향상된 2.4kbps 하모닉 스토케스틱 여기 음성 부호화 방법

김종학, 신경진, 이인성
충북대학교 전파공학과

Enhanced 2.4kbps Harmonic Stochastic Excitation Coding

Jonghark Kim, KyungJin Shin, Insung Lee
Dept. of Radio Engineering, Chungbuk National Univ.

요 약

본 논문은 주파수 전이신호와 시간 전이 신호에 대해서 고조파 잡음 여기 방법과 시간 분리 여기 방법을 적용한 2.4kbps 음성부호화 방법을 제안한다. 혼합 여기 부호화 방법은 주기 신호와 비 주기 신호를 효과적으로 표현하기 위해 하모닉 잡음 모델을 사용한다. 혼합신호에 대한 잡음 성분은 캡스트럴 분석 방법을 사용함으로써 추출되고, AR(Autoregressive Model) 모델에 의해 표현된다. 시간 전이구간 신호에서의 모호한 음성을 효과적으로 제거하기 위한 또 다른 방법이 제안된다. 제안된 시간 분리 방법은 시간 에너지 변화정도를 관찰함으로써 전이 시점을 감지하고 다른 시간 길이를 가지는 두 블록으로 분리하여 분석한다. 시간 분리 방법은 분석을 위한 비대칭 윈도우와 합성에서의 위상 합성 방법을 포함한다. 제안된 방법을 사용한 2.4kbps 음성부호화 방법은 주관적 음질 평가에서 전이구간에서의 지각적 음질의 향상을 보여주었으며, 원본 음성 스펙트럼과의 고조파 비 매칭에 의한 윙윙거리는 기계적인 잡음을 감소시킨다.

1. 서 론

최근 저 전송률 음성부호화 시스템은 STC(Sinusoidal Excitation Coding)나 MBE(Multi Band Excitation Coding)에 바탕을 두고 있다[1][2]. 이러한 스펙트럴 영역 코더들은 저 전송률에서 주기적인 신호를 표현하는데 효과적인 것으로 알려져 있으며, 고음질을 제공하는 것으로 알려져 있다. 그러나, 하모닉 코딩은 전이구간에서의 비 주기 신호나 유/무성음 혼합신호를 표현하기에는 불충분한 모델이기 때문에 자연스런 음질을 만들어내기는 어렵다. 이러한 면에서, 새로운 음성부호화기들이 저 전송률에서 음질을 향상시키기 위해 제안되었다.

본 연구는 한국과학재단 산학협력연구지원(과제번호 9820703012)으로 수행되었음.

새로운 U.S. 연방 표준 음성부호화기인 2.4kbps MELP(Mixed Excited Linear Prediction)를 비롯한 HSX(Harmonic Stochastic Excitation)은 이러한 접근의 예로 보여질 수 있다. 이러한 음성 부호화기들은 하모닉 및 스토케스틱 부호화의 혼합 기술을 사용한다. 그 혼합 방법은 유성음 세기 정도에 따라 정현파 및 잡음 발생정도를 달리하는 방법에 바탕을 둔다. 그러나, 이러한 방법은 고정 대역분석에 의한 스펙트럴 왜곡을 발생시키고 전 대역에 걸친 정확한 잡음 스펙트럴 정보 표현하기에는 불충분하다. 그러므로, 음성 스펙트로그램으로부터 추출된 잡음 성분을 나타내기 위한 새로운 혼합 방법이 필요로 된다. 또 다른 중요한 점은 저 전송률에서 하모닉 부호화는 상대적으로 긴 프레임 분석 요구에 의해 합성시에 선 에코(pre-echo)나 갑작스런 피치 변화에 적절한 적응능력을 갖지 못한다는 것이다. 선 에코(pre-echo)나 피치 왜곡은 고정 프레임 분석, 선형 보간 합성, 갑작스런 피치구간에 대한 단일분석 등에 기인한다. 이러한 효과는 윙윙거리는 소리나 기계적인 소리를 유발시킨다. 시간영역 상에서의 갑작스런 전이 구간은 onset 시간에서의 음소의 갑작스런 변화를 뜻하기 때문에 시간 전이로 구분될 수 있으며, 반면 주파수상에서의 유/무성음의 스펙트럴 혼합구간은 주파수 전이로 구분될 수 있다.

본 논문에서는 주파수 전이를 위한 캡스트럴-LPC 혼합 방법과 시간 전이를 위한 시간 분리 방법을 제안한다. 캡스트럴-LPC 방법은 원본 음성 신호의 스펙트럴값으로부터 캡스트럴 값을 구하고, 이를 이용하여 잡음 성분을 추출한 다음 LPC 분석과정으로 잡음 스펙트럴 포곡선을 예측하는 방법이다. 시간 분리 방법은 전이 시점을 감지하고, 비대칭 윈도우를 사용하여 양 분리구간에 대해 분석을 한다. 이러한 방법을 적용한 효과적인 구조를 유도하기 위해 우리는 2.4kbps 혼합 시간 분리 하모닉 스토케스틱 부호화기를 제안한다. 제안된 기본 구조는 2장에서 소개되며, 혼합 코딩은 3장에서 설명된다. 보다 상세히, 주기/비 주기 영역문제와 혼합 부호화 방법이 3장에 설명된다. 그런 다음, 시간 분리 부호화를 4장에 비트 할당을 포함한 양자화과정에 대한 설명은 5장에 설명된다.

2. MTHSX 의 기본구조

제안된 인코더와 디코더가 그림 1과 2에 나타나있다. 그림 1에 보여진 것처럼 인코더는 잔여신호를 얻기 위해 선형예측(Linear Prediction) 모델을 사용한다. 그런 다음, 잔여신호는 유성음과 무성음에 대해서 각각 하모닉 부호화기와 스토캐스틱 부호화기를 사용하여 부호화 된다. 하모닉 부호화기는 정현파 모델에 바탕을 둔다. 하모닉 부호화기의 주요특징은 선형 위상합성, 샘플 율 선형 보간, 파형 선형 보간 등으로 구성된다.

제안된 음성 부호화기에서, 하모닉 부호화기는 해밍윈도우의 DFT에 의해 얻어진 기본함수를 사용하여 스펙트럴 크기를 예측하고 동시에 스펙트럴 예측 오차를 최소화하도록 하는 A_l, w_0을 발견한다. 또한, 식 (3)은 IFFT 합성법을 사용하여 구현될 수 있다. 무성음 신호는 잡음 특성을 가지며, 피치 정보는 필요로 하지 않는다. 따라서, 피치 주기에 대한 비트 할당이 없는 스토케스틱 부호화방법이 적절한 방법이 될 수 있다. 그 스토케스틱 부호화는 CELP 부호화 방법과 유사하다. 새로이 제안된 부호화기는 혼합부호화 방법과 시간분리 부호화 방법을 포함한다. 혼합 부호화기는 하모닉 부호화기와 잡음 부호화기의 혼합형태로 제안된다. 하모닉 부호화기는 유성음에서의 부호화기와 같은 파라미터를 가지며, 잡음 부호화기는 비 주기 LSP 파라미터와 비주기 이득 값을 파라미터로 가진다. 반면, 시간 전이 프레임에 대한 시간 분리 부호화기는 전이시점, 2개의 하모닉 스펙트럴 크기 값들, 피치 값 파라미터들을 가진다. 상세한 설명은 다음 2개의 장에 걸쳐 설명된다.

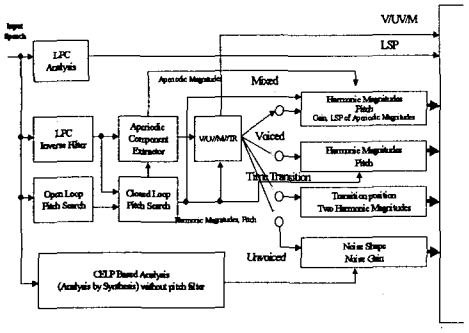


그림 1. MTHSX 인코더의 블록다이어그램

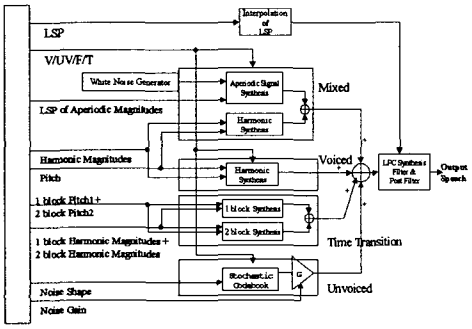


그림 2. MTHSX 디코더의 블록다이어그램

3. 유/무성음 혼합신호에 대한 하모닉 노이즈 코딩

음성 신호에서, 잡음 성분들은 하모닉 성분과 혼합된다. 특별히 단어의 중간이나 끝지점에서 발생하는 잡음 성분들은 대개 시간에 따라 부드럽게 변화하는 특성을 지닌다. 이러한 경우에는 시간영역에서의 변화보다 주파수 영역에서의 스펙트럴이 더욱 중요한 정보이기 때문에 그 스펙트럴 포곡선을 예측하는 것이 적절하다. 혼합신호를 주파수 전이로 간주하며, 주파수 영역 분석을 사용한다.

하모닉 부호화는 합성된 스펙트럼과 원본 스펙트럼 사이의 MSE(Mean Square Error)를 최소화 하도록 결정하는 폐루프 피치 검색에 의해 추출된 기본 주파수에 의존한다. 함께 계산된 스펙트럴 크기 값들은 하모닉 잡음 부호화에서의 하모닉 성분으로 간주된다. 잡음 성분들은 기본주파수를 사용함으로써 추출된다. 이러한 두 가지 성분들은 비 주기 에너지 단계에 따라 2개의 다른 구조를 적용함으로써 부호화된다.

3.1 캡스트럼-LPC 잡음 스펙트럴 추정

음성 신호는 여기 신호와 보컬 트랙의 임펄스 응답의 길쌈과정으로 표현된다. 구체적으로 여기신호는 의사 주기부분과 비 주기 부분으로 구성되며, 의사주기 부분은 피치주기의 글로탈 펄스 열을 의미하고, 비 주기 부분은 폐로부터의 공기흐름이나 입으로부터의 방사과정에 의한 잡음유사 신호를 의미한다.

$$s(t) = e(t) * h(t) = (v(t) + u(t)) * h(t) \quad (4)$$

$$c(t) = \text{IDFT}[\log |V(w) + U(w)| + \log |H(w)|]$$
$$= \text{IDFT}[\log |V(w) + U(w)|] + \text{IDFT}[\log |H(w)|] \quad (5)$$

여기서 $s(t)$은 음성 신호이며, $h(t)$은 보컬 트랙 시스템의 임펄스 응답이고, $e(t)$은 여기 신호를 뜻한다. $v(t)$은 여기신호의 의사 주기 부분이며, $u(t)$은 여기신호의 비 주기 부분이다. $S(w), U(w), V(w)$과 $H(w)$은 각각 $s(t), u(t), v(t)$과 $h(t)$의 푸리에 변환 값이다. 식 (5)에서 보여진 것처럼 큐프런시 영역에서 피치주기의 좌측 부분은 스펙트럴 포곡선을 가지는 보컬 트랙 응답에 의한 성분으로 분류될 수 있으며, 피치 주기의 오른쪽 큐프런시 영역 부분은 여기 신호 성분으로 분류될 수 있다. 특히 피치 주기에서의 피크주변의 값은 하모닉들이 기본주파수의 배수 주변에 집중되어 있기 때문에 하모닉 성분으로 간주 될 수 있다. 따라서, 피치주기에서의 피크 주변 캡스트럼이 리프팅되고 그림 4에서 보여지는 것처럼 로그 크기 스펙트럼으로 변환된다. 잡음성분 영역은 로그 스펙트럼의 음의 부분으로 정의된다. 이러한 과정은 순환 주기/비 주기 분해 방법과 유사하다[7]. 그러나, 이러한 잡음 추정 방법은 너무나 많은 FFT/IFFT를 적용하기 때문에 그 복잡도 면에서 저 전송률 부호화에 적용하기 어려우며, 또한 비 주기 성분 추출에 있어 작은 하모닉 왜곡에도 순환 FFT/IFFT 방법으로 인해 큰 왜곡을 발생시킨다. 이러한 문제를 해결하기 위해 순환 방법을 쓰지 않고 LPC 방법을 이용하여 그 잡음 스펙트럴 포곡선을 예측하는 방법을 적용한다. 잡음 성분에 대한 LPC 분석은 그림 5에 나타내었다. 그 과정은 원본 음성 신호에 대한 하모닉 스펙트럴 포곡선을 추정하는 과정과 유사하다.

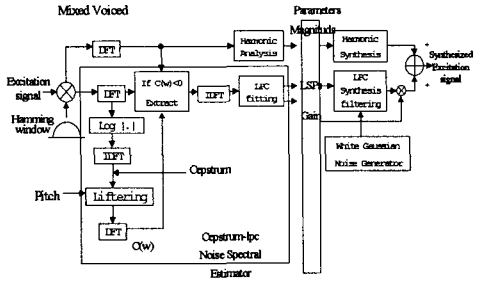
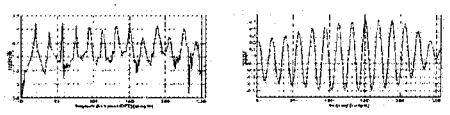
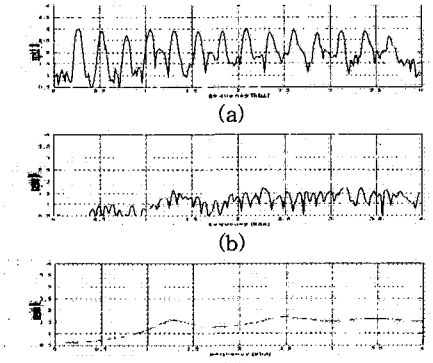


그림 3. 혼합 부호화기의 블록다이어그램



(a) (b)

그림 4. 유성음 프레임의 캡스트럼 분석. (a) 여기신호 로그 크기 스펙트럼. (b) 하모닉 성분의 로크 크기 스펙트럼.



(a)

(b)

(c)

그림 5. 비 주기 성분의 여기신호 및 합성. (a) 원본 여기신호 스펙트럼. (b) 비 주기신호의 LPC 고정된 주파수 응답. (c) 합성된 비 주기 신호의 스펙트럼.

3.2 잡음 성분의 코딩

부분 잡음 성분들은 스펙트럴 포곡선을 LP 파라미터로 고정하는 과정을 거친다. 그 LP 파라미터는 효율적인 양자화를 위해 LSP(Line Spectrum Pair) 파라미터로 변환된다. 변환된 LSP파라미터는 벡터양자화를 이용해 양자화 된다. 디코더에서의 합성 과정은 각 프레임사이의 위상 일치과정 없이 가우시안 백색잡음을 LP 필터링 시킴으로써 간단히 구현된다. 제안된 부호화기에 대한 LP모델의 차수는 6차를 적용하였다. LP파라미터들과 이득 파라미터들은 6, 3비트 양자화기에 의해 양자화 된다.

4. 시간영역 전이구간에 대한 시간 분리 코딩

시간 분리 부호화는 전이 프레임에서의 갑작스런 변화를 나타내기 위해 사용된다. 전이 시점은 대개 음성의 시작 지점에 있고, 전이 시점에 의해 분리된 근방 블록은 다른 특성을 지닌다. 그러므로, 양 블록을 분리하여 분석하는 과정이 적절하다.

4.1 전이구간 및 전이시점 결정

전이 시점 결정은 전이 시점의 좌측신호와 우측신호가 큰 에너지 차이를 지닌다는 개념에 바탕을 둔다. 그 결정방법은 그림 6에서 보여진 것처럼 좌/우측 에너지 비율 $E_{rate}(n)$을 정의함으로써 구현된다. 다시 말해, $E_{rate}(n)$의 최고 값에서의 n이 전이 시점으로 결정된다. 좌-우측 에너지 비율 값은 다음과 같다.

$$E_{\min}(n) = \min[\sum_{i=0}^{P} s^2(n+i), \sum_{i=0}^{P} s^2(n-i)]$$

$$E_{\max}(n) = \max[\sum_{i=0}^{P} s^2(n+i), \sum_{i=0}^{P} s^2(n-i)] \quad (6)$$

$$E_{rate}(n) = \left(\frac{E_{\max} - E_{\min}}{E_{\max}}\right)^2 \quad (7)$$

여기서 $s(n)$은 입력신호이고 P는 피치주기이다. $E_{rate}(n)$은 160샘플에 대해 계산되고 가장 큰 에너지 변화를 가지는 위치가 선택된다. 여기서, 피치주기 값이 피치주기에서의 피크들에 대한 영향을 감소시키기 위한 샘플 제한 개수 값으로 사용된다. 전이 프레임은 2개의 다음 제한 값에 의해 결정된다.

$$E_{rate}(n) > T_1$$
$$E_{\max}(n) - E_{\min}(n) > T_2 \quad (8)$$

여기서 제안된 부호화기에서 T_1, T_2은 각각 0.55, 1.5×10^6로 주어진다.

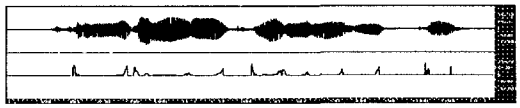


그림 6. 10^4에 의해 스케일된 좌우 에너지 비율 값의 결과

4.2 시간 분리 코딩

전이시점의 위치 값은 동일한 길이를 가진 4개의 블록으로 160샘플을 나누어 표현한다. 그리고, 각 중심점은 전이시점 위치 값에 의해 분리된다. 전이 시점은 256샘플의 분석 프레임에서 80, 112, 144, 176으로 정의되며, 2비트에 의해 디코더에 전달된다. 그 중앙 위치 값은 변형된 해밍윈도우가 신호의 에너지를 중앙 위치 값에 집중시키도록 적절히 시간 휨(Time Warping) 윈도우 함수를 사용한다. 그 윈도우 함수는 다음과 같이 나타난다.

$$w(c,n) = \begin{cases} w_h(c,n) & ;\ 0 \le c \le (N-1)/2 \\ w_h(128-c, 128-n) & ;\ (N-1)/2 \le c \le N-1 \\ 0 & ;\ \text{otherwise} \end{cases} \quad (9)$$

$$w_h(c,n) = 0.54 - 0.46\cos\left(2\pi \frac{f_w(c,n)}{N-1}\right)$$

$$f_w(c,n) = \frac{(N-1)}{2\log\left(\frac{N-1-c}{c}\right)} \log\left(1 + \frac{N-1-2c}{c} n\right) \quad ;\ c \le (N-1)/2 \quad (10)$$

여기서, c은 중앙 위치 값이며, N은 분석 프레임 번호이다. 각 블록에 대한 윈도우 샘플은 스펙트럴 크기 값들과 최종 피치 값을 예측하는 하모닉 분석에 대한 입력으로 사용된다. 스펙트럴 크기 값, 최종 피치 값, 전이시점 파라미터들은 합성과정을 위해 디코더에 전달되며, 합성과정 시 빠른 IFFT 합성방법을 이용하여 합성을 하게 된다. 이때, IFFT 합성과정에서 프레임간의 위상을 맞추기 위한 오프셋 조정 값은 그 값이 변하는 변이 값인 전이 시점 값에서의 위상을 맞추도록 조정되어야 한다. 그 오프셋 조정 값은 전이시점에 순환파형의 끝 지점을 위치시키도록 하는 값으로 계산된다. 이러한 합성과정의 결과로써, 그림 7 (e),(f),(g) 는 분석 중앙 위치 값에 따라 변화하는 시간 훰 윈도우를 사용한 시간 분리 부호화의 효과를 보여준다.

5. 비트 할당 및 실험 결과

제안된 MTHSX 부호화기는 20ms 프레임 길이와 미래 샘플 12ms를 사용하여 2.4kbps에서 구현되었다. MTHSX의 비트할당을 표 1에 나타내었다. 피치 2는 전 이프레임으로부터 차분 값을 사용하여 양자화된다. 2.4kbps MTHSX, 2.4kbps MELP 를 포함하는 주관적인 평가인 MOS 테스트가 제안된 부호화기의 성능을 측정하기 위해 수행되었다. MOS 테스트로부터 2.4kbps MTHSX 부호화기는 2.4kbps MELP 보다 더 좋은 음질을 나타내었다. 특히 제안된 MTHSX 부호화기는 여자음성에서 더욱큰 향상으 보였으며, 이는 큰 하모닉 구간과 잡음 성분을 지니는 여자음성에 대해 제안된 캡스트럼 LPC 잡음 부호화기로 인한 향상으로 볼 수 있다. 또한 그림 7으로부터 각 새로이 제안된 방법에 대한 향상된 결과를 도식적으로 관찰할 수 있다.

표 1. 2.4kbps MTHSX 부호화기의 비트할당

Parameter	Voiced	Mixed	Time Transition	Unvoiced
LSP	16			
V/UV/M	2			
Pitch	7	7	0	0
Magnitudes	23	14	0	0
Noise LSPs and Gains	0	9	0	0
Time Domain Shape	0	0	0	30
Transition Point & VQ Select	0	0	3	0
Frame1 Magnitudes & Pitch1	0	0	8+7(Pitch1) =15	0
Frmae2 Magnitudes & Pitch2	0	0	8+4(Pitch2) =12	0
Total	48/20ms			

표 2. 2.4kbps MTHSX 부호화기의 MOS 테스트

Classification	Girl	Man	Total
Original speech	4.47	4.61	4.54
8kbps CS-ACELP	3.86	4.01	3.94
2.4kbps MELP	2.52	3.49	3.00
2.4kbps MTHSX without proposed method	2.57	3.33	2.95
2.4kbps MTHSX with proposed method	2.94	3.44	3.19

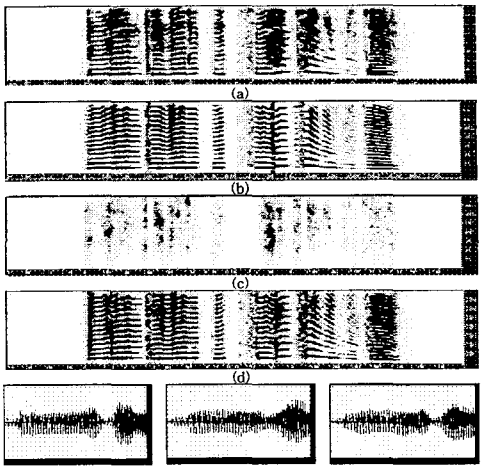


그림 7. 합성된 여기신호와 음성 신호의 비교. (a) 원본 여기신호. (b) 제안된 방법을 적용치 않은 합성 여기신호. (c) 캡스트럼-LPC 잡음 추정기를 사용한 합성 잡음 성분. (d) 새로운 하모닉 잡음 및 시간 분리 개념을 적용한 합성 여기 신호. (e) 시간 전이를 가지는 원본 파형. (f) 시간 분리 부호화를 적용하지 않은 합성 파형. (g) 시간 분리 부호화를 적용한 합성 파형.

6. 결 론

본 논문에서 혼합 잡음 신호와 시간 전이 신호에 대한 새로운 두가지 효율적인 부호화 방법을 제안하였다. 그 부호화 방법은 캡스트럼 LPC 잡음 스펙트럴 예측기와 전이시점 검축기, 시간 훰 윈도우를 사용한 분석방법등 새로운 기술을 포함하고 있다. 이러한 구조는 하모닉 부호화와 스토케스틱 부호화 방법에대해 접목되어 그 단점을 보완하도록 효과적으로 적용되었으며, 제안된 2.4kbps MTHSX 음성 부호화기는 향상된 음질을 갖는 혼합구조를 보여주었다.

참고문헌

[1] R. V. Cox, "Speech Coding Standards", *Speech Coding and Synthesis, Chapter 2*, W. B. Kleijn, and K. K. Paliwell Eds., Elsevier, 1995

[2] A. M. Kondoz, "Coding Strategies and Standards", *Digital Speech, Chapter 5*, John Wiley, 1994.

[3] A. V. McCree, K. Trung, E, B, George, T.P.Banwell and V. Viswanathan, " A 2.4kbit/s MELP Coder Candidate for the New U.S. Federal Standard", in *Proc IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol.1 ,pp 200-203,May 1996.

[4] C. Laflamme, R. Slami, R.Matmi, J-P. Adoul, "Harmonic-Stochastic Excitation (HSX) Speech Coding Below 4Kbbit/s", in *Proc IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol.1 ,pp 204-207,May 1996.

[5] R. J. McAulay, T. F. Quartieri, "Sinusoidal coding", *Speech Coding and Synthesis, Chapter 4*, W. B. Kleijn, and K. K. Paliwell Eds., Elsevier, 1995.

[6] Masayuki, Nishiguchi and Jun Matsumotor, "Harmonic and Noise Coding of LPC Residuals with Classified Vector Quantization"in *Proc IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.484-487, 1995.

[7] B. Yegnanarayana, Christophe d'Alessandro and Vassilis Darsinos, "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", *IEEE Transaction on Speech and Audio processing*, vol. 6, NO. 1,pp. 1-11,1998.