

# 잡음 환경에서의 전송률 감소를 위한 G.723.1 음성활동 검출기 성능 개선에 관한 연구

## Improvement of VAD Performance for the Reduction of the Bit Rate Under the Noise Environment in the G.723.1

장 경 아\*, 김 정 진\*\*, 배 명 진\*\*\*  
(Kyung A Jang\*, Jeong Jin Kim\*\*, Myung Jin Bae\*\*\*)

\*송실대학교 대학원 정보통신공학과, \*\*LG전자 중앙연구소 이동멀티미디어실, \*\*\*송실대학교 정보통신공학과 교수  
(접수일자: 2000년 11월 1일; 채택일자: 2001년 2월 27일)

본 논문에서는 인터넷 폰 및 화상통신 등을 위해 개발된 G.723.1 부록 A 6.3kbps/5.3kbps 이중 전송률 음성 코덱의 음성활동 검출기의 성능을 개선한다. 본 논문에서는 에너지 레벨을 크게 3가지 범위로 구분하여 음성활동 유무 판정을 하도록 한다. 제한한 알고리즘에 대한 처리시간, 음질평가 및 전송률 감소량을 측정할 결과 처리시간의 경우 묵음구간에 대한 판정의 정확성에 기인하여 G.723.1에 비해 감소하고 주관적인 음질평가의 경우 G.723.1과 거의 차이가 없다. 전송률 측정을 위해 VAD=1로 판정한 프레임의 수를 측정할 결과 묵음 구간이 많이 나타날수록 뚜렷한 전송률 감소효과를 얻을 수 있다.

**핵심용어:** 음성활동 검출기, 에너지, 피치이득, LSP

**투고분야:** 음성처리 분야 (2.1)

This paper improves the performance of VAD (Voice Activity Detector) in G.723.1 Annex A 6.3kbps/5.3kbps dual rate speech coder, which is developed for Internet Phone and videoconferencing. The VAD decision is based on a three-level energy threshold. We evaluate for processing time, speech quality, and bit rate. The processing time is reduced due to the accuracy of VAD decision on the silence period. On subjective quality test there is almost no difference compared with the G.723.1. In order to measure the bit rate we count the active speech frame (VAD=1) and we can reduce more bit rate as silence periods are shown.

**Keywords:** Voice activity detection, Energy, Pitch gain, LSP (Linear Spectrum Pairs)

**Ask subject classification:** Speech signal processing (2.1)

## I. 서론

최근 십여 년간의 사회 동향을 볼 때 정보통신분야, 특히 멀티미디어와 이동 통신으로 대표되는 두 분야의 발전은 이루 헤아릴 수 없이 급속한 발전의 양상을 보이고 있다. 이들 분야는 여러 가지 표현 수단을 가지고 최적의 조건하에서 정확한 정보의 전달을 목적으로 하고

있다. 최근 디지털 이동통신 및 개인 통신망의 눈부신 발전으로 이동통신 사용자보다 많이 수용하고 양질의 서비스를 제공하기 위한 음성부호화 연구가 진행되고 있다. 가입자의 수용율을 증가시키는데 있어서 효과적인 방법 중의 하나는 부호화기의 전송율을 낮추는 방법이다. 이론적으로 디지털 셀룰라 망에서 보코더의 전송율이 1/2로 낮아질 경우 약 2배의 가입자를 수용할 수 있다고 알려져 있다[1]. 하지만 전송률 감소에 따른 음질저하의 발생은 필수적이므로 선진국에서는 다음과 같은 음성부호화기의 요구조건을 충족시키는 보코더 개발을 진행하고 있다[1].

책임저자: 장경아 (kajang@hotmail.com)  
156-743 서울시 동작구 상도 5동  
송실대학교 정보통신공학과 음성통신연구실  
(전화: 017-610-3410; 팩스: 820-0018)

첫째, 가능한 원음의 음질을 유지하면서 낮은 전송율로 전송이 가능해야 하고 둘째, 간단한 알고리즘으로 하드웨어 구현시 복잡도가 낮아야 하고 저전력 소비로 단말기 이용시간을 증대해야 한다. 셋째, 부호기/복호기의 지연이 적어야 하고 에러에 강인해야 한다. 현재까지 발표된 음성 부호화기 중 가장 많은 연구가 이루어지고 있는 방식이 CELP (Code Excited Linear Prediction) 방식으로 4.8Kbps 내외의 전송율에서 양호한 음질을 얻을 수 있으며 특히 국내에서는 인터넷폰과 PCS (personal communication system) 등의 출현으로 CELP 부호화기에 대해 많은 관심이 집중되고 있다[2]. 이런 CELP 계열 부호화기 중 인터넷폰과 화상회의를 목적으로 개발된 G.723.1은 전송환경에 따라 전송율을 달리하는 이중 전송률 구조 (Dual rate speech coder: 5.3kbps/6.3kbps)를 가지고 있다. G.723.1의 전송율은 여기신호의 압축방법에 따라 전송율에 차이가 나며 저 전송율에서는 ACELP (Algebraic CELP) 방식으로, 고 전송율에서는 MP-MLQ (Multi-Pulse Maximum Likelihood Quantization) 방식을 사용한다[3]. 1996년도에 표준화가 이루어진 G.723.1과 G.729나 현재 국내의 디지털 셀룰러 서비스에 사용되고 있는 Qualcomm사의 QCELP 등에서는 묵음 구간에서의 전송율을 감소시키기 위해 음성활동 검출기 (Voice Activity Detection)를 사용하고 있다. 이 중 인터넷폰 및 화상회의를 목적으로 개발된 G.723.1의 음성활동 검출기는 현재 프레임의 음성신호 존재유무를 판정을 위해 에너지 파라미터를 사용하고 있으며, 연속적인 프레임의 피치값과 정현파 검출 여부를 이용하여 에너지 임계값을 설정한다[3].

일반적으로 음성활동 검출기의 가장 큰 문제점은 어떠한 배경 잡음에 대해서도 음성 신호를 검출할 수 있어야 한다는 것이다. 이런 문제를 해결하기 위해서 G.723.1에서는 스펙트럼 특성과 입력신호의 주기성을 이용하고 있다. 하지만 이런 파라미터는 판정에 직접적으로 영향을 미치는 것이 아니기 때문에 신호대 잡음비 (SNR: signal-to-noise-ratio)이 낮은 신호에 대해 정확한 판정을 한다는 상당히 어렵다.

따라서 본 논문에서는 처리시간의 불필요한 소모를 방지하기 위해 미리 설정된 에너지 임계값을 사용하여 아주 큰 에너지를 갖거나 아주 낮은 에너지를 갖는 경우에는 간단한 결정논리를 사용하고, 현재 프레임의 입력신호의 에너지가 임계값을 넘는 경우에는 묵음구간에서 구한 LSP (Line Spectrum Pairs)계수와 입력신호에서 구한 LSP 계수 사이의 거리값을 이용하여 스펙트럼 특성을 고

려하고 피치이득 값을 이용하여 주기성을 고려하여 음성활동 유무판정을 하는 알고리즘을 제안한다.

## II. 음성활동 검출 알고리즘

### 2.1. 에너지, 피치 이득 및 LSP 계수를 이용한 음성활동 검출 알고리즘

묵음 구간에서의 전송율을 낮추는 위한 G.723.1 음성활동 검출기는 판정의 안정성과 연속성을 위해서 위에서 설명한 바와 같이 여러 가지 파라미터를 사용하고 있다. 또한 신호대 잡음비가 낮은 신호에 대한 정확한 판정을 위해 스펙트럼 특성을 이용하고 있다. 하지만 위와 같은 방법은 실제 임계값이 현재의 에너지 레벨에 도달할 때까지 정확한 판정을 하는 것이 불가능하며 또한 신호대 잡음비가 낮은 신호가 입력으로 들어올 때 묵음이 존재하는 프레임에 대해서도 음성이 존재하는 프레임으로 판별함으로써 전송율을 낮추는데 기여하지 못하고 있다. 따라서 본 논문에서는 기존의 방법이 가지는 장점은 해치지 않는 범위 내에서 묵음이 존재하는 프레임을 보다 정확히 검출함으로써 전송율을 낮추는 방법에 대해 제안한다.

본 논문에 쓰이는 파라미터는 에너지, 피치 이득 및 LSP 계수이다. 입력 신호의 처음 3프레임 동안은 묵음이라는 가정하에 에너지 임계값을 구하고 묵음 구간에서의 평균 LSP 계수들을 선정하게 된다. 그런 다음 음성활동 구간 판정을 3가지 경우로 나누어 판정을 수행하게 된다. 먼저 현재 프레임의 에너지가 에너지 최대 임계값을 넘는 경우와 에너지 임계값을 넘지 않는 경우에는 각각 음성활동 구간과 음성 비활동 구간으로 판정하고 현재 프레임의 에너지가 에너지 임계값을 넘는 경우에는 신호대 잡음비가 낮은 신호의 경우를 고려하기 위해 피치이득과 LSP 파라미터를 이용하게 된다[5,6].

또한 묵음구간으로 판정된 프레임에 대해서는 합성을 위한 최소한의 파라미터만을 전송하게 되므로 파라미터 추출을 위한 계산시간을 단축할 수 있게 된다. 따라서 본 논문에서 제안한 방법을 이용할 경우 전송율 감소 효과 뿐만 아니라 부가적으로 처리시간 감소 효과도 얻을 수 있게 된다.

### 2.2. 에너지 임계값 및 묵음 구간의 LSP 계수 설정

본 논문에서는 처음의 입력 3프레임 동안은 묵음이라는

가정하에 3프레임 동안 다음과 같이 평균에너지와 평균 LSP계수들을 구하게 된다[5,7].

$$Ene_i = \sum_{n=0}^{N-1} s_i^2[n] / N, \quad i=0,1,2 \quad (1)$$

$$NLSP_k = \sum_{i=0}^2 LSPvect_{i,k}, \quad k=1,2,\dots,10 \quad (2)$$

여기서  $Ene_i$ 는 3프레임 동안 구해진 평균 에너지이고,  $NLSP_k$ 는 3프레임 동안의 구해진 LSP 계수값이다.  $N$ 은 240이며  $s_i[n]$ 은 현재 프레임  $t$ 의 입력신호이며  $LSPvect$ 는 현재 프레임에서 구한 LSP 계수들이다. 위의 파라미터를 이용하여 다음과 같은 에너지 임계값과 목음의 평균 LSP 계수들을 계산한다.

$$EneThr = mean(Ene) + 1.3 \times StdDev(Ene) \quad (3)$$

$$LSPave_k = NLSP_k / 3, \quad k=1,2,\dots,10 \quad (4)$$

위에서 구한 에너지 문턱값  $EneThr$ 은 [512, 131072]의 경계값을 갖는다.  $LSPave_k$ 은 3프레임 동안의 LSP 계수의 평균값이다.

### 2.3. 음성 존재 유무 판별

본 논문에서는 음성의 존재 유무 판정을 위해 크게 3가지로 나누어 판정을 한다. i) 현재 프레임  $t$ 에서 구한 에너지가 최대 에너지 임계값을 넘는 경우 ii) 현재 프레임  $t$ 에서 구한 에너지가 에너지 임계값을 넘지 않는 경우 iii) 현재 프레임  $t$ 에서 구한 에너지가 에너지 임계값을 넘는 경우로 나누어 판정하며, i), ii)의 경우 각각 음성이 활동하는 프레임, 음성이 활동하지 않는 프레임으로 판정을 한다. 마지막 iii)의 경우 입력 신호가 낮은 신호대 잡음비를 갖는 경우를 고려하기 위해 피치 이득과 LSP 거리 파라미터를 이용하여 판정을 수행한다. 즉 에너지가 임계치를 넘는다고 하더라도 신호대 잡음비가 낮은 신호의 경우 목음 구간에 존재하는 잡음에 의한 경우를 배제하기 위해 피치 이득과 LSP 거리가 미리 설정된 각각의 피치이득 임계치와 LSP 거리 임계치를 넘는 경우에만 음성이 존재하는 것으로 판정한다.

#### 2.3.1. 현재 프레임의 에너지가 에너지 최대 임계값을 넘는 경우

피치이득이나 LSP 거리와 상관없이 항상 음성활동 구간으로 설정한다 ( $VAD=1$ ). 또한  $EneThr$ 은 다음과 같이 갱신된다.

$$EneThr_t = EneThr \cdot (1025/1024) \quad (5)$$

#### 2.3.2. 현재 프레임의 에너지가 에너지 임계값을 넘지 않는 경우

목음구간으로 설정한다 ( $VAD=0$ ). 또한  $EneThr$ 은 다음과 같이 갱신되며 LSP 계수 또한 식 (2-4)를 이용하여 갱신된다.

$$EneThr_t = EneThr_{t-1} \cdot (31/32) \quad (6)$$

#### 2.3.3. 현재 프레임의 에너지가 에너지 임계값을 넘는 경우

##### 2.3.3.1. 피치 이득 계산

피치 이득  $\beta_t$ 은 다음과 같이 구한다.

$$\beta_t = \frac{C_{max}}{Ene_t} \quad (7)$$

여기서 상관관계 관계값  $C_{max}$ 는 다음 식의 상관관계 계수  $C_b$ 를 최대로 하는 값이다.  $C_b$ 은 현재 프레임내 샘플간의 상관관계도를 의미한다.

$$C_b(j) = \frac{(Cor(j))^2}{\sum_{n=b}^{24} s_i[n-j] \cdot s_i[n-j]}, \quad 18 \leq j \leq 142 \quad (8)$$

$$Cor(j) = \sum_{n=0}^{N-1} s_i[n] \cdot s_i[n-j], \quad 18 \leq j \leq 142 \quad (9)$$

##### 2.3.3.2. LSP (Line Spectrum Pairs) 거리 계산

목음 구간의 LSP 계수들 사이에는 일반적으로 등간격을 가지고 있지만 음성이 존재하는 경우는 포만트가 위치하는 주파수영역에 LSP 계수들이 많이 존재하는 특징이 있다[9]. 즉 목음구간에서 구한 LSP 계수들과 음성이 존재하는 LSP 계수들 사이의 오차를 구하면 그 값이 크게되지만 목음구간의 LSP 계수들 사이의 오차는 상당히 적게 된다. 따라서 LSP 계수들 사이의 오차를 이용하면 음성의 존재유무를 판정할 수 있게 된다. LSP 계수들 사이의 거리는 다음과 같이 구할 수 있다[5].

$$LSPDist = \sqrt{\sum_{k=1}^{10} (LSP(k) - LSPave(k))^2} \quad (10)$$

$LSPDist$ 은 목음구간에서 얻어진 LSP계수의 평균값과 음성구간의에서 얻어진 LSP계수와의 오차를 구한 값이다.

##### 2.3.3.3. 음성활동 검출

위에서 구한 피치이득 값 LSP 거리값이 미리 설정된

각각의 임계값  $bthr$ ,  $LSPThr$  보다 작은 경우 묵음 구간으로, 그렇지 않은 음성활동 구간으로 설정하게 된다.

$$VAD = \begin{cases} 0, & \text{if } b < bthr \text{ and } LSPdist < LSPThr \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

### 2.4. 판정의 지속성 유지

판정의 지속성을 위해 본 논문에서는 다음과 같은 기능을 추가하였다.

$$Vcnt = \begin{cases} Vcnt + 2, & \text{if } Ene_t \geq EneThr \\ Vcnt - 1, & \text{if } Ene_t < EneThr \end{cases} \quad (12)$$

비록 제안한 알고리즘이 묵음구간으로 판정하더라도 판정의 급격한 변화를 방지하기 위해 음성활동 검출 카운터,  $Vcnt$ 가 0보다 큰 경우에는 음성활동 구간으로 설정하게 된다.

$$VAD = \begin{cases} 1, & \text{if } Vcnt > 0 \\ 0, & \text{if } Vcnt = 0 \end{cases} \quad (13)$$

## III. 실험 및 결과

제안한 알고리즘을 실험하기 위한 장비는 IBM-PC 586 (333MHz)에 상용화된 AD/DA (Analog to Digital/Digital to Analog)컨버터를 인터페이스한 시스템을 사용하였다. 입력신호는 G.723.1의 입력신호와 같이 8kHz로 표본화하고 16bit로 양자화한 음성을 입력으로 하며 각 음성시료

에 대해 한 프레임의 길이를 240표본, 부프레임의 길이를 60샘플로 하여 처리하였다. 처리결과의 성능을 측정하기 위해 다음의 음성 시료에 대해 AWGN (Additive White Gaussian Noise)과 실제 실험실 환경의 잡음을 첨가하여 사용하였다. 본 논문에서는 음성시료의 신호대 잡음비가 각각 -5, 0, 10, 20dB가 되도록 AWGN을 Clean Speech에 첨가한 음성과 실제 실험실 환경에서 접할 수 있는 잡음을 마이크를 통하여 입력받은 다음 Clean Speech에 더한 음성시료를 사용하였다. 또한 묵음구간의 검출 결과를 알아보기 위해 각 어절마다 묵음구간을 길게 삽입하여 묵음의 검출여부를 관찰하였으며 대표적인 문장을 연령층이 다른 남녀화자가 발성한 음성을 음성시료로 사용하였다.

발성1: /인수네 꼬마는 천재소년을 좋아한다./

발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./

발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다./

발성4: /숭실대학교 정보통신과 음성통신 연구팀이다./

발성5: /공일이삼사오육칠팔구/

제안한 알고리즘의 시뮬레이션은 C-언어로 구현하여 수행하였다. 성능 비교는 G.723.1 부록 A를 통과한 음성과 제안한 알고리즘을 통과한 음성을 비교하였다. 전송률 감소량을 측정하기 위해 전체 프레임 중에서 VAD=1로 판정한 프레임 수를 비교하였으며 처리시간 감소량 측정은 C 알고리즘이 제공하는 clock 함수로 측정하였다. 실험 환경은 MP-MLQ 모드에서 실험실 환경의 잡음이 첨가된

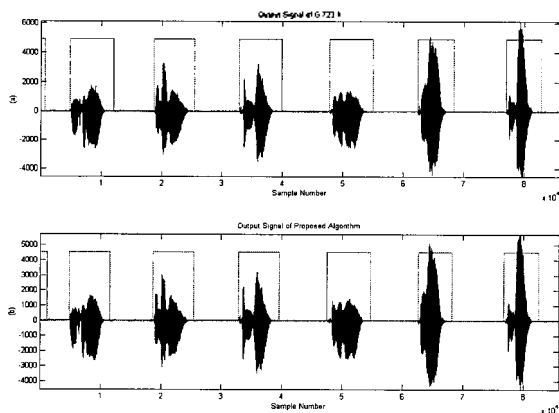


그림 1. Clean Speech (남성화자) /인수네 꼬마는 천재 소년을 좋아한다/  
(a) G.723.1 알고리즘 (b) 제안한 알고리즘

Fig. 1. Clean speech (male speaker); /Insune komaneun cheonjae sonyuneul joahanda/,  
(a) G.723.1 algorithm, (b) proposed algorithm.

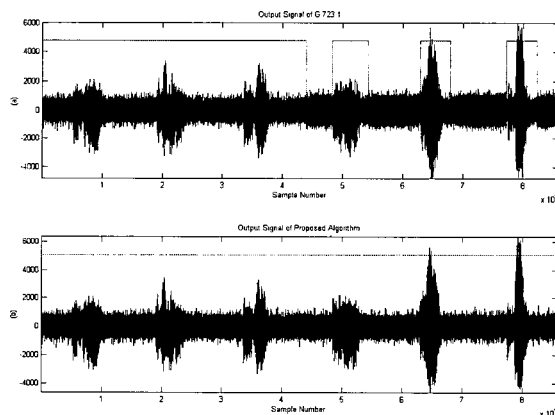


그림 2. Clean Speech + AWGN (SNR 0dB, 남성화자) /인수네 꼬마는 천재 소년을 좋아한다/  
(a) G.723.1 알고리즘 (b) 제안한 알고리즘

Fig. 2. Clean speech + AWGN (SNR 0dB, male speaker); /Insune komaneun cheonjae sonyuneul joahanda/,  
(a) G.723.1 algorithm, (b) proposed algorithm

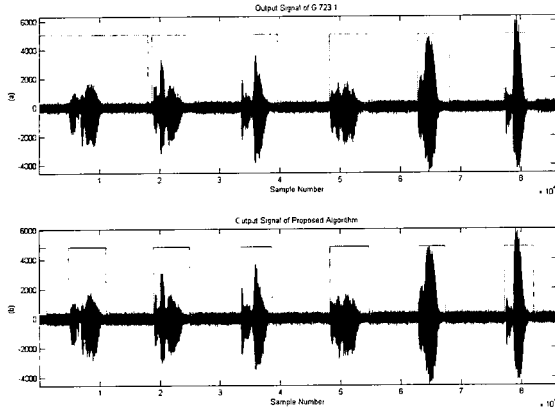


그림 3. Clean Speech + AWGN (SNR 10dB, 남성화자) /인수네 꼬마는 천재 소년을 좋아한다/ (a) G.723.1 알고리즘 (b) 제안한 알고리즘  
 Fig. 3. Clean speech + AWGN (SNR 10dB, male speaker): /Insune komaneun cheonjae sonyuneul joahanda/, (a) G.723.1 algorithm, (b) proposed algorithm.

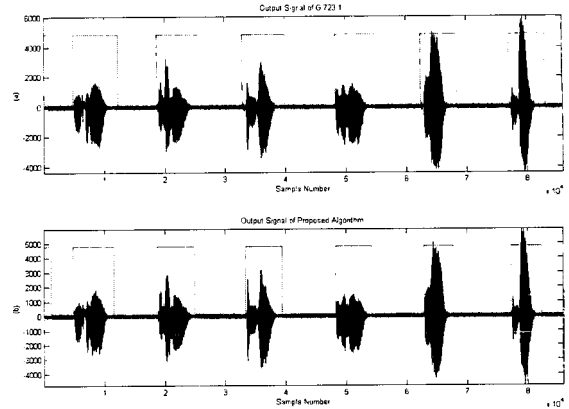


그림 4. Clean Speech + AWGN (SNR 20dB, 남성화자) /인수네 꼬마는 천재 소년을 좋아한다/ (a) G.723.1 알고리즘 (b) 제안한 알고리즘  
 Fig. 4. Clean speech + AWGN (SNR 20dB, male speaker): /Insune komaneun cheonjae sonyuneul joahanda/, (a) G.723.1 algorithm, (b) proposed algorithm.

표 1. VAD=1로 판정한 프레임의 수 (남성화자-전체 358 프레임, 음성자료 1, 여성화자-전체 425 프레임, 음성 자료 3)  
 Table 1. The number of active speech frame (male speaker-total 368 frames, speech data1, female speaker-total 425 frames, speech data3).

|        | G.723.1 VAD 알고리즘 |     | 제안한 VAD 알고리즘 |     | 감소율 (%) |      |
|--------|------------------|-----|--------------|-----|---------|------|
|        | 남                | 여   | 남            | 여   | 남       | 여    |
| clean  | 169              | 231 | 166          | 235 | 1.7     | -1.7 |
| -5dB   | 358              | 263 | 358          | 207 | 0       | 21.3 |
| 0dB    | 250              | 259 | 358          | 170 | -43.2   | 34.4 |
| 10dB   | 203              | 233 | 144          | 199 | 29.1    | 14.6 |
| 20dB   | 171              | 254 | 150          | 223 | 12.3    | 12.2 |
| 실험실 잡음 | 245              | 295 | 149          | 190 | 39.2    | 35.6 |

음성을 사용하여 5번 측정된 다음 평균값을 사용하였다. 음질 측면에서는 MOS test를 사용하였다. 그림 1~4에서 점선으로 나타난 부분은 음성 활동 구간을 검출한 것을 나타낸 것으로, 음성 활동 검출 장치를 실행한 결과 0의 판정을 받은 구간은 그림의 점선 또한 0의 값으로 나타내고, 1의 판정 받은 구간은 음성이 있다고 간주한 것이다. 남성화자가 음성자료 1을 발성한 것으로 clean speech의 경우 기존의 G.723.1과 제안한 알고리즘에는 거의 차이가 없음을 알 수 있으며, 0dB 음성의 경우 제안한 알고리즘에서 전송율은 증가하나 묵음구간의 잡음레벨이 큰 신호에 대해 CNG를 사용하여 음성을 합성한 경우 상당한 음질저하를 초래하게 되므로 이런 신호의 경우 묵음구간에 대해서도 음성활동 구간으로 설정하는 것이 급격한 음질저하를 방지하게 된다. 그 외의 경우에는 잡음의 에너지 레벨에 따라서 전송율이 감소함을 알 수 있다. 그림 4은 clean

표 2. 처리시간의 비교 (sec, 실험실 환경의 잡음을 첨가한 경우)  
 Table 2. Comparison of processing time (sec, in case of adding a background noise of laboratory).

|                  | 발성 1 | 발성 2  | 발성 3  | 발성 4  | 발성 5  | 평균   |
|------------------|------|-------|-------|-------|-------|------|
| G.723.1 VAD 알고리즘 | 8.90 | 10.26 | 10.69 | 10.22 | 11.93 | 10.4 |
| 제안한 VAD 알고리즘     | 7.89 | 9.59  | 9.83  | 9.32  | 11.80 | 9.69 |
| 감소율 (%)          | 11.3 | 6.5   | 8.1   | 8.8   | 1.1   | 7.16 |

표 3. 주관적 음질의 비교 (MOS, 실험실 잡음을 첨가한 경우)  
 Table 3. Comparison of subjective quality (MOS, in case of adding a background noise of laboratory).

|          | 발성 1 | 발성 2 | 발성 3 | 발성 4 | 발성 5 | 평균   |
|----------|------|------|------|------|------|------|
| G.723.1  | 3.7  | 3.8  | 3.7  | 3.9  | 3.8  | 3.78 |
| 제안한 알고리즘 | 3.7  | 3.7  | 3.7  | 3.9  | 3.8  | 3.76 |

speech에 실험실 환경에서 채집한 잡음을 첨가한 신호로 제안한 알고리즘에서는 에너지가 큰 잡음구간과 음성활동 구간을 정확하게 판정하는 반면 기존의 알고리즘에서는 정확한 검출이 불가능하다는 것을 알 수 있다. 표 1은 전송율 감소량을 측정하기 위해 VAD=1로 판정한 프레임 수를 비교한 것으로 신호대 잡음비 5dB 이상의 음성신호에서 잡음의 에너지가 클수록 전송율 감소효과가 큰 것을 알 수 있다.

표 2는 처리시간을 비교한 것으로 음성활동 검출 알고리즘에 의해 묵음 구간으로 판정한 경우 파라미터 추출 과정을 생략함으로써 부가적으로 처리시간 감소효과를 얻을 수 있으므로 제안한 알고리즘이 기존의 알고리즘에 비해 처리시간이 감소하는 것을 알 수 있다. 표 3은 음질을

비교한 것으로 기존의 알고리즘에 비해 전송율은 감소하는 반면 음질에는 거의 차이가 없음을 알 수 있다. 실험 결과 전반적으로 제안한 알고리즘이 기존의 G.723.1 보다 월등한 성능을 보이는 것을 알 수 있다. 신호대 잡음비가 5dB~20dB사이의 음성신호의 경우 제안한 알고리즘이 기존의 G.723.1 음성활동 검출기보다 전송율 감소 측면에서 보다 나은 결과를 얻었다. 처리시간의 경우 제안한 알고리즘이 신호대 잡음비가 낮은 환경에서 보다 정확한 음성활동 유무판정을 함으로써 고정코드북 검색시간 및 이득 계산에 소요되는 처리시간을 절약함으로써 전체 부호화기의 처리시간을 감소시키는 결과를 얻을 수 있었다. 주관적 음질 평가의 경우 전송율은 감소되는 반면 음질에는 거의 차이가 없음을 알 수 있었다.

#### IV. 결론

CELP 계열 음성 부호화가 중에서 인터넷 폰 및 화상통신을 위해 개발된 G.723.1 6.3/5.3Kbps 이중 전송율 음성 코덱은 묵음 구간에서의 전송율을 감소시키기 위해 VAD (Voice Activity Detection)를 사용하고 있으며, 음성활동 유무판정에 의해 묵음으로 간주된 프레임에 대해서는 신호합성에 필요한 최소한의 파라미터만을 전송하여 전송율을 감소와 실시간 구현을 위한 처리시간 감소 효과도 더불어 얻을 수 있다. 일반적으로 음성활동 검출기의 가장 큰 문제점은 어떠한 배경 잡음에 대해서도 음성 신호를 검출할 수 있어야 하는데, 해결방안으로 G.723.1에서는 스펙트럼 특성과 입력신호의 주기성을 이용하고 있다. 하지만 이런 파라미터는 판정에 직접적으로 영향을 미치는 것이 아니기 때문에 이런 과정을 거치는 것으로 신호대 잡음비가 낮은 신호에 대해 정확한 판정을 한다는 상당히 어렵다. 따라서 처리시간의 불필요한 소모를 방지하기 위해 미리 설정된 에너지 임계값을 사용하여 간단한 결정을 사용하고 현재 프레임의 입력신호의 에너지가 경계 조건 안에 해당하는 경우 잡음 구간에서 구한 LSP 계수와 입력신호에서 구한 LSP 계수 사이의 거리값을 이용하여 스펙트럼 특성을 고려하고, 피치이득 값을 이용하여 주기성을 고려하여 음성활동 유무판정을 하는 알고리즘을 제안한다. 실험 결과 신호대 잡음비가 5dB에서 10dB 사이의 음성신호의 경우 최적의 전송율 감소 효과를 얻을 수 있었으며 처리시간의 비교 결과 평균 7% 정도의 처리시간 감소효과를 얻을 수 있었다. 주관적 음질 평가의 결과 음질 열하는 거의 발생하지 않았다.

#### 참고 문헌

1. W. B. Kleijn and K. K. Paliwal, "Speech Coding and Synthesis", ELSEVIER SCIENCE B. V., pp. 6-10, 1995.
2. A. M. Kondoz, Digital Speech-Coding for Low Bit Rate Communication System, Wiley, 1994.
3. ITU-T Recommendation G.723.1, March, 1996.
4. 민병준, 강병준, "EVRC패킷에서 LSP거리를 이용한 음성 품질 검출", 한국 음향학회지, 18권 6호, 1999.
5. 김정진, 장경아, 배명진, 홍성훈, 심민규, 성유나, "LSP 정보와 에너지 파라미터를 이용한 G.723.1 음성 활동 검출 장치", 신호처리합동학술발표대회 논문집, October, 1999.
6. L. R. Rabiner, R. W. Schafer, Digital Processing of Speech Signal, Prentice Hall, 1978.
7. 나덕수, 노원석, 함명규, 배명진, "피치동기에 의한 음성신호의 전이구간 검출," 한국음향학회, 제15회음성통신 및 신호처리워크샵 논문집, Vol. 15, No. 1, pp. 454-459, 1998.
8. 나덕수, 정찬중, 박영호, 배명진, "LSP를 이용한 음성신호 성분분리에 의한 CELP 보코더의 전송률 감소에 관한 연구", 한국음향학회, 학술발표대회논문집, August, 1999.
9. Sadaaki Furui, Digital Speech Processing, Synthesis, and Recognition, Marcel Dekker, INC., 1992.
10. P. E. Papamichalis, Practical Approaches to Speech Coding, Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
11. Steve Ammon and Jianxin Liu, "Implementation of the ITU-T G.723.1 Speech Coding Algorithm on the Hitachi SH-DSP," ICSPAT, pp. 456-460, 1995.

#### 저자 약력

##### ● 장 경 아 (Kyung A Jang)



1998년 2월: 동신대학교 정보통신공학과 졸업 (학사)  
 2001년 3월: 숭실대학교 대학원 정보통신공학과 졸업 (석사)  
 2001년 3월: 숭실대학교 대학원 정보통신공학과 재학중 (박사)  
 \* 주관심분야: 음성 신호처리, 음성코딩, 실시간 DSP 응용

##### ● 김 정 진 (Jeong Jin Kim)



1998년 2월: 동신대학교 정보통신공학과 졸업 (학사)  
 2000년 8월: 숭실대학교 대학원 정보통신공학과 졸업 (석사)  
 2000년 6월~현재: LG전자 중앙연구소 이동멀티미디어 연구원  
 \* 주관심분야: 음성 신호처리, 음성코딩, 실시간 DSP 응용

##### ● 배 명 진 (Myung Jin Bae)



한국 음향학회지 제 14권 제2호 초조