

가변전송률 음성부호화기 적용을 위한 음성활성도 측정 알고리즘 설계

정회원 김재원*

Design of Voice Activity Detection Algorithm for Variable Rate Speech Coders

Jae-won Kim* *Regular Member*

요약

디지털 이동통신 시스템에서 가장 빈번하게 발생하는 음성 서비스의 궁극적인 목표는 양호한 음성 품질과 높은 주파수 효율의 제공에 있다. 음성은 묵음 구간에 의하여 구분되어진 짧고 간헐적인 음성 에너지의 반복으로 표현 가능하며 실제 음성 통화중 활성 음성이 존재하는 구간은 약 40%, 나머지 60% 구간은 묵음 또는 상대방의 음성을 듣는 구간이다. 이 묵음 구간을 효율적으로 활용함에 의해 시스템의 스펙트럼 이득을 얻을 수 있다. 본 논문에서는 디지털 이동통신 시스템과 같이 다양하게 변화하는 주변 잡음 환경에서도 강건하게 동작 가능하여 10msec 프레임 크기를 갖는 음성부호화기에 적용 가능한 음성 활성도 측정 방안을 설계하였다. 설계된 알고리즘은 음성 에너지, 스펙트럼 분포, 영교차율, 그리고 LPC 잔여신호의 Peakiness 측정값을 이용하였다.

ABSTRACT

The goals of voice service in digital mobile communication systems are to provide high speech quality and high spectral efficiency. Speech can be modeled as short bursts of vocal energy separated by silence gaps. During typical conversations, talks-spurts comprise only 40% of each party speech and remaining 60% is silence. Communication systems can achieve spectral gain by disconnecting the users from the spectral resource during silence periods. This paper develops an efficient voice activity detection (VAD) algorithm to work under a digital mobile communication environments exhibiting dynamically varying background noise. The VAD uses a classification method by frame energy, spectral distribution, zero-crossing rate of speech samples, and peakiness measure of LPC residual signals.

1. 서론

디지털 이동통신 시스템 적용을 위하여 설계된 음성부호화기들은 양호한 음성 품질과 높은 스펙트럼 효율을 만족시켜야 한다^[1]. 이의 달성을 위해서는 낮은 평균전송률을 갖는 가변전송률 음성부호화기의 개발이 요구되어지는데^[2], 이는 새로운 코딩 모델의 설계와 아울러 활성 음성(Active Speech)이 존재하지 않는 묵음 구간의 효율적 활용에 의해서

가능하다^[3]. 묵음 구간에서는 신호의 Stationary한 특성에 의하여 적은 비트로서도 표현이 가능하여 절약된 트래픽 링크 자원을 다른 이용자에게 활용함에 의하여 시스템의 주파수 효율을 증대시킬 수 있다^[4]. 음성 활성도 측정은 활성 음성의 존재 유무를 판별하는 기능으로서 Acoustic 배경잡음이 존재하지 않는 환경에서는 음성신호의 에너지에 근거한 단순한 방법으로 높은 성능을 얻을 수 있지만 이동통신 환경과 같이 다양하게 변화하는 배경 잡음 하

* 경주대학교 컴퓨터전자공학부(jwkim@kyongju.ac.kr),
논문번호: K01102-0314, 접수일자: 2001년 3월 14일

에서는 활성 음성과 배경 잡음의 특성 분석에 근거한 세련된 음성 활성화 측정 방안이 요구되어진다^[5]. 본 논문에서는 디지털 이동통신 시스템에서 주파수 효율 향상을 위하여 요구되어지는 효율적인 VAD 알고리즘을 설계하였다. 2장에서는 음성 신호와 이동통신 배경잡음의 특징과 급격히 변화하는 배경잡음하에서의 음성 존재여부 판별의 문제점을 언급하고, 잡음 구간의 예측과 다수 임계치 판별에 근거한 VAD 알고리즘을 설명하였으며, 3장에서는 설계된 알고리즘의 성능 평가를 위하여 다양한 배경잡음 환경에서의 감지된 활성도와 Clipped Frame Rate를 측정하였으며, 그 결과를 동일한 프레임 크기를 갖는 ITU-T G.729B VAD^[6] 성능과 비교 검토하였다. 결론에서는 설계된 알고리즘의 성능을 토대로 가변 전송률 음성부호화기의 음성 활성화 측정 방안으로 제안하고자 한다.

II. 설계된 VAD 알고리즘

설계된 음성활성도 측정 알고리즘의 구성은 음성을 구성하는 유/무성음과 이동통신 환경에서 발생 가능한 배경잡음의 특징을 간략히 언급하고, 음성과 배경잡음의 특징에 근거하여 효율적으로 동작하는 음성활성도 측정 방법에 대하여 기술하였다.

1. 음성과 배경잡음의 특징

음성 발생체계는 Sound Source Generation, Vocal Tract에 의한 Articulation, 입술을 통한 방사과정으로 요약되는데, 이러한 특성을 분석하기 위한 방법으로 시간에 따른 음성파형의 변화, 주파수 영역에 따른 에너지 분포, 공진에 의한 Formants 주파수 분포, 여기신호 형태등이 있으며 이를 토대로 유성음(Voiced Speech), 무성음(Unvoiced Speech), 그리고 묵음(Silence)으로 구분 가능하다^[7]. 유성음은 성대의 진동에 의해 발생한 주기성을 갖는 공기의 흐름이 성도를 통과하면서 발생하는 소리이며, 무성음은 성대의 진동없이 빠른 속도의 공기 흐름이 수축된 성도를 지나면서 생성된 소리이다. 그리고 묵음은 인간의 음성 발생체계에 의하여 생성된 소리가 존재하지 않는 구간으로 Stationary 특징을 가진다. 하지만 이동통신 통화 환경에서는 다양한 종류의 배경잡음에 의하여 음성과 묵음 구간의 특성이 영향을 받게 된다. 이러한 배경잡음은 두 가지 형태로 구분 가능한데, 첫째는 건물내의 통화환경에서 발생하는 비교적 낮은 에너지 레벨과 Stationary

한 특징을 갖는 Office Noise, 둘째는 차량 내부, 거리, 많은 사람이 모이는 대중 공간등과 같이 높은 에너지 레벨과 시간에 따른 특징의 변화가 큰 Dynamic Noise로 구분 가능하다. 전자는 단순한 방법에 의해서도 대처 가능하지만 후자는 실제 이동통신 환경에서 빈번히 발생하는 형태이며 파라미터의 특성이 광범위하게 변화하기 때문에 예측하기가 매우 까다롭다^[8]. 이의 제어를 위해서는 다양한 방법에 의한 많은 연구가 요구되어진다^[9,10]. 표 1은 활성음성과 이동통신 배경잡음에 대한 특징을 요약하였다.

표 1. 음성과 배경잡음의 일반적 특징

파라미터	프레임 에너지	저주파수 에너지분포	영교차율	PM 값
유성음	고	고	저	고
무성음	중-고	저	고	저-중
Dynamic 잡음	중-고	중-고	중	저-중
Office 잡음	저	중-고	중	저

2. 음성활성도 측정 알고리즘의 동작

본 논문에서 설계된 VAD 알고리즘의 전체적인 흐름은 그림 1에 나타난 바와 같이 입력음성 80샘플(10msec)를 수신하여 실제 음성의 존재 여부를 결정하는 프레임 단위의 음성 활성화 측정을 수행한다. 그 과정은 오프셋이 제거된 음성 신호 생성을 위하여 Pre-process 시킨 필터의 cut-off 주파수 만큼의 음성 대역 에너지, 1kHz 이하의 저주파 대역에서의 에너지 분포, 음성 신호의 영교차율을 계산하게 된다. 그리고 원래 음성 신호에서 LPC 상관계수가 제거된 신호를 이용한 Peakiness 측정값을 구하는 과정을 수행한다.

음성을 구성하는 유성음과 일부 무성음 구간은 큰 에너지 분포를 가지기 때문에 음성과 묵음 구간의 판별을 위하여 가장 기본적이고 효율적인 방법으로 프레임 에너지를 이용하게 된다. 식 (1)의 음성 대역의 프레임 에너지($E_f(n)$)는 음성 분석 과정에서 생성되는 음성 신호의 자기 상관계수 $R(0)$ 의 정규화된 값을 이용하여 다음과 같이 정의된다.

$$E_f(n) = 10 * \log_{10}[R(0)] \tag{1}$$

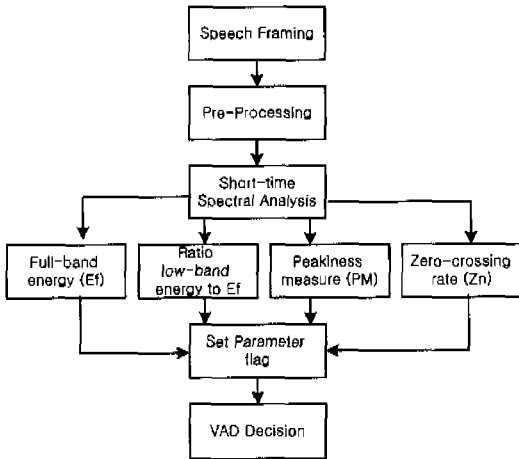


그림 1. 설계된 음성활성도 측정 알고리즘의 동작 흐름도

추출된 음성프레임 에너지와 다양한 배경잡음 레벨에서의 시뮬레이션을 통하여 추출된 에너지 임계치 값 ($T_e(B)$)에 근거하여 묵음 구간이라고 판단된 프레임들을 나타내기 위한 플래그 값 (f_{e_sil})은 식 (2)와 같이 정의된다.

$$f_{e_sil} = \begin{cases} 1, \dots E_f(n) < T_e(B) \\ 0, \dots otherwise \end{cases} \quad (2)$$

그러나 이동통신 통화환경에서는 설정된 Call들이 뿐만 아니라 동일한 Call상에서도 배경잡음이 다양하게 변화하기 때문에 입력 음성신호의 에너지 측정만으로는 정확한 음성 활성도 측정에 문제가 발생하게 되는데^[8,11], 이의 보안을 위하여 음성의 여기신호가 성문(Glottis)과 입술 사이의 Vocal Tract를 통과하면서 발생하는 공진특성에 의하여 결정되는 한 세트의 Formants를 분석함에 의하여 성능의 보장이 가능하다. 첫 번째 Formant 주파수에 많은 에너지 분포가 이루어지며, 이 주파수의 분포가 음성음의 경우 1kHz 이하에, 무성음의 경우 높은 주파수 대역에 분포하게 되어 배경잡음에 의하여 손상된 음성신호의 유/무성음 판별을 위하여 주파수에 따른 에너지 분포가 음성 활성도 측정의 파라미터로 활용 가능하다. 1kHz 이하의 주파수 대역에 분포된 에너지 계산은 식 (3)과 같이 정의된다.

$$E_1(n) = 10 * \log_{10} \left[\frac{1}{N} h^T R \right] \quad (3)$$

여기서 h는 1kHz에서 Cutoff 주파수를 갖는 FIR

필터의 임펄스 응답을 나타내며, T는 전치행렬, R은 13*13의 크기를 갖는 Toeplitz 자기상관 계수 행렬을 나타낸다. 그리고 음성 전체 대역에 대한 1kHz 이하의 대역에서의 에너지 분포는 $R_{lb} = E_1(n)/E_f(n)$ 와 같이 정의된다. 프레임 에너지에 대한 저주파 대역 에너지 분포를 통하여 음성음과 무성음을 배경잡음으로부터 구분하기 위한 플래그 값 (f_{lb})은 식 (4)와 같이 정의되어 진다.

$$f_{lb} = \begin{cases} 0, \dots (T_{lb1} \leq R_{lb} \leq T_{lb2}) \\ 1, \dots otherwise \end{cases} \quad (4)$$

여기서 T_{lb1} , T_{lb2} 는 무성음, 음성음 저주파 대역의 에너지 분포를 위한 실험적 임계치를 나타낸다. 그리고 오프셋이 제거된 음성신호의 인접 샘플간의 부호 변화 즉 영교차율은 음성 신호의 Spectral Distribution을 반영하는데, 단순한 연산 과정 만으로도 유용하게 활용될 수 있는 파라미터가 된다. 시간에 따른 진폭의 변화가 음성음 구간에서는 주기적이고 높은 진폭 레벨을 갖기 때문에 낮은 영교차율 값을 가지며, 낮은 신호레벨과 Random Noise와 같은 특성을 갖는 무성음에서는 높은 값을 갖게 된다. 음성신호의 영교차율 ($Z_x(n)$)은 식 (5)와 같이 정의되어 진다.

$$Z_x(n) = \frac{1}{2} \sum_{k=1}^N |\text{sgn}[X(n-k)] - \text{sgn}[X(n-k-1)]| \quad (5)$$

여기서 $\text{sgn}(X)$ 는 X가 0보다 클 때 1의 값, 그렇지 않은 경우에는 0의 값을 가지게 되며, 유/무성음 판단을 위한 두 플래그 값 (f_{z_vce} , f_{z_unv})이 식 (6)과 같이 각각 정의된다.

$$f_{z_vce} = \begin{cases} 1, \dots Z_x(n) < T_{z1} \\ 0, \dots otherwise \end{cases}$$

$$f_{z_unv} = \begin{cases} 1, \dots Z_x(n) < T_{z2} \\ 0, \dots otherwise \end{cases} \quad (6)$$

여기서 T_{z1} , T_{z2} 는 음성음, 무성음 구간의 영교차율 임계치를 각각 나타내고 있다.

이러한 3개의 파라미터들은 음성 생성 모델에 근거한 신호의 특징을 나타내는 파라미터이기 때문에 이동통신망에서 이용되는 가변 전송률 음성부호화기의 음성 활성도 측정 방안으로서 많이 활용되어 왔

으나 높은 레벨의 배경잡음이 존재하는 환경에서 동일 프레임내에 나머지 샘플들에 비하여 훨씬 큰 절대값을 갖는 몇 개만의 음성샘플들이 존재하는 경우에 활성음성과 배경잡음의 판별에 어려운 점이 발생하였다. 이러한 현상은 유성음 구간의 시작과 끝부분에 해당하는 프레임, Burst Energy 부분에 이어지는 Short Silence가 존재하는 파열음 구간에서 빈번히 발생하게 된다. 이러한 구간에서의 효율적인 동작 방안으로는 큰 진폭을 갖는 소수의 샘플들의 효과를 크게 반영함에 의하여 가능하게 되는데 이러한 개념은 원 음성 신호에서 단구간 상관관계가 제거된 LPC 잔여신호의 RMS (Root Mean Square) 값을 이용함에 의하여 가능하게 된다. LPC 잔여신호를 이용한 peakiness 측정값 (PM)은 식 (7)과 같이 정의된다.

$$PM(n) = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N r^2(n+k)}}{\frac{1}{N} \sum_{k=1}^N |r(n+k)|} \quad (7)$$

여기서 $r(n)$ 은 음성 신호에서 단구간 상관 관계인 LPC가 제거된 LPC 잔여신호를 나타내며, N 은 프레임 크기를 나타내며 그 값은 80이다. PM 값을 이용한 두 플래그 값 (f_{pm_u} , f_{pm_l})이 식 (8)와 같이 정의된다.

$$f_{pm_u} = \begin{cases} 1, \dots PM(n) > T_{pm2} \\ 0, \dots otherwise \end{cases}$$

$$f_{pm_l} = \begin{cases} 1, \dots T_{pm1} < PM(n) < T_{pm2} \\ 0, \dots otherwise \end{cases} \quad (8)$$

여기서 T_{pm1} , T_{pm2} 는 유성음, 파열음을 포함한 무성음의 PM값의 상한치, 하한치 임계값을 각각 나타내고 있다.

식(2), 식(4), 식(6), 식(8)을 통하여 추출되어진 파라미터 각각에 대한 플래그 값들을 활성음성 판별에 대하여 보수적인 방법으로 최종적인 음성 활성화 측정치 식 (9)와 같이 이루어진다. 음성 프레임 에너지, 영교차율을 통한 무성음 판별, PM의 상한치를 통한 유성음을 추출하여 활성음성으로 판단하고 나머지 구간의 보정을 위하여 각 파라미터의 조합을 통하여 부가적인 활성 음성 프레임을 판단하게 된다.

$$VAD(n) = (f_{e_sil}) | (f_{z_unv}) | (f_{pm_u}) | (f_{e_sil} \& f_{lb} \& f_{z_voe} \& f_{pm_l}) \quad (9)$$

여기서 '!', '|', '&'는 논리적인 'Not', 'Or', 'And'을 나타내고 있으며, VAD 결과 값이 1인 경우 활성음성, 0인 경우 배경잡음 구간으로 판단하게 된다.

III. 실험 및 결과

VAD 알고리즘의 성능 확인을 위하여 이용된 음성 샘플은 잡음이 존재하지 않는 환경에서 남녀 각 2명의 화자에 의하여 입력 신호레벨이 -12dBov에서 녹음된 1,145 프레임의 음성 데이터베이스를 이용하였으며 이 음성샘플들의 Voice Activity Factor는 약 0.47이었다. 입력 음성 샘플의 구성은 표 2에 나타나 있다.

표 2. 설계된 VAD 알고리즘의 성능 평가를 위한 clean speech 샘플

음성샘플		음성 샘플수	음성 프레임 수
남성	KM1	20,800	260
	KM2	25,360	317
여성	KF1	19,840	248
	KF2	25,600	320
Total		91,600	1,145

- KM1 : “이런 겨울은 예년과 달리 포근합니다.”
- KM2 : “개인통신시대가 조만간에 개막될것입니다.”
- KF1 : “미는 피부 한점질 차이입니다.”
- KF2 : “지나친 흡연은 건강을 해칩니다.”

그리고 이동통신 환경에서 발생하는 잡음을 고려하여 차량잡음, 거리잡음, Babble 잡음등을 채취하여 신호대 잡음비를 50dB ~ 10dB구간에서 10dB 간격을 갖게 디지털 합성방법으로 noisy 음성 샘플을 구성하였다.

알고리즘 성능 항목은 감지된 활성화도(Detected Activity), 음질 저하 프레임 비율 (Clipped Frame Rate) 등 2가지 객관적인 평가 기준을 사용하였다^[12]. 전자는 음성 활성화 측정 알고리즘에 의하여 활성 음성으로 판단된 프레임의 비율을 나타내는데 동일 음성 샘플의 Clean Speech와 Noisy Speech에

서의 감지된 활성음성 프레임의 변화를 반영하여 실제 이동통신 시스템 적용시 주파수 효율의 변화를 나타내게 된다. 후자의 Clipped Frame Rate는 Clean Speech 샘플에서는 실제 음성이 존재하는 구간이었으나 배경잡음에 의하여 음성의 특징 파라미터가 손상되어 VAD 기능에 의하여 비활성음성 구간으로 판단되는 경우로서 이러한 프레임이 증가되면 음성부호화기의 복호기에 의하여 재생된 음성 품질이 손상되게 된다. 이와 같이 주파수 효율과 음성 품질의 변화를 객관적인 평가 기준으로 평가하였다.

그림 2는 설계된 알고리즘에서 이용된 파라미터들의 적합성 여부를 위하여 높은 배경잡음이 존재하는 환경에서 각 파라미터 값의 변화를 나타내고 있다. (a)는 차량잡음에 의하여 신호대 잡음비가 10dB 인 Noisy 음성 샘플을 나타내고 있으며, (b)는 0 ~ 4kHz 대역에서의 프레임 에너지를 나타내고 있는데, 배경잡음에 의하여 활성 음성 구간의 판단이 어려운 구간이 존재함을 볼 수 있다. (c)는 전체 음성 대역에너지에 대한 1kHz 이하의 저주파 대역의 에너지 분포의 비를 나타내고 있는데, 묵음 구간과 유성음 구간은 비교적 높은 값을 갖고 있지만 무성음 구간에서는 낮은 값을 갖고 있음을 확인 가능하다. (d)는 영교차율을 나타내고 있다. 저주파 대역의 에너지 분포와는 반대로 무성음 구간에서 높은 값을 유지함을 알 수 있으며, (e)는 LPC 잔여신호의 Peakiness 측정값을 표현하고 있는데, 유성음 구간과 묵음으로 판별될 수 있는 일부 무성음 구간에서 큰 값을 나타냄을 알 수 있다.

그림 3은 차량잡음이 존재하는 음성 샘플에서 신호대 잡음비에 따른 설계된 알고리즘과 ITU-T G.729B VAD와의 성능을 비교한 것이다. (a)는 감지된 활성도의 변화를 나타내고 있는데 50 ~ 40dB SNR 영역에서는 두 알고리즘이 Clean Speech의 VAF 0.47과 큰 차이를 보이지 않지만 잡음 레벨이 증가됨에 따라 30, 20dB SNR 영역에서는 G.729B VAD가 설계된 알고리즘에 비하여 높은 활성도를 보였으며, 10dB SNR에서 G.729B는 활성도의 급격한 감소를 나타내었다. 이는 (b)에서 볼 수 있는 바와 같이 활성음성 구간이 배경잡음 구간으로 판단되는 경우가 다수 발생함에 따른 것이다. (b)는 SNR에 따른 Clipped Frame Rate의 변화를 나타내고 있는데 50 ~ 20dB SNR 영역에서 두 알고리즘은 유사한 성능을 나타내지만 10dB SNR 음성 샘플에서 G.729B VAD는 설계된 알고리즘에 비하여

음질 저하 프레임의 비율이 급격히 증가함을 알 수 있다.

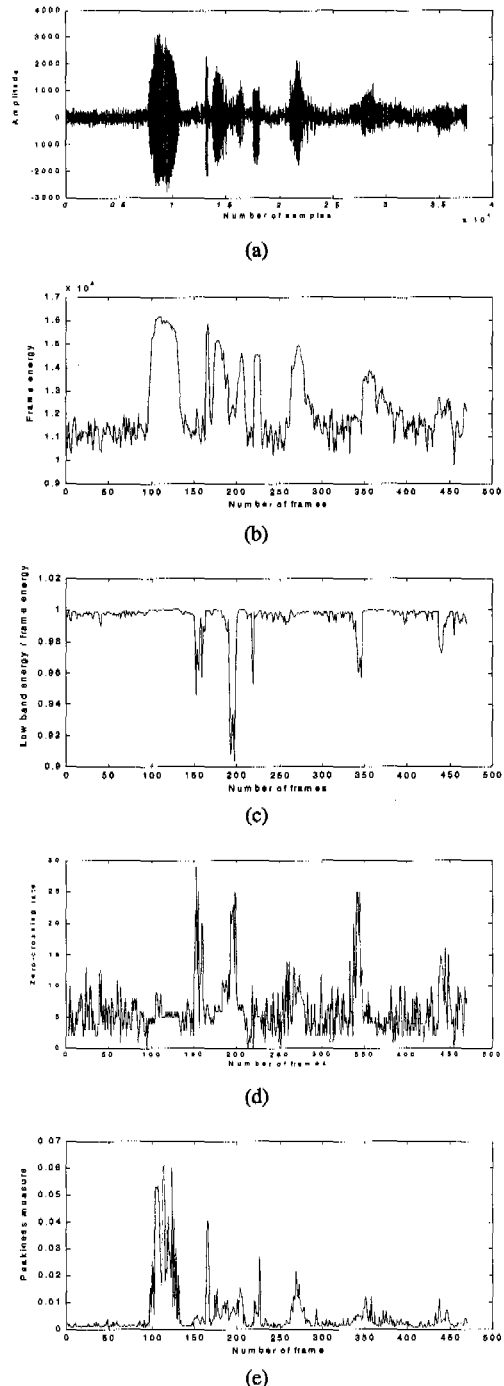


그림 2. 10dB SNR 차량잡음 환경의 파라미터특성
 (a) 10dB noisy 샘플 (b) 프레임 에너지 (c) 저주파 대역 에너지 분포 (d) 영교차율 (e) 잔여신호 PM값

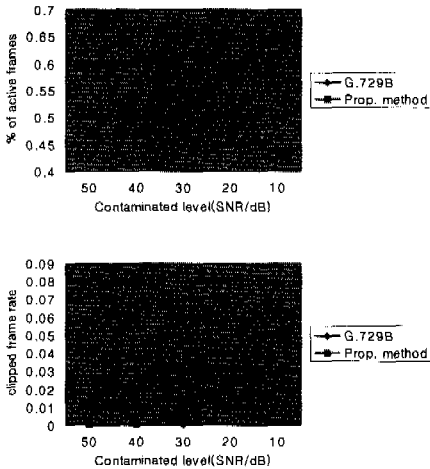


그림 3. 차량잡음 환경하에서의 감지활성도, 음절저하 프레임 변화

그림 4는 차량, 거리, babble 잡음하에서의 잡음 레벨에 따른 평균 감지 활성도와 Clipped Frame Rate의 변화를 나타내고 있다. 30, 20dB SNR에서 거의 동일한 Clipped Frame Rate에서 G.729B VAD는 높은 감지 활성도를 나타내고 있으며, 10dB SNR에서는 동일한 값의 감지 활성도에서 설계된 알고리즘이 G.729B VAD에 비하여 낮은 Clipped Frame Rate를 나타내어 설계된 음성 활성도 알고리즘이 높은 배경잡음이 존재하는 환경에서 양호한 음성 성능을 얻을 수 있음을 확인하였다.

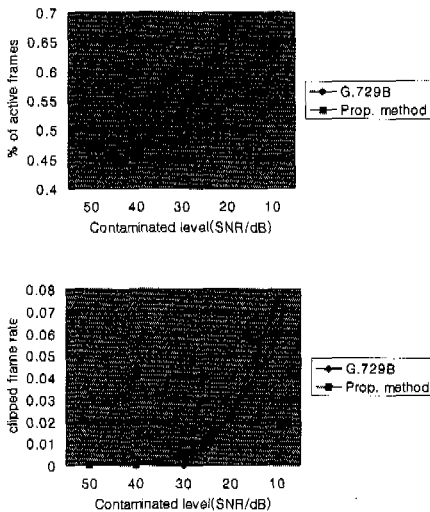


그림 4. 3가지 잡음 환경하에서의 감지 활성도, 음절저하 프레임 성능 변화

그림 5는 설계된 알고리즘과 G.729B VAD 알고리즘의 음성 활성도 측정 결과의 세밀한 관찰을 위하여 Clean Speech와 10dB SNR Noisy Speech 환경하에서 두 알고리즘의 프레임 단위 측정 결과를 나타내고 있다. 설계된 알고리즘이 낮은 에너지 레벨을 갖는 무성음 구간에서 G.729B VAD에 비하여 활성음성 프레임에 대한 판별이 양호하게 동작함을 알 수 있다.

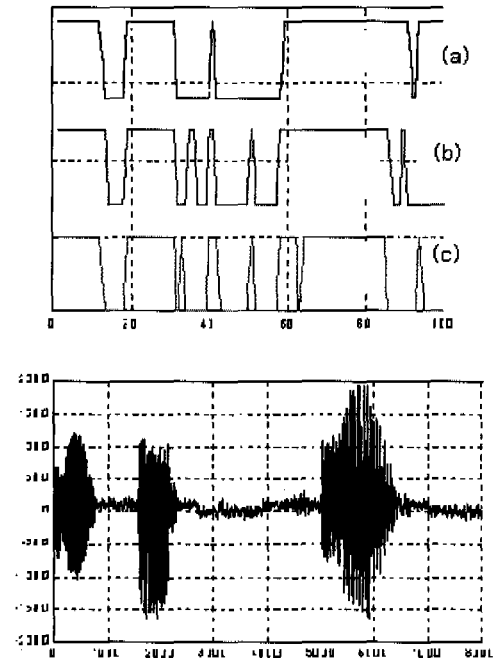


그림 5. 100 프레임 동안의 음성 활성도 측정 결과 (a) G.729B (b) Clean speech (c) 제안방법

표 3과 표 4는 20dB, 10dB SNR의 잡음 레벨을 갖는 차량, 거리, Babble잡음 음성샘플에서의 설계된 알고리즘과 G729B VAD의 감지 활성도와 Clipped Frame Rate 변화를 나타내고 있다. 20dB SNR에서는 차량 잡음과 Babble 잡음 환경에서 제안된 알고리즘이 G.729B VAD에 비하여 감지 활성도와 Clipped Frame Rate에서 양호한 특성을 나타내고 있으며, 3가지 잡음환경에 대한 평균 성능 또한 동일한 Clipped Frame Rate에서 G.729B VAD에 비하여 약 10% 낮은 감지 활성도를 나타냄을 알 수 있다. 그리고 10dB SNR에서는 설계된 알고리즘이 Babble 잡음 환경에서 양호한 성능을 얻을 수 있었으며, 평균적인 성능에서는 동일한 감지 활성도에서 G.729B VAD에 비하여 약 1.6% 낮은 Clipped Frame Rate를 나타냄을 확인하였다.

표 3. 20dB SNR 음성 샘플에서의 두 알고리즘의 비교 성능

잡음	증가된 Activity(%)		Clipped Rate(%)	
	설계방법	G.729B	설계방법	G.729B
차량	11.57	8.47	2.04	2.42
Babble	17.89	20.79	1.39	2.77
거리	3.35	22.36	3.04	1.25
평균	10.60	20.54	2.16	2.15

표 4. 10dB SNR 음성 샘플에서의 두 알고리즘의 비교 성능

잡음	증가된 Activity(%)		Clipped Rate (%)	
	설계방법	G.729B	설계방법	G.729B
차량	17.96	12.53	6.21	7.87
Babble	24.85	20.38	4.26	8.04
거리	8.03	16.96	7.05	6.38
평균	16.95	16.62	5.84	7.43

설계 알고리즘과 G.729B VAD 알고리즘의 성능을 비교 검토한 결과, 배경잡음이 적은 음성 샘플에서는 모두 양호한 성능을 나타내었으나 높은 배경잡음 레벨을 갖는 음성샘플에서는 G.729B VAD는 급격히 성능이 열화되는 특성을 나타내었으며, 설계된 알고리즘은 완전한 성능 저하를 나타내어 이동통신과 같이 다양하게 변화하는 배경잡음을 갖는 환경에서 강건하게 동작할 수 있는 음성 활성도 측정 방법임을 확인하였다.

IV. 결론

본 논문에서는 디지털 이동통신 시스템의 높은 음성품질 제공과 주파수 자원의 효율적 활용을 위하여 요구되어지는 가변 전송률 음성부호화기의 음성 활성도 측정방안을 연구하여 10msec 프레임 크기에 적용 가능한 음성 활성도 측정 알고리즘을 설계하였다. 설계된 알고리즘은 객관적인 평가항목에 의하여 그 성능을 평가하였으며, ITU-T G.729B VAD의 결과와 비교 검토하였다. 높은 배경잡음을 갖는 음성 샘플에서 설계된 알고리즘이 G.729B VAD 알고리즘 보다 양호한 성능을 가짐을 확인하여 이동통신망에서의 음성 활성도 측정 알고리즘으로의 활용이 가능함을 확인하였다.

참고 문헌

- [1] R. V. Cox and P. Kroon, "Low bit-rate speech coders for multimedia communications," *IEEE Communications Mag.*, vol. 34, no. 12, pp. 34-41, December 1996.
- [2] A. Das, E. Parksoy and A. Gersho, "Multimode and Variable-Rate Coding of Speech," in *Speech Coding and Synthesis*, pp. 257-288, Elsevier 1995.
- [3] M. Rodriguez and H. Tawfik, "Outage performance in cellular systems with discontinuous transmission," *IEE Electronics Letters*, vol. 29, no. 15, pp. 1329-1330, July 1993.
- [4] K. Sririam, P. K. Varshney and J. G. Shanthikumar, "Discrete-time analysis of integrated voice/data multiplexers with and without speech activity detectors," *IEEE J. Selected Area in Communications*, vol. SAC-1, pp. 1124-1132, December 1983.
- [5] Jaewon Kim, Minsik Seo, Byungsik Yoon, Songin Choi and Younggap You, "A voice Activity Detection Algorithm for Wireless Communication Systems with Dynamically Varying Background Noise," *IEICE Trans. on Communications*, Vol. E83-B, no. 2, pp. 414-418, Feb. 2000.
- [6] ITU-T, Draft Recommendation G.729, Annex B : Voice Activity Detection, 1996.
- [7] E. Paksoy, K. Srinivasan and A. Gersho, "Variable bit-rate CELP coding of speech with phonetic classification," *Eur. Transactions on Telecommunications*, vol. 5, no. 9, pp. 591-601, September 1994.
- [8] K. Srinivasan and A. Gersho, "Voice activity detection for cellular networks," in *Proc. IEEE Workshop Speech Coding for Telecommunications*, pp. 85-86, October 1993.
- [9] J. Stegmann and G. Schroder, "Robust voice activity detection based on the wavelet transform," *IEEE VTC97*, pp. 99-100, 1997.
- [10] F. Beritelli, S. Casale and A. Cavallaro, "A voice activity detector for mobile communications based on fuzzy logic," in *Proc.*

European Symposium intelligent Techniques (ESIT97), pp. 91-95, March 1997.

- [11] R. A. Goubran and H. M. Hafez, "Background acoustic noise reduction in mobile telephony," *Proc. 36th IEEE Soc. Conf. on Vehicular Technology*, pp. 72-76, May 1986.
- [12] D. K. Freeman et al., "The voice activity detector for the Pan-European digital cellular mobile telephone service," in *Proc. Intl. Conf. Acoustic, Speech, Signal Processing*, pp. 369-372, May 1989.

김 재 원(Jae-won Kim)

정회원



1986년 2월 : 경북대학교
전자공학과 공학사
1988년 2월 : 경북대학교
전자공학과 공학석사
2001년 2월 : 충북대학교
정보통신공학과 공학박사

1991년 1월~2000년 2월 : 한국전자통신연구원
무선방송기술연구소 선임연구원

2000년 3월~현재 : 경주대학교 컴퓨터전자공학부
전임강사

<주관심 분야> 디지털 이동통신, 디지털 신호처리,
음성 신호처리, 인터넷폰