 객관적인 평가 척도인 SNRseg와 주관적인

평가 척도인 MOS(Mean Opinion Score) 평가를 동시에 수행한다. 본 논문에서 제시한 방법은 V/S/TSIUVC 형태의 새로운 음성부호화 방식에 적용할 수 있는 근사합성의 방법론을 제공하고, 파형의 일그러짐을 제어함으로써 음성부호화 방식의 SNRseg와 MOS를 개선할 수 있을 것으로 기대된다.

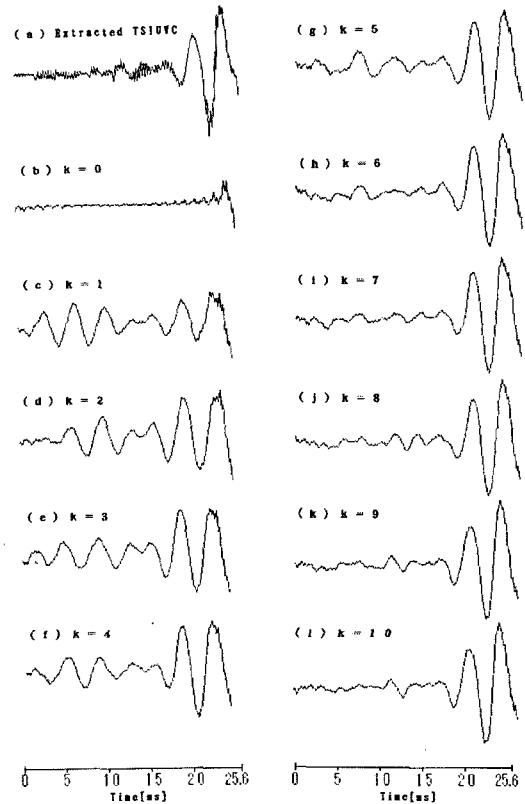


그림 4. TSIUVC 근사합성 파형

#### 5. 결론

유성음/무성음(Voiced/Unvoiced)의 음성부호화 방식에서는 프레임 안에 유성음과 무성음이 같이 존재하는 프레임의 경우, 이를 무성음원과 유성음원의 어느 한쪽의 음원을 선택하여 음성신호를 재생함으로써 음질저하의 요인으로 작용한다. 즉, 유성음과 무성음이 같이 존재하는 프레임의 경우에 나타나는 TSIUVC는 유성음도 무성음도 아닌 과도기적인 음성신호이다. 때문에 TSIUVC를 유성음원 또는 무성음원 어느 한쪽의 음원으로 재생하기 곤란하다는 결론을 얻을 수 있

며, 이를 본 연구에서는 연속음성에서 추출한 TSIUVC를 FFT하여 얻은 주파수 신호에 최소 자승법(LMS)을 적용하여 오차신호가 최대인 위치에 있는 주파수신호(k)를 사용하여 파형의 일그러짐을 제어하는 방법을 제시하였다. 최소 자승법의 근사합성 과정에서 나타나는 오차신호를 적절히 사용함으로써 근사합성 파형의 일그러짐의 정도를 파악할 수 있었다. 이 방법은 새로운 형태의 V/S/TSIUVC 음성부호화 방식에 적용할 수 있을 것으로 기대하며, V/S/TSIUVC 음성부호화 방식에 적용하였을 경우의 음질을 SNRseg와 MOS로 정량화하는 과제가 남아 있다.

### 참고문헌

[1] CHONG KWAN UN, HYEONG HO LEE: "Voiced/Unvoiced/Silence Discrimination of Speech by Delta Modulation", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-28, No.4, August 1980

[2] HIDEFUMI KOBATAKE: "Optimization of Voiced/Unvoiced Decisions in Non stationary Noise Environments", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No.1, January 1987

[3] Juan M. López-Soler and Victoria Sánchez: " Linear inter-frame dependencies for very low bit-rate speech coding", Speech Communication, Volume 34, Issue 4, Pages 333-349, July 2001

[4] Weiran Lin, Soo Ngee Koh and Xiao Lin: "An 8.0-/8.4-kbps wide band speech coder based on mixed excitation linear prediction Signal Processing, Volume 81, Issue 7, Pages 1437-1448, July 2001

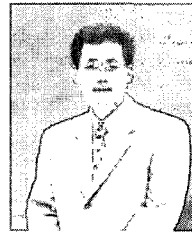
[5] Jürgen Schnitzler and Peter Vary: "Trends and perspectives in wide band speech coding Signal Processing, Volume 80, Issue 11, 1, Pages 2267-2281, November 2000

[6] Widrow B. and Hoff M. E.: "Adaptive Switching Circuit", IRE WESCON Conv. Rec, pp.96-104, June 2000

[7] Clark G.A., Mitra S.K. and Parker S.R.: "Block Implementation of Adaptive Digital Filters" J.Acoust. Soc. Am.57.1985

이 시 우(See-Woo Lee)

[정회원]



- 1987년 : 동국대학교 전자공학과 (공학사)
- 1990년 : 日本大學(Nihon Univ) 전자공학과 (공학석사)
- 1994년 : 日本大學(Nihon Univ) 전자공학과 (공학박사)
- 1994년 ~ 1998년 : (주)삼성전자 통신연구소/멀티미디어 연구소

• 1998년 ~ 현재 : 상명대학교 정보통신학부 교수

<관심분야>

음성신호처리, 음주판독시스템, 유무선통신