

# 최소 자승법을 이용한 TSIUVC 근사합성법에 관한 연구

이 시 우<sup>†</sup>

## 요 약

유성음원과 무성음원을 사용하는 음성부호화 방식에 있어서, 같은 프레임 안에 모음과 무성자음이 있는 경우에 음성 파형에 일그러짐이 나타난다. 본 논문에서는 최소 자승법을 적용한 새로운 TSIUVC 근사합성법을 제시하였다. 이 방법에서 TSIUVC의 추출에는 zero crossing rate과 FIR-STREAK 필터의 잔차신호를 이용한 개별피치 추출 알고리즘을 사용하였다. 실험결과, 최소 자승법을 적용하여 양호한 TSIUVC 근사합성 파형을 얻을 수 있었다. 또한 최대 오차신호에 있어서의 주파수 신호가 일그러짐이 적은 근사합성 파형에 중요한 역할을 한다는 것이다. 이 방법은 음성합성, 음성분석, 새로운 Voiced/Silence/TSIUVC의 음성부호화 방식에 활용할 수 있을 것으로 기대된다.

## A Study on TSIUVC Approximate-Synthesis Method using Least Mean Square

See Woo Lee<sup>†</sup>

## ABSTRACT

In a speech coding system using excitation source of voiced and unvoiced, it would be involves a distortion of speech waveform in case coexist with a voiced and an unvoiced consonants in a frame. This paper present a new method of TSIUVC (Transition Segment Including UnVoiced Consonant) approximate-synthesis by using Least Mean Square. The TSIUVC extraction is based on a zero crossing rate and IPP (Individual Pitch Pulses) extraction algorithm using residual signal of FIR-STREAK Digital Filter. As a result, This method obtain a high quality approximation-synthesis waveform by using Least Mean Square. 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 speech synthesis.

**키워드 :** 음성신호처리(Speech Signal Processing), 근사합성(Approximate-Synthesis)

### 1. 서 론

근래, 셀룰러폰과 PCS폰의 보급률은 세계적 수준에 이르렀으며 컴퓨터와 인터넷의 보급률 또한 급속도로 성장하고 있는 단계이다. 아울러, 통신의 기초 기술력은 외국 기술의 의존도가 높다고 할 수 있으나 통신장비의 제조 기술력은 상당한 수준에 도달하였다고 해도 과언이 아닐 것이다.

유선통신의 회선 이용자가 급증함에 따라 통신회선의 용량을 불가피하게 증설해야하는 경우에 물리적으로 통신용량을 늘리는 것 보다 신호를 압축/복원하는 음성부호화 방식에 의하여 통신용량을 늘리는 것이 보다 경제적이고 효과적일 것이다.

음성부호화 방식에서 음성신호를 처리하기 위해서는 일

반적으로 수십ms의 프레임으로 분할하여 처리한다. 특히, 유성음원과 무성음원으로 음성신호를 재생하는 방식에서는 피치정보나 프레임의 유성음(Voiced)/무성음(Unvoiced) 선택정보가 중요한 파라미터가 될 것이다. 음성신호를 유성음(Voiced)/무성음(Unvoiced) 혹은 유성음(Voiced)/무성음(Unvoiced)/무음(Silence)과 같은 선택정보에 의해서 유성음원과 무성음원을 구동하여 음성신호를 재생하는 방식 [1-5]에서는 음성신호를 수십ms의 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 일반적으로 사용하여 음성신호를 재생하는데 문제점이 있다. 또한 음성신호는 연속적이며 전후의 모음과 자음의 결합에 의하여 과도기적으로 변

<sup>†</sup> 정 회 원 : 상명대학교 컴퓨터정보통신공학부 교수  
논문접수 : 2001년 7월 5일, 심사완료 : 2002년 2월 8일

화하게 되는 천이구간내의 음성신호를 유성음원 또는 무성음원으로 재생하는 것도 한계가 있다고 볼 수 있다. 이러한 문제점을 해결하는 방법으로는 유성음과 무성자음이 같은 프레임에 존재하지 않도록 프레임의 길이를 동적으로 할당하는 것도 고려해 볼 수 있으나, 이것은 디지털 신호처리의 특성상 상당히 어려운 처리과정이라 생각된다.

본 연구에서는 길이가 일정한 프레임에서 연속음성을 처리함에 있어서 유성음, 무음, TSIUVC가 같은 프레임에서 존재하지 않도록 유성음부, 무음부, TSIUVC부를 탐색/추출한 다음 프레임을 재구성하고, 최소 자승법(Least Mean Square)에 의하여 TSIUVC 재생에 유효한 주파수 신호를 탐색/추출하여 근사합성하는 방법에 관하여 기술하고자 한다. 이 방법은 특성을 달리하는 유성음과 TSIUVC를 특성에 맞게 각기 다른 방법으로 음성신호를 처리하는 V/S/TSIUVC 음성부호화 방식에 응용하기 위한 방법이다.

**2. 최소 자승법에 의한 근사합성 모델**

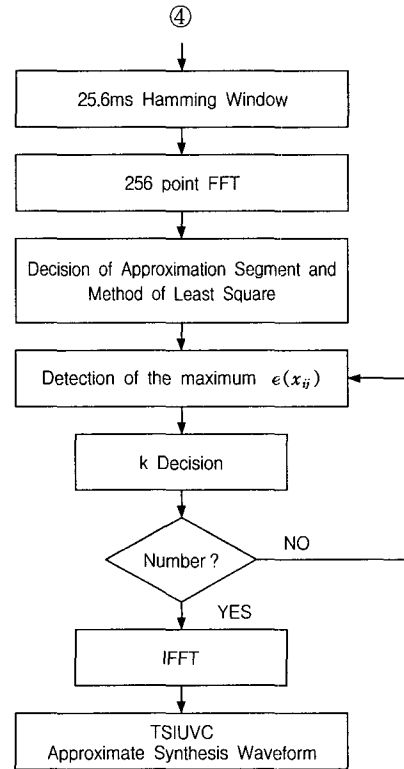
TSIUVC 근사합성은 음성부호화 방식에 있어서의 재생에 해당하는 처리과정이다. 이러한 처리과정에서 사용하는 파라미터의 수와 할당 bit수에 의하여 구현하고자 하는 음성부호화 방식의 bit rate에 많은 영향을 미치게 된다. 이번 실험에서는 낮은 bit rate의 V/S/TSIUVC 음성부호화 방식을 새롭게 구현하기 위한 구체적인 부호화 bit의 할당보다는 일그러짐은 적은 TSIUVC의 근사합성 과정을 재생하는 방법에 연구의 초점을 맞추었다.

기존의 Voiced/Unvoiced 형태의 음성신호처리[6-8] 방식에서는 음성신호를 Clustering하여 생성된 프레임 안에 유성음 혹은 무성음이 같이 존재할 수 있으며, 이러한 경우에 TSIUVC를 유성음원 혹은 무성음원의 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 재생하는데는 한계가 있다.

최소 자승법을 적용한 TSIUVC 근사합성의 방법(Approximate-Synthesis Method)을 (그림 1)에 제시하였다. 10kHz로 표본화된 연속음성신호에서 추출한 TSIUVC를 스펙트럼 상에서 신호처리하기 위해 256 point Hamming Window와 FFT를 사용하였다. 따라서 25.6ms 프레임에 256개의 신호가 존재하며, 이 주파수 신호에 최소 자승법을 적용하게 된다. 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 달라지는데, 이에 관한 실험데이터는 6절에서 기술하고자 한다. 아울러, 3절에서는 연속음성신호에서 TSIUVC를 탐색하고 추출하는 방법, 4절에서는 최소 자승법에 의한 신호처리, 5절에서는 오차신호에 의한 보정방법에 관하여 기술하고자 한다.



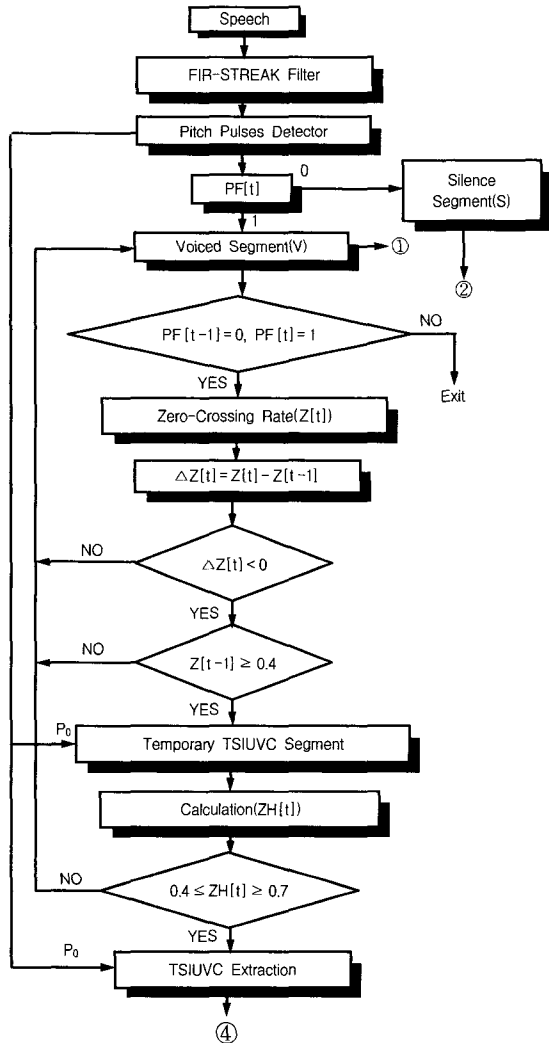
(그림 1) 최소 자승법을 이용한 TSIUVC 근사합성법

**3. 연속음성신호에서 TSIUVC의 추출**

TSIUVC에 최소 자승법을 적용하기 위해 연속음성신호에서 TSIUVC를 탐색하고 추출하여야 한다. 이를 위해 연속음성에서 유성음, 무음, TSIUVC를 판독할 수 있는 파라미터가 필요한데, 이에 적절한 파라미터로 피치정보와 zero crossing rate이 있다. 피치정보를 추출하는 방법에는 프레임 단위로 정규화된 피치정보를 추출하는 방법[9-11]과 프레임 단위로 정규화하지 않고 각 프레임마다 여러 피치펄스를 추출하는 방법이 있다[12]. 전자는 프레임마다 하나의 피치정보를 추출할 수 있기 때문에 적은 bit의 정보전송으로 음성신호를 재생할 때 유효한 방법이다. 반면, 후자는 여러 피치펄스를 추출하기 때문에 bit의 정보전송의 측면에서는 불리하나 음성신호의 추이에 따라서 피치정보를 추출할 수 있다는 장점이 있다. 따라서, TSIUVC와 같이 모음과 자음의 결합에 의하여 비교적 불규칙적으로 변화하는 음성신호의 추이에 따라서 피치정보를 추출할 수 있는 후자의 방법이 TSIUVC의 위치를 탐색·하는데 유

효할 것이다.

연속음성신호에서 TSIUVC를 탐색·추출하기 위해서는 천이구간과 무성자음이 갖는 신호의 특징을 분석하여야 한다. 우선, 음성신호가 없는 무음 구간에서 무성자음 구간으로 변위하는 부분에서는 zero crossing rate이 상승하는 현상이 나타나며 피치펄스 정보를 갖지 않는다. 아울러, 무성자음 구간에서 천이구간으로 변위하는 부분에서는 zero crossing rate이 감소하는 현상이 나타나며, 첫 번째 피치펄스 정보가 나타나는 유성음의 시작위치를 TSIUVC의 종료위치로 간주할 수 있다. 이와 같은 특징을 고려하여 연속음성에서 TSIUVC를 자동으로 탐색하여 추출하는 방법을 (그림 2)에 나타냈다.



(그림 2) TSIUVC의 탐색과 추출법

(그림 2)를 간략히 설명하면, 우선 음성신호를 3.4kHz LPF로 주파수 대역을 제한한 다음 10kHz, 12bit로 표본화 및 양자화하고, 프레임의 길이가 25.6ms인 음성신호를 FIR 필터와 STREAK필터를 혼합한 형태의 FIR-STREAK 디

지털 필터로 처리하여 (그림 3)와 같은 잔차신호를 얻는다. 이 잔차신호에는 잡음성 잔차신호와 펄스성 잔차신호( $R_p$ )가 있는데, 개별 피치 추출법[12]에 의하여 잡음성 잔차신호로부터 펄스성 잔차신호를 추출하고 보완처리하여 피치펄스의 위치를 추정한다. 이러한 피치펄스는 프레임 안에 여러 개의 피치펄스가 있으며, 정규화된 피치정보가 아니라 음성파형의 추이에 따라서 피치의 위치와 개수가 변하게 된다. 또한, 피치정보가 유성음에만 존재하고 무성자음과 무음에는 존재하지 않기 때문에 프레임 안에 개별 피치정보가 한개라도 존재하면 프레임 안의 음성신호를 유성음( $V : PF[t] = 1$ )으로 간주하고, 그렇지 않으면 무음( $S : PF[t] = 0$ )으로 간주한다. 이렇듯 프레임 내의 음성신호를 무음(Silence), 유성음부(Voiced), 그리고 본 연구에 의한 방법으로 TSIUVC부로 판독하여, 이들 신호들이 각기 독립적으로 존재하도록 프레임을 재구성함으로써 Voiced/Silence/TSIUVC 형태의 음성부호화 방식에 적용할 수 있다.

무성자음에는 피치정보는 없으나 잡음성 신호인 까닭에 높은 zero crossing rate을 나타낼 것이다. 또한 음성학적으로 무성자음 후에 유성음이 나타나는 까닭에 무성자음이 포함된 프레임의 다음에 지속되는 프레임의 zero crossing rate은 낮은 값을 갖게 될 것이다. 따라서, 유성음으로 판단된 프레임과 다음 프레임간의 zero crossing rate( $Z[t]$ ,  $t$ : 프레임 번호)의 차 ( $\Delta Z[t]$ ,  $t$ : 프레임 번호)가  $\Delta Z[t] < 0$ 인 경우에 TSIUVC가 존재하는 프레임일 가능성이 높을 것이다. 또한,  $\Delta Z[t] < 0$ 인 경우에 있어서 이전 프레임이  $Z[t-1] \geq 0.4$ 의 조건을 만족하면 현재 분석중인 프레임에 TSIUVC가 존재한다고 가정하였다. 특히, TSIUVC로 판단된 프레임의 전반부의 zero crossing rate을 재차 측정함으로써 TSIUVC 탐색·추출의 신뢰성을 높이고자 하였다. 실험결과에 의하면 TSIUVC 전반부의 zero crossing rate( $ZH[t]$ )은  $0.4 \leq ZH[t] \leq 0.7$  범위 내에 있는 것을 알 수 있었다.

이와 같은 조건에서 남여 9명의 연속음성(73문장, 모음 수: 609개, 무성자음수: 195개)에서 본래 TSIUVC가 존재함에도 불구하고 추출되지 않았을 경우( $b_j$ )와 본래의 TSIUVC가 존재하지 않는데도 불구하고 추출된 경우( $c_j$ )를 TSIUVC 추출 오류로 규정한 다음 식에 의하여 TSIUVC 추출률을 구하였다.

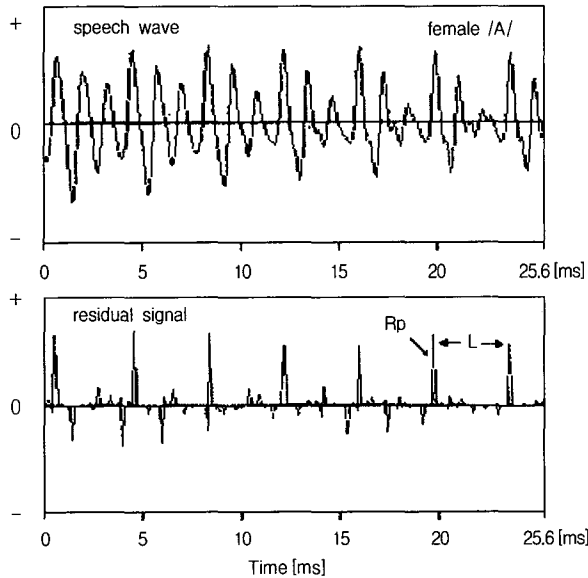
$$R_{TSIUVC} = \frac{\sum_{j=1}^m [a_j - (b_j + c_j)]}{\sum_{j=1}^m a_j} \quad (1)$$

$a_j$ : 실제 관찰된 TSIUVC 수  $m$ : 음성샘플 수

실험결과, TSIUVC 추출율은 남자음성에서 96.2%, 여자음성에서 91%를 얻었다.

zero crossing rate 정보와 피치정보를 이용하여 추출한

TSIUVC신호를 FFT하여 주파수상의 신호로 변환하면, 이때 주파수 신호의 주파수 간격이  $\Delta f = 39.0625\text{Hz}$ 이기 때문에 3.4kHz 주파수 대역에 87개의 주파수 신호가 존재하게 된다. 이 87개의 주파수 신호에 최소 자승법을 적용하는 방법에 관해서는 다음절에서 기술하고자 한다.



(그림 3) 개별 피치 펄스

4. 최소 자승법에 의한 신호처리

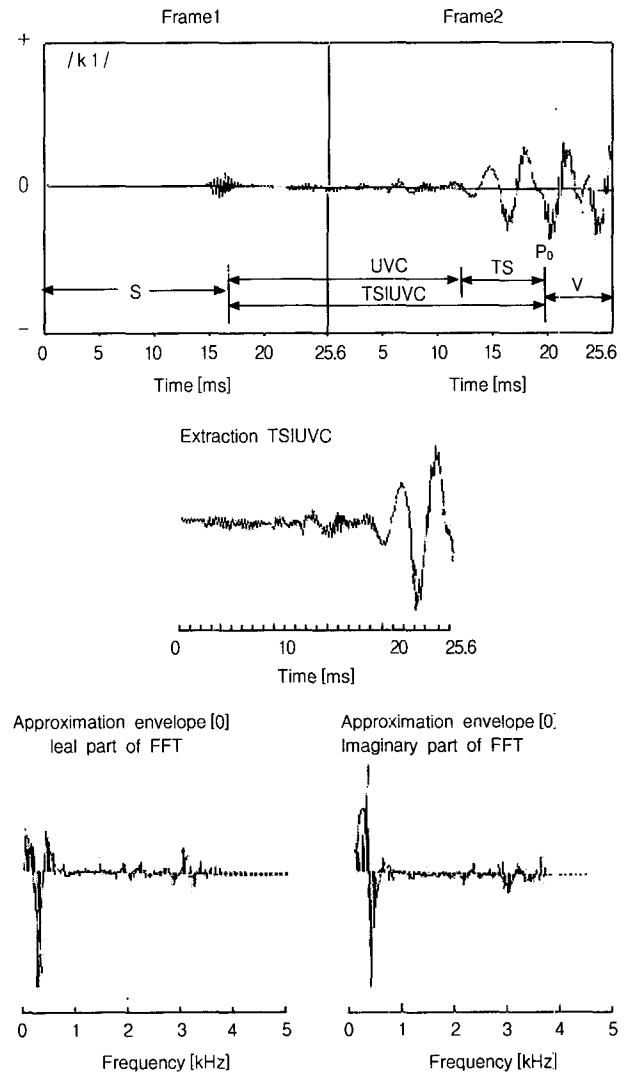
최소 자승법의 대표적인 적응 알고리즘인 Widrow-Hoff의 LMS(Least Mean Square)는 분석하고자 하는 파라미터가 시간과 함께 변동한다는 가정하에서 추정한다[13]. 이때, 파라미터 변화의 추이에 따라서 LMS 추정속도가 급격히 떨어지는 경우가 있는데, 이것을 보완하기 위하여 고안된 방법이 재귀 최소 자승법(Recursive Least Square)이다. 이 방법은 분석하고자 하는 파라미터가 시간과 함께 변동하지 않는 것으로 간주하여 분석하는 기법으로 파라미터를 안정적으로 분석할 수 있으나 파라미터의 계산 량이 증가하는 것으로 알려져 있다[14]. 이 문제를 Clark. G. A는 입력한 데이터를 블록으로 처리하여 신호처리 속도의 향상을 꾀하였다[15].

본 연구에서는 연속음성에서 추출한 TSIUVC를 FFT하여 얻은 주파수 신호를 블록화한 후 최소 자승법을 적용하고자 하는데, 이때 신호의 블록화의 의미는 신호처리 속도의 향상보다는

급격히 변화하는 TSIUVC 신호의 특성을 고려하여 적절한 근사신호를 얻기 위하여 블록화 하였다. 연속음성을 프레임처리 할 경우에 (그림 4)의 상부와 같이 Frame1과 Frame2에 무성자음과 천이구간 그리고 유성음 부분이 있는 것을 알 수 있다. 이러한 경우에 TSIUVC 추출 알고리

즘에 의하여 추출한 TSIUVC를 (그림 4)의 중앙에 나타내었다. (그림 4)의 하단은 5[kHz] 주파수 대역에서의 real part와 imaginary part의 주파수 신호에 대한 근사곡선을 나타낸 것이다.

TSIUVC 신호를 FFT하면 3.4kHz 주파수 대역에 존재하는 87개 주파수신호가 존재하게 되며, 이들 신호에는 천이구간(TS : Transition Segment)과 무성자음(UnVoiced Consonant) 신호의 특성을 갖고 있다.



(그림 4) TSIUVC 추출과 최소 자승법에 의한 신호처리

87개의 주파수 신호  $x_{ij}$ 를 다음 식과 같이 블록화한다고 하자.

$$\begin{aligned}
 x_{ij} = & (x_{00}, x_{01}, \dots, x_{0M}) \\
 & + (x_{10}, x_{11}, \dots, x_{1M}) \\
 & \dots \dots \dots \\
 & + (x_{m0}, x_{m1}, \dots, x_{mM})
 \end{aligned}
 \tag{2}$$

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

$$y(x) = a_{00} + a_{01}x + a_{02}x^2 + \dots + a_{0M}x^n + a_{10} + a_{11}x + a_{12}x^2 + \dots + a_{1M}x^n + \dots + a_{m0} + a_{m1}x + a_{m2}x^2 + \dots + a_{mM}x^n \quad (3)$$

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 될 것이다.

$$e(x_{ij}) = y(x_{ij}) - f_{ij} \quad (4)$$

그리고, 식 (3)의 계수  $a_{ij}$ 는 식 (4)의 오차신호 평가값(A)이 최소가 되도록  $a_{ij}$ 에 대하여 편미분하여 얻어지는 연립 방정식으로부터 구한다.

$$A = \sum_{i=0}^m \sum_{j=0}^M e^2(x_{ij}) \quad (5)$$

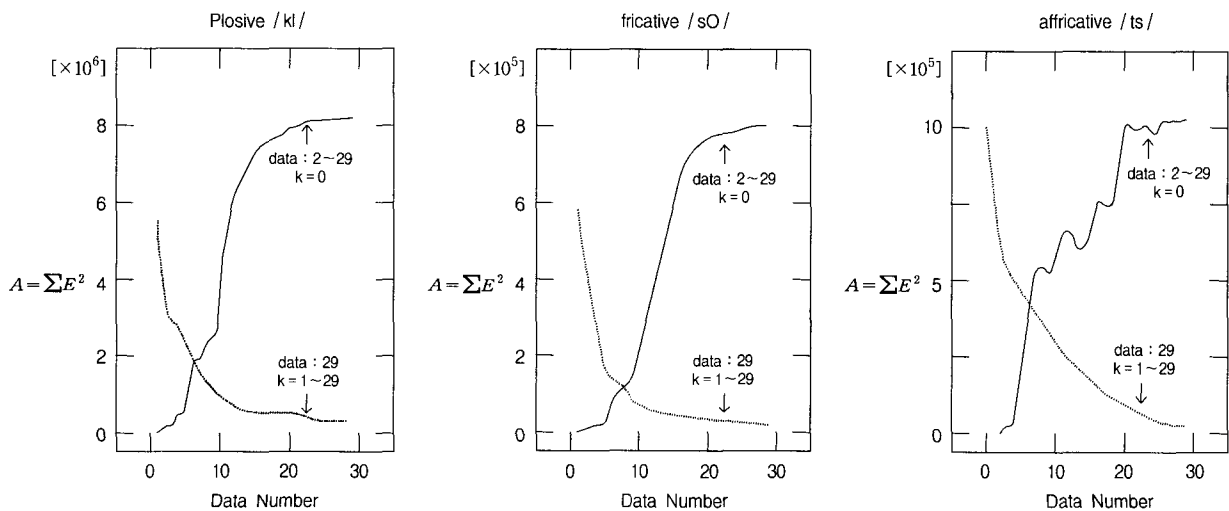
아울러, 계수  $a_{ij}$ 를 식 (3)에 대입함으로써 블록화 한 신호의 근사신호를 구할 수 있는데, 이때 오차신호가  $e = 0$ 이라면 측정신호( $f_{ij}$ )에 대하여 오차 없이 근사신호( $y(x_{ij})$ )를 얻은 결과로서, 오차신호( $e(x_{ij})$ )에는 근사신호를 보정하기 위한 정보가 포함되어 있다고 볼 수 없다. 만약,  $e \neq 0$ 이라면 오차신호( $e(x_{ij})$ )에는 근사신호를 보정할 수 있는 정보가 포함되어 있다고 볼 수 있다. 이것은, 최소 자승법에 의하여 구한 오차신호가 크다고 가정하면 근사신호를 보정하기 위한 정보가 오차신호에 많이 포함되어 있다고 가정할 수 있으며, 이 오차신호의 정보를 이용하여 오차신

호의 평가값(A)을 제어할 수 있을 것으로 생각된다. 결과적으로 최소 자승법에 의하여 TSIUVC의 신호를 처리하는 과정에서 발생하는 오차신호를 사용하여 TSIUVC 근사합성 파형의 일그러짐을 보상할 수 있을 것으로 생각된다.

### 5. 오차신호에 의한 보정

TSIUVC를 FFT하여 얻은 3.4kHz 주파수 대역에 존재하는 87개 주파수신호에 있어서 천이구간의 신호는 비교적 낮은 주파수에 분포하나, 무성자음의 경우는 높은 주파수 대역에 분포하고 있다. 따라서, 87개 주파수신호에 최소 자승법을 적용하여 오차신호의 평가값을 추정하는 실험을 하였다. 실험에 의하면, 낮은 차수의 근사 다항식으로 추정하면 연립방정식의 계수행렬이 감소하여 적은 계산량으로 처리할 수 있으며, 근사 다항식의 차수  $n \geq 3$ 에서는 거의 같은 근사값을 얻을 수 있었다. 때문에 본 연구에서는 근사 다항식을  $n = 3$ 으로 하고 최소 자승법에 의하여 측정치에 대한 근사 값을 구한 후 이들 신호들을 연결하면 (그림 4)의 하부에 있는 형태의 근사곡선을 구할 수 있다. 특히, 근사 곡선을 얻기 위해서는 최소한 2개 이상의 주파수 신호가 필요하며, TSIUVC 신호의 특성을 고려하여 최소 자승법에 적용하는 데이터 수는 최대 29개로 하였다. 근사 신호의 수를 최소 2개에서 최대 29개로 하였을 경우에 분할 가능한 블록 수는 최소 3블록에서 최대 43블록이 된다. 여기에서 최소 자승법을 적용하는 주파수 신호의 수(데이터 수) 또는 블록수가 클수록 TSIUVC 신호의 재생에 필요한 파라미터가 적어져 낮은 bit rate의 음성부호화 방식의 구현에 유리할 것이다.

연속음성에서 자동 추출한 195개의 TSIUVC 신호를 사용하여 오차의 평가 값을 측정하였는데, 한 예로 (그림 5)



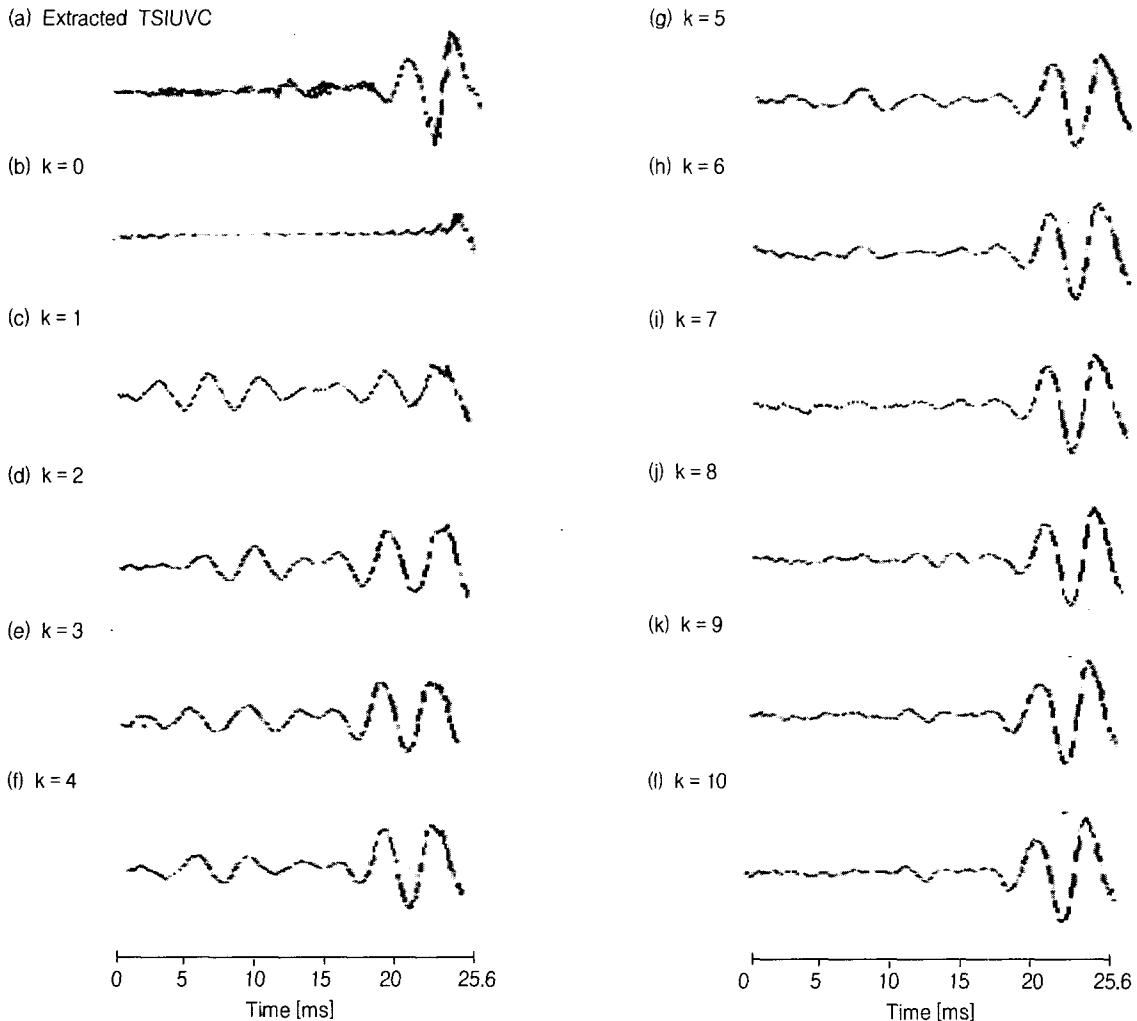
(그림 5) 최소 자승법의 오차 평가

에는 무성 파열자음/*kI*/, 무성 마찰자음/*sO*/, 무성 파찰자음/*ts*/이 포함된 TSIUVC의 오차 평가 값을 나타냈다. 여기에서 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 *k*를 사용하는 개수에 따라서 변화하는 오차의 평가 값을 알아보기 위하여 29개의 TSIUVC 주파수 신호에 대하여 *k*를 1~29개 적용한 경우의 오차의 평가 값을 파선으로 나타냈다. 실험결과, *k*를 사용하지 않고 주파수 신호의 수를 증가시키에 따라서 오차의 평가 값이 증가하는 것을 알 수 있는데, 이것은 데이터의 수가 커질수록 측정신호와 근사신호의 차가 커지는 것을 나타내는 것이다. 반면에 *k*를 점차 늘릴수록 오차의 평가 값이 현저히 감소하는 것을 알 수 있다. 오차의 평가 값이 큰 경우 /*kI*/에서 약  $8.5 \times 10^5$ , /*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. 결국, *k*를 사용하지 않는 경우에 비해 사용한 경우의 오차 평가 값이 /*kI*/에서 약 93.5%, /*sO*/에서 약 51.9%, /*ts*/에서 약 36.7% 정도 감소한 것을 알 수 있었다. 그리고, 이번 실험에서 최대 29개의 신호를 사용한 경우, 특히 0~1.13kHz의 낮은 주파수 신호에서 근사곡선의 발산하는 현상이 가끔 보여지는데, 이것은 낮은 주파수 대역에서는 주파수 신호의 진폭이 급격하게 변화하기 때문에 발생하는 것으로 추정된다.

6. 근사합성

실제 TSIUVC 파형을 근사합성함에 있어서 주파수신호(*k*)를 사용하지 않은 경우와 *k*를 0~10개 사용한 경우의 근사합성 파형을 (그림 6)에 나타냈다. (그림 6)에서 (a)는 연속음성에서 추출한 TSIUVC 파형이고 (b)는 *k*=0인 경우의 근사합성파형, (c)~(l)은 각기 *k*=1~10을 사용하였을 경우의 근사합성 파형이다. 195개의 샘플 실험을 통하



(그림 6) TSIUVC 근사합성 파형

여  $k=0$ 인 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 주파수신호( $k$ )의 수를 늘릴수록 파형의 일그러짐이 개선되는 것을 알 수 있었으며, 천이구간의 파형은 적은 수의  $k$ 로도 적절하게 근사합성할 수 있으나 무성자음부의 파형은 비교적 많은 수의  $k$ 를 사용하여야만 원래에 가까운 근사합성 파형을 얻을 수 있다. 특히, 천이구간의 파형을 원래의 파형에 가깝게 재생하기 위해서는 5개 이상의  $k$ 를 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  $k$ 를 늘리는 것보다 적절한 수의  $k$ 를 사용하는 것이 유리하다고 볼 수 있다.

음성부호화 방식에서의 음질 평가는 주로 객관적인 SNRseg 평가와 주관적인 MOS(Mean Opinion Score) 평가를 동시에 수행하는 것이 일반적이다. 일반적으로 파형의 일그러짐을 개선함으로써 SNRseg를 개선할 수 있으며, SNRseg의 개선은 MOS 평가를 높이는 효과를 얻을 수 있다. 특별한 경우를 제외하고는 SNRseg가 개선되면 MOS 평가도 좋아지는 것을 고려하면 본 논문에서 제시하는 방법을 V/S/TSIUVC 음성부호화 방식에 적용함으로써 SNRseg와 MOS를 개선할 수 있을 것으로 기대되며, 이를 정량화 하는 과제가 남아있다.

## 7. 결 론

Voiced/Unvoiced의 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 있으며, 이를 본 연구에서는 연속음성에서 TSIUVC를 탐색·추출한 다음 프레임 안에 음성신호가 유성음/무음/TSIUVC가 되도록 프레임을 재구성한 다음, TSIUVC 신호를 FFT하여 얻은 주파수 신호에 최소 자승법을 적용하여 오차신호가 최대인 위치에 있는 주파수신호( $k$ )를 사용함으로써 원래의 파형에 근접한 근사합성 파형을 얻을 수 있었다. 이때,  $k$ 의 수를 늘릴수록 파형의 일그러짐이 개선되는데, 특히 천이구간의 경우에는  $k \leq 5$  조건으로 원래의 파형에 근접한 근사합성 파형을 얻을 수 있는 반면, 무성자음의

경우에는  $k > 5$  조건에서 양호한 근사합성 파형을 얻을 수 있었다.

본 연구는 연속음성에서 흔히 볼 수 있는 TSIUVC와 같은 과도기적인 음성 파형을 재생하는데 유효한 주파수 신호를 최소 자승법에 의하여 탐색하고 이러한 주파수 신호가 양호한 근사합성 파형을 얻는데 유효한 파라미터임을 입증하였다. 향후, 본 논문에서 제시한 방법을 V/S/TSIUVC 음성부호화 방식에 적용하여 개선되는 음질의 정도를 SNRseg와 MOS로 정량화 하고자 한다.

## 참 고 문 헌

- [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 Nonstationary Noise Environments," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.ASSP-35, No.1, January 1987.
- [3] 武田 昌一他, "殘差音源利用分析合成方式とマルチパルス法の基本特性の比較検討", 電子情報通信學會論文誌, Vol.J73-A, No.11, 1990.
- [4] 小澤 一範, 荒關 卓, "ピッチ情報を用いる9.6~4.8kbit/sマルチパルス音聲符號化方式", 電子情報通信學會論文誌, Vol. J72-D2, No.8, 1989.
- [5] 眞野 淳, 小澤 慎治, "LPC有聲音殘差のピッチ同期メルLSP分析合成方式", 電子情報通信學會論文誌, Vol.J71-A, No.3, 1988.
- [6] LEAH J. SIEGEL, ALAN C.BESSEY, "Voiced/ Unvoiced/Mixed Excitation Classification of Speech," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-30, No.3, June, 1982.
- [7] LEAH J. SIEGEL, "A Procedure for Using Pattern Classification Techniques to Obtain a Voiced/Unvoiced Classifier," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.ASSP-27, No.1, February, 1979.
- [8] SIEGFRIED G, KNORR, "Reliable Voiced/Unvoiced Decision," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol.ASSP-27, No.3, June, 1979.
- [9] 藤井 健作, "自己相關法による電話帶域音聲のピッチ抽出法", 電子情報通信學會 技術報告書, pp.87-65. 1987.
- [10] Chong Kwan Un, Shin-Chien Yang, "A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF," IEEE, Vol.ASSP-39, Feb, 1991.
- [11] Carol A. McGonegal, Lawrence R. Rabiner, Aaron E.

Rosenberg : "Subjective Evaluation of Pitch Detection Methods Using LPC Synthesized Speech," IEEE. Vol. ASSP-25, June, 1997.

- [12] 이시우, "FIR-STREAK 디지털 필터를 사용한 피치추출 방법에 관한 연구", 정보처리학회논문지, 제6권 제1호, pp.247-252, 1999.
- [13] Widrow B. and Hoff M. E., "Adaptive Switching Circuit," IRE WESCON Conv. Rec., pp.96-104, 1960.
- [14] 相良, 筋夫, 秋月, 影雄, "システム同定", 計測自動制御學會, 1981.
- [15] 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

e-mail : swlee@smuc.ac.kr

1987년 동국대학교 전자공학과 졸업(학사)

1990년 日本大學(Nihon Univ) 대학원  
전자공학과(공학석사)

1994년 日本大學(Nihon Univ) 대학원  
전자공학과(공학박사)

1994년~1995년 삼성전자 통신연구소

1995년~1997년 삼성전자 멀티미디어 연구소

1997년~1998년 삼성전자 정보통신본부

1998년~현재 상명대학교 컴퓨터정보통신공학부 정보통신전공  
교수

관심분야 : 음성신호처리, 음성부호화 방식, 멀티미디어통신 등